

**Dipl.-Ing. Johannes Hofrichter**

# **Change Point Detection in Generalized Linear Models**

**DISSERTATION**

zur Erlangung des akademischen Grades  
Doktor der technischen Wissenschaften



**Technische Universität Graz**

Betreuer:

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Herwig Friedl

Institut für Statistik

Graz, Jänner 2007

# Abstract

In this thesis, we discuss linear regression type models allowing for one change as well as for multiple changes in the mean structure. Such a change is located at a specific point of the data, the so-called change point. Models with both, discontinuous or continuous changes at the change points are considered.

First, we introduce into the topic of ordinary linear models with one change point. Then we gradually extend this rather simple concept onto the class of generalized linear models (GLMs). In addition to some classical approaches for testing on the necessity of a change point we consider a test which is based on recursive residuals. After introducing recursive residuals for GLMs we investigate the performance of this test by a Monte Carlo (MC) simulation study. GLMs with multiple change points are then presented and two methods to estimate these change points are introduced. Again, some MC simulations are performed to compare the two methods. In a next step, we examine GLMs with a continuity constraint at the change points. For this class of models we apply a grid search and an exact method to estimate the change points.

Finally, we introduce change point models for panel data. In particular, we assume that the slope parameters of the last segments of all panels are the same. A new iterative algorithm is presented that allows to estimate the change points under these special assumptions. We then investigate the performance of this new algorithm by means of a further MC simulation study.

The new method is applied on hydrological data describing the discharges of a river after several rainstorms.

To apply this new algorithm and several other methods to estimate change points to real data, these methods are implemented in the R system for statistical computing. All these functions are combined in a new R-package called CpInGLM.

# Acknowledgment

The first person I would like to thank is my direct supervisor Herwig Friedl. Especially the strict and extensive comments and the many discussions and the interactions with Herwig Friedl had a direct impact on the final form and quality of this thesis. A special thanks goes to my co-advisor, Ernst Stadlober, for helping me to complete the writing of this work. I would like to thank Klemens Fuchs who kept an eye on the progress of my work and was always available when I needed his advises.

I feel a deep sense of gratitude for my late father and my mother who formed part of my vision and taught me the good things that really matter in life. The happy memory of my father still provides a persistent inspiration for my journey in this life.

I am very grateful for my partner Andrea, for her love and patience during the PhD period, for listening to my complaints and frustrations, and for believing in me.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	2
1.2	Change Points in Linear Models . . . . .	2
1.3	Implementation in R . . . . .	9
1.4	Overview . . . . .	9
<b>2</b>	<b>Theory</b>	<b>10</b>
2.1	Discontinuous Models with One Change Point . . . . .	10
2.1.1	Simple Linear Regression . . . . .	10
2.1.2	Multiple Linear Regression . . . . .	15
2.1.3	Generalized Linear Model . . . . .	18
2.1.4	Recursive Residuals . . . . .	22
2.2	Discontinuous Change Points in GLMs . . . . .	34
2.2.1	Multiple Discontinuous Change Points . . . . .	35
2.2.2	Fitting Algorithms . . . . .	36
2.2.3	Determining the Number of Change Points . . . . .	47
2.3	Continuous Change Points in GLMs . . . . .	48
2.3.1	One Continuous Change Point . . . . .	50
2.3.2	Multiple Continuous Change Points . . . . .	61
2.4	Continuous Change Points for Panel Data . . . . .	68
2.4.1	The Independent Case . . . . .	69
2.4.2	Common Slope in the Last Segment . . . . .	72
2.4.3	Fitting Procedure . . . . .	74
2.4.4	Simulation Study . . . . .	80
<b>3</b>	<b>Application</b>	<b>91</b>
3.1	Hydrological Runoff Model . . . . .	91
3.2	Data Description . . . . .	94
3.2.1	Extracting of Recession Curves . . . . .	95
3.2.2	Data Reduction . . . . .	96
3.2.3	Hydrological Assumptions . . . . .	98
3.2.4	Methods for Recession curve analysis . . . . .	99
3.3	Results . . . . .	100
<b>A</b>	<b>Generalized Linear Models</b>	<b>104</b>

<b>B Help R-Package CpInGLM</b>	<b>107</b>
CpInGLM.internal . . . . .	107
HSvsDP . . . . .	108
ccp . . . . .	108
ccpComSlope . . . . .	112
ccpComSlope.control . . . . .	114
dcp . . . . .	115
dcpBinarySplit . . . . .	118
dcpDynamicProg . . . . .	120
exmpl1 . . . . .	121
exmpl2 . . . . .	122
panel.dat . . . . .	122
plot.ccpComSlope . . . . .	123
plot.ccp . . . . .	124
plot.dcp . . . . .	126
quandt . . . . .	127
recres . . . . .	128
residuals.ccp . . . . .	129
residuals.dcp . . . . .	130
uniqueMatrix . . . . .	131
<b>Index</b>	<b>133</b>
References . . . . .	135

# List of Tables

1.1	Quandt's data. . . . .	4
2.1	Log likelihood (2.3) and (2.4) for all reasonable change points $\tau$ , assuming equal and unequal variances, respectively. . . . .	15
2.2	Percentage of finding the true change points for both algorithms in the three settings based on 2000 replications. . . . .	42
2.3	Comparison of the mean, median and standard deviation of the last slope estimates and last change point estimates, with respect to the three different methods. . . . .	81
3.1	Start- and endpoints and duration of each of the 13 recessions. . . . .	96
3.2	Estimated slopes of the third segments using four different methods. . . . .	101

# List of Figures

1.1	The two fitted models for Quandt's data. . . . .	5
1.2	Fitted change point model with two discontinuous change points at $\hat{\tau}_1 = 61$ and $\hat{\tau}_2 = 132$ for the crime data of Redding. . . . .	6
1.3	The fitted continuous change point model for the lung capacity data set with an estimated change point at the age of $\hat{\gamma} = 22$ . . . . .	7
1.4	The fitted continuous change point model for the runoff of the river Sulm with estimated change points (dotted lines) at $\hat{\gamma}_1 = 0.33$ and $\hat{\gamma}_2 = 1.71$ days. . . . .	8
2.1	Maximized log likelihood $\ell_e(\tau \hat{\beta}, \hat{\sigma}^2, \mathbf{y})$ for several change points, $\tau = 3, \dots, 17$ , assuming equal variances. . . . .	15
2.2	Relative frequencies of the change point estimate for the first model. . . . .	31
2.3	Cumulative relative frequencies of the change point estimate for the first model. . . . .	31
2.4	Relative frequencies of the change points estimate for the second model. . . . .	32
2.5	Cumulative relative frequencies of the change point estimate for the second model. . . . .	33
2.6	Relative frequencies of the change point estimate for the OLM. . . . .	34
2.7	Possible models for the 1st segment. . . . .	39
2.8	Possible combinations for the first two segments. . . . .	40
2.9	Possible combinations for a GLM with three segments for the first $m = 9, \dots, 12$ observations, in particular, for the entire data set of size $n = 12$ . . . . .	41
2.10	Relative frequency of the cumulative difference of the change points for $\varepsilon \stackrel{iid}{\sim} N(0, 1)$ . . . . .	43
2.11	Relative frequency of the cumulative difference of the change points for $\varepsilon \stackrel{iid}{\sim} N(0, 1.5)$ . . . . .	44
2.12	Fitted linear model with three change points $\hat{\tau} = (15, 30, 37)^T$ using the HS algorithm. . . . .	46
2.13	Fitted linear model with three change points $\hat{\tau} = (15, 30, 45)^T$ using the DP algorithm. . . . .	47
2.14	Comparison of the two information criteria AIC and BIC. . . . .	49

2.15	Maximized log likelihood with respect to the number of change points. . . . .	49
2.16	Cumulative relative frequencies of the change point estimate. . .	56
2.17	Relative frequencies of the change point estimate for Model 1. . .	56
2.18	Relative frequencies of the change point estimate for Model 2. . .	57
2.19	Normal plot of the deviance residuals. . . . .	59
2.20	Deviance residuals versus fitted values of a GLM with one continuous change point for the vital capacity data. . . . .	59
2.21	Curve of the difference between the maximized log likelihood at $\hat{\gamma} = 22$ and the profile log likelihood of the change point. The dashed line at 3.84 corresponds to the 95% quantile of the $\chi_1^2$ -distribution. . . . .	60
2.22	Example for a graph, where the first change point is fixed at $\gamma_1 = x_s$ and the second change point lies in between $x_r$ and $x_{r+1}$	65
2.23	Six different pattern which can occur. . . . .	66
2.24	Contour plot of the profile log likelihood depending on two continuous change points. . . . .	68
2.25	Illustration of the <i>glm</i> method to estimate a common slope for the last segments. . . . .	78
2.26	Comparison of the MC frequencies of the common last slope estimates with respect to the three different methods. . . . .	82
2.27	Comparison of the MC frequencies of the change point estimates of the first model, with respect to the three different methods. . .	83
2.28	Comparison of the MC frequencies of the change point estimates of the second model, with respect to the three different methods.	84
2.29	Comparison of the MC frequencies of the change point estimates of the third model, with respect to the three different methods. .	85
2.30	Frequencies of the number of iterations until the algorithm converged, with respect to the three considered methods. . . . .	86
2.31	Simulated panel data, where the explanatory variables of each panel are centered at $x = 8.2$ , the change point of the second panel. . . . .	87
2.32	Evolution of the estimated last slope, last change points and index of the last change points. . . . .	89
2.33	Fitted model for the panel data set, with a common slope for the last segment. . . . .	90
2.34	Some diagnostic plots for the fitted model of the first panel. . .	90
3.1	Direct, surface and base flow of the theoretical runoff model . .	92
3.2	Discharge of Sulm at Leibnitz from 01/01/1999 until 01/01/2000.	95
3.3	Discharge of the Sulm at Leibnitz from 02/06/1999 until 01/07/1999, showing five recession curves, where only the first and the fifth recession curve satisfy the restrictions for the subsequent analysis.	97
3.4	Hourly recorded discharge of the fifth recession curve of Sulm at Leibnitz. . . . .	97



3.5	Step length depending on the index for the fifth recession curve.	98
3.6	Reduced data of the fifth recession curve of the Sulm. . . . .	99
3.7	13 recessions of Sulm at Leibnitz after various rainfall events. . .	100
3.8	Evolution of the common slope $\hat{\delta}^{(t)}$ for all recession curves. . . .	102
3.9	Evolution of the second change points for all recession curves. .	102
3.10	Comparison between the fitted models of a single analysis of the eighth recession (red) and an analysis within the panel data set assuming a common slope for the last segment (black). . . . .	103

# Chapter 1

## Introduction

In ordinary linear models (OLMs) and generalized linear models (GLMs), it is usually assumed that the considered model holds for the whole data. In contrast, in this thesis, we consider linear models, where the structure of the model changes. There might be a single change or multiple changes in the structure of the model. Such a change is determined by a specific point of the data. In the remainder of this thesis we denote such a specific point as a change point. In general, change points can be known or unknown. The latter case is of main interest in this thesis. Thus, we are interested in estimating the unknown change points under different assumptions.

First, we consider a simple linear regression with one change point. Then we gradually extend this to the cases of OLMs and GLMs. In addition to classical approaches to test the necessity of a change point we consider a test based on recursive residuals. Therefore, first we describe this test in the case of OLMs. Then we introduce recursive residuals for GLMs and investigate the performance of this test applied on GLMs via Monte Carlo (MC) simulations. GLMs with multiple change points are then presented. Moreover, two different methods to estimate the change points are introduced. Again, MC simulations are performed to compare these two methods.

So far, the models of different segments are not restricted to each other. Consequently a discontinuous change of the models at the change points are allowed. In a next step, GLMs with a continuous change at the change point are considered. Therefore, a continuity constraint at the change points is mandatory. GLMs with one change point, as well as those with multiple change points are considered. Furthermore, a grid search and an exact method to estimate the change points are presented.

Finally, change point models for panel data are considered. In particular, it is assumed that the slope of the last segment of each panel is the same. To estimate the change points under these special models a new algorithm is presented. The performance of this new algorithm is investigated with MC simulations.

To apply this new algorithm and several other methods to estimate change points to real data, these methods are implemented in the R system (Ihaka &

Gentleman, 1996) for statistical computing. Moreover, all these functions are combined in a new R-package called **CpInGLM**.

In this work, matrices and vectors are denoted by boldface letters like  $\mathbf{X}$  and  $\mathbf{x} = (x_1, \dots, x_n)^T$ , respectively, where vectors are usually defined as column vectors.

## 1.1 Motivation

This thesis is motivated by a real problem in hydrology. The general question is to find a suitable statistical model that allows one to describe some properties of the catchment of a river. Such a catchment is defined as the area of the landscape where all the rain falling in this area discharges into the river. One of the properties of interest is the groundwater storage. Information about the groundwater storage can be obtained by analyzing the runoff of a river after a rainstorm. The runoff after a rainstorm is the discharge measured at a gauging station at the river. Under certain circumstances, two changes can be recognized in the time dependent behavior of the runoff. A main part of the hydrological analysis is to determine these points of change that divide the runoff into three segments. The determination of these change points is usually done by an experienced hydrologist applying graphical methods combined with a simple linear regression. Thus, results are often subjective and depend on the experience of the hydrologist. The goal is to establish a method for detecting the change points based on the data themselves. As the runoff can be described by a linear model (for details see Section 3.1), the challenge is to detect the change points in such a model. Usually there exists data from more than one runoff due to the fact that more than one rainstorm occurs over the entire observation period. Thus, observing more than one runoff yields a data set that can be interpreted as a panel data set where each panel represents one runoff after a rainstorm. According to hydrological considerations about the runoff behavior, a common slope for the last segment of each runoff is mandatory (Fank et al., 1993). Consequently, this common slope must be considered in the analysis. Thus, the statistical challenge is to detect the change points in a linear model for panel data sets where the slope in the last segment is the same for each panel.

## 1.2 Change Points in Linear Models

Linear models are widely used in statistics to describe the relationship between a response variable and one or more explanatory variables. Usually it is assumed, that the same linear model holds for the whole data set, but this is not always the fact. In contrast, the model may change after a specific point, which may or may not be known. Such a change point partitions the data into two segments where the models of these segments are different. Thus, a linear model with a change point is appropriate for data sets, where the structure

of the linear model changes after a specific point. The data must have some natural ordering to be able to find such a change point. This ordering can be determined by the time of observation or by some other observed variable (e.g. time, space, age, etc.). In general, there may be multiple change points that partition the data into several segments with the structure of the model changing after each change point. The change in the structure can be in the mean structure, in the variance structure, or in both. The most common case is a change only in the mean structure, assuming that the variance structure is the same for the whole data. In this thesis, a change in the mean structure provides a change in the parameters describing the mean, but the function, which connects the explanatory variables and the response, remains the same for all observations. A less frequent case is a change in the variance structure without a change in the mean structure.

In general, change point models can be divided into two groups. Models with a discontinuous change at the change point (denoted as discontinuous change point models) and models with a continuous change at the change point (denoted as continuous change point models). Regardless of this classification, change point models involve three issues: First, the choice of the location of the change points; second, the choice of an appropriate number of change points; and third, the choice of a suitable model for each segment. The main aim in this work is to develop a method for estimating the change points, especially under some restrictions mentioned later. The second question will be only briefly discussed here (Subsection 2.2.3) and we assume, that for each segment the same type of model holds but we allow for different parameter values.

In the remaining part of this section various types of linear models with one or multiple change points are illustrated to give a short overview on different change point models.

The first change point model under consideration is a simple linear regression with one discontinuous change point. Consider  $n$  pairs of observations  $(x_i, y_i)$ , where  $y$  is the response variable and  $x$  is the explanatory variable. The index  $i = 1, \dots, n$  describes the ordering of these observations. Note, this induces that the explanatory variable  $x_i$  need not to be ordered. Furthermore it is assumed that the relationship between  $x$  and  $y$  can be described by a simple linear regression, where the structure of the mean changes after a change point  $\tau \in \{i | 3 \leq i \leq n - 2\}$ . This restriction on  $i$  is needed to ensure that the parameters in the model are estimable. Thus, the observations  $(x_i, y_i)$  follow a linear model for  $i \leq \tau$  and another linear model for  $i > \tau$ . In addition, there is no continuity constraint for the models at the change point, which allows a discontinuous change at the change point. Therefore, the model is given by

$$y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_{1i} & i = 1, \dots, \tau \\ \alpha_2 + \beta_2 x_i + \varepsilon_{2i} & i = \tau + 1, \dots, n, \end{cases} \quad (1.1)$$

where  $\alpha_d$  and  $\beta_d$ ,  $d = 1, 2$ , are the unknown parameters of interest and  $\varepsilon_{di}$  are independent errors, with  $\varepsilon_{di} \stackrel{iid}{\sim} N(0, \sigma_d^2)$ . Note, the case of different variances

$(\sigma_1^2 \neq \sigma_2^2)$  indicates an additional change in the variance structure. In contrast, for equal variances only the mean structure changes at the change point. Indeed, model (1.1) can be extended in several ways. First, there might be multiple changes in the linear model. Second, the response  $y$  can depend on more than one explanatory variable. These explanatory variables may or may not include the variable which describes the natural ordering. Again, if it is not included, the index  $i$  describes the ordering.

**Example 1.1** *This example was first illustrated by Quandt (1958). A simple linear regression model with one discontinuous change point is considered. The artificial data was generated in the following way. For the explanatory variable  $x$ , a single independent variable was generated by a random permutation of the first 20 integers. The errors  $\varepsilon_i$  were generated as 20 iid normal variates with zero mean and variance equal to one. These errors were added to a mean model that changed after the first 12 observations, giving*

$$y_i = \begin{cases} 2.5 + 0.7x_i + \varepsilon_i & i = 1, \dots, 12 \\ 5.0 + 0.5x_i + \varepsilon_i & i = 13, \dots, 20. \end{cases}$$

The data, listed in Table 1.1, and the fitted model with a change point at  $\hat{\tau} = 12$  are plotted in Figure 1.1. Note, that in this data the ordering is determined

$i$	$x_i$	$y_i$	$i$	$x_i$	$y_i$
1	4	3.473	11	15	13.036
2	13	11.555	12	11	8.264
3	5	5.714	13	3	7.612
4	2	5.710	14	14	11.802
5	6	6.046	15	16	12.551
6	8	7.650	16	10	10.296
7	1	3.140	17	7	10.014
8	12	10.312	18	19	15.472
9	17	13.353	19	18	15.650
10	20	17.197	20	9	9.871

Table 1.1: Quandt's data.

by the index  $i$  of the observation. Consequently, the explanatory variable need not to be ordered, which can be seen in Table 1.1. The results are two different fitted lines for the first 12 and last 8 observations, respectively (Figure 1.1).

If the change point is a nuisance parameter and not of main interest, a discontinuous change point model can be interpreted as a mixture of two linear regressions. This was first considered by Quandt and Ramsey (1978). Leisch (2004) introduced a general framework for estimating finite mixtures of linear models, which is based on the EM-algorithm. In addition, he developed the R-package `flexmix` for estimating such mixture models.

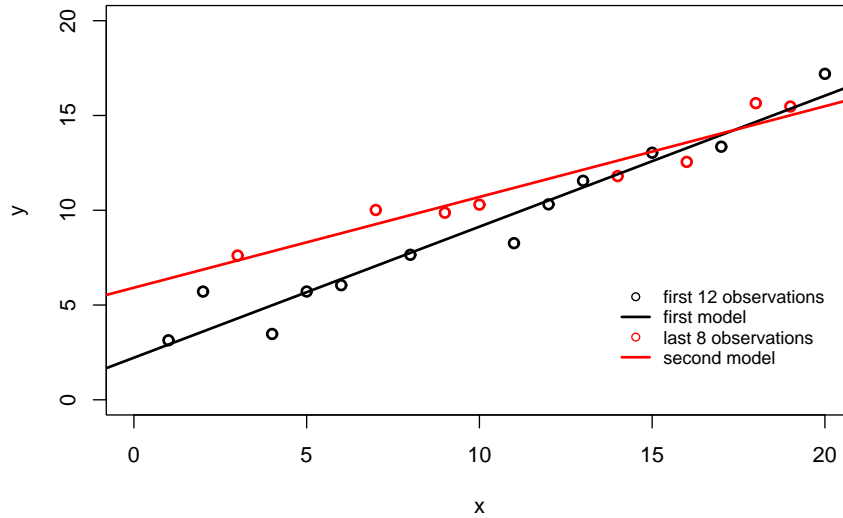


Figure 1.1: The two fitted models for Quandt's data.

For another type of change point model, model (1.1) is again considered, but now the explanatory variable describes the ordering. This is often the case when the time of observation is important. Let the explanatory variable  $x$  be the time of observation, then  $x_i \leq x_{i+1}$  for all  $i \in \{1, \dots, n-1\}$ . In general, it is assumed, that the change point lies within the interval  $[\min x_i; \max x_i]$  and can be any value within this interval. However, without additional information about the change point, it is not possible to determine the exact change point. In contrast, it is only possible to determine the two consecutive observations in between which the change point is located. Thus, regardless of whether the index or an explanatory variable describes the ordering, the index  $i$  is used to describe the change point.

**Example 1.2** *The data we considered in this example is published at the website (<http://reddingpolice.org/>) of the police department located in the City of Redding, California. On pure chance we found this data set on the web and decided that this is a good example for a change point model where the time of observation describing the ordering is the explanatory variable. Here monthly data about the number of assaults of the years 1989-2004 was analyzed. Between 1989 and 1999, simple assaults were included in the assault totals. Since 2000, only aggravated assaults were included for the totals. This change can be clearly seen in Figure 1.2, where the data and the fitted linear model with two discontinuous change points are plotted. In addition, there is a second change detected at January 1994. Up until this date, the number of assaults follow an increasing trend, and afterwards, a decreasing one.*

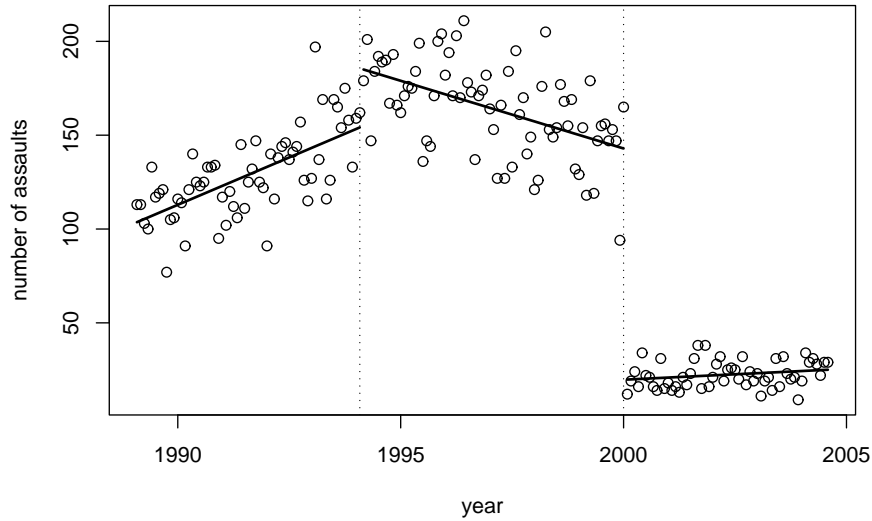


Figure 1.2: Fitted change point model with two discontinuous change points at  $\hat{\tau}_1 = 61$  and  $\hat{\tau}_2 = 132$  for the crime data of Redding.

The next type of a change point model is again a linear model, but with a continuity constraint at the change point. Consider again  $n$  pairs of observations  $(x_i, y_i)$ , with a linear model up to an unknown change point and another linear model afterwards. In contrast to model (1.1), a continuity constraint at the change point is assumed. Thus, both models must predict the same mean value at the change point, which results in a continuous transition of the two models. Furthermore, it is assumed that the explanatory variable is sorted in ascending order, i.e.  $x_i \leq x_{i+1}$ ,  $i = 1, \dots, n-1$ . Now the location of the change point is no longer restricted to an observed  $x_i$ . Instead, it can be any value from within the interval  $[a; b]$ , where  $a = \min x_i = x_1$  and  $b = \max x_i = x_n$ . To distinguish the continuous case from the discontinuous one, we now denote the change point by  $\gamma$ . The model of a linear regression with one continuous change point can then be stated as

$$y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_{1i} & a \leq x_i \leq \gamma \\ \alpha_2 + \beta_2 x_i + \varepsilon_{2i} & \gamma < x_i \leq b, \end{cases} \quad (1.2)$$

with continuity constraint

$$\alpha_1 + \beta_1 \gamma = \alpha_2 + \beta_2 \gamma.$$

Again  $\alpha_d$  and  $\beta_d$  are the unknown parameters of interest and  $\varepsilon_{di}$  are iid errors, with  $\varepsilon_{di} \stackrel{iid}{\sim} N(0, \sigma_d^2)$ . Of course, this model can also be extended to the case of multiple continuous change points. A multiple linear regression model with

continuous change points in more than one explanatory variable is imaginable, but is not considered in this work.

**Example 1.3** *The data is from a health study in Styria (Austria) of the lung capacity of people living in two regions and may be found in Friedl (1998). One point of interest was the dependence of vital capacity VC on the age of the person. In this study, 79 male persons between the ages of 16 years and 56 years were examined. The data shows a clear, increasing trend for younger people and a decreasing trend for older people (Figure 1.3). Assuming that both trends are linear, a simple linear regression with one change point seems to be appropriate for this data set. Of course, only a continuous change of the model at the change point makes sense. Thus, the two models must predict the same mean value at the change point and a continuity restriction is mandatory. The fitted model with an estimated continuous change point at the age of  $\hat{\gamma} = 22$  is plotted in Figure 1.3.*

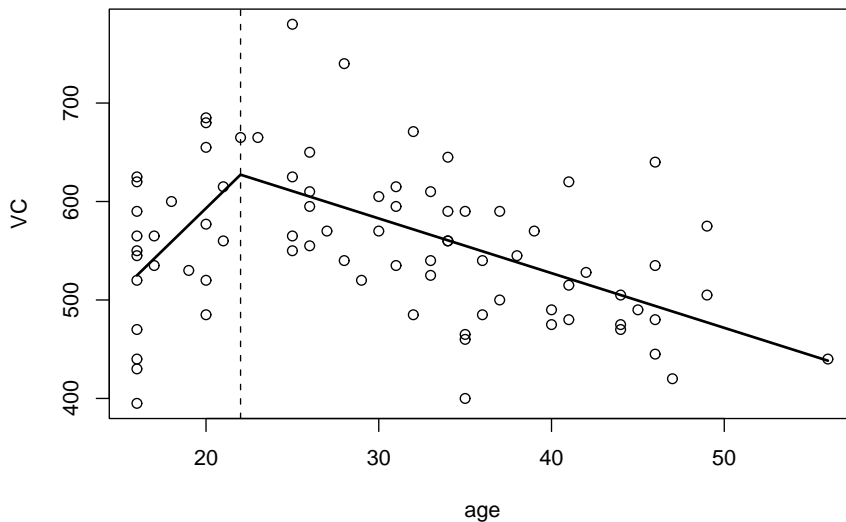


Figure 1.3: The fitted continuous change point model for the lung capacity data set with an estimated change point at the age of  $\hat{\gamma} = 22$ .

The last type of change point model considered in this section is the extension of change point models to the wide class of GLMs. For a detailed description of GLMs see Appendix A. In GLMs, a change in the mean structure can be either a change in the parameters or a change in the link function. Furthermore, a change in the variance structure is either due to a change in the dispersion parameter or because the variance function is different in each segment. In this thesis, only change point models are considered where the



link function and the variance function are the same for all segments. Thus, only the linear parameters change at the change points. Again, discontinuous or continuous change points are considered in the GLM framework.

GLMs with multiple continuous change points play a main role in this thesis. This is due to the fact that the runoff can be modeled in the GLM framework and the changes in the runoff model must be continuous. In the next example, the mean runoff is modeled through a GLM with two continuous change points.

**Example 1.4** *The data of a runoff recorded at a gauging station located on the river Sulm in Styria (Austria) is analyzed. The runoff was recorded in summer 1999, starting on 08/08/99 and continued for almost 5 days. The discharge was measured in  $m^3/sec$  and is modeled as a function of time using a GLM with two continuous change points. It is assumed that the response follows a Normal distribution and the link function is  $g(\mu) = \mu^{-1/2}$ . Furthermore, it is assumed that the dispersion parameter is the same for all observations. The data and the fitted model are plotted in Figure 1.4. The estimated change points are at  $\hat{\gamma}_1 = 0.33$  and  $\hat{\gamma}_2 = 1.71$  days.*

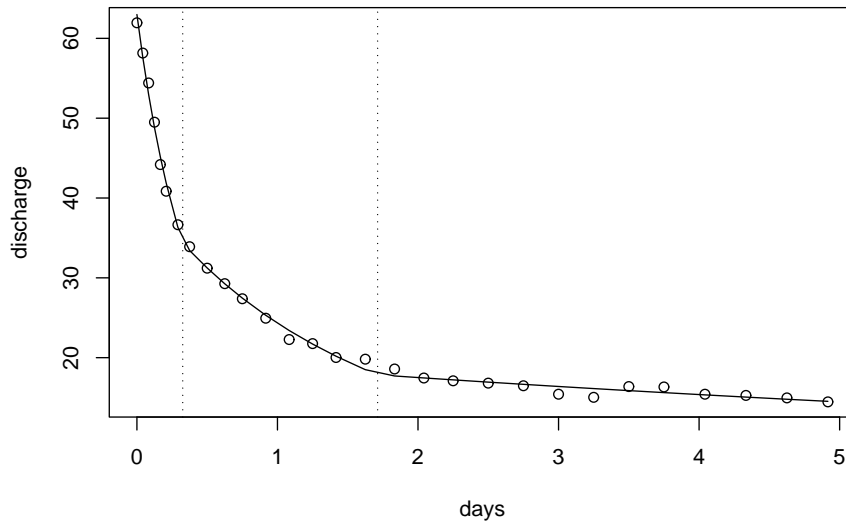


Figure 1.4: The fitted continuous change point model for the runoff of the river Sulm with estimated change points (dotted lines) at  $\hat{\gamma}_1 = 0.33$  and  $\hat{\gamma}_2 = 1.71$  days.

### 1.3 Implementation in R

All the calculation and simulation in this work was done within the R system. Therefore, several functions to estimate change points under the assumptions mentioned above are implemented in R. These functions are summarized in a new package called `CpInGLM`. The help documentation of this package is included in Appendix B.

With the use of some examples, the usage of these functions is demonstrated. To highlight the R-code, it is printed in `verbatim`. Lines beginning with the prompt `>` are the input to R, e.g.

```
> 2 + 3
```

and the resulting output are lines without a prompt, e.g.

```
5
```

The source files of R and a detailed description can be obtained from the Home Page <http://www.r-project.org/>. As R is an open source project which its syntax very similar to S, the script language of S-plus, the book of Venables and Ripley (2002) provides a comprehensive description of many statistical applications using S.

### 1.4 Overview

The remainder of this thesis is organized as follows. Section 2.1 describes change point models with a discontinuous change point. They are considered for OLMs and GLMs. In addition, recursive residuals for both type of models are introduced. Section 2.2 presents the extension to the case of multiple discontinuous change points in GLMs. Furthermore, two algorithms for estimating these change points are presented. These two algorithms are compared using an MC simulation. Section 2.3 gives an introduction to continuous change point models in the GLM framework. Models with either one or multiple continuous change points are discussed. In Section 2.4 a new method for estimating continuous change points for panel data sets is introduced. This new method considers continuous change points in GLMs where the slope of the last segment is the same for all panels. Finally, an MC simulation demonstrates nice properties of the performance of this new algorithm

In Chapter 3, the new method is applied to a hydrological data set. Section 3.1 provides a short discussion on a theoretical hydrological runoff model and how this model fits into GLM framework. Section 3.2 describes the data and the data preprocessing. Section 3.3 presents the results of this application.

Appendix A gives a short overview on GLMs. Appendix B describes R-functions for fitting change points in the GLM context.

# Chapter 2

## Theory

### 2.1 Discontinuous Models with One Change Point

Discontinuous change point models are models with no continuity constraints at the change points. Thus, the models of all segments are not restricted to common values at the change points. Consequently, for a known change point, the models of each segment are autonomous and all parameters in the linear predictor can be estimated separately. An unknown change point can be either estimated by a simple grid search over all feasible possibilities or by analyzing recursive residuals (see Subsection 2.1.4). The second method is appropriate if there is only one change point in the model, or the number of change points is small with respect to the sample size.

This section considers change point models with one discontinuous change point. After introducing such a model for OLMs, it is then generalized for the wider class of GLMs. Finally, recursive residuals for OLMs and GLMs are introduced, which can be used to estimate the change point as well as to test the necessity of a change point.

#### 2.1.1 Simple Linear Regression

Consider a simple linear regression model with a discontinuous change at a fixed but unknown change point. Let  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , denote pairs of observations, where  $y_i$  is the response and  $x_i$  some explanatory variable. Let us further assume that such  $n$  pairs  $(x_i, y_i)$  of observations can be arranged in some natural ordering. In this thesis, if not quoted otherwise, the index  $i$  describes this kind of order. Thus, the change point  $\tau$  is given by any index  $i$  and determines the observation  $x_\tau$ , after which the structural change in the relationship between  $x_i$  and  $y_i$  might occur. The change point  $\tau$  partitions the data into two separate segments, in which the mean structure as well as the variance may be different. In fact, the first  $\tau$  observations in a sample of size  $n$  follow one OLM and the last  $n - \tau$  observations follow another OLM. The

linear parameters of these two models are  $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11})^T$  and  $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21})^T$ , respectively. Then, such an OLM can be written as

$$y_i = \begin{cases} \beta_{10} + x_i \beta_{11} + \varepsilon_{1i} & i = 1, \dots, \tau \\ \beta_{20} + x_i \beta_{21} + \varepsilon_{2i} & i = \tau + 1, \dots, n, \end{cases} \quad (2.1)$$

where the errors  $\varepsilon_{di}$  are independent random variables and follow a normal distribution with zero mean and variance  $\sigma_1^2$  for  $i \leq \tau$  and variance  $\sigma_2^2$  for  $i > \tau$ , i.e.  $\varepsilon_{1i} \stackrel{iid}{\sim} N(0, \sigma_1^2)$  and  $\varepsilon_{2i} \stackrel{iid}{\sim} N(0, \sigma_2^2)$ , respectively. Such a model was first considered by Quandt (1958). He introduced a maximum likelihood (ML) method for estimating the unknown parameters  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ ,  $\sigma^2 = \sigma_1^2 = \sigma_2^2$  and  $\tau$ , which we describe in the following paragraph. Quandt's data and the corresponding change point model are illustrated in Example 1.1.

The parameters of interest are the linear parameter  $\boldsymbol{\beta}$  and the change point  $\tau$ . To guarantee the estimability of the parameters  $\boldsymbol{\beta}$  and  $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)^T$ , possible values of  $\tau$  are restricted to  $\{3, 4, \dots, n-3\}$ . To estimate these parameters with the ML method, we have to take a closer look to the log likelihood according to model (2.1). The log likelihood of a simple linear regression is

$$\ell(\alpha, \beta, \sigma^2 | \mathbf{y}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2,$$

where  $\alpha$  and  $\beta$  are the intercept and slope of the simple linear regression, respectively. Then, in the case where  $\tau$  is known, the log likelihood under model (2.1) is

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \tau, \mathbf{y}) &= -\frac{\tau}{2} \log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2} \sum_{i=1}^{\tau} (y_i - \beta_{10} - \beta_{11}x_i)^2 \\ &\quad - \frac{n-\tau}{2} \log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2} \sum_{i=\tau+1}^n (y_i - \beta_{20} - \beta_{21}x_i)^2, \end{aligned}$$

or

$$\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \tau, \mathbf{y}) = \ell(\beta_{10}, \beta_{11}, \sigma_1^2 | y_1, \dots, y_\tau) + \ell(\beta_{20}, \beta_{21}, \sigma_2^2 | y_{\tau+1}, \dots, y_n). \quad (2.2)$$

The first term on the right hand side of (2.2) is the log likelihood of the first  $\tau$  observations and the second term is the log likelihood of the last  $n - \tau$  observations. For  $\tau$  known, both terms are mutually independent. Thus, the ML estimates for  $\boldsymbol{\beta}$  and  $\boldsymbol{\sigma}^2$  are the ML estimates of the two separate models.

In the case of  $\tau$  unknown, the change point has to be estimated. The main problem in estimating the change point is that there is no solution in closed form for estimating the parameter  $\boldsymbol{\beta}$  and  $\tau$  simultaneously. This is due to the fact that the ML estimate of the parameter  $\boldsymbol{\beta}$  is a function of  $\tau$ . It is only possible for a given value of  $\tau$  to derive the ML estimate of  $\boldsymbol{\beta}$ . Therefore, the only feasible way to estimate the change point is to apply a grid search over a

set of all possible values of  $\tau$ . Now, for an arbitrary  $\tau \in \{3, 4, \dots, n-3\}$  the log likelihood given the ML estimates  $\hat{\boldsymbol{\beta}}$  and  $\hat{\sigma}^2$  is

$$\ell_u(\tau|\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{\tau}{2} \log \hat{\sigma}_1^2 - \frac{n-\tau}{2} \log \hat{\sigma}_2^2 - \frac{n}{2} \quad (2.3)$$

for  $\sigma_1^2 \neq \sigma_2^2$  and

$$\ell_e(\tau|\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\tau \hat{\sigma}_1^2 + (n-\tau) \hat{\sigma}_2^2) - \frac{n}{2} \quad (2.4)$$

for  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , where the subscripts  $u$  and  $e$  stand for unequal and equal variances, respectively. The ML estimate  $\hat{\tau}$  is the value of  $\tau$  that maximizes (2.3) respectively (2.4).

Next we consider testing whether there is a change in the regression regime or not. A very common method for testing hypothesis is the likelihood ratio (LR) test. It is applicable for testing nested models and the test statistic is defined as

$$\lambda(\mathbf{y}) = \frac{\sup_{\Theta_0} L(\boldsymbol{\theta}|\mathbf{y})}{\sup_{\Theta} L(\boldsymbol{\theta}|\mathbf{y})},$$

where  $L(\boldsymbol{\theta}|\mathbf{y})$  is the likelihood function of the parameter vector  $\boldsymbol{\theta}$  for the given data  $\mathbf{y}$  and  $\Theta$  is the entire parameter space. The set  $\Theta_0$  is the parameter space restricted under  $H_0$  and is necessarily a subset of  $\Theta$ , i.e.  $\Theta_0 \subset \Theta$ . Using the ML method for estimating the parameter  $\boldsymbol{\theta}$ , the LR test statistic can be written as

$$\lambda(\mathbf{y}) = \frac{L(\hat{\boldsymbol{\theta}}_0|\mathbf{y})}{L(\hat{\boldsymbol{\theta}}|\mathbf{y})},$$

where  $\hat{\boldsymbol{\theta}}$  is the unrestricted ML estimate of  $\boldsymbol{\theta}$  which can be realized in the entire parameter space  $\Theta$ , and  $\hat{\boldsymbol{\theta}}_0$  is the restricted ML estimate where the maximization is restricted to  $\Theta_0$ . Under some regularity conditions, minus twice the LR test statistic, i.e.

$$\Lambda(\mathbf{y}) = -2 \log \lambda(\mathbf{y}),$$

follows asymptotically a  $\chi^2$ -distribution with  $q$  degrees of freedom, where  $q$  is the difference of the number of parameters in the models under  $H_0$  and  $H_1$ , respectively (see Casella & Berger, 2002, for a detailed discussion).

To test whether there is a change in the regression regime or not, and considering model (2.1), the hypothesis is

$$\begin{aligned} H_0 &: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 \\ H_1 &: \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2. \end{aligned}$$

An assumption for applying the LR test is that the models under  $H_0$  and  $H_1$  are nested. For model (2.1) it is not obvious that a simple linear regression without a change point is nested in model (2.1). To see this let  $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1 + \boldsymbol{\delta}$  with  $\boldsymbol{\delta} = (\delta_0, \delta_1)^T$  and

$$z_i = \begin{cases} 0 & i = 1, \dots, \tau \\ 1 & \text{otherwise.} \end{cases}$$

Then (2.1) can be written as

$$y_i = \beta_{10} + \beta_{11}x_i + z_i(\delta_0 + \delta_1x_i) + \varepsilon_i \quad i = 1, \dots, n, \quad (2.5)$$

and it can be clearly seen, that an OLM without a change point, given by

$$y_i = \beta_{10} + \beta_{11}x_i + \varepsilon_i \quad i = 1, \dots, n,$$

is nested in (2.5). Thus, this assumption for the LR test is satisfied.

Recall that the maximized log likelihood according to a simple linear regression is

$$\ell(\hat{\boldsymbol{\beta}}, \tilde{\sigma}^2 | \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\tilde{\sigma}^2) - \frac{n}{2}, \quad (2.6)$$

where  $\tilde{\sigma}^2$  is the usual ML estimate of  $\sigma^2$  based on all observations. Then  $\Lambda_u(\mathbf{y})$  is obtained by subtracting (2.4) from (2.6) as

$$\Lambda_u(\mathbf{y}) = n \log(\tilde{\sigma}^2) - \tau \log(\hat{\sigma}_1^2) - (n - \tau) \log(\hat{\sigma}_2^2).$$

In case of equal variances this becomes

$$\Lambda_e(\mathbf{y}) = n \log(\tilde{\sigma}^2) - n \log(\tau \hat{\sigma}_1^2 - (n - \tau) \hat{\sigma}_2^2).$$

As mentioned above, under standard regularity conditions  $\Lambda_e(\mathbf{y})$  is asymptotically  $\chi^2$ -distributed. However, as Seber and Wild (1989) noted, standard asymptotical theory does not apply here because  $\tau$  takes only discrete values and  $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$  is also true if the change point lies outside the range of the data. Moreover, Hawkins (1980) showed that the LR test statistic tends to infinity as  $n$  increases. Therefore, the LR test can only be used as an approximative device. Another test was introduced by Chow (1960). He assumed that the change point is known and uses the usual F-test statistic for testing two nested models in linear regression. As usually the change point is unknown it is taken to be  $\tau = n/2$ . The problem that arises here is, that either the model on the left hand side or the model on the right hand side of the change point contains observations of the other regime. Thus, this test only provides satisfactory results if the true change point is  $n/2$ .

Farley and Hinich (1970) presented another test statistic for testing a change point in an OLM based on a Bayesian approach. They considered the model

$$y_i = \begin{cases} \alpha + \beta x_i + \varepsilon_i & i = 1, \dots, \tau - 1 \\ \alpha - \delta x_\tau + (\beta + \delta)x_i + \varepsilon_i & i = \tau, \dots, n, \end{cases} \quad (2.7)$$

where  $\delta$  determines the shift at the change point and  $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ . Using the notation from above and defining

$$z_i = \begin{cases} 0 & i = 1, \dots, \tau - 1 \\ x_i - x_\tau & \text{otherwise,} \end{cases}$$

then (2.7) can be written as

$$y_i = \alpha + \beta x_i + \delta z_i + \varepsilon_i \quad i = 1, \dots, n,$$

and the hypothesis for testing a shift  $\delta$  in the OLM at the change point is

$$\begin{aligned} H_0 : \delta &= 0 \\ H_1 : \delta &\neq 0. \end{aligned}$$

Farley and Hinich (1970) suggested, that a priori every value of  $\tau$  is equally likely, i.e.

$$P(\tau = i) = 1/n \quad \text{for} \quad i = 1, \dots, n. \quad (2.8)$$

Then under  $H_0$  the marginal response mean is assumed to follow

$$E_0[y_i] = \alpha + \beta x_i. \quad (2.9)$$

Under the alternative, i.e. if a shift of size  $\delta$  occurs at the change point  $\tau = i^*$ , we have the conditional mean model

$$E_\delta[y_i | \tau = i^*] = \alpha + \beta x_i + \delta z_i.$$

which is

$$E_\delta[y_i | \tau = i^*] = \begin{cases} \alpha + \beta x_i & i = 1, \dots, i^* \\ \alpha + \beta x_i + \delta(x_i - x_{i^*}) & \text{otherwise.} \end{cases}$$

Using (2.8) yields the marginal mean

$$E_\delta[y_i] = \frac{1}{n} \sum_{j=1}^n E_\delta[y_i | \tau = j]$$

which is

$$E_\delta[y_i] = \begin{cases} \alpha + \beta x_i & i = 1 \\ \alpha + \beta x_i + \delta \frac{1}{n} \sum_{j=1}^i (x_i - x_j) & \text{otherwise.} \end{cases} \quad (2.10)$$

Farley and Hinich (1970) substituted (2.9) and (2.10) in the likelihood function of the OLM with and without a change point respectively, and gave a first order approximation of the LR test statistic. Furthermore, they mentioned that for  $\sigma^2$  known, this statistic follows a normal distribution.

**Example 2.1 (Example 1.1 continued)** *The change point  $\tau$  for Quandt's data is estimated by a simple grid search over all reasonable change points  $\tau \in \{3, 4, \dots, 17\}$ . The values of the log likelihood (2.3) and (2.4) under both assumptions on the variances ( $\sigma_1^2$  and  $\sigma_2^2$ ) are given in Table 2.1. In both cases, the maximum is at  $\hat{\tau} = 12$ , which equals the true value. For the case of equal variances the values of  $\ell_e(\tau | \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y})$  are plotted in Figure 2.1. This plot shows that the maximum of  $\ell_e(\tau | \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y})$  at  $\hat{\tau} = 12$  seems to be well defined.*

$\tau$	$\ell_e(\tau \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y})$	$\ell_u(\tau \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y})$	$\tau$	$\ell_e(\tau \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y})$	$\ell_u(\tau \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y})$
3	-32.05	-62.12	11	-30.49	-60.81
4	-34.51	-64.92	12	<b>-27.63</b>	<b>-57.89</b>
5	-33.89	-63.97	13	-31.66	-62.38
6	-33.37	-63.32	14	-31.55	-62.04
7	-31.90	-61.86	15	-30.09	-61.20
8	-31.25	-61.30	16	-30.55	-61.37
9	-30.79	-61.04	17	-31.93	-63.29
10	-30.65	-60.72			

Table 2.1: Log likelihood (2.3) and (2.4) for all reasonable change points  $\tau$ , assuming equal and unequal variances, respectively.

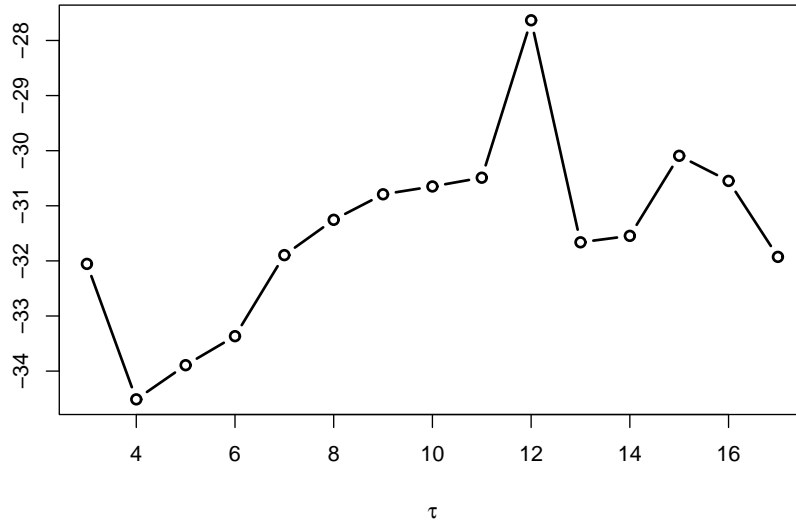


Figure 2.1: Maximized log likelihood  $\ell_e(\tau|\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y})$  for several change points,  $\tau = 3, \dots, 17$ , assuming equal variances.

### 2.1.2 Multiple Linear Regression

Next we consider an OLM with more than one explanatory variable, commonly known as multiple linear regression. Again, let  $y_i$ ,  $i = 1, 2, \dots, n$  denote observations on the response variable. In contrast to Subsection 2.1.1, let  $\mathbf{x}_i \in \mathbb{R}^{p \times 1}$  denote the column vector of  $p$  independent explanatory variables, i.e.  $\mathbf{x}_i = (1, x_{i2}, \dots, x_{ip})^T$ , with  $x_{i1} = 1$  for all  $i$ , to include an intercept in the model. Then an OLM with one discontinuous change point can be written as

$$y_i = \begin{cases} \mathbf{x}_i^T \boldsymbol{\beta}_1 + \varepsilon_{1i} & i = 1, \dots, \tau \\ \mathbf{x}_i^T \boldsymbol{\beta}_2 + \varepsilon_{2i} & i = \tau + 1, \dots, n, \end{cases} \quad (2.11)$$



where  $\boldsymbol{\beta}_d$ ,  $d = 1, 2$ , are  $p \times 1$  vectors of unknown parameters and  $\varepsilon_{di}$  are iid errors with  $\varepsilon_{di} \stackrel{iid}{\sim} N(0, \sigma_d^2)$ . To ensure valid estimates for  $\boldsymbol{\beta}_d$  and  $\sigma_d^2$  the possible values of  $\tau$  are restricted to  $\{p + 1, \dots, n - p - 1\}$ . Moreover, it is assumed that the first  $p + 1$  and the last  $n - p - 1$  vectors of  $\mathbf{x}_i$  are linearly independent.

In matrix representation, model (2.11) can be written as two separate OLMs

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \\ \mathbf{y}_2 &= \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \end{aligned} \quad (2.12)$$

where  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are both column vectors of the first  $\tau$  and the last  $n - \tau$  observations of the response variable, respectively. The matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are the first  $\tau$  and the last  $n - \tau$  row vectors of the design matrix respectively, and hence given by  $\mathbf{X}_1 = (\mathbf{x}_1, \dots, \mathbf{x}_\tau)^T$  and  $\mathbf{X}_2 = (\mathbf{x}_{\tau+1}, \dots, \mathbf{x}_n)^T$ . Furthermore, the error vectors follow a Normal distribution, i.e.  $\boldsymbol{\varepsilon}_d \sim N(\mathbf{0}, \sigma_d^2 \mathbf{I}_d)$ , where  $\mathbf{I}_d$  is the identity matrix with rank  $\tau$  for  $d = 1$  and rank  $n - \tau$  for  $d = 2$ .

As there is no continuity constraint for the two models at the change point, the two models of (2.12) are autonomous and can be written as

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{pmatrix}. \quad (2.13)$$

Note that the design matrix in (2.13) is block diagonal, which indicates independence between the estimates of  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$ . Thus, if the change point  $\tau$  is known, the log likelihood can be partitioned into two terms, namely

$$\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \tau, \mathbf{y}) = \ell(\boldsymbol{\beta}_1, \sigma_1^2 | y_1, \dots, y_\tau) + \ell(\boldsymbol{\beta}_2, \sigma_2^2 | y_{\tau+1}, \dots, y_n), \quad (2.14)$$

with  $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)^T$ . These two terms correspond to the log likelihood of the first  $\tau$  observations and the last  $n - \tau$  observations and both terms are mutually independent. Thus, the ML estimates for  $\boldsymbol{\beta}$  and  $\boldsymbol{\sigma}^2$  are the ML estimates of the two separate models and are given by

$$\hat{\boldsymbol{\beta}}_d = (\mathbf{X}_d^T \mathbf{X}_d)^{-1} \mathbf{X}_d^T \mathbf{y}_d, \quad d = 1, 2$$

and

$$\hat{\sigma}_1^2 = \frac{1}{\tau} \hat{S}_1^2, \quad \hat{\sigma}_2^2 = \frac{1}{n - \tau} \hat{S}_2^2,$$

where

$$\hat{S}_d^2 = (\mathbf{y}_d - \mathbf{X}_d \hat{\boldsymbol{\beta}}_d)^T (\mathbf{y}_d - \mathbf{X}_d \hat{\boldsymbol{\beta}}_d)$$

is the residual sum of squares for the model in the  $d$ th segment.

In the case of  $\tau$  unknown, however, the change point has to be estimated. The ML estimate of  $\tau$  is again the value which maximizes the log likelihood (2.14) at the given ML estimates  $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T)^T$  and  $\hat{\boldsymbol{\sigma}}^2$  of the two models. Note that  $\hat{\boldsymbol{\beta}}$  and  $\hat{\boldsymbol{\sigma}}^2$  are again functions of  $\tau$  and have to be estimated for each  $\tau$  separately. The log likelihood (2.14) at these ML estimates is given by

$$\begin{aligned} \ell(\tau | \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2, \mathbf{y}) &= -\frac{n}{2} \log(2\pi) - \frac{\tau}{2} \log \hat{\sigma}_1^2 - \frac{n - \tau}{2} \log \hat{\sigma}_2^2 \\ &\quad - \frac{1}{2\hat{\sigma}_1^2} \tau \hat{\sigma}_1^2 - \frac{1}{2\hat{\sigma}_2^2} (n - \tau) \hat{\sigma}_2^2. \end{aligned}$$

Thus,  $\hat{\tau}$  is obtained by maximizing

$$\ell(\tau|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2, \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{\tau}{2} \log \hat{\sigma}_1^2 - \frac{n-\tau}{2} \log \hat{\sigma}_2^2 - \frac{n}{2}, \quad (2.15)$$

with respect to  $\tau = p+1, \dots, n-p-1$ .

Under the assumption  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  the ML estimate of  $\sigma^2$  is

$$\hat{\sigma}^2 = \frac{1}{n} \left( \hat{S}_1^2 + \hat{S}_2^2 \right)$$

and (2.15) reduces to

$$\ell(\tau|\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2}.$$

This means, that in the case of equal variances,  $\hat{\tau}$  minimizes  $\hat{S}_1^2 + \hat{S}_2^2$ .

For testing the necessity of a change in an OLM, consider again the LR test statistic and Chow's test. The quantity  $\Lambda_u(\mathbf{y})$  based on the LR test statistic for model (2.12) is

$$\Lambda_u(\mathbf{y}) = \left[ n \log \frac{\tilde{S}^2}{n} - \tau \log \frac{\hat{S}_1^2}{\tau} - (n-\tau) \log \frac{\hat{S}_2^2}{n-\tau} \right]_{\tau=\hat{\tau}}, \quad (2.16)$$

where

$$\tilde{S}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

is the residual sum of squares for the model assumed under the null hypothesis. This statistic can be used to tests for a change in the variance as well as for a change in the regression coefficients (Worsley, 1983). In the case of equal variances,  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , we have

$$\Lambda_e(\mathbf{y}) = n \log \left[ \frac{\tilde{S}}{\hat{S}_1^2 + \hat{S}_2^2} \right]_{\tau=\hat{\tau}}.$$

For  $\tau$  known and  $\sigma_1^2 = \sigma_2^2$  the usual  $F$ -test statistic for  $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$  is

$$F_\tau = \frac{\left[ \tilde{S}^2 - (\hat{S}_1^2 + \hat{S}_2^2) \right] / p}{(\hat{S}_1^2 + \hat{S}_2^2) / (n-2p)},$$

which under  $H_0$  follows an  $F$ -distribution with  $p$  and  $n-2p$  degrees of freedom. Worsley (1983) and Beckman and Cook (1979) suggested to use a generalized  $F$ -test statistic, namely

$$F_{max} = \max_{p < \tau < n-p} F_\tau$$

for testing  $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ . They gave an approximation to the distribution of  $F_{max}$  under the null hypothesis based on the Bonferroni inequality. As the

distribution of the  $F$ -test statistic depends on the configuration of the design matrix, Beckman and Cook (1979) simulated four OLMs with different design matrices to investigate the influence of the design on the distribution of the  $F$ -test statistic. They showed that there is a non-negligible influence. Furthermore, they gave approximative upper bounds for the 90% percentiles of the  $F_{max}$  distribution based on these simulations. The bounds were conservative when testing for a change in the linear regression or if the variability of the explanatory variable is large. Therefore, if the variability of the explanatory variable is greater than in the considered design of Beckman and Cook (1979), they recommended to apply the usual Bonferroni inequality instead of the simulated values. Worsley (1983) introduced upper bounds for the percentiles of the  $F_{max}$  distribution, based on an improved Bonferroni inequality (Worsley, 1982). Furthermore, to avoid the integration for calculating these bounds, he approximated these bounds using the MacLaurin series. He showed that both, the exact and the approximated bounds are more accurate than the bounds calculated with the usual Bonferroni inequality.

Farley, Hinich, and McGuire (1975) introduced a simpler interpretation of the test presented by Farley and Hinich (1970). Furthermore, they compared the power of the three methods, the Chow test, the approach based on  $F_{max}$  and the method introduced by Farley and Hinich (1970). Their results, based on a few simulations were that Chow's test using  $\tau = n/2$  is most powerful if the change point lies in the middle of the data. In this case, the method introduced by Farley and Hinich (1970) has less power than that of Chow, but performs better than the LR test. In contrast, if the change point lies near the left or right extremes of the data, the LR test is most powerful.

Esterby and El-Shaarawi (1981) considered a linear regression with one change point, where the explanatory variables are polynomials of unknown degree  $p_1$  and  $p_2$  for the first and second segment, respectively. They showed that the maximum likelihood for the assumed change point model is proportional to  $\hat{\sigma}_1^{-\tau} \hat{\sigma}_2^{-(n-\tau)}$  assuming equal variances, and proportional to  $(\tau - p_1 - 1)\hat{\sigma}_1^2 + (n - \tau - p_2 - 1)\hat{\sigma}_2^2$  assuming unequal variances, where  $\hat{\sigma}_d^2$  are the ML estimates of  $\sigma_d^2$ . Thus, in the case of equal variances, maximizing the log likelihood corresponds to minimizing the residual sums of squares. Furthermore, they proposed an iterative method for estimating simultaneously the degrees of the polynomials and the change point.

Tests for general hypotheses, where the variance additionally changes at the change point, were first introduced by Brown, Durbin, and Evans (1975) using recursive residuals. These residuals will be considered in Subsection 2.1.4. A more detailed discussion on testing a change point in OLMs is given in Seber and Wild (1989).

### 2.1.3 Generalized Linear Model

In this section GLMs with one discontinuous change point are considered. A short introduction into GLMs is given in Appendix A. GLMs are a gener-

alization of OLMs. First, the response variable must be no longer normal distributed, but can follow any distribution from the linear exponential family. Second, in GLMs the mean structure is determined by a continuous link function  $g(\cdot)$  and an unknown parameter vector  $\boldsymbol{\beta}$ , namely,

$$g(\mu) = \eta = \mathbf{x}^T \boldsymbol{\beta},$$

where  $\eta$  is the so called linear predictor. Third, it follows that the response variance is the product of a so-called dispersion parameter  $\phi$  and the variance function  $V(\cdot)$ , which is allowed to depend on  $\mu$ , i.e.

$$\text{Var}(y) = \phi V(\mu).$$

In general, a change in the mean structure as well as a change in the variance structure is imaginable. A different mean structure for both segments may be due to different link functions as well as different linear predictors, where the difference of the linear predictors may be due to different sets of explanatory variables or different values of the linear parameter  $\boldsymbol{\beta}$ . A change in the variance structure can be due to different probability models for each segment, which indicates different variance functions  $V(\cdot)$  or different dispersion parameters specific for each segment. However, in this work only a change in the linear parameters is considered. Moreover, the probability model is the same for all segments and we assume that the dispersion parameter is constant for all observations and segments. Thus, in the remainder of this work a common dispersion parameter  $\phi$  is considered. It is important to note that the variance of the observation  $y$  is a function of the mean  $\mu$ . Thus, a change in the mean indicates a change in the variance of  $y$ , as well, even if the dispersion parameter is constant for all observations and the variance function is the same for all segments.

As GLMs are generalizations of OLMs, the model (2.11) with one discontinuous change point is extended to GLMs as

$$g(\mu_i) = \begin{cases} \mathbf{x}_i^T \boldsymbol{\beta}_1 & i = 1, \dots, \tau \\ \mathbf{x}_i^T \boldsymbol{\beta}_2 & i = \tau + 1, \dots, n, \end{cases} \quad (2.17)$$

where both parameter vectors  $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{1p})^T$  and  $\boldsymbol{\beta}_2 = (\beta_{21}, \dots, \beta_{2p})^T$  are of the same dimension ( $p \times 1$ ). Again, the change point  $\tau$  is an index  $i$  and determines the observation  $\mathbf{x}_\tau$  after which the relationship between the response and the explanatory variable changes. Under the assumptions about a unique link function and a unique variance function for both segments, (2.17) can be partitioned into two autonomous GLMs, which can be written as

$$\begin{aligned} g(\boldsymbol{\mu}_1) &= \mathbf{X}_1 \boldsymbol{\beta}_1 \\ g(\boldsymbol{\mu}_2) &= \mathbf{X}_2 \boldsymbol{\beta}_2, \end{aligned} \quad (2.18)$$

where  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$  are both column vectors containing the first  $\tau$  and last  $n - \tau$  values of the mean  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ , and the matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are build up by the first  $\tau$  and the last  $n - \tau$  row vectors of the design matrix, respectively.

To derive the ML estimates of  $\beta_d$ ,  $d = 1, 2$ ,  $\phi$ , and  $\tau$ , again a closer look at the log likelihood is necessary. The log likelihood of GLMs without a change point of a sample  $\mathbf{y} = (y_1, \dots, y_n)^T$  is

$$\ell(\boldsymbol{\theta}, \phi | \mathbf{y}) = \sum_{i=1}^n \left[ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right], \quad (2.19)$$

where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$  is the vector of the canonical parameter of the exponential family. Usually in GLMs,  $\boldsymbol{\beta}$  is the parameter of interest. Thus, it is common to write the log likelihood in terms of  $\boldsymbol{\beta}$ , i.e.  $\ell(\boldsymbol{\beta}, \phi | \mathbf{y})$ .

First, consider the case where  $\tau$  is known. The log likelihood for a GLM with one discontinuous change point  $\tau$  and the parameter of interest  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$  is given by

$$\ell(\boldsymbol{\beta}, \phi | \tau, \mathbf{y}) = \sum_{i=1}^{\tau} \left[ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right] + \sum_{i=\tau+1}^n \left[ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right]. \quad (2.20)$$

Note that  $b'(\theta_i) = \mu_i$ , and (2.17) holds. As the  $y_i$ 's are independent, again, both terms on the right hand side are autonomous (see Subsection 2.1.1). Consequently the ML estimates  $\hat{\beta}_d$  are the ML estimates of the two models of (2.18) corresponding to the first  $\tau$  and last  $n - \tau$  observations, respectively. The dispersion parameter  $\phi$  is estimated by the usual Pearson statistic (A.3) based on all observations.

In the case where  $\tau$  is unknown, the change point has to be estimated. As the estimates of the parameters  $\beta_d$  and  $\phi$  depend on the change point  $\tau$ , the same problem arises as in OLMs. That is, there is no closed form solution of the estimates  $\hat{\tau}$ ,  $\hat{\beta}_d$ , and  $\hat{\phi}$ . Hence, a grid search over all reasonable change points is applied to find the global maximum of the log likelihood

$$\ell(\boldsymbol{\beta}, \phi, \tau | \mathbf{y}) = \ell(\boldsymbol{\beta}_1, \phi | y_1, \dots, y_{\tau}) + \ell(\boldsymbol{\beta}_2, \phi | y_{\tau+1}, \dots, y_n).$$

To guarantee the estimability of the parameters  $\beta_d$  and  $\phi$ , the reasonable values of  $\tau$  are restricted to  $\{p + 1, \dots, n - p - 1\}$ .

A common quantity to evaluate the goodness-of-fit of a GLM is the deviance. As the fitted value  $\hat{\mu}_i$  of a GLM is a function of the explanatory variables and the estimated linear parameter  $\hat{\boldsymbol{\beta}}$ , we reparameterize the log likelihood. Thus, in what follows, we denote the log likelihood in terms of  $\hat{\boldsymbol{\mu}}$  instead of  $\hat{\boldsymbol{\beta}}$ . In a GLM without a change point, the deviance is defined as

$$D = D(\mathbf{y}, \hat{\boldsymbol{\mu}}, \phi) = 2\phi [\ell(\mathbf{y}, \phi | \mathbf{y}) - \ell(\hat{\boldsymbol{\mu}}, \phi | \mathbf{y})],$$

where  $\ell(\mathbf{y}, \phi | \mathbf{y})$  is the log likelihood of the saturated model with  $\hat{\boldsymbol{\mu}} = \mathbf{y}$ . As for a given  $\phi$  and given data set,  $\ell(\mathbf{y}, \phi | \mathbf{y})$  is a constant, maximizing the log likelihood is equivalent to minimizing the deviance. Besides applying the deviance to evaluate the goodness-of-fit of GLMs, it is widely used to compare nested models. This is done by considering the difference between the deviances of

the two models under consideration. In particular, differences between the deviances are used to decide if some additional explanatory variables improve the fit of the model. In general, the difference of the deviance of two nested GLMs equals the LR test statistic. Therefore, under certain regularity conditions, it follows asymptotically a  $\chi^2$ -distribution with  $q$  degrees of freedom, where  $q$  is the difference of the number of parameters of these two models.

As mentioned in Subsection 2.1.1, an OLM without a change point can be considered as nested in a OLM with a change point. This holds for GLMs if the structure of the variance is the same over the entire model and because the design matrix for a GLM with a change point is the same as for an OLM with a change point. Hence, an intuitive and obvious method to compare a GLM with a change point to a GLM without a change point is to analyze the difference between the deviances of these two models. The deviances of the two submodels of (2.18) are

$$\begin{aligned} D(\mathbf{y}_1, \hat{\boldsymbol{\mu}}_1, \phi) &= 2\phi \left[ \sum_{i=1}^{\tau} \ell(y_i, \phi|y_i) - \sum_{i=1}^{\tau} \ell(\hat{\boldsymbol{\mu}}_i, \phi|y_i) \right] \\ D(\mathbf{y}_2, \hat{\boldsymbol{\mu}}_2, \phi) &= 2\phi \left[ \sum_{i=\tau+1}^n \ell(y_i, \phi|y_i) - \sum_{i=\tau+1}^n \ell(\hat{\boldsymbol{\mu}}_i, \phi|y_i) \right], \end{aligned}$$

where  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are column vectors of the first  $\tau$  and last  $n - \tau$  observations, respectively. As the deviance of two autonomous models are additive, the deviance of a GLM with a change point is

$$D^{cp} = D^{cp}(\mathbf{y}, \hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \phi) = D(\mathbf{y}_1, \hat{\boldsymbol{\mu}}_1, \phi) + D(\mathbf{y}_2, \hat{\boldsymbol{\mu}}_2, \phi),$$

where the superscript denotes that this is the deviance corresponding to a GLM with a change point. Then the difference between the deviance  $D$  of a GLM without change point and  $D^{cp}$  is

$$D - D^{cp} = -2\phi [\ell(\hat{\boldsymbol{\mu}}, \phi|\mathbf{y}) - \ell(\hat{\boldsymbol{\mu}}_1, \phi|y_1, \dots, y_{\tau}) - \ell(\hat{\boldsymbol{\mu}}_2, \phi|y_{\tau+1}, \dots, y_n)].$$

By definition, this is minus twice the LR test statistic of a GLM with and without a change point. For normal errors, identity link function and equal variances, this difference is

$$\begin{aligned} D - D^{cp} &= (\mathbf{y} - \hat{\boldsymbol{\mu}})^T (\mathbf{y} - \hat{\boldsymbol{\mu}}) \\ &\quad - [(\mathbf{y}_1 - \hat{\boldsymbol{\mu}}_1)^T (\mathbf{y}_1 - \hat{\boldsymbol{\mu}}_1) + (\mathbf{y}_2 - \hat{\boldsymbol{\mu}}_2)^T (\mathbf{y}_2 - \hat{\boldsymbol{\mu}}_2)] \\ &= \tilde{S}^2 - [\hat{S}_1^2 + \hat{S}_2^2], \end{aligned}$$

which is the difference between the residual sum of squares of the two models. As the same deviations from the regularity conditions mentioned in Subsection 2.1.1 arises, the difference  $D - D^{cp}$  does not follow a  $\chi^2$ -distribution, even in the case of a common dispersion parameter for all observations. Thus, this difference can only be used as an approximative test. Further work in this area has to be done.

### 2.1.4 Recursive Residuals

In general, residuals can be used to detect departures from a specified model. Brown et al. (1975) first introduced recursive residuals to detect the departure from an OLM after a change point. In this section, we discuss recursive residuals in OLMs and present an extension suitable for GLMs.

In general, raw residuals are defined as the difference between the observed values  $y_i$  and the fitted values  $\hat{\mu}_i$  (for example see Davison & Snell, 1991). Here it is important to note, that the fitted values  $\hat{\mu}_i$  are usually based on estimates from the entire data. This is in contrast different for recursive residuals where the fitted model is only based on part of the data. In particular, the recursive residual of the  $r$ th observation is defined as the difference between  $y_r$  and the fitted mean  $\hat{\mu}_{r-}$ , where  $\hat{\mu}_{r-}$  is only based on the first  $r - 1$  observations. Note, that for this definition an ordering of the data is mandatory.

First, consider an OLM where some raw recursive residuals are defined as

$$u_r = y_r - \hat{\mu}_{r-} = y_r - \mathbf{x}_r^T \hat{\boldsymbol{\beta}}_{r-}, \quad r = p + 1, \dots, n.$$

Here  $\hat{\boldsymbol{\beta}}_{r-}$  is the ML estimate of  $\boldsymbol{\beta}$  based on only the first  $r - 1$  observations and  $p$  is the number of parameters in the linear predictor. The expectation  $E[u_r]$  of the recursive residual is

$$\begin{aligned} E[u_r] &= E \left[ y_r - \mathbf{x}_r^T \hat{\boldsymbol{\beta}}_{r-} \right] \\ &= E[y_r] - \mathbf{x}_r^T E \left[ \hat{\boldsymbol{\beta}}_{r-} \right] \\ &= \mu_r - \mathbf{x}_r^T \boldsymbol{\beta} \\ &= 0, \end{aligned}$$

as  $\hat{\boldsymbol{\beta}}_{r-}$  is of course also an unbiased estimate of  $\boldsymbol{\beta}$ . In order to achieve a suitable standardization of the recursive residuals, the variance of  $u_r$  has to be calculated. The ML estimate  $\hat{\mu}_{r-}$  is based on the first  $r - 1$  observations and therefore independent from the remaining observations, in particular, from  $y_r$ . Consequently, the two components in  $u_r$  are independent and the variance of  $u_r$  is the sum of the variances of  $y_r$  and  $\hat{\mu}_{r-}$ . The variance of  $y_r$  is  $\sigma^2$  by assumption and the variance of  $\hat{\mu}_{r-}$  is

$$\text{Var}(\hat{\mu}_{r-}) = \text{Var}(\mathbf{x}_r^T \hat{\boldsymbol{\beta}}_{r-}) = \mathbf{x}_r^T \text{Var}(\hat{\boldsymbol{\beta}}_{r-}) \mathbf{x}_r = \sigma^2 \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{x}_r, \quad (2.21)$$

where  $\mathbf{X}_{r-1}$  is the design matrix containing the first  $r - 1$  row vectors of  $\mathbf{X}$ . Thus,

$$\text{Var}(u_r) = \sigma^2 (1 + \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{x}_r)$$

and standardized recursive residuals are defined as

$$u_r^* = \frac{y_r - \mathbf{x}_r^T \hat{\boldsymbol{\beta}}_{r-}}{\sqrt{1 + \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{x}_r}}.$$

Both,  $y_r$  and  $\mathbf{x}_r^T \hat{\boldsymbol{\beta}}_{r-}$  are normally distributed, hence  $u_r^* \sim N(0, \sigma^2)$ . Moreover, Brown et al. (1975) provided that under the assumption that there is no change in the regime in a homoscedastic OLM, these recursive residuals are independent, hence  $u_r^* \stackrel{iid}{\sim} N(0, \sigma^2)$ . As the  $u_r$  are normally distributed it is only necessary to show that the covariance  $Cov[u_r, u_s]$ ,  $r \neq s$ , of two arbitrary recursive residuals is zero in order to obtain this result. That is, we are interested in

$$Cov[u_r, u_s] = E[u_r u_s] - E[u_r]E[u_s]. \quad (2.22)$$

As  $E[u_r] = 0$ , the second term on the right hand side of (2.22) is zero. Next, we consider  $E[u_r u_s]$ . Without loss of generality, it is assumed that  $r < s$ . To calculate  $E[u_r u_s]$ , the recursive residuals are expressed in terms of  $\varepsilon_i$ . Therefore,  $\hat{\boldsymbol{\beta}}_{r-}$  is rewritten as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{r-} &= (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T \mathbf{y}_{r-1} \\ &= (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T (\mathbf{X}_{r-1} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{r-1}) \\ &= \boldsymbol{\beta} + (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T \boldsymbol{\varepsilon}_{r-1}, \end{aligned}$$

where  $\mathbf{y}_{r-1}$  and  $\boldsymbol{\varepsilon}_{r-1}$  are column vectors of the first  $r - 1$  components of the response variable and the error vector, respectively. The recursive residual  $u_r$  can then be expressed as

$$\begin{aligned} u_r &= y_r - \mathbf{x}_r^T \hat{\boldsymbol{\beta}}_{r-} \\ &= y_r - \mathbf{x}_r^T \boldsymbol{\beta} - \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T \boldsymbol{\varepsilon}_{r-1} \\ &= \varepsilon_r - \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \sum_{i=1}^{r-1} \mathbf{x}_i \varepsilon_i \end{aligned}$$

and

$$\begin{aligned} E[u_r u_s] &= E[\varepsilon_r \varepsilon_s] \\ &- E \left[ \varepsilon_s \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \sum_{i=1}^{r-1} \mathbf{x}_i \varepsilon_i \right] \\ &- E \left[ \varepsilon_r \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \sum_{i=1}^{s-1} \mathbf{x}_i \varepsilon_i \right] \\ &+ E \left[ \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \sum_{i=1}^{r-1} \mathbf{x}_i \varepsilon_i \times \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \sum_{i=1}^{s-1} \mathbf{x}_i \varepsilon_i \right]. \end{aligned} \quad (2.23)$$

As the  $\varepsilon_i$ 's are independent and  $E[\varepsilon_i] = 0$ , the first term above is zero. Second,  $r < s$  indicates that  $\varepsilon_s$  is independent of  $\varepsilon_1, \dots, \varepsilon_{r-1}$ , thus all components of the second term in (2.23) are also independent and, hence, this term is zero as well. Next, consider the third term in (2.23). By defining a set of indices



$A = \{1, \dots, r-1, r+1, \dots, s-1\}$ , then

$$E \left[ \varepsilon_r \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \sum_{i=1}^{s-1} \mathbf{x}_i \varepsilon_i \right] = E \left[ \varepsilon_r \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_r \varepsilon_r \right] + E \left[ \varepsilon_r \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \sum_{i \in A} \mathbf{x}_i \varepsilon_i \right].$$

The components of the second term of the right hand side are independent and hence the expectation of this term is zero. Thus, the third term in (2.23) is the product of three scalars and can be represented as

$$E \left[ \varepsilon_r^2 \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_r \right] = \sigma^2 \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_r.$$

The fourth term in (2.23) can be split up in the same way, i.e.

$$\begin{aligned} E \left[ \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \sum_{i=1}^{r-1} \mathbf{x}_i \varepsilon_i \times \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \sum_{i=1}^{s-1} \mathbf{x}_i \varepsilon_i \right] = \\ E \left[ \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \sum_{i=1}^{r-1} \mathbf{x}_i \varepsilon_i \times \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \sum_{i=1}^{r-1} \mathbf{x}_i \varepsilon_i \right] \\ + E \left[ \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \sum_{i=1}^{r-1} \mathbf{x}_i \varepsilon_i \times \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \sum_{i=r}^{s-1} \mathbf{x}_i \varepsilon_i \right], \end{aligned} \quad (2.24)$$

where the components in the second term on the right hand side in (2.24) are independent and thus this second term is zero. The first term on the right hand side in (2.24) can be written as

$$E \left[ \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T \boldsymbol{\varepsilon}_{r-1} \times \mathbf{x}_s^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{X}_{r-1}^T \boldsymbol{\varepsilon}_{r-1} \right]. \quad (2.25)$$

As the two terms in this expectation are scalars, the second term can be transposed and (2.25) becomes

$$E \left[ \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T \boldsymbol{\varepsilon}_{r-1} \times \boldsymbol{\varepsilon}_{r-1}^T \mathbf{X}_{r-1} (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_s \right]. \quad (2.26)$$

In this equation, the terms  $\mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T$  and  $\mathbf{X}_{r-1} (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_s$  are constant, thus (2.26) can be written as

$$\mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T E \left[ \boldsymbol{\varepsilon}_{r-1} \boldsymbol{\varepsilon}_{r-1}^T \right] \mathbf{X}_{r-1} (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_s, \quad (2.27)$$

consequently, we are interested in  $E \left[ \boldsymbol{\varepsilon}_{r-1} \boldsymbol{\varepsilon}_{r-1}^T \right]$ . This product,  $\boldsymbol{\varepsilon}_{r-1} \boldsymbol{\varepsilon}_{r-1}^T$ , is a matrix of dimension  $(r-1) \times (r-1)$  with diagonal elements  $\varepsilon_i^2$  and off diagonal elements  $\varepsilon_i \varepsilon_j$ ,  $i \neq j$ . Taking the expectation of this matrix results in

$$E \left[ \boldsymbol{\varepsilon}_{r-1} \boldsymbol{\varepsilon}_{r-1}^T \right] = \sigma^2 \mathbf{I}_{r-1},$$

since  $E[\varepsilon_i^2] = \sigma^2$  and  $E[\varepsilon_i \varepsilon_j] = 0$ , for all  $i \neq j$ . Consequently, (2.27) becomes

$$\mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{X}_{r-1}^T \sigma^2 \mathbf{I}_{r-1} \mathbf{X}_{r-1} (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_s$$

and simplifies to

$$\sigma^2 \mathbf{x}_r^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_s.$$

Applying these results to equation (2.23) yields the desired result

$$E[u_r u_s] = 0 - 0 - \sigma^2 \mathbf{x}_r^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_s + \sigma^2 \mathbf{x}_r^T (\mathbf{X}_{s-1}^T \mathbf{X}_{s-1})^{-1} \mathbf{x}_s = 0,$$

i.e.  $Cov[u_r, u_s] = 0$  and the recursive residuals are uncorrelated.

To determine the location of the change point and to test if the change point is necessary or not, Brown et al. (1975) did not use the pure standardized recursive residuals, but instead used the quantities

$$U_r = \frac{1}{\sigma} \sum_{j=p+1}^r u_j^*, \quad r = p+1, \dots, n. \quad (2.28)$$

If  $\sigma$  is unknown, it can be estimated by standard deviation based on all data, i.e.  $\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) / (n - p)$ .

As long as  $\boldsymbol{\beta}$  is constant, i.e. one OLM is appropriate for the entire data, the mean of  $U_r$  is zero. From that point onwards, where the relationship between response and explanatory variable changes,  $U_r$  will have non-zero mean. Thus, the sequence  $U_{p+1}, \dots, U_n$  drifts away from zero. Therefore, this sequence can be used to test the necessity of a change point. The considered null hypothesis,  $H_0$ , is that there is no change in the regime. Under this  $H_0$ ,  $U_{p+1}, \dots, U_n$  is a sequence of normal variates with

$$E[U_r] = 0, \quad Var[U_r] = r - p$$

and with  $r < s$

$$Cov[U_r, U_s] = r - p.$$

To test for change, we need to define boundaries such that the probability that the path of  $U_{p+1}, \dots, U_n$  crosses these boundaries is  $\alpha$ . As for an arbitrary  $p+1 \leq t \leq n$ , the variance of  $U_t$  is  $t - p$ , these boundaries should be curves of the form  $\pm \lambda \sqrt{t - p}$ , where  $\lambda$  is a constant. For simplicity, however, we follow the approach of Brown et al. (1975) and consider straight lines. These straight lines should be tangents to these curves at the point  $t^* = p + (n - p)/2$ , which lies half a way between  $p$  and  $n$ . We assume that the probability that the path crosses both boundaries is negligible. Therefore, only the positive curve  $\lambda \sqrt{t - p}$  is considered and  $\alpha/2$  is taken as crossing probability instead of  $\alpha$ . For such a line we consider the function

$$\xi(t) = a + bt. \quad (2.29)$$

First, we calculate the slope  $b$  of this function, which is the derivative of the function  $f(t) = \lambda \sqrt{t - p}$  at the point  $t^*$ . The derivative is

$$f'(t) = \frac{\lambda}{2} \frac{1}{\sqrt{t - p}},$$

which, at the point of interest, is

$$\begin{aligned}
 b = f'(t^*) &= \frac{\lambda}{2} \frac{1}{\sqrt{\frac{n-p}{2} + p - p}} \\
 &= \frac{\lambda}{2} \frac{\sqrt{\frac{n-p}{2}}}{\frac{n-p}{2}} \\
 &= \frac{\lambda}{n-p} \sqrt{\frac{n-p}{2}} \\
 &= \frac{\lambda\sqrt{2}}{2} \frac{\sqrt{n-p}}{n-p}.
 \end{aligned} \tag{2.30}$$

Next, the intercept  $a$  has to be calculated. For that reason consider  $f(t^*)$ , which is

$$f(t^*) = \frac{\lambda\sqrt{2}}{2} \sqrt{n-p}. \tag{2.31}$$

Substituting (2.30) and (2.31) in (2.29) yields

$$\begin{aligned}
 \frac{\lambda\sqrt{2}}{2} \sqrt{n-p} &= a + \frac{\lambda\sqrt{2}}{2} \frac{\sqrt{n-p}}{n-p} \left[ p + \frac{n-p}{2} \right] \\
 \frac{\lambda\sqrt{2}}{2} \sqrt{n-p} &= a + \frac{\lambda\sqrt{2}}{2} \sqrt{n-p} \left[ \frac{p}{n-p} + \frac{n-p}{2(n-p)} \right] \\
 \frac{\lambda\sqrt{2}}{2} \sqrt{n-p} &= a + \frac{\lambda\sqrt{2}}{2} \sqrt{n-p} \left[ \frac{n+p}{2(n-p)} \right] \\
 a &= \frac{\lambda\sqrt{2}}{2} \sqrt{n-p} \left[ 1 - \frac{n+p}{2(n-p)} \right] \\
 a &= \frac{\lambda\sqrt{2}}{2} \sqrt{n-p} \left[ \frac{n-3p}{2(n-p)} \right].
 \end{aligned}$$

The function  $\xi(t)$  then becomes

$$\xi(t) = \frac{\lambda\sqrt{2}}{4} \frac{\sqrt{n-p}}{n-p} (n - 3p + 2t), \quad p \leq t \leq n. \tag{2.32}$$

While this function is not very well aesthetically appealing, it can be described by two well defined points. Therefore, consider (2.32) at the points  $p$  and  $n$ . Substituting  $d = \frac{\lambda\sqrt{2}}{4}$  these two points are  $(p, d\sqrt{n-p})$  and  $(n, 3d\sqrt{n-p})$ . To obtain a general straight line for any arbitrary  $n$  and  $p < n$ , we transform  $t$  to the interval  $[0; 1]$ . In this way, any index  $p \leq i \leq n$  can be given by  $i = \lfloor t(n-p) + p \rfloor$  for  $0 \leq t \leq 1$ . Thus, these points become  $(0, d\sqrt{n-p})$  and  $(1, 3d\sqrt{n-p})$ . From these two points, it is obvious that by substituting  $\nu = d\sqrt{n-p}$ , function (2.32) can be expressed as

$$\xi(t) = \nu + 2\nu t \quad \text{but with} \quad 0 \leq t \leq 1.$$

To calculate  $\nu$  for a given crossing probability, we rewrite (2.28) as

$$U_n(t) = \frac{1}{\sigma} \sum_{j=p+1}^{\lfloor p+t(n-p) \rfloor} u_j^*, \quad 0 \leq t \leq 1.$$

Krämer, Plomberger, and Alt (1988) showed that for  $n \rightarrow \infty$

$$U_n(t) \rightarrow B(t)$$

in distribution, where  $B(t)$  is the standard Brownian motion. In addition, Durbin (1971) examined that the probability that such a Brownian motion crosses a line  $a + bt$  once is

$$P(B(t) > a + bt \text{ for some } 0 \leq t \leq 1) = 1 - \Phi(a + b) + \exp(-2ab)\Phi(b - a),$$

where  $\Phi(\cdot)$  is the standard normal distribution function. Thus, the asymptotic probability for  $n \rightarrow \infty$  that  $U_n(t)$  crosses the line  $\xi(t)$  is

$$P(U_n(t) > \nu + 2\nu t \text{ for some } 0 \leq t \leq 1) = 1 - \Phi(3\nu) + \exp(-4\nu^2)\Phi(\nu). \quad (2.33)$$

Solving equation (2.33) for  $\alpha/2$  yields  $\nu = (1.143, 0.948, 0.85)$  for common  $\alpha = (0.01, 0.05, 0.1)$ . Note, here it is assumed that the probability that  $U_n(t)$  crosses the lines more than once is zero. However, Zeileis (2004) showed, that for small  $\alpha$  (2.33) is a rather good approximation even for the case of multiple crossings.

It is also important to note that this method tests a general departure from the assumed model after the change point and is not restricted to a change only in the parameter  $\beta$ . Furthermore, the observation after which  $U_n(t)$  fall outside these lines provides an estimate of the change point. The properties of this estimate are investigated in an MC simulation at the end of this section.

Considering a GLM, the raw recursive residuals are

$$u_r = y_r - \hat{\mu}_{r-} = y_r - g^{-1}(\hat{\eta}_{r-}), \quad r = p + 1, \dots, n,$$

where  $\hat{\eta}_{r-} = \mathbf{x}_r^T \hat{\beta}_{r-}$ . Again,  $\hat{\beta}_{r-}$  is the ML estimate of  $\beta$  based on the first  $r-1$  observations. For a suitable standardization of the raw recursive residuals, the variance of  $u_r$  has to be calculated. As  $y_r$  and  $\hat{\mu}_{r-}$  are independent,  $Var(u_r)$  is the sum of  $Var(y_r) = \phi V(\mu_r)$  and  $Var(\hat{\mu}_{r-})$ . Therefore, we are interested in a method for calculating  $Var(\hat{\mu}_{r-})$ . Recall, that standard results about inference in GLMs (see e.g. Fahrmeir & Kaufmann, 1985) provide

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(\mathbf{0}, n\phi(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1})$$

in distribution, where  $\mathbf{W}$  is a diagonal matrix with

$$1/w_i = V(\mu_i)[g'(\mu_i)]^2.$$

In addition,  $\hat{\mu}_{r-}$  is a function of  $\hat{\beta}_{r-}$ . These two results make it obvious to apply the Delta Method to approximate  $Var(\hat{\mu}_{r-})$ . The Delta Method is appropriate to calculate an approximative variance of some function of a random variable if the variance of the random variable itself is known, or at least approximately known. Two things have to be satisfied to apply the Delta Method. First, a sequence of random variables  $Y_n$  satisfies  $\sqrt{n}(Y_n - \theta) \rightarrow N(0, \sigma^2)$  in distribution and, second, there exists a function  $h(\cdot)$  with  $h'(\theta) \neq 0$ . Then

$$\sqrt{n}(h(Y_n) - h(\theta)) \rightarrow N(0, \sigma^2 h'(\theta)^2)$$

in distribution (see Casella & Berger, 2002, for a detailed discussion). Here,  $Y_n = \hat{\eta}_{r-}$ ,  $\theta = \eta_r = \mathbf{x}_r^T \boldsymbol{\beta}$ , and the function  $h(\cdot) = g^{-1}(\cdot)$  is the inverse link function. However, the Delta Method provides only an asymptotic variance for  $n \rightarrow \infty$ . In contrast,  $\hat{\mu}_{r-}$  is based on a finite sample. Moreover, in most cases it is only based on a few observations. Thus, from a theoretical point of view, the Delta Method is not appropriate to calculate the variance of  $\hat{\mu}_{r-}$ . However, practical results showed that it provides at least a useful approximation of  $Var(\hat{\mu}_{r-})$ .

First, consider the derivative of  $h(\cdot)$  at an arbitrary  $\eta_r$ , which is

$$h'(\eta_r) = \frac{d\mu_r}{d\eta_r} = \frac{d\mu_r}{dg(\mu_r)} = \frac{1}{g'(\mu_r)}.$$

Then, applying the Delta Method to calculate  $Var(\hat{\mu}_{r-})$  yields

$$Var(\hat{\mu}_{r-}) = \frac{1}{g'(\mu_r)^2} Var(\hat{\eta}_{r-}). \quad (2.34)$$

Recall, that the asymptotic variance of  $\hat{\eta}_{r-}$  is

$$Var(\hat{\eta}_{r-}) = \phi \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{W}_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r,$$

where  $\mathbf{W}_{r-1}$  is the matrix of the first  $r-1$  rows of  $\mathbf{W}$ . Then, the variance of  $\hat{\mu}_{r-}$  is

$$Var(\hat{\mu}_{r-}) = \frac{1}{g'(\mu_r)^2} \phi \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{W}_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r. \quad (2.35)$$

As mentioned above,  $Var(u_r) = Var(y_r) + Var(\hat{\mu}_{r-})$ . Thus, using the Delta Method to calculate the variance of  $\hat{\mu}_{r-}$  yields

$$Var(u_r) = \phi \left[ V(\mu_r) + \frac{1}{g'(\mu_r)^2} \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{W}_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r \right]. \quad (2.36)$$

Using (2.36) to standardize the raw recursive residuals  $u_r$  leads to the standardized recursive residuals for GLMs, which are defined as

$$u_r^* = \frac{y_r - \hat{\mu}_{r-}}{\sqrt{\phi \left[ V(\mu_r) + \frac{1}{g'(\mu_r)^2} \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{W}_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r \right]}}. \quad (2.37)$$

For a canonical link function  $g'(\cdot) = 1/V(\cdot)$  and (2.35) simplifies to

$$\text{Var}(\hat{\mu}_{r-}) = \phi V(\mu_r)^2 \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{W}_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r$$

and (2.37) becomes

$$u_r^* = \frac{y_r - \hat{\mu}_{r-}}{\sqrt{\phi V(\mu_r) [1 + V(\mu_r) \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{W}_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r]}}.$$

In the case of normal distributed  $y_i$  and identity link function we have  $1/w_i = 1$ ,  $\phi = \sigma^2$ , and  $g'(\mu) = 1$ . Hence (2.35) becomes

$$\text{Var}(\hat{\mu}_{r-}) = \sigma^2 \mathbf{x}_r^T (\mathbf{X}_{r-1}^T \mathbf{X}_{r-1})^{-1} \mathbf{x}_r$$

which equals the exact result for  $\text{Var}(\hat{\mu}_{r-})$  in (2.21). Thus, for normal distributed responses, the variance approximated by the Delta Method equals the exact variance.

As  $\hat{\beta}_{r-}$  and therefore  $\hat{\mu}_{r-}$  are asymptotically unbiased estimators for  $\beta$  and  $\mu_r$  respectively, the asymptotical expectation of  $u_r^*$  is zero. Thus, the standardized recursive residuals are at least asymptotically  $(0, 1)$  variates.

Deletion residuals can also be taken as recursive residuals. For this reason we consider the first  $r$  observations. Then, we fit a model based on these data, but leave the  $r$ th observation out. The deletion residual of the  $r$ th observation can then be interpreted as recursive residual. Pregibon (1979) first introduced a one step approximation for the deletion residuals  $r_{Gi}$ . This approximation is given by

$$r_{Gi} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{(1 - h_i) r_{Di}^2 + h_i r_{Pi}^2}, \quad (2.38)$$

where  $r_{Pi}$  are the standardized Pearson residuals,  $r_{Di}$  are the standardized deviance residuals, and  $h_i$  is the  $i$ th diagonal element of the generalized Hat-Matrix  $\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$  (see Appendix A). In general, the deletion residuals are closer to the standardized deviance residuals than to the standardized Pearson residuals. This is due to the fact that, for most observations the quantity  $h_i$  is close to  $p/n$ , which is the average of the diagonal elements of  $\mathbf{H}$ . This gives, for most observations, more weight to the  $r_{Di}$  than to the  $r_{Pi}$ . Only for points which are extreme in the domain of  $x$ , the  $h_i$ 's are markedly greater than  $p/n$  and therefore weights the  $r_{Pi}$  more than the  $r_{Di}$ . However, here, the deletion residuals of interest corresponds to an explanatory variable at an extreme in the domain of  $x$ . Therefore, the deletion residuals are close to the standardized Pearson residuals and inherit their properties. In the case of non-normal responses this means that the deletion residuals inherit the skewness of the standardized Pearson residuals.

Now, two different types of standardized recursive residuals are described, i.e.  $u_r^*$  and  $r_{Gi}$ . Both types are at least asymptotically  $(0, 1)$  variates but do not necessarily follow a normal distribution. However, we try to establish, whether the method proposed by Brown et al. (1975) can be either used to determine the location of the change point or as a test statistic, and to establish whether the change point is necessary or not. These two questions are investigated in the following MC simulation.

### Simulation Study

In this MC simulation, two different GLMs are considered both with one change point. The response variable in both models follows a Poisson distribution, hence the dispersion parameter is known and equals 1 for all segments. These two models differ only in the magnitude of change at the change point.

Consider first a GLM and a data set of sample size  $n = 40$  and a given discontinuous change point at  $\tau = 20$ . The explanatory variable  $x$  is given by  $x_i = i/40$ ,  $i = 1, \dots, 40$ , and the supposed model is defined as

$$g(\mu_i) = \begin{cases} 5.0 - 2.0x_i & i = 1, \dots, 20 \\ 3.8 - 1.5x_i & i = 21, \dots, 40, \end{cases}$$

where  $g(\mu) = \log \mu$  is the canonical link function. Based on this model 10000 samples of Poisson distributed response vectors  $\mathbf{y} = (y_1, \dots, y_{40})^T$  are generated. For each generated sample the standardized recursive residuals  $u_r^*$  (2.37) and the standardized deletion residuals  $r_{Gi}$  (2.38) are calculated. Of course, using the first few observations for fitting the model can lead to an insufficient fit and hence to huge residuals. Therefore, we consider the cumulative sum of these residuals after the fifth observation. Thus, for both types of residuals, the quantities

$$U_r^* = \sum_{j=3}^r u_j^* \quad \text{and} \quad U_{Gr} = \sum_{j=3}^r r_{Gj} \quad r = 6, \dots, n \quad (2.39)$$

are calculated and compared to the bounds described above with  $\alpha = 5\%$ . The main interest of this simulation study is, whether the paths of the sequences (2.39) cross these bounds and if so, after which index the crossing can be observed. This index determines the estimate  $\hat{\tau}$  of the change point.

The second model considered differs from the first model only in the intercept of the second segment and is defined as

$$g(\mu_i) = \begin{cases} 5.0 - 2.0x_i & i = 1, \dots, 20 \\ 4.2 - 1.5x_i & i = 21, \dots, 40, \end{cases}$$

Again, 10000 Poisson distributed response vectors are generated.

The MC results of the first model show that the change point is detected with a lag of a few observations (Figure 2.2). The mode of the simulated distribution of the estimated change point  $\tau$  is 22 for both types of residuals. Thus the most frequently observed lag with respect to the true change point is two. The relative frequency of this mode is 28.25% for using  $r_{Gi}$  and 25.2% for using  $u_r^*$ . Furthermore, using the  $r_{Gi}$  leads to slightly better results than the use of  $u_r^*$ . This can be seen in the higher bars for  $\hat{\tau} = 21, 22$  in Figure 2.2 and in the faster increase of the cumulative relative frequencies of the change point estimate in Figure 2.3. Besides, in 99.6% of the replications using  $r_{Gi}$  and 98.9% of the replications using  $u_r^*$  the quantities (2.39) fall outside the

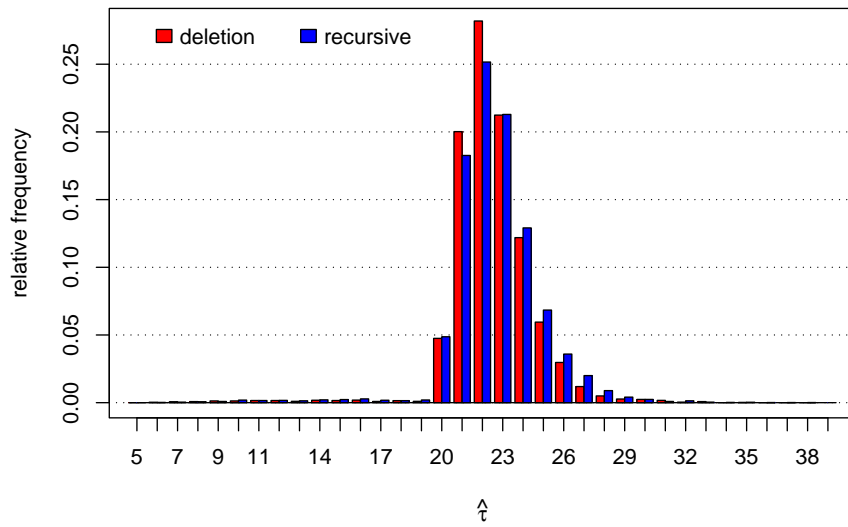


Figure 2.2: Relative frequencies of the change point estimate for the first model.

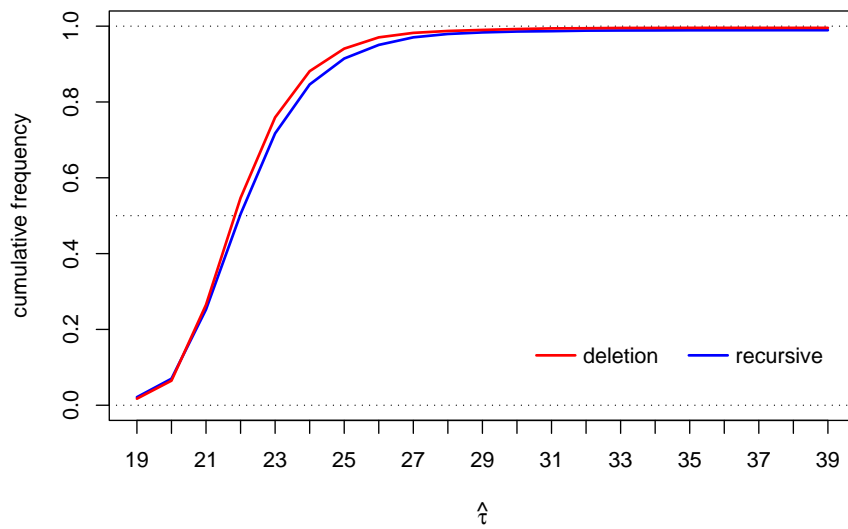


Figure 2.3: Cumulative relative frequencies of the change point estimate for the first model.



bounds (Figure 2.3). Thus, in almost all replications, the necessity of a change point is found.

The MC simulation under the second model leads to similar results. Again, for both types of residuals, the change point is detected after a lag of a few observations. Here, the mode of the simulated distribution of the estimated change points is 23. Note, that the relative frequencies at the mode are markedly lower than in the first model. They are only 10.4% for using  $r_{Gi}$  and 9.0% for using  $u_r^*$  (Figure 2.4). Furthermore, the variability of the estimated change point is greater. Again, the standardized deletion residuals provide better results. The lower magnitude of change means that only 64.7% of the change points are detected using  $r_{Gi}$  and 57.4% using  $u_r^*$  (Figure 2.5).

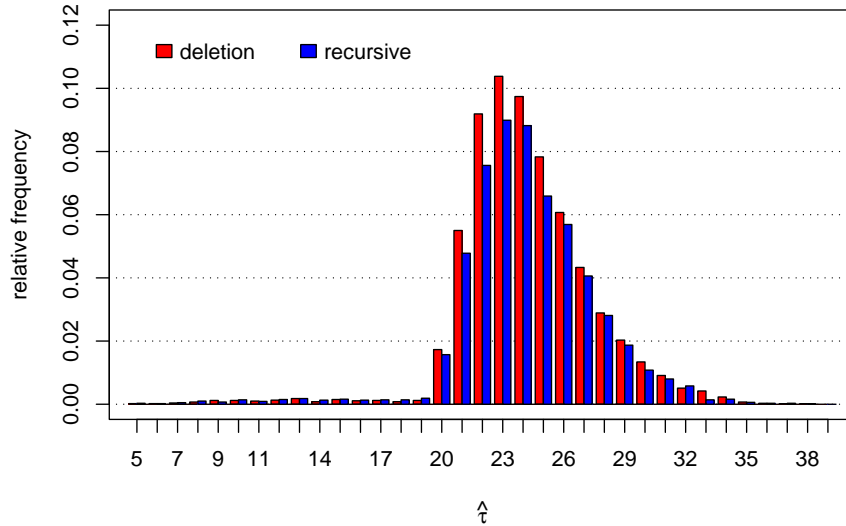


Figure 2.4: Relative frequencies of the change points estimate for the second model.

These MC simulations give a first indication that the standardized deletion residuals should probably be preferred to the standardized recursive residuals for detecting the change point using the method of Brown et al. (1975).

Furthermore, the location of the change point is detected with a lag of a few observations. This lag depends on the deviation between the two models partitioned by the change point. In addition, the frequency of detecting a change point, again, depends strongly on this deviation. Therefore, in most cases, this method detects the change point with a lag of a few observations. Moreover, it is a conservative test for testing the significance of the change point.

One question that arises is whether such a lag is due to the fact that the considered recursive residuals are asymptotic  $(0, 1)$  variates even though do

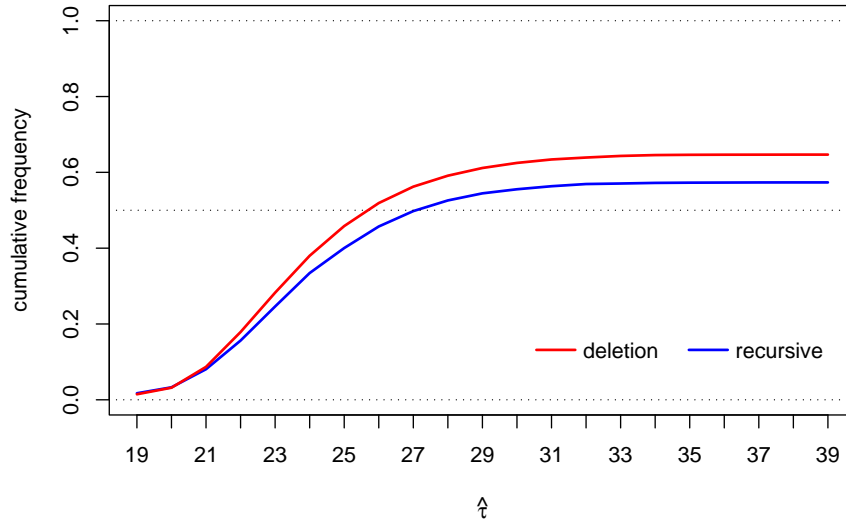


Figure 2.5: Cumulative relative frequencies of the change point estimate for the second model.

not necessarily follow a normal distribution. Hence, the considered bounds are not appropriate to detect a change point as well as to test the necessity of the change point. To investigate this, a second MC simulation based on an OLM with one change point is performed. To enable a comparison of the results, similar assumptions for the OLM with one change point are necessary. That is, the magnitude of the change at the change point, as well as the variance of the observations at the change point, should be similar. Therefore, the considered model is

$$y_i = \begin{cases} 9 + 22x_i + \varepsilon_i & i = 1, \dots, 20 \\ -45 + 180x_i + \varepsilon_i & i = 21, \dots, 40 \end{cases}$$

with  $\varepsilon_i \stackrel{iid}{\sim} N(0, 36)$ . 10000 normal distributed response vectors are generated. For each of the 10000 replications the quantities (2.28) are calculated. We are particularly interested in investigating, whether the path of these sequences crosses the considered bounds and if so, after which index this crossing can be observed. Using this index as an estimate for the change point, this MC simulation again yields a lag of the estimated change point between zero and five observations (Figure 2.6). The mode of the frequency of the estimated change point is 21 with a relative frequency of about 31%. Thus the most observed lag is one observation. This is quite smaller than in the case of GLMs, but a lag can be still observed. Thus even in the case of iid normal distributed recursive residuals, the method introduced by Brown et al. (1975) yields an estimated change point with a lag of up to five observations. Consequently, regardless of whether this method is applied to OLMs or GLMs, it provides an

estimate of the change point with a lag of up to a few observations. Therefore, in the case of a marked change at the change point this method can be applied on GLMs as well.

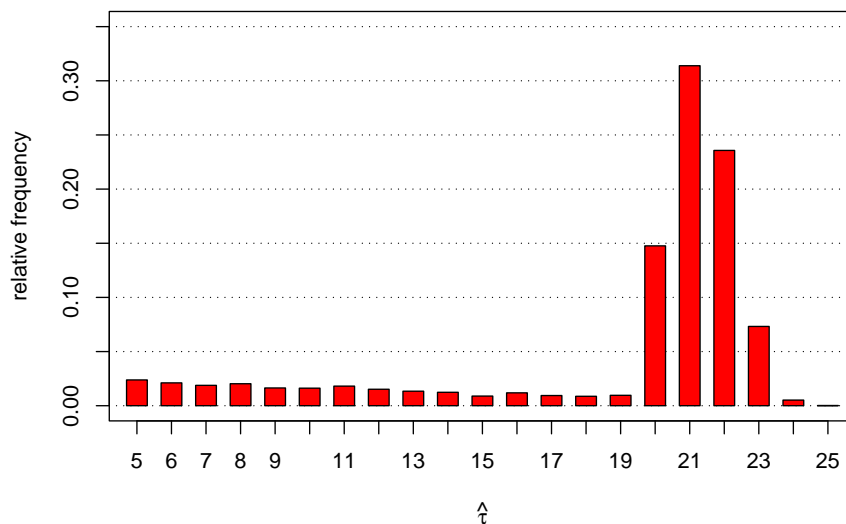


Figure 2.6: Relative frequencies of the change point estimate for the OLM.

## 2.2 Discontinuous Change Points in GLMs

In this section, the change point model previously considered is extended to a linear model with multiple discontinuous change points. Therefore, it is assumed that there are at least two changes in the regime of the linear model. OLMs are a special case of GLMs, hence, in the remainder of this thesis, only change point models for GLMs are considered. As in the case of linear models with one change point, multiple change points partition the data into several segments. It is assumed that the number of change points is known. Moreover, it is assumed that for all segments, the same type of model holds but with different parameters. The challenge of change point models with more than one discontinuous change point is to find an appropriate method for estimating the change points. Two different methods are discussed, the hierarchic binary splitting algorithm (Hawkins, 2001) and a method based on a dynamic programming algorithm (Hawkins, 2001). Furthermore, these two methods are compared using an MC simulation.

### 2.2.1 Multiple Discontinuous Change Points

In this section, GLMs with multiple discontinuous change points are considered where the change points are fixed but unknown. In particular,  $d - 1$  change points partition the data into  $d$  segments. Again only a change in the linear parameter  $\boldsymbol{\beta}_k = (\beta_{k1}, \dots, \beta_{kp})^T$ ,  $k = 1, \dots, d$ , is considered, whereas all other quantities defining a GLM, i.e.  $g(\cdot)$ ,  $V(\cdot)$ , and  $\phi$ , are the same for all segments.

In the case of multiple discontinuous change points, the vector of change points is denoted as  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{d-1})^T$  where  $\tau_k$  is any index. For simplicity, these change points are augmented with  $\tau_0 = 0$  and  $\tau_d = n$  which are artificial change points. To guarantee the estimability of  $\boldsymbol{\beta}_k$  these change points must fulfill  $\tau_k - \tau_{k-1} > p$  for all  $k = 1, \dots, d$ . A GLM with  $d - 1$  discontinuous change points can then be given by

$$g(\mu_i) = \begin{cases} \mathbf{x}_i^T \boldsymbol{\beta}_1 & \tau_0 < i \leq \tau_1 \\ \mathbf{x}_i^T \boldsymbol{\beta}_2 & \tau_1 < i \leq \tau_2 \\ \vdots & \vdots \\ \mathbf{x}_i^T \boldsymbol{\beta}_d & \tau_{d-1} < i \leq \tau_d, \end{cases} \quad (2.40)$$

where  $\boldsymbol{\beta}_k$  are the parameter vectors of interest. As there are no continuity constraints at the change points, (2.40) can be separated in  $d$  autonomous GLMs, which can be written in matrix representation as

$$\begin{aligned} g(\boldsymbol{\mu}_1) &= \mathbf{X}_1 \boldsymbol{\beta}_1 \\ g(\boldsymbol{\mu}_2) &= \mathbf{X}_2 \boldsymbol{\beta}_2 \\ &\vdots \\ g(\boldsymbol{\mu}_d) &= \mathbf{X}_d \boldsymbol{\beta}_d, \end{aligned} \quad (2.41)$$

where  $\boldsymbol{\mu}_k = (\mu_{\tau_{k-1}+1}, \dots, \mu_{\tau_k})^T$  and  $\mathbf{X}_k$  contains the corresponding row vectors of the design matrix.

First, consider the case where all change points are known. Then, the log likelihood under model (2.40) can be separated into  $d$  terms, namely

$$\ell(\boldsymbol{\beta}, \phi | \boldsymbol{\tau}, \mathbf{y}) = \sum_{k=0}^{d-1} \ell(\boldsymbol{\beta}_{k+1}, \phi | y_{\tau_k+1}, \dots, y_{\tau_{k+1}}), \quad (2.42)$$

with  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_d^T)^T$  and  $l = k + 1$ . The ML estimates of  $\boldsymbol{\beta}$  and  $\phi$  are the ML estimates of the GLMs of each segment and can therefore be calculated with the common IWLS method and the Pearson statistic, respectively. If the change points are unknown, the ML estimate  $\hat{\boldsymbol{\tau}}$  is obtained by maximizing

$$\ell(\boldsymbol{\beta}, \phi, \boldsymbol{\tau} | \mathbf{y}) = \sum_{k=0}^{d-1} \ell(\boldsymbol{\beta}_{k+1}, \phi | y_{\tau_k+1}, \dots, y_{\tau_{k+1}}). \quad (2.43)$$

Again, the ML estimates of  $\boldsymbol{\beta}_k$  and  $\phi$  depend on  $\boldsymbol{\tau}$ , hence these parameters can not be estimated simultaneously. A grid search over all possible combinations of change points seems appropriate to solve this problem. However,

as  $\tau$  contains  $d - 1$  unknown change points, a grid search over all possible combinations of change points is computationally burdensome. The order of a standard grid search is  $O(n^d)$  (Bai & Perron, 2003). Thus, the computational complexity increases rapidly with the number of change points and effective algorithms are necessary.

## 2.2.2 Fitting Algorithms

In this section, two algorithms for finding multiple change points are presented:

- the hierarchic binary splitting (HS) algorithm
- and a dynamic programming (DP) algorithm.

Furthermore, an MC simulation is performed to compare these two algorithms.

### Hierarchic Binary Splitting Algorithm

The HS algorithm to find multiple change points in a linear model is a Greedy type algorithm. It splits the data in a hierarchic way like in regression trees (Breiman, Friedman, Olshen, & Stone, 1984). The first step is to split the data into two segments. These two segments are partitioned by that change point which maximizes the sum of the log likelihood of both segments. The second step is to find a potential change point in each of these two segments. From these new potential change points, choose that one which increases the sum of the log likelihood of all three segments the most. Thus, one step of the HS algorithm consists of two tasks,

- find a potential change point in each segment,
- given these potential change points in each segment, choose that one which increases the log likelihood the most.

This procedure is repeated until the final number of change points is achieved. This final number can be limited for practical reasons or determined by using any information criterion such as Akaike's information criterion (AIC) or Bayesian information criterion (BIC) (e.g. see Burnham & Anderson, 1998).

The advantage of this algorithm is its simplicity and speed. At each step, the maximum number of possible change points is  $n$ . Thus, the computational order of this algorithm is  $O(n(d - 1))$ , as for each of these  $d - 1$  steps, at most  $n$  different potential change points have to be investigated. In addition, it is very easy to implement it in any commonly used programming language. Furthermore, the algorithm also performs well in problems with large data sets. The disadvantage is, however, that a determined change point is fixed for all subsequent steps. This can lead to a suboptimal solution for models with more than one change point. The first change point maximizes the log likelihood for a model with one change point, but for a model with multiple

change points this primarily found change point may no longer be an optimal choice.

So it is desirable to be able to find different change points for models with different numbers of change points. One method to achieve this goal is the DP algorithm (Hawkins, 2001).

### Dynamic Programming Algorithm

The DP algorithm allows the determination of different change points if the number of change points increases. It is based on Bellman's principle of optimality (Bellman & Dreyfus, 1962) and is built up recursively. Consider a GLM and a data set of size  $n$  with  $d - 1$  change points ( $d$  segments). Then, knowing the rightmost change point, the problem can be reduced to split the remaining data on the left hand side of this change point into  $d - 1$  segments. This is repeated until all  $d$  segments are determined. A prerequisite for this algorithm is that, for known change points, the models in the segments must be autonomous, i.e. at least the parameters  $\beta_k$  are unrestricted for each segment (see Subsection 2.2.1).

We consider a GLM with  $d - 1$  inner change points where  $\beta_k$ ,  $k = 1, \dots, d - 1$ , changes at each change point. Again, the extended set of change points is denoted by  $\tau = (\tau_0, \dots, \tau_d)^T$ . Then, the log likelihood for known change points is given by (2.42). For an arbitrary pair of consecutive change points  $(\tau_k, \tau_l)$ , with  $l = k + 1$ , let  $Q(\tau_k, \tau_l)$  be the maximized log likelihood for the data  $y_{\tau_k+1}, \dots, y_{\tau_l}$ , namely

$$Q(\tau_k, \tau_l) = \ell \left( \hat{\beta}_{k+1}, \hat{\phi} | y_{\tau_k+1}, \dots, y_{\tau_l} \right),$$

with  $\tau_l - \tau_k > p$  to guarantee the estimable of the parameters  $\beta_{k+1}$  and  $\phi$ . Then, the overall maximized log likelihood can be written as

$$\begin{aligned} \ell(\hat{\beta}, \hat{\phi}, \hat{\tau} | \mathbf{y}) &= \max_{\tau} \left[ \max_{\beta, \phi} \ell(\beta, \phi | \tau, \mathbf{y}) \right] \\ &= \max_{\tau} \sum_{k=0}^{d-1} Q(\tau_k, \tau_l). \end{aligned} \quad (2.44)$$

Thus, finding the maximum of (2.44) can be split into two problems. The first problem is to find the ML estimate  $\hat{\tau}$ . The second problem is to find the ML estimate  $\hat{\beta}$  for given change points  $\hat{\tau}$ . Hawkins (2001) denoted these as the *outer* and *inner* problem respectively. We start with searching the ML estimate  $\hat{\tau}$ , the *outer* problem.

The main idea of solving this *outer* problem is to build up a GLM with  $d$  segments recursively based on the maximized log likelihood of different sub-models. That is, first calculate the maximized log likelihoods of GLMs based on all feasible combinations of observations for the first segment. In particular, these are the first  $m$  observations, where the possible range of  $m$  is specified

later on. Second, based on these GLMs, calculate all feasible GLMs with two segments, again based on the first  $m$  observations. This is repeated until a GLM with  $d$  segments is determined. The cleverness of the DP algorithm is to do this in a very efficient way. Finally, if a GLM with  $d$  segments is determined, the estimates  $\hat{\tau}_k$  can then be obtained by back-tracing the DP algorithm.

To describe this algorithm in detail, let us define  $F(k, m)$  as the maximized log likelihood of a GLM with  $k$  segments for the first  $m$  observations. Moreover, let  $p_k = k(p + 1)$  be the smallest possible endpoint of the  $k$ th segment, i.e. the smallest possible value for the  $k$ th change point, where  $p$  is the number of the unknown linear parameters. Furthermore, let  $n_k = n - (d - k)(p + 1)$  be the largest possible endpoint of the  $k$ th segment. Thus, the largest possible range for the  $k$ th segment is  $p_k - p, \dots, p_k, \dots, n_k$ . This is due to the fact, that at least  $k(p + 1)$  observations are mandatory for the first  $k$  segments and for the remaining  $d - k$  segments on the right hand side of the  $k$ th segment,  $(d - k)(p + 1)$  observations are mandatory.

As the maximized log likelihoods for all possible submodels are the basis for this algorithm, first calculate  $Q(r, s)$  for all possible combinations of  $0 \leq r < s - p \leq n$ . Then we consider  $Q(0, m)$  for  $m = p_1, \dots, n_1$ , which is the maximized log likelihood for a GLM without a change point based on these first  $m$  observations, i.e.  $Q(0, m) = F(1, m)$ .

To illustrate this algorithm with the help of an example, consider a data set of size  $n = 12$ . Furthermore, we are interested to fit a GLM with  $d = 3$  segments to this data, where the number of linear parameters is  $p = 2$ . Then, the first segment can be based on the first  $m = 3, \dots, 6$  observations, as for the remaining two segments at least six observations are mandatory, i.e.  $y_7, \dots, y_{12}$ . These four different cases for the first segment are plotted in Figure 2.7. The length of the lines are drawn with respect to the corresponding observations. They are colored to distinguish them in the following steps. The red vertical line determines the largest possible observation for the first segment.

In a next step, we are interested in the maximized log likelihood of a GLM with two segments for the first  $m = p_2, \dots, n_2$  observations, i.e.  $F(2, m)$ . For an arbitrary  $m$  this is

$$\begin{aligned} F(2, m) &= \max_{p_1 \leq \tau_1 < m - p} [Q(0, \tau_1) + Q(\tau_1, m)] \\ &= \max_{p_1 \leq \tau_1 < m - p} [F(1, \tau_1) + Q(\tau_1, m)]. \end{aligned} \quad (2.45)$$

For back-tracing the DP algorithm, it is necessary to store the change point, that yields the maximized log likelihood at each step. Let us denote  $H(k, m)$  as that change point that yields the maximum of  $F(k - 1, \tau_{k-1}) + Q(\tau_{k-1}, m)$ . Then, for (2.45),  $H(2, m)$  is that change point that maximizes the log likelihood of a GLM with one change point based on the first  $m$  observations.

In Figure 2.8, all possible combinations to obtain  $F(2, m)$ ,  $m = 6, \dots, 9$ , are illustrated. For  $m = 6$ , the only possible combination to obtain the maximum  $F(2, 6)$  is a change point at  $\tau_1 = 3$ , which is denoted as  $H(2, 6)$ . In the case of  $m = 7$ , there are two possible combinations, namely  $\tau_1 = 3$  and  $\tau_1 = 4$ . In

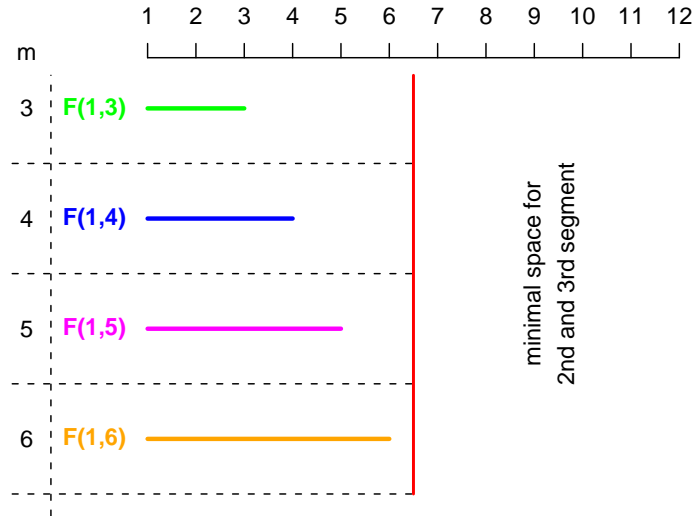


Figure 2.7: Possible models for the 1st segment.

this example, it is assumed that  $\tau_1 = 4$  yields the maximum, i.e.  $H(2, 7) = 4$ . Thus,  $F(2, 7) = F(1, 4) + Q(4, 7)$  and the corresponding lines in Figure 2.8 are colored blue and red respectively. In general, a red line on the right hand side of the potential change point shows that this combination yields the maximum. For  $m = 8$  and  $m = 9$ , three and four combinations have to be investigated respectively. The maximum is obtained by a change point at  $H(2, 8) = 4$  and  $H(2, 9) = 5$  respectively.

After  $F(2, m)$  is obtained for all possible  $m$ 's, we are interested in  $F(3, m)$  for  $m = p_3, \dots, n_3$  and so on. This is repeated until  $F(d, m)$  is determined. Thus, this algorithm can be expressed as following recursion

$$\begin{aligned}
 F(1, m) &= Q(0, m) \\
 F(k + 1, m) &= \max_{p_k \leq \tau_k < m-p} [F(k, \tau_k) + Q(\tau_k, m)] , \quad (2.46)
 \end{aligned}$$

with  $k = 1, \dots, d - 1$  and  $m = p_k, \dots, n_k$ . Hence, the  $(k + 1)$ th step of the algorithm consists of finding the change point  $\tau_k$  in the remaining  $p_k, \dots, m - p - 1$  observations, where, for the first  $p_k - p$  observations,  $k - 1$  change points are already determined. This can be repeated until the largest achievable number of change points is determined which is, at most,  $d_{max} = \lfloor n/p \rfloor$ . The big advantage of this recursion is, that at the  $(k + 1)$ th step, the maximization in (2.46) is autonomous of the previous steps. Thus, for the maximization of  $F(k, \tau_k) + Q(\tau_k, m)$  it is irrelevant which  $(\tau_1, \dots, \tau_{k-1})^T$  yields the maximum of  $F(k, \tau_k)$ . Hence, at the  $(k + 1)$ th step, only  $n_k - p_k$  possible values for  $\tau_k$  have to be investigated. Therefore, once all possible values of  $Q(r, s)$  are calculated,  $F(k, \tau_k)$  are easily obtained. Finally, the maximized log likelihood of the full



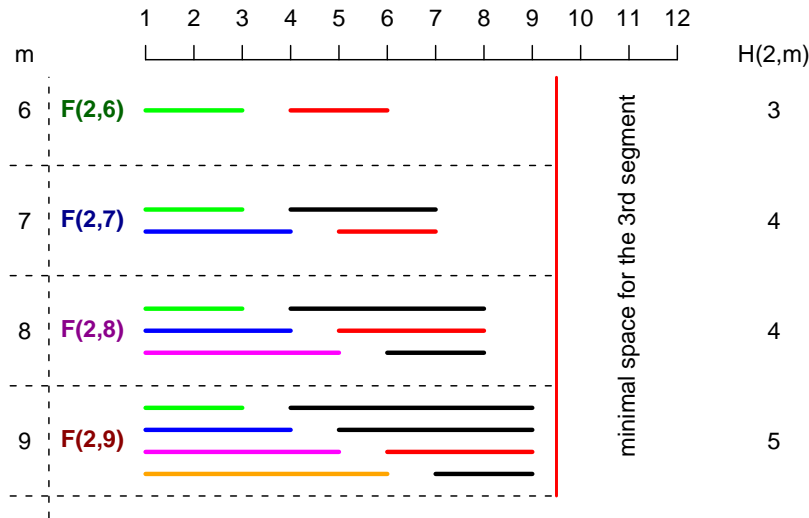


Figure 2.8: Possible combinations for the first two segments.

data with  $d - 1$  change points is  $F(d, n)$ .

In the considered example, this is  $F(3, 12)$ . Figure 2.9 shows all possible combinations that can lead to  $F(3, 12)$ . Moreover, for completeness, the cases for  $m = 9, 10, 11$  are also illustrated in this figure even though they are not necessary for the remaining analysis. Here it is assumed that  $F(2, 8) + Q(8, 12)$  yields the maximum. Thus  $H(3, 12) = 8$  and, again, the corresponding line on the right hand side of this change point in Figure 2.9 is red.

Given  $F(d, n)$  the estimate of the change points can be obtained by backtracing the algorithm, i.e. solving the *inner* problem. Note that the rightmost change point for all observations is the artificial change point  $\tau_d = n$  and is, of course, always known. Knowing this rightmost change point, the second rightmost change point  $\tau_{d-1}$  is given by  $H(d, n)$ . This is the change point which maximizes  $F(d - 1, \tau_{d-1}) + Q(\tau_{d-1}, n)$ . Consequently, the remaining change points are given by  $\hat{\tau}_k = H(k + 1, \hat{\tau}_{k+1})$ . Thus, following this backtracing, the maximized log likelihood  $F(d, n)$  can be split in its  $d$  autonomous components, namely

$$F(d, n) = \sum_{k=0}^{d-1} Q(\hat{\tau}_k, \hat{\tau}_{k+1}).$$

Finally, the ML estimate of  $\beta_k$  is the parameter which maximizes  $Q(\hat{\tau}_k, \hat{\tau}_{k+1})$ .

To solve the *inner* problem in the example, we start with  $H(d, n)$ , i.e.  $H(3, 12) = 8$ . Thus,  $\hat{\tau}_2 = 8$  is the estimate of the second rightmost change point. Knowing this change point, the remaining eight observations have to be partitioned into two segments. The maximized log likelihood for a GLM with one change point (two segments) for this remaining data is  $F(2, 8)$ . This

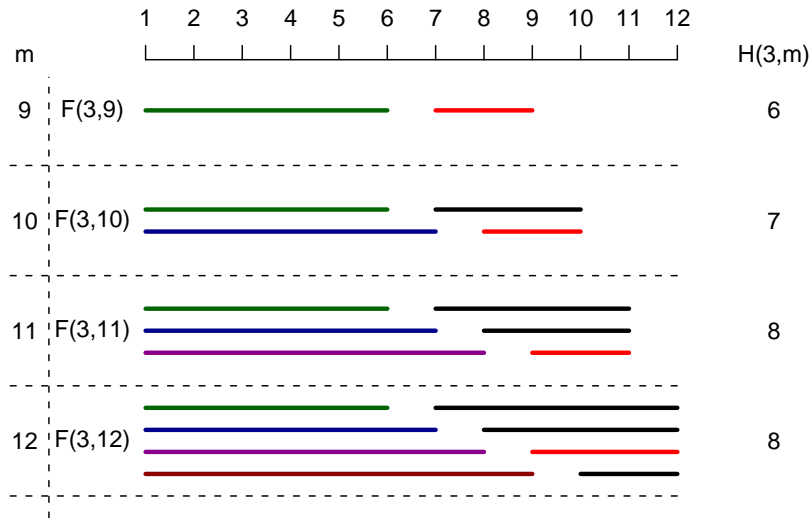


Figure 2.9: Possible combinations for a GLM with three segments for the first  $m = 9, \dots, 12$  observations, in particular, for the entire data set of size  $n = 12$ .

is represented by the dark magenta line in Figure 2.9. The corresponding case with  $m = 8$  in Figure 2.8 is also dark magenta (i.e.  $F(2, 8)$ ). The change point that yields this maximum is  $H(2, 8) = 4$ . Thus, the estimated change points are  $\hat{\tau} = (4, 8)^T$ .

Due to computational efficiency reasons we restrict the calculation  $F(k, m)$  on  $m = p_k, \dots, n_k$ . Hence, we implicitly assume that the number of change points is known. If this is not the case, one can calculate  $F(k, m)$  for  $m = p_k, \dots, n$ , with  $k = 1 \dots, d_{max}$ . Then, the  $d_{max}$  maximized log likelihoods,  $F(k, n)$ , can be used to decide how many change points are appropriate. To take the number of estimated parameters into account, this is usually done by comparing some information criterion of models with different numbers of change points.

The big advantage of the DP algorithm is its flexibility in determining the change points. The change points of a model with  $r$  change points are not necessarily a subset of the change points of a model with  $s > r$  change points. For instance, a model with three change points may have completely different change points than a model with two change points. This is in contrast to the HS algorithm where the change points of a model with  $r$  change points is always a subset of the change points of a model with  $s > r$  change points. The cost is the greater computational complexity but this becomes only a problem when analyzing a huge amount of data. Therefore, in most cases the DP algorithm should be the preferred algorithm for detecting multiple change points. This is confirmed by the following MC simulation.

### Simulation Study

In this MC simulation, both algorithms are compared with respect to their bias finding the true change points. The focus lays on the frequency of finding the true change points under different variance assumptions. Therefore, three different settings are considered where the mean structure of the model is the same for all settings but the variance varies between the settings. For all three settings, the results are based on 2000 replications.

Consider a simple linear regression with three discontinuous change points. The explanatory variable  $x$  is given by  $x_i = i/10$ , with  $i = 1, \dots, 60$ . The change points are located at the indices  $\tau_1 = 15$ ,  $\tau_2 = 30$  and  $\tau_3 = 45$ . Then, the linear model

$$y_i = \begin{cases} x_i + \varepsilon_i & 0 < i \leq 15 \\ 5 + x_i + \varepsilon_i & 15 < i \leq 30 \\ 18 - 0.8x_i + \varepsilon_i & 30 < i \leq 45 \\ 1 + 2x_i + \varepsilon_i & 45 < i \leq 60 \end{cases} \quad (2.47)$$

is used to generate artificial data sets. The variances of the errors,  $\varepsilon_i$ , vary over the settings and are  $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.5^2)$ ,  $\varepsilon_i \stackrel{iid}{\sim} N(0, 1)$ , and  $\varepsilon_i \stackrel{iid}{\sim} N(0, 1.5^2)$  for the first, second and third setting, respectively. The MC simulation is done in R applying the function `dcp` of the package `CpInGLM`. For details, see Appendix B.

Of course, the DP algorithm is expected to perform better than the HS algorithm as it is the more flexible algorithm. The questions arising are:

- how often do the two algorithms find the true change points and
- does the DP algorithm clearly perform better than the HS algorithm.

From Table 2.2, it can be seen that the frequency depends on the variance of the errors. For a small variance  $\sigma^2 = 0.5^2$ , the DP algorithm finds the true change points in 99.9% of the replications, whereas the HS algorithm in only 88.7% of the replications. The ratio between these percentages is 1.12 and is listed in the last column of Table 2.2.

setting	$\sigma$	DP	HS	ratio
1	0.5	99.9%	88.7%	1.12
2	1	73.7%	35.1%	2.10
3	1.5	26.4%	9.2%	2.65

Table 2.2: Percentage of finding the true change points for both algorithms in the three settings based on 2000 replications.

If  $\sigma^2 = 1$ , the DP algorithm performs well, finding the true change points in 73.7% of the replications, which is more than twice the frequency of the HS algorithm, as shown by the ratio of 2.1. If the variance of the errors increases to  $\sigma^2 = 1.5^2$  both algorithms perform poorly.

To measure the deviation of the estimated change points from the true ones, we define the following quantity

$$D_\tau = \sum_{k=1}^3 |\hat{\tau}_k - \tau_k|.$$

This quantity,  $D_\tau$ , is plotted for the second setting in Figure 2.10 and for the third setting in Figure 2.11. Both figures show a better performance of the DP algorithm. Furthermore, a second peak of the relative frequencies of  $D_\tau$  around 8 can be observed. This can be interpreted as evidence that, if one change point is not found at the true location, the others are not estimated at the true location as well.

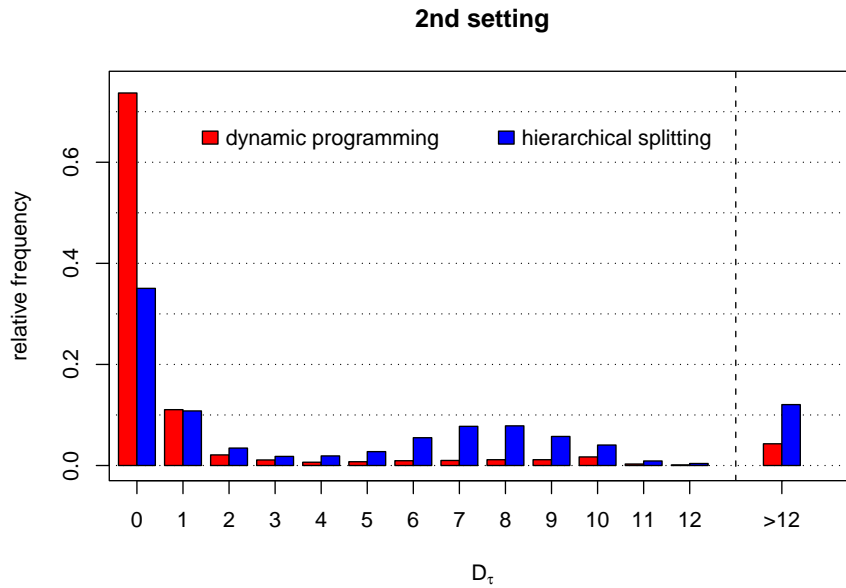


Figure 2.10: Relative frequency of the cumulative difference of the change points for  $\varepsilon \stackrel{iid}{\sim} N(0, 1)$ .

To resume, this MC simulation shows that a DP algorithm should be applied to detect multiple change points. In particular, if the variance of the errors are not a negligible quantity, the HS solution is outperformed by the DP result.

**Example 2.2 Simple linear model with three change points**

*In this example, a generated data set (Figure 2.12) of the model (2.47) considered in the simulation study above, is analyzed. In particular, a data set from the second setting with  $\sigma^2 = 1$  is considered. Both algorithms to detect the change points are used to compare their results.*

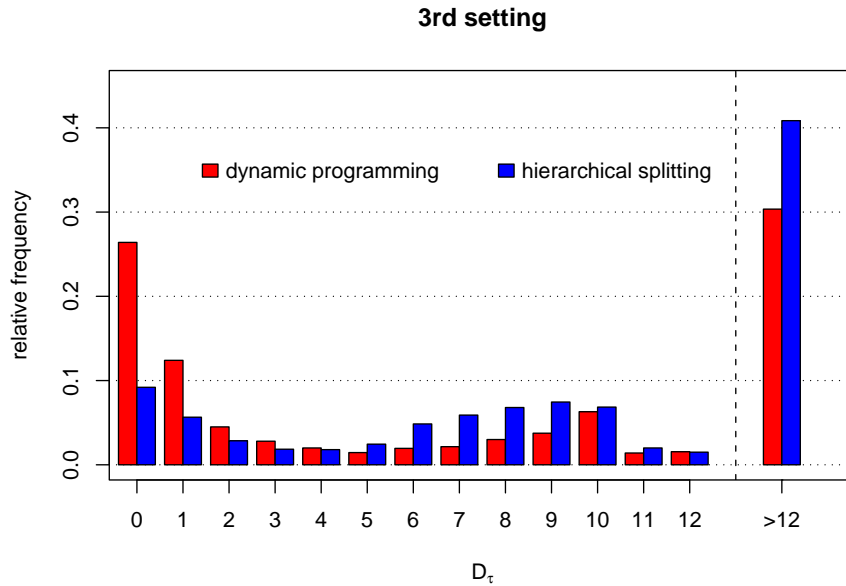


Figure 2.11: Relative frequency of the cumulative difference of the change points for  $\varepsilon \stackrel{iid}{\sim} N(0, 1.5)$ .

*In the first step, the HS algorithm finds a change point at 37. This change point maximizes the sum of the log likelihood of both segments. The second change point is found in the first segment at 30 and the third change point is found at 15. Thus, the estimated change points are  $\hat{\tau} = (15, 30, 37)^T$  (Figure 2.12), which are different to the true change points,  $\tau = (15, 30, 45)^T$ . This is due to the fact that once a change point is determined, it is fixed for the remaining part of the algorithm.*

*To find discontinuous change points in a GLM, the function `dcp` was implemented in the programming language R and is used in this example. First, a GLM is fitted with the usual function `glm`. The output of this function is an object of class `glm`. Whereas several functions in R can handle an object either of class `glm` or of class `lm`, this function are designed to work only with objects of class `glm`. This object is the basis of the `dcp` function call. Two additional arguments of the function call of `dcp` are `nOfCp` and `dps`. These arguments determine the number of change points and the minimum number of observations in each segment respectively. The fitting algorithm used, is determined by the argument `method` and is set to `'hs'` for the HS algorithm. Then, the R-code for fitting a GLM with three change points and its results are given as*

```
> data(HSvsDP)
> m1 <- glm(y~x, data=HSvsDP)
> m1.hs <- dcp(m1, nOfCp=3, dps=5, method="hs")
> m1.hs
```

Call:

```
dcp(object = m1, nOfCp = 3, dps = 5, method = "hs")
```

Coefficients:

	(Intercept)	x
segment1	0.2815	1.1461
segment2	3.7080	1.2672
segment3	20.6459	-1.1850
segment4	-3.9691	2.5309

Degrees of Freedom: 59 Total (i.e. Null); 49 Residual

Null Deviance: 2417

Residual Deviance: 58.61

AIC : 186.2

BIC : 209.3

Indices of Discontinuous Change Points : 15 30 37

```
> m1.hs$logLikelihood
```

```
[1] -82.10936
```

The output of the function is an object of class `dcp` with several elements. The generic function `print` can be used to obtain a short overview of the fitted model. This overview includes the estimates of the linear parameter and the change points and some information about the goodness-of-fit. As the estimate of the third change point is not the true change point, the parameter estimates of the third and fourth segment differ more from the true values than that one from the first and second segment. The null deviance is related to a GLM without a change point and one intercept. Therefore, the corresponding degrees of freedom are 59. The residual deviance corresponds to the fitted change point model with four intercepts, four slopes, and three change points. Hence the number of estimated parameters is eleven and the degrees of freedom reduces to 49. An additional element of this class is the maximized log likelihood, which is  $-82.1$  for this model. Using the definitions of the information criteria described in Subsection 2.2.3 results in an AIC of 186.2 and a BIC of 209.3.

Next, the DP algorithm is applied. For a model with one change point, the change point is the same as determined by the HS algorithm, namely 37. The change points for a model with two change points are  $(30, 45)^T$ , which are different from the change points determined by the HS algorithm. Accordingly, the change point 37 for the model with one change point, is not a subset of the change points  $(30, 45)^T$  of the model with two change points. Finally, the change points of the model with three change points are  $\hat{\tau} = (15, 30, 45)^T$  (Figure 2.13), which match the true ones exactly.

The discontinuous change points can be again estimated with the function `dcp`, but with `method='dp'`. Again, the GLM without a change point, i.e. model `m1`, is the basis for the `dcp` function call. Then the R-code and its results are as follows:

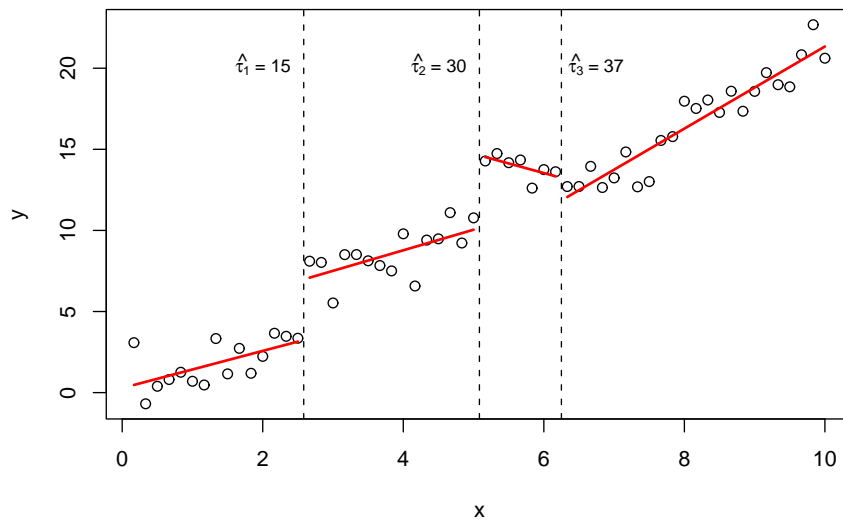


Figure 2.12: Fitted linear model with three change points  $\hat{\tau} = (15, 30, 37)^T$  using the HS algorithm.

```
> m1.dp <- dcp(m1, nOfCp=3, dps=5, method="dp")
> m1.dp
```

Call:

```
dcp(object = m1, nOfCp = 3, dps = 5, method = "dp")
```

Coefficients:

	(Intercept)	x
segment1	0.2815	1.1461
segment2	3.7080	1.2672
segment3	16.6363	-0.4869
segment4	-1.1808	2.2342

Degrees of Freedom: 59 Total (i.e. Null); 49 Residual

Null Deviance: 2417

Residual Deviance: 52.6

AIC : 180.7

BIC : 203.7

Indices of Discontinuous Change Points : 15 30 45

```
> m1.dp$logLikelihood
[1] -79.32898
```

*As the estimate of the first and the second change point are the same using*

the HS algorithm, the estimates of the linear parameter of the first and second segments are the same, as well. In contrast, the estimates of the linear parameter of the third and fourth segment differ and are closer to the true values. Furthermore, the maximized log likelihood is  $-79.3$ , which is greater than that one found by the HS algorithm. Moreover, both information criteria are smaller than that one of the HS algorithm, which also indicates, that the DP algorithm yields the better fit.

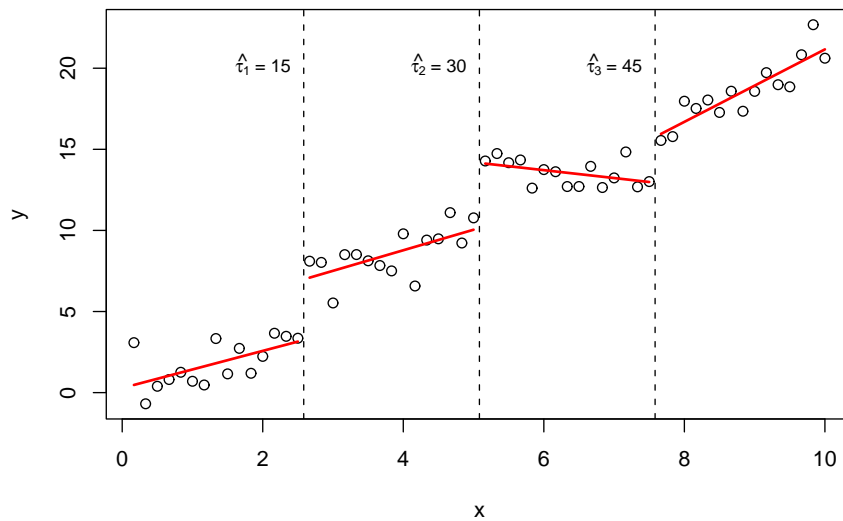


Figure 2.13: Fitted linear model with three change points  $\hat{\tau} = (15, 30, 45)^T$  using the DP algorithm.

### 2.2.3 Determining the Number of Change Points

Of course one important question in the case of multiple change points is still to be answered. *“How many change points are necessary to describe the data”*. As for the data considered in this work, where the number of change points is fixed to two, only a brief overview on this topic is given.

Recall the results for testing the necessity of a change point in an OLM. Even in this very simple case no exact method exists. All methods considered are based on approximations of quantiles of the distribution of the test statistic. Besides, Bai and Perron (2003) introduced methods for testing no change versus a fixed number of change points and for testing  $k$  change points versus  $k+1$  change points. Both tests are considered for an OLM with several change points. These tests can be used for sequential testing whether an additional change point is necessary or not. Again, the Bonferroni inequality has to be applied to correct for the true significance level. Thus, even for these simple



circumstances no exact methods are available. So, one can expect that for estimating the number of change points in a GLM, more technical difficulties will arise. Therefore, in both cases (OLMs and GLMs), the use of an information criterion is a good alternative. In general, there are two well known information criteria used for model selection, i.e. Akaike's information criterion (AIC) and Bayesian information criterion (BIC) (see Burnham & Anderson, 1998, for a detailed discussion), which are defined as

$$\begin{aligned} AIC &= -2\log(L) + 2p \\ BIC &= -2\log(L) + p\log(n), \end{aligned}$$

where  $p$  is the number of estimated parameters, i.e. the linear parameters and the change points,  $n$  is the sample size, and  $\log(L)$  is the maximized log likelihood. Yao (1988) and Chen and Gupta (1997), among others, suggested to use the BIC for estimating the number of change points. This is confirmed by the suggestion of Bai and Perron (2003). They found that the AIC performs very badly and tends to overfit the model. Chen and Gupta (2000) recommended the use of the BIC in several parametric change point models, in particular for OLMs. Thus, it also seems appropriate if the number of change points in a GLM has to be estimated. Nevertheless, Hawkins (2001) recommended to additionally use a graphical method to figure out an appropriate number of change points. He suggested to plot the maximized log likelihoods of consecutive models, i.e. the maximized log likelihoods of a GLMs with  $k = 1, \dots, d$  change points, and looking for an "elbow" in the plot. Though these methods provide good approximate results, further work has to be done in this area, especially for estimating the number of change points in GLMs.

**Example 2.3 (Continuation of Example 2.2)** *Recall the artificial data set of Example 2.2, where an OLM with three change points is considered. The DP algorithm was used to estimate change point models with zero to eight change points. The AIC, as well as the BIC, of these nine models are plotted in Figure 2.14. Using the BIC,  $k = 3$  change points is the best solution, which is also the true number of change points. In contrast, using the AIC would result in  $k = 6$  change points, which is twice the true number. Thus, this again confirms the recommendation that the BIC should be used instead of the AIC. Next, we consider the plot of the maximized log likelihood with respect to the number of change points (Figure 2.15). This plot does not establish a clearly formed "elbow". While using this kind of plot it is not easy to decide how many change points are appropriate.*

## 2.3 Continuous Change Points in GLMs

In this section, GLMs with a continuous change at each change point are considered. Consequently, the predicted means of consecutive models are the same at the change point. Thus, the parameter estimates of  $\beta$  in each model are

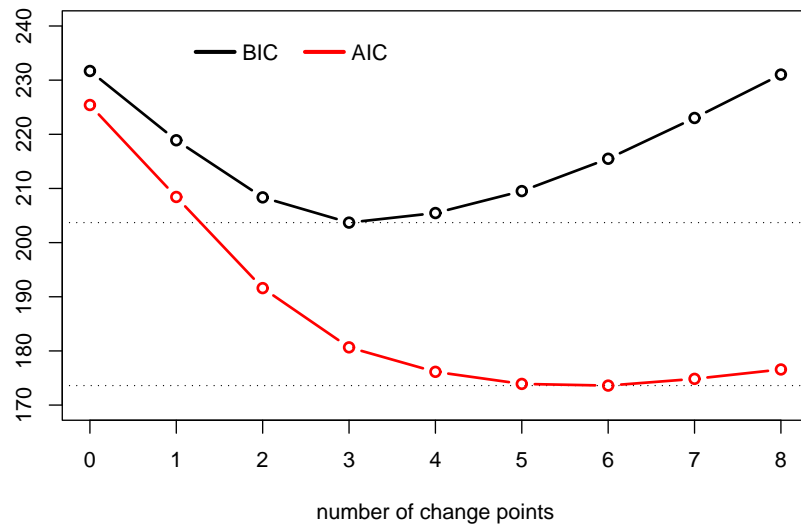


Figure 2.14: Comparison of the two information criteria AIC and BIC.

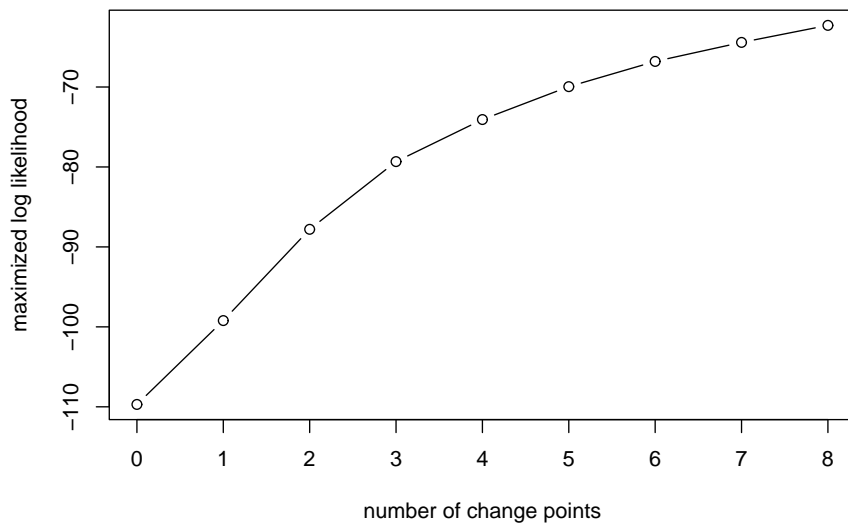


Figure 2.15: Maximized log likelihood with respect to the number of change points.

restricted to this continuity constraint and the estimates of the parameters in each segment are no longer autonomous. Therefore, the algorithms introduced in Subsection 2.2.2 for fitting multiple change points can not be applied and other methods have to be considered.

Note that because of the continuity constraint, the exact location of the change point can be estimated. Moreover, the change point is no longer restricted to any observed value of the explanatory variable but can be any value inside the interval  $[\min x_i; \max x_i]$  and is not an index as in Section 2.1 and 2.2. Therefore, the continuous change point is denoted as  $\gamma$  to distinguish it from a discontinuous change point.

Subsection 2.3.1 presents GLMs with one continuous change point and an MC simulation shows, that recursive residuals can be applied to get a bound for a continuous change point in GLMs. Furthermore, two algorithms to estimate the change point are presented, one is based on an ordinary grid search and the other one is an exact method. In Subsection 2.3.2, GLMs with multiple change points are considered and both algorithms are adapted for these models.

### 2.3.1 One Continuous Change Point

In this section, GLMs with response variable  $y_i$ , a single explanatory variable  $x_i$ ,  $i = 1, \dots, n$ , and one continuous change point  $\gamma$  are considered. Then, the model can be written as

$$g(\mu_i) = \begin{cases} \beta_{10} + \beta_{11}x_i & a \leq x_i \leq \gamma \\ \beta_{20} + \beta_{21}x_i & \gamma < x_i \leq b, \end{cases} \quad (2.48)$$

with  $a = \min x_i$  and  $b = \max x_i$ , where  $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11})^T$  and  $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21})^T$  are the parameters of interest. Furthermore, it is assumed that there is a continuous transition between the two models at the change point. Thus, the model must satisfy a continuity constraint at the change point.

In general, in GLMs the model for the mean  $\mu_i = E[y_i|\mathbf{x}_i]$  is given by

$$g(E[y_i|\mathbf{x}_i]) = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Using this notation for model (2.48) gives

$$g(E[y_i|x_i]) = \begin{cases} \beta_{10} + \beta_{11}x_i & a \leq x_i \leq \gamma \\ \beta_{20} + \beta_{21}x_i & \gamma < x_i \leq b. \end{cases}$$

The continuity constraint at the change point  $\gamma$  implies that  $E[y_i|x_i]$  at the change point is the same for the model of both segments, i.e.

$$E[y_i|\gamma] = g^{-1}(\beta_{10} + \beta_{11}\gamma) = g^{-1}(\beta_{20} + \beta_{21}\gamma). \quad (2.49)$$

As the link function is a monotonic differentiable function and is the same for both segments, the continuity constraint (2.49) can be simplified to

$$\beta_{10} + \beta_{11}\gamma = \beta_{20} + \beta_{21}\gamma. \quad (2.50)$$

Without loss of generality, the intercept of the second model can be written as a function of the intercept and slope of the first model and the change point  $\gamma$ , i.e.

$$\beta_{20} = \beta_{10} + \beta_{11}\gamma - \beta_{21}\gamma. \quad (2.51)$$

Thus, model (2.48) can be rewritten as

$$g(\mu_i) = \begin{cases} \beta_{10} + \beta_{11}x_i & a \leq x_i \leq \gamma \\ \beta_{10} + \beta_{11}\gamma + \beta_{21}(x_i - \gamma) & \gamma < x_i \leq b. \end{cases} \quad (2.52)$$

Therefore, a change point model with one continuous change point can be interpreted as a model with one intercept and two different slopes for the different segments. Due to the continuity constraint, the intercept of the second segment is determined by the parameters  $\beta_{10}$ ,  $\beta_{11}$  and  $\beta_{21}$  and the change point  $\gamma$ . It can be calculated through equation (2.51).

Again, the change point  $\gamma$  divides data into two segments with respect to the explanatory variable. Without loss of generality, it is assumed, that  $x_i \leq x_j$  for  $i < j$  and  $x_s \leq \gamma < x_{s+1}$ , then  $\mathbf{x}^{(1)} = (x_1, \dots, x_s)^T$  and  $\mathbf{x}^{(2)} = (x_{s+1}, \dots, x_n)^T$ . Consequently, GLMs with one continuous change point can be written in matrix notation as

$$g(\boldsymbol{\mu}) = \begin{pmatrix} 1 & x_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_s & 0 \\ 1 & \gamma & x_{s+1} - \gamma \\ \vdots & \vdots & \vdots \\ 1 & \gamma & x_n - \gamma \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{21} \end{pmatrix}, \quad (2.53)$$

with  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ . As it was very easy to include the continuity constraint into the structure of the model, the big disadvantage of this structure is that model (2.53) can only be fitted with the standard IWLS method if the change point is known. Of course, usually the change point is unknown. Nevertheless, consider the case where the change point  $\gamma$  is known with  $x_s \leq \gamma < x_{s+1}$ . Then, the log likelihood of model (2.52) with  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$  is

$$\ell(\boldsymbol{\beta}, \phi|\gamma, \mathbf{y}) = \ell(\boldsymbol{\beta}_1, \phi|\gamma, y_1, \dots, y_s) + \ell(\boldsymbol{\beta}_1, \beta_{21}, \phi|\gamma, y_{s+1}, \dots, y_n). \quad (2.54)$$

Note, the log likelihood can be separated with respect to the first  $s$  and the last  $n - s$  observations, but the estimates of the parameters depend on both sets. This dependency can be seen from the parameter score functions under this model, which are given by

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}, \phi|\gamma, \mathbf{y})}{\partial \beta_{10}} &= \sum_{i=1}^s \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \right] + \sum_{i=s+1}^n \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \right] \\ \frac{\partial \ell(\boldsymbol{\beta}, \phi|\gamma, \mathbf{y})}{\partial \beta_{11}} &= \sum_{i=1}^s \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_i \right] + \sum_{i=s+1}^n \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \gamma \right] \\ \frac{\partial \ell(\boldsymbol{\beta}, \phi|\gamma, \mathbf{y})}{\partial \beta_{21}} &= \sum_{i=s+1}^n \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} (x_i - \gamma) \right], \end{aligned}$$

with

$$\eta_i = \begin{cases} \beta_{10} + \beta_{11}x_i & a \leq x_i \leq \gamma \\ \beta_{10} + \beta_{11}\gamma + \beta_{21}(x_i - \gamma) & \gamma < x_i \leq b. \end{cases}$$

Thus,  $(\beta_{10}, \beta_{11})^T$  and  $(\beta_{20}, \beta_{21})^T$  can not be estimated separately as it is the case for discontinuous change point models. However, the usual IWLS method based on model (2.53) can be used to estimate the parameters  $\beta_{10}$ ,  $\beta_{11}$  and  $\beta_{21}$ . The ML estimate of  $\beta_{20}$  can then be obtained by plugging in the ML estimates  $\hat{\beta}_{10}$ ,  $\hat{\beta}_{11}$  and  $\hat{\beta}_{21}$  into equation (2.51).

Next, consider the case where the change point  $\gamma$  is unknown and has to be estimated. The ML estimate of  $\gamma$  is the value, which maximizes

$$\ell(\boldsymbol{\beta}, \phi, \gamma | \mathbf{y}) = \ell(\boldsymbol{\beta}_1, \phi | \gamma, y_1, \dots, y_s) + \ell(\boldsymbol{\beta}_1, \beta_{21}, \phi | \gamma, y_{s+1}, \dots, y_n), \quad (2.55)$$

where, again, the ML estimates of  $\boldsymbol{\beta}$  and  $\phi$  depend on the location of the change point. Using the same model structure as in (2.52), one difficulty arises. The design matrix in (2.53) depends on the continuous change point  $\gamma$  for two reasons. First, the separation of the rows in it's two components representing the two models depends on  $\gamma$ . Second, the value of  $\gamma$  is an element of the design matrix. Thus, no analytical solution for estimating  $\gamma$  is available. Stasinopoulos and Rigby (1992) suggested to use a grid search for estimating the change point. Note that the change point is not restricted to an observed value of the explanatory variable, hence the grid can be freely chosen. Regardless of this possibility, an intuitive grid search seems to be a search over all observed  $x_i$ , in particular, in the case of equidistant  $x_i$ . Then, the ML estimate  $\hat{\gamma}$  is that observed  $x_a$  with

$$a = \arg \max_{2 \leq s \leq n-2} [\ell(\boldsymbol{\beta}_1, \phi | \gamma, y_1, \dots, y_s) + \ell(\boldsymbol{\beta}_1, \beta_{21}, \phi | \gamma, y_{s+1}, \dots, y_n)].$$

After a grid search over all feasible  $x_i$ , the grid search can be refined on a region around  $x_a$ , with lower bound  $x_{a-1}$  and upper bound  $x_{a+1}$ . Regardless of the refinement of the grid size, the solution of the grid search will always be suboptimal. Thus, it would be desirable to define a method for an exact estimation of the change point.

For such an exact method, consider the change point model (2.48) with continuity constraint (2.50). This exact method consists of two steps. First, we assume that the change point lies in an arbitrary open interval  $(x_s; x_{s+1})$ ,  $s = 3, \dots, n - 3$ , and calculate a candidate of the ML estimate of the change point. If this candidate lies in this open interval, it is the ML estimate. Second, for all intervals where this candidate is not an element of the corresponding interval, the endpoint of the interval, i.e.  $x_s$  and  $x_{s+1}$  are considered.

Therefore, we first assume that the the change point  $\gamma$  lies in the interval  $(x_s; x_{s+1})$ . Then, the log likelihood of (2.48) is given by

$$\ell(\boldsymbol{\beta}, \phi, \gamma | \mathbf{y}) = \ell(\boldsymbol{\beta}_1, \phi, \gamma | y_1, \dots, y_s) + \ell(\boldsymbol{\beta}_2, \phi, \gamma | y_{s+1}, \dots, y_n), \quad (2.56)$$

where both terms on the right hand side are the log likelihood functions of the GLMs of the first  $s$  and last  $n - s$  observations respectively. Given the two

observations between which the change point is located, i.e.  $x_s \leq \gamma < x_{s+1}$ , these two log likelihood functions do not depend on the value  $\gamma$  of the change point. This independency can be again seen from the parameter score functions which are given by

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}, \phi, \gamma | \mathbf{y})}{\partial \beta_{10}} &= \sum_{i=1}^s \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \right] \\ \frac{\partial \ell(\boldsymbol{\beta}, \phi, \gamma | \mathbf{y})}{\partial \beta_{11}} &= \sum_{i=1}^s \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_i \right] \\ \frac{\partial \ell(\boldsymbol{\beta}, \phi, \gamma | \mathbf{y})}{\partial \beta_{20}} &= \sum_{i=s+1}^n \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \right] \\ \frac{\partial \ell(\boldsymbol{\beta}, \phi, \gamma | \mathbf{y})}{\partial \beta_{21}} &= \sum_{i=s+1}^n \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_i \right], \end{aligned}$$

with

$$\eta_i = \begin{cases} \beta_{10} + \beta_{11}x_i & a \leq x_i \leq \gamma \\ \beta_{20} + \beta_{21}x_i & \gamma < x_i \leq b. \end{cases}$$

Consequently, the ML estimates  $\hat{\boldsymbol{\beta}}_1$  and  $\hat{\boldsymbol{\beta}}_2$  are the usual ML estimates of these two GLMs. Given the continuity constraint (2.50), the change point  $\gamma$  can be expressed as a function of the components of the parameter vector  $\boldsymbol{\beta}$ , namely

$$\gamma = \frac{\beta_{20} - \beta_{10}}{\beta_{11} - \beta_{21}}.$$

Due to the invariance property of ML estimators, i.e. if  $\hat{\theta}$  is the ML estimate of an unknown parameter  $\theta$ , then for any function  $h(\cdot)$ , the ML estimate of  $h(\theta)$  is  $h(\hat{\theta})$  (for details see for example Casella & Berger, 2002), the ML estimate of  $\gamma$  is

$$\hat{\gamma} = \frac{\hat{\beta}_{20} - \hat{\beta}_{10}}{\hat{\beta}_{11} - \hat{\beta}_{21}}. \quad (2.57)$$

Let  $\hat{\beta}_{10}, \hat{\beta}_{11}$  be the usual ML estimates of the GLM of the first  $s$  observations and  $\hat{\beta}_{20}, \hat{\beta}_{21}$  be the ML estimates of the GLM of the last  $n - s$  observations. Then a candidate of the ML estimate is given by (2.57). Note, however that because the change point is restricted to be within the interval  $(x_s; x_{s+1})$ , this is only the ML estimate of  $\gamma$ , if  $\hat{\gamma} \in (x_s; x_{s+1})$ . If this is not the case, then the ML estimate of  $\gamma$  is either  $x_s$  or  $x_{s+1}$ . Thus, finding a continuous change point between two consecutive observations can lead to two different cases. Either the continuous change point is found in the open interval between these two observations or it is one of these two observations. Regardless which of these two cases appears, this is the ML estimate of  $\gamma$  restricted to the interval  $[x_s; x_{s+1}]$ . To obtain an unrestricted change point, i.e.  $\gamma \in [x_3; x_{n-2}]$ , this has to be repeated for each interval  $[x_s; x_{s+1}]$  with  $s = 3, \dots, n - 3$ .

Küchenhoff (1997) first proposed this method for an exact estimation of the change point. He recommended to derive (2.57) for each open interval  $(x_s; x_{s+1})$  with  $s = 3, \dots, n - 3$ . If  $\hat{\gamma} \in (x_s; x_{s+1})$ , then there exists a local maximum of the log likelihood and hence the ML estimate of  $\gamma$  is given by (2.57). If  $\hat{\gamma} \notin (x_s; x_{s+1})$ , then there is no local maximum in  $(x_s; x_{s+1})$ , but it is either  $x_s$  or  $x_{s+1}$ . This second step is to maximize the log likelihood at the endpoint  $x_s$  of each interval, which is nothing else than the grid search over all observed  $x_i$  mentioned above. The global maximum is then given by the maximum of this final number of local maxima of both steps.

For a huge data set, this exact method can be very time consuming. One way to save time is to apply a grid search with a larger grid size, e.g. use only every fifth observation to get an indication where the change point might be. Then the exact method can be applied in the vicinity of this first clue. A second way is to use recursive residuals to find the change point. The simulation study in Subsection 2.1.4 demonstrated, that the method of Brown et al. (1975), which was developed to find change points in OLMs, can also be applied to find discontinuous change points in GLMs. Furthermore, the MC simulation showed that the performance of this method depends on the amount of the difference of the two GLMs at the change point. In contrast, in GLMs with a discontinuous change point, this amount is determined by the difference between the intercept and the slope of the two GLMs. In GLMs with a continuous change point this amount is only determined by the difference of the slopes of these two models. Thus, one would expect that the performance of Brown's method depends on this difference. To establish this dependency and how useful it is to apply recursive residuals to find a continuous change point in GLMs, the following MC simulation is performed.

### Simulation Study

In this MC simulation, the usefulness of the method proposed by Brown et al. (1975) (see Subsection 2.1.4) for detecting a continuous change point in GLMs is examined. For that reason, two different GLMs with one continuous change point are considered. As in Subsection 2.1.4, the performance of this method using either the standardized recursive residuals (2.37) or the deletion residuals (2.38) are of main interest.

For both models Poisson responses are considered. The explanatory variable  $x$  is  $x_i = i/40$ ,  $i = 1, \dots, 40$  and a continuous change point is located at  $\gamma = 0.5125$ , which is  $(x_{20} + x_{21})/2$ . Furthermore, it is assumed that the link function  $g(\mu) = \log(\mu)$  is the same for both segments. The first GLM considered, denoted as Model 1, is given by

$$g(\mu_i) = \begin{cases} 4 + 0.5x_i & 1/40 \leq x_i \leq \gamma \\ 3.4875 + 1.5x_i & \gamma < x_i \leq 1 \end{cases}$$

and the second GLM, denoted as Model 2, is

$$g(\mu_i) = \begin{cases} 4 + 0.5x_i & 1/40 \leq x_i \leq \gamma \\ 3.23125 + 2x_i & \gamma < x_i \leq 1. \end{cases}$$

Thus, the two models differ in the slope of the second segment and, consequently, also in the intercept of the second segment to guarantee a continuous transition at the change point. Based on these two models, 10000 samples of Poisson responses for both models were generated. These samples are analyzed in the same way as in the MC simulation in Subsection 2.1.4.

One result of this MC simulation is, that there is a large lag between the true change point (between the 20th and 21th observation) and the detected change point in the simulation study. Second, there is no marked difference in the results between the two considered types of residuals, i.e. standardized recursive residuals and deletion residuals.

For Model 1, the change point was detected in only 45.5% of the 10000 replications using the deletion residuals and in 43.4% of the replications using the recursive residuals (see Figure 2.16). This was done with a lag of up to 20 observations (Figure 2.17). For Model 2, the change point was detected in 91.9% of the replications using the deletion residuals and in 91.5% of the replications using the recursive residuals (Figure 2.16), but also with a large lag with its mode at 13 observations for both types of residuals (Figure 2.18). The relative frequency of detecting the change point at these modes are 9.8% and 9.6% using deletion residuals or recursive residuals, respectively. As it was expected, the lag of detecting the change point depends strongly on the difference between the slope of the models in the two segments, i.e.  $\beta_{21} - \beta_{11}$ , which is 1 for Model 1 and 1.5 for Model 2.

Thus, finding a continuous change point in GLMs using Brown's method can be applied if the difference between the two slopes of the two GLMs is a marked quantity. The power of this method decreases as the difference between the two slopes decreases. In the case where this difference is small with respect to the variance of the data at the change point, this method can only be applied as a first screening device. Regardless of the difference between the two slopes, this method can only be used to obtain an upper bound for the true continuous change point. This upper bound has to be investigated carefully, however, as in some cases this method provides smaller estimates than the true change point.

To get a lower bound for the continuous change point, one can start the analysis from the other side of the data, in this case the right hand side. However, this is only valid if it can be assumed that there is only one continuous change point. This lower and upper bound can be used as restrictions for a final grid search or the exact method suggested by Küchenhoff (1997). The computational expense for this combined method is greater than for an ordinary grid search. Nevertheless, in this thesis the recursive residuals are applied in Section 2.4 as an initial screening device to provide a first guess of the true continuous change point.



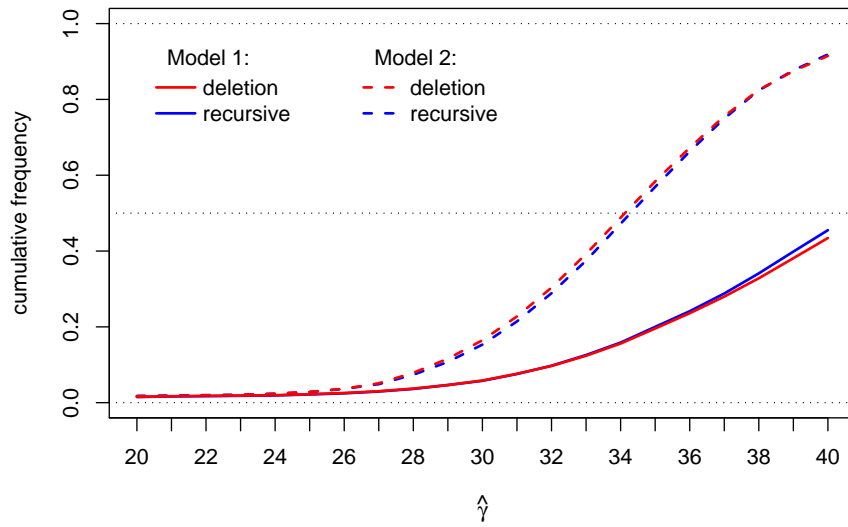


Figure 2.16: Cumulative relative frequencies of the change point estimate.

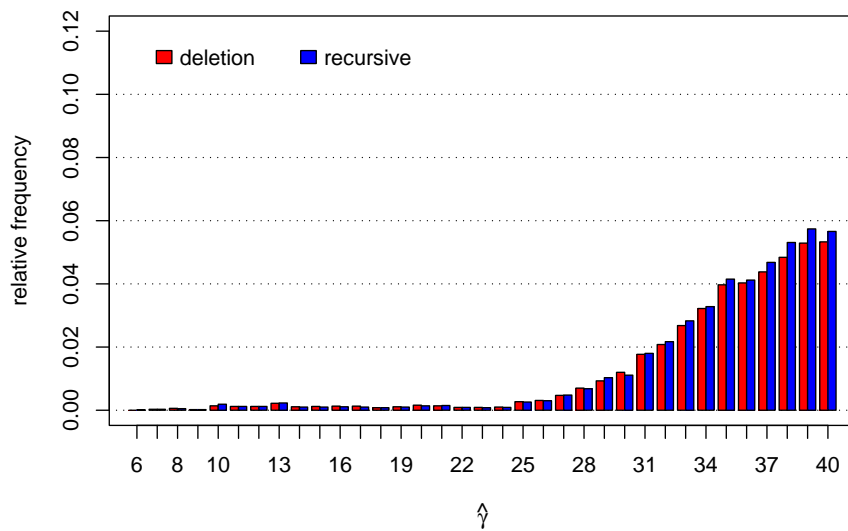


Figure 2.17: Relative frequencies of the change point estimate for Model 1.

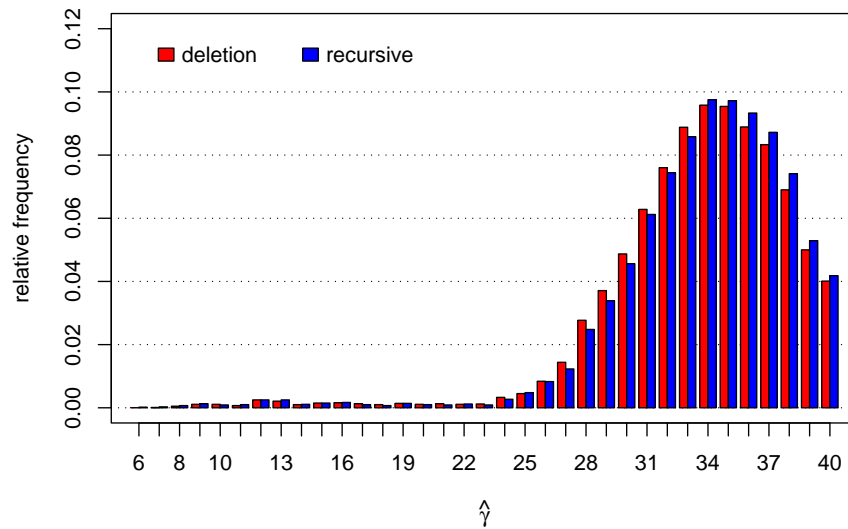


Figure 2.18: Relative frequencies of the change point estimate for Model 2.

#### Example 2.4 (*Continuation of Example 1.3*)

Recall, that for the data of the vital capacity an OLM with one continuous change point was assumed. To fit this model the function `ccp` can be used. As the call of this function requires an object of class `glm`, first, a GLM without a change point, Gaussian distribution for the response, and identity link function is fitted to the data. The result is stored in the object `aimu.glm`. Then the function `ccp` applied to `aimu.glm` fits a GLM with a continuous change point. The arguments `nOfCp` and `method` specify the number of change points and the method, either a grid search or the exact method, to estimate the change point, respectively. The result, stored as `aimu.ccp` is an object of class `ccp`. There exist a few generic accessor functions to extract some useful features of the fitted model. One of these is the function `print`, which gives a short overview of the fitted model. This includes the call of the function, the estimated coefficients and change points and some information about the goodness-of-fit.

```
> aimu.glm <- glm(vc~age, data=aimu)
> aimu.ccp <- ccp(aimu.glm, nOfCp=1, method="exact")
> print(aimu.ccp) # or only: aimu.ccp
```

Call:

```
ccp(object = aimu.glm, nOfCp = 1, method = "exact")
```

Coefficients:

```
(Intercept)    age.seg1    age.seg2
      252.154      17.050      -5.556
```

Continuous Change Point at :

cp.1  
22

Degrees of Freedom: 78 Total (i.e. Null); 75 Residual  
Null Deviance: 453700  
Residual Deviance: 323000  
AIC : 891.2  
BIC : 897

Another useful generic function is the `plot` function, which plots the fitted model (Figure 1.3), produces a normal plot of the deviance residuals (Figure 2.19), plots the fitted values versus the deviance residuals (Figure 2.20), and the profile log likelihood of the change point (Figure 2.21).

```
> plot(aimu.ccp, which=2)
> plot(aimu.ccp, which=3)
> plot(aimu.ccp, which=4)
```

The normal plot (Figure 2.19) shows a fairly good alignment of the deviance residuals to the Gaussian distribution. Moreover, there appears to be no pattern in Figure 2.20, hence the change point model seems appropriate for this data. This is confirmed by the goodness-of-fit criterion  $AIC = 891$ , which is quite smaller than that one of an OLM with a linear trend ( $AIC = 906$ ) and with a quadratic trend ( $AIC = 901$ ).

In Figure 2.21 the curve of the difference between the maximized log likelihood at  $\hat{\gamma} = 22$  and the profile log likelihood is plotted. Under certain regularity conditions the profile log likelihood with respect to one parameter follows asymptotically a  $\chi^2$ -distribution with one degree of freedom. Thus, the points where the curve intersects a line, that corresponds to the 95% quantile of a  $\chi^2_1$ -distribution, i.e.  $\chi^2_{1;0.95} = 3.84$ , provides approximative bounds of a 95% confidence interval for the change point. This 95% quantile is plotted as a dashed line in Figure 2.21. As, to the left hand side of  $\hat{\gamma} = 22$ , this curve does not intersect the dashed line the lower bound for the confidence interval is the smallest possible value for the change point, that is 18.85. The upper bound is 25, hence an approximative 95% confidence interval for the change point is [18.85; 25].

Next, consider an extension of GLMs with one continuous change point where additional explanatory variables are augmented to the linear predictor, but with a constant parameter for all observations for these additional explanatory variables. Then the model is given by

$$g(\mu_i) = \begin{cases} \beta_{10} + \beta_{11}x_i + \mathbf{z}_i^T \boldsymbol{\vartheta} & a \leq x_i \leq \gamma \\ \beta_{20} + \beta_{21}x_i + \mathbf{z}_i^T \boldsymbol{\vartheta} & \gamma < x_i \leq b, \end{cases} \quad (2.58)$$

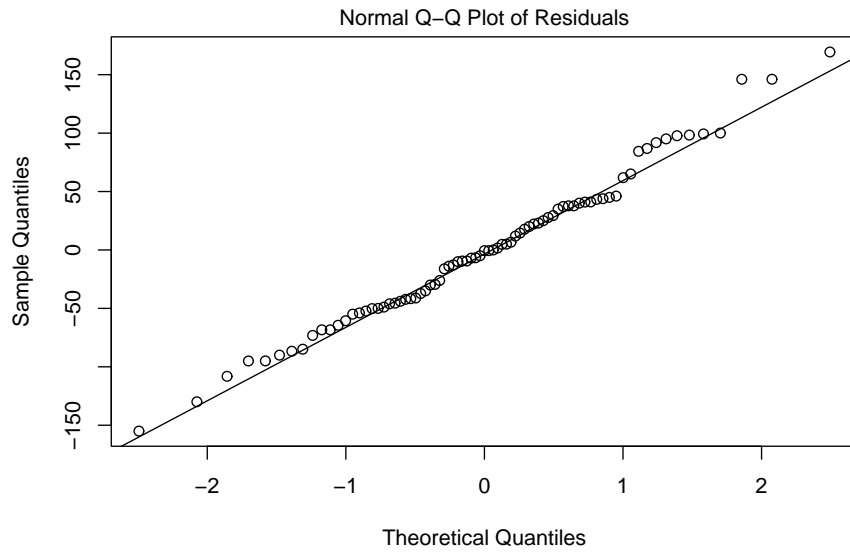


Figure 2.19: Normal plot of the deviance residuals.

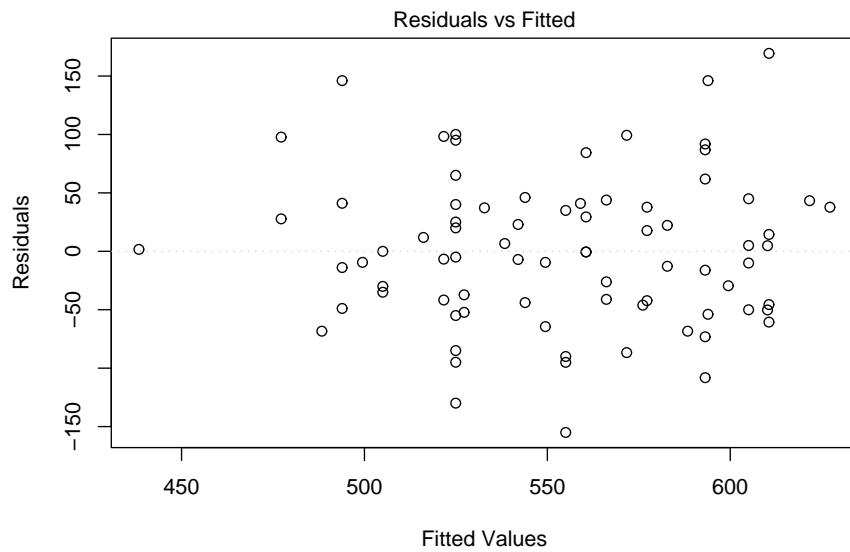


Figure 2.20: Deviance residuals versus fitted values of a GLM with one continuous change point for the vital capacity data.

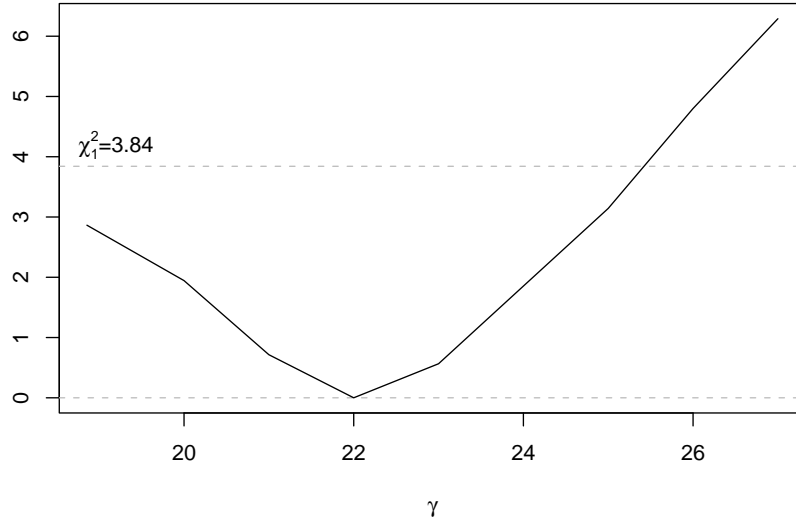


Figure 2.21: Curve of the difference between the maximized log likelihood at  $\hat{\gamma} = 22$  and the profile log likelihood of the change point. The dashed line at 3.84 corresponds to the 95% quantile of the  $\chi_1^2$ -distribution.

where  $\mathbf{z}_i = (z_{i1}, \dots, z_{i,p-2})^T$  are the additional explanatory variables without a change point and  $\boldsymbol{\vartheta}$  is a  $(p - 2) \times 1$  column vector of unknown parameters which has to be estimated. Thus, assuming that the continuous change point  $\gamma$  lies between  $x_s$  and  $x_{s+1}$ , this model can be written in matrix representation as

$$g(\boldsymbol{\mu}) = \begin{pmatrix} 1 & x_1 & 0 & 0 & z_{11} & \cdots & z_{1,p-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_s & 0 & 0 & z_{s1} & \cdots & z_{s,p-2} \\ 0 & 0 & 1 & x_{s+1} & z_{s+1,1} & \cdots & z_{s+1,p-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & x_n & z_{n1} & \cdots & z_{n,p-2} \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{20} \\ \beta_{21} \\ \boldsymbol{\vartheta} \end{pmatrix}. \quad (2.59)$$

The log likelihood under this model is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \phi | \gamma, \mathbf{y}) = \ell(\boldsymbol{\beta}_1, \boldsymbol{\vartheta}, \phi | \gamma, y_1, \dots, y_s) + \ell(\boldsymbol{\beta}_2, \boldsymbol{\vartheta}, \phi | \gamma, y_{s+1}, \dots, y_n),$$

where again both terms on the right hand side are the log likelihood functions of the first  $s$  and last  $n - s$  observations, respectively. Now, however, both terms are functions of the parameter vector  $\boldsymbol{\vartheta}$ , which is the same for both segments. Thus, the ML estimate of  $\boldsymbol{\vartheta}$  cannot be calculated by analyzing the two models of (2.58) separately. This is confirmed by the score function with

respect to any element  $\vartheta_j$ ,  $j = 1, \dots, p - 2$ , of  $\boldsymbol{\vartheta}$ , which is given by

$$\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \phi | \gamma, \mathbf{y})}{\partial \vartheta_j} = \sum_{i=1}^s \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} z_{ij} \right] + \sum_{i=s+1}^n \left[ \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} z_{ij} \right].$$

Consequently, the same algorithm presented above must be used for estimating the unknown parameters  $\boldsymbol{\beta}$ ,  $\boldsymbol{\vartheta}$ ,  $\phi$ , and  $\gamma$ . That is, first calculate the solution (2.57) of each interval  $(x_s; x_{s+1})$  and verify whether this is the ML estimate of the change point in this interval. If this is not the case, then establish which of the endpoints of the interval maximizes the log likelihood and is therefore the ML estimate of the change point in this interval. For the first step of this algorithm, model (2.58) can be used to fit the unknown parameters, in particular, the four components of  $\boldsymbol{\beta}$ , which are necessary to calculate the ML estimate of  $\gamma$ . For the second step, the grid search over those observations where a valid solution for the change point was not found between their corresponding intervals has to be applied. For that reason, model (2.58) is rewritten as

$$g(\boldsymbol{\mu}) = \begin{pmatrix} 1 & x_1 & 0 & z_{11} & \cdots & z_{1,p-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_s & 0 & z_{s1} & \cdots & z_{s,p-2} \\ 1 & \gamma & x_{s+1} - \gamma & z_{s+1,1} & \cdots & z_{s+1,p-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \gamma & x_n - \gamma & z_{n1} & \cdots & z_{n,p-2} \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{21} \\ \boldsymbol{\vartheta} \end{pmatrix}, \quad (2.60)$$

which is a generalization of model (2.53). Due to the continuity constraint, the intercept  $\beta_{20}$  of the second model is given by equation (2.51) and the ML estimate can be obtained by plugging in the ML estimate  $\hat{\beta}_{10}$ ,  $\hat{\beta}_{11}$  and  $\hat{\beta}_{21}$  into this equation.

The final ML estimates of the unknown parameters are those local ML estimates corresponding to the local maximum of the log likelihood, which is the global maximum.

### 2.3.2 Multiple Continuous Change Points

In this section, GLMs with one continuous change point are generalized to GLMs with at least two continuous change points. Therefore, consider a GLM with  $d - 1$  continuous change points  $\gamma_k$ ,  $k = 1, \dots, d - 1$ , with  $\gamma_k < \gamma_{k+1}$  and one explanatory variable  $x_i$  with  $a = \min x_i$  and  $b = \max x_i$ . These  $d - 1$  change points partition the data into  $d$  segments and the considered model is

$$g(\mu_i) = \begin{cases} \beta_{10} + \beta_{11}x_i, & a \leq x_i \leq \gamma_1 \\ \beta_{20} + \beta_{21}x_i, & \gamma_1 < x_i \leq \gamma_2 \\ \vdots & \vdots \\ \beta_{d0} + \beta_{d1}x_i, & \gamma_{d-1} < x_i \leq b, \end{cases} \quad (2.61)$$

where at each change point  $\gamma_k$ , a continuity constraint

$$g(E[y_i|\gamma_k]) = g(\beta_{k0} + \beta_{k1}\gamma_k) = g(\beta_{k+1,0} + \beta_{k+1,1}\gamma_k), \quad k = 1, \dots, d-1$$

is considered. As the link function is the same for all segments and is a monotonic differentiable function, these continuity constraints can be simplified to

$$\beta_{k0} + \beta_{k1}\gamma_k = \beta_{k+1,0} + \beta_{k+1,1}\gamma_k, \quad k = 1, \dots, d-1.$$

Then model (2.61) can be rewritten as a model with one intercept  $\beta_{10}$  and  $d$  different slopes  $\beta_{k1}$ ,  $k = 1, \dots, d$ . The intercept of each segment, except the first one, is then given by

$$\beta_{k+1,0} = \beta_{k0} + \beta_{k1}\gamma_k - \beta_{k+1,1}\gamma_k, \quad k = 1, \dots, d-1.$$

As the intercept of the succeeding segment is a function of the intercept and slope of the previous segment and the corresponding change point, the intercept of each segment can be built up recursively. Thus, (2.61) can be written as

$$g(\mu_i) = \begin{cases} \beta_{10} + \beta_{11}x_i & a \leq x_i \leq \gamma_1 \\ \beta_{10} + \beta_{11}\gamma_1 + \beta_{21}(x_i - \gamma_1) & \gamma_1 < x_i \leq \gamma_2 \\ \beta_{10} + \beta_{11}\gamma_1 + \beta_{21}(\gamma_2 - \gamma_1) + \beta_{31}(x_i - \gamma_2) & \gamma_2 < x_i \leq \gamma_3 \\ \vdots & \vdots \\ \beta_{10} + \beta_{11}\gamma_1 + \beta_{21}(\gamma_2 - \gamma_1) + \dots + \beta_{d1}(x_i - \gamma_{d-1}) & \gamma_{d-1} < x_i \leq b, \end{cases}$$

which in matrix notation is

$$g(\boldsymbol{\mu}) = \begin{pmatrix} \mathbf{e}_1 & \mathbf{x}^{(1)} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{e}_2 & \gamma_1 \mathbf{e}_2 & \mathbf{x}^{(2)} - \gamma_1 \mathbf{e}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{e}_3 & \gamma_1 \mathbf{e}_3 & (\gamma_2 - \gamma_1) \mathbf{e}_3 & \mathbf{x}^{(3)} - \gamma_2 \mathbf{e}_3 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{e}_d & \gamma_1 \mathbf{e}_d & (\gamma_2 - \gamma_1) \mathbf{e}_d & (\gamma_3 - \gamma_2) \mathbf{e}_d & \dots & \mathbf{x}^{(d)} - \gamma_{d-1} \mathbf{e}_d \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \vdots \\ \beta_{d1} \end{pmatrix}, \quad (2.62)$$

where the elements of the design matrix are column vectors of the same length for each row. In particular,  $\mathbf{x}^{(k)}$  is a column vector of elements given by  $\{x_i | \gamma_{k-1} < x_i \leq \gamma_k\}$ , with  $k = 1, \dots, d$ ,  $\gamma_d = b$  and  $\gamma_0 = a - \xi$  with  $\xi > 0$ . Furthermore,  $\mathbf{e}_k = (1, \dots, 1)^T$  is a column vector of ones with length  $|\mathbf{x}^{(k)}|$ . Thus, the length of these column vectors is determined by the location of the change points  $\gamma_k$ . This notation of a change point model is practicable if the change points are known. Thus, for a given vector of change points  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{d-1})^T$  the ML estimates of  $\beta_{10}$  and  $\beta_{k1}$ , can be derived with the common IWLS method applied on (2.62).

If the change points are unknown (as it is usually the case) then they have to be estimated. As it can be seen in the design matrix of (2.62), the models of each segment are no longer autonomous. Thus, the algorithms for

determining multiple discontinuous change points (Section 2.2) are not appropriate. One way to estimate the change points is to apply a grid search over all possible cases. However, for an efficient grid search, an appropriate grid has to be defined and, regardless of the defined grid, the solution of the search will always be suboptimal. Moreover, the computational complexity increases rapidly with the number of change points and the required degree of accuracy. An exact method can be obtained by extending the exact method for one continuous change point (Subsection 2.3.1) to the case of multiple continuous change points. Küchenhoff (1997) showed this for models with two continuous change points. As a GLM with two continuous change points is of main interest in this thesis, this exact method, first introduced by Küchenhoff (1997), is described in detail in the remainder of this section.

Consider a GLM with one explanatory variable  $x$  and two continuous change points, say  $\gamma_1$  and  $\gamma_2$ . It is assumed that  $x_s \leq \gamma_1 < x_{s+1}$  and  $x_r \leq \gamma_2 < x_{r+1}$  for some  $s + 2 < r$ . Then the considered GLM with two continuous change points equals

$$g(\mu_i) = \begin{cases} \beta_{10} + \beta_{11}x_i & a \leq x_i \leq \gamma_1 \\ \beta_{20} + \beta_{21}x_i & \gamma_1 < x_i \leq \gamma_2 \\ \beta_{30} + \beta_{31}x_i & \gamma_2 < x_i \leq b, \end{cases} \quad (2.63)$$

with continuity constraints at the change point given by

$$\begin{aligned} \beta_{10} + \beta_{11}\gamma_1 &= \beta_{20} + \beta_{21}\gamma_1 \\ \beta_{20} + \beta_{21}\gamma_2 &= \beta_{30} + \beta_{31}\gamma_2. \end{aligned} \quad (2.64)$$

To write this model in matrix notation the following notation is considered. To simplify the notation of the design matrix let

$$\mathbf{X}_{i|j} = \begin{pmatrix} 1 & x_i \\ \vdots & \vdots \\ 1 & x_j \end{pmatrix}$$

be the design matrix of the GLM for the observations  $y_i, \dots, y_j$  and  $\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k1})^T$ ,  $k = 1, 2, 3$ , the parameters of interest. Then (2.63) can be rewritten as

$$g(\boldsymbol{\mu}) = \begin{pmatrix} \mathbf{X}_{1|s} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{s+1|r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_{r+1|n} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{pmatrix}. \quad (2.65)$$

To estimate the unknown change points  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^T$ , we assume without loss of generality, that  $\gamma_1 \in (x_s; x_{s+1})$  and  $\gamma_2 \in (x_r; x_{r+1})$ . Then, the log likelihood with  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T)^T$  is

$$\begin{aligned} \ell(\boldsymbol{\beta}, \phi | \boldsymbol{\gamma}, \mathbf{y}) &= \ell(\boldsymbol{\beta}_1, \phi | \boldsymbol{\gamma}, y_1, \dots, y_s) \\ &+ \ell(\boldsymbol{\beta}_2, \phi | \boldsymbol{\gamma}, y_{s+1}, \dots, y_r) \\ &+ \ell(\boldsymbol{\beta}_3, \phi | \boldsymbol{\gamma}, y_{r+1}, \dots, y_n). \end{aligned}$$



For the same reason as in the case of one continuous change point (compare Subsection 2.3.1, equation (2.56)), the terms on the right hand side are autonomous. Hence, the ML estimates of  $\beta_k$  are the ML estimates of the GLMs based on the corresponding observations, i.e.  $(y_1, \dots, y_s)^T$ ,  $(y_{s+1}, \dots, y_r)^T$ , and  $(y_{r+1}, \dots, y_n)^T$ . Given these ML estimates  $\hat{\beta}_k$ , the candidates for the ML estimates of  $\gamma_1$  and  $\gamma_2$  can be derived through

$$\hat{\gamma}_1 = \frac{\hat{\beta}_{20} - \hat{\beta}_{10}}{\hat{\beta}_{11} - \hat{\beta}_{21}} \quad \text{and} \quad \hat{\gamma}_2 = \frac{\hat{\beta}_{30} - \hat{\beta}_{20}}{\hat{\beta}_{21} - \hat{\beta}_{31}}. \quad (2.66)$$

If  $\hat{\gamma}_1 \in (x_s; x_{s+1})$  and  $\hat{\gamma}_2 \in (x_r; x_{r+1})$ , then  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$  are ML estimates because they are restricted to the rectangle  $(x_s; x_{s+1}) \times (x_r; x_{r+1})$ . If this is not the case, the ML estimates lie on the boundaries of this rectangle.

In the next step, each of these four boundaries are investigated. For that reason, one change point is fixed at an observed value  $x_i$  and the other one varies in between the two corresponding observations. For instance and without loss of generality, hold the first change point fixed, i.e.  $\gamma_1 = x_s$ . Then the second change point varies in between  $(x_r; x_{r+1})$ . This corresponds to a GLM with one continuous change point with a slightly different design matrix. To simplify notation, let

$$\mathbf{Z}_{i|j|k} = \begin{pmatrix} 1 & x_i & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_j & 0 \\ 1 & x_j & x_{j+1} - x_j \\ \vdots & \vdots & \vdots \\ 1 & x_j & x_k - x_j \end{pmatrix}$$

be the design matrix for a change point model with a continuous change point at  $x_j$ . Then the model for the first change point fixed at  $x_s$  and the second change point may vary between  $x_r$  and  $x_{r+1}$  is

$$g(\boldsymbol{\mu}) = \begin{pmatrix} \mathbf{Z}_{1|s|r} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{r+1|n} \end{pmatrix} \begin{pmatrix} \beta_1^* \\ \beta_3 \end{pmatrix}, \quad (2.67)$$

with  $\beta_1^* = (\beta_{10}, \beta_{11}, \beta_{21})^T$ . As the second change point is restricted to  $(x_r; x_{r+1})$ , the ML estimates of  $\beta_1^*$  and  $\beta_3$  do not depend on the value of the change point but only on it's location and are therefore autonomous. Thus, the ML estimates of  $\beta_1^*$  and  $\beta_3$  are the ML estimates of the GLMs with the design matrix  $\mathbf{Z}_{1|s|r}$  and  $\mathbf{X}_{r+1|n}$  respectively. After fitting these two models, it has to be again verified, whether the candidate of the second change point

$$\hat{\gamma}_2 = \frac{\hat{\beta}_{20} - \hat{\beta}_{10}}{\hat{\beta}_{11} - \hat{\beta}_{21}} \quad (2.68)$$

with

$$\hat{\beta}_{20} = \hat{\beta}_{10} + (\hat{\beta}_{11} - \hat{\beta}_{21})\gamma_1 = \hat{\beta}_{10} + (\hat{\beta}_{11} - \hat{\beta}_{21})x_s$$

lies in the interval  $(x_r; x_{r+1})$ . If this is the case, then (2.68) is the ML estimate of the second change point. In contrast, if the second change point is fixed to a certain observation, i.e.  $\gamma_2 = x_r$ , then model (2.67) changes to

$$g(\boldsymbol{\mu}) = \begin{pmatrix} \mathbf{X}_{1|s} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{s+1|r|n} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2^* \end{pmatrix}, \quad (2.69)$$

with  $\boldsymbol{\beta}_2^* = (\beta_{20}, \beta_{21}, \beta_{31})^T$ .

This investigation of the boundary has to be done for all four boundaries of the rectangle  $[x_s; x_{s+1}] \times [x_r; x_{r+1}]$ , i.e. subsequently the first change point is fixed to  $x_s$  and  $x_{s+1}$  and then the second change point is fixed to  $x_r$  and  $x_{r+1}$ . The result of the investigation of these four boundaries can yield six different patterns. To illustrate these different patterns, the rectangle  $[x_s; x_{s+1}] \times [x_r; x_{r+1}]$  is defined as a graph where each corner of the rectangle is represented by a node. The edges between the nodes indicate that a maximum of the log likelihood between these two nodes exists. For instance, the case where the first change point is fixed at  $x_s$  and the second change point lies between  $x_r$  and  $x_{r+1}$  is illustrated in Figure 2.22. The six different patterns

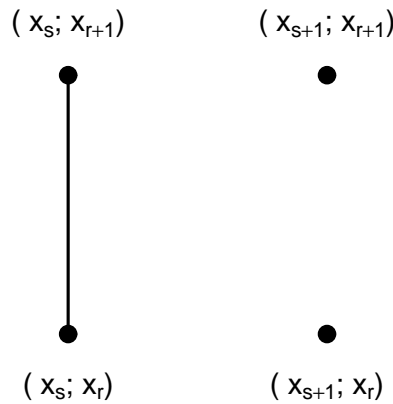


Figure 2.22: Example for a graph, where the first change point is fixed at  $\gamma_1 = x_s$  and the second change point lies in between  $x_r$  and  $x_{r+1}$

using this representation are illustrated in Figure 2.23.

If pattern (a) is the result of the investigation of the boundaries, i.e. a local maximum of the log likelihood is obtained at each boundary, then the ML estimate of the continuous change point lies on that boundary that corresponds to the maximum of these four maximized log likelihoods. For pattern (b) and (c), the ML estimate of  $\boldsymbol{\gamma}$  is also the value, that yields the maximum log likelihood of these three and two boundaries, respectively. This is due to the fact that, if there is a local maximum of the log likelihood for one change point fixed and the other one restricted to an open interval between two consecutive observations, then the maximum of the log likelihood can not be at the corners

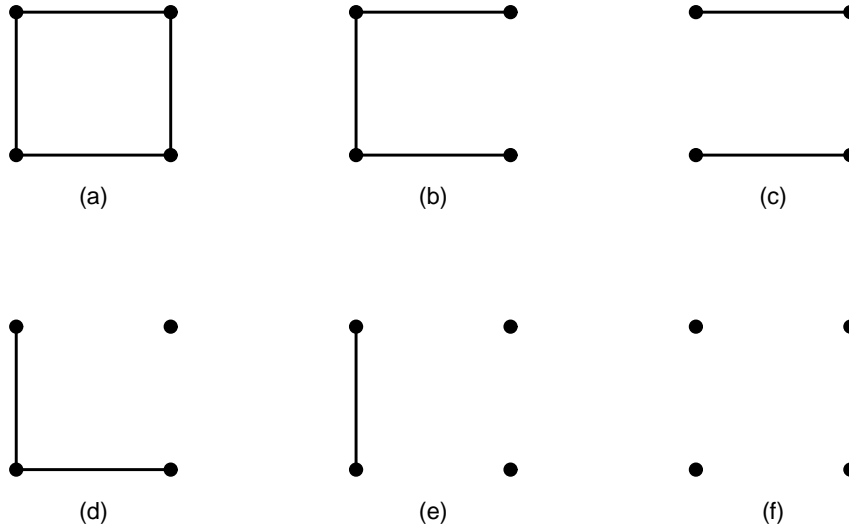


Figure 2.23: Six different pattern which can occur.

of the rectangle. Hence, in the case of pattern (b) and (c) the maximum log likelihood lies on the boundaries and not at these two observations, i.e. at one of the corners. For pattern (d) - (f), a subsequent analysis of the corners with no edge connecting this corner to another one is necessary, i.e. the two change points are fixed to the corresponding observations. Without loss of generality it is assumed that the change points are fixed to  $\gamma_1 = x_s$  and  $\gamma_2 = x_r$ . Then the considered model can be written as

$$g(\boldsymbol{\mu}) = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_s & 0 & 0 \\ 1 & x_s & x_{s+1} - x_s & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_s & x_r - x_s & 0 \\ 1 & x_s & x_r - x_s & x_{r+1} - x_r \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_s & x_r - x_s & x_n - x_r \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{21} \\ \beta_{31} \end{pmatrix}, \quad (2.70)$$

which can be fitted with the usual IWLS method. Again the intercepts of the second and third segment are obtained by plugging in the ML estimates of  $\beta_{10}$ ,  $\beta_{11}$ ,  $\beta_{21}$  and  $\beta_{31}$  into the corresponding continuity constraints (2.64).

Summing up, the exact method for estimating the change points of a GLM with two continuous change points consists of three steps.

1. For each rectangle  $(x_s; x_{s+1}) \times (x_r; x_{r+1})$ , model (2.65) is fitted to the

data and subsequently it has to be investigated if the candidates (2.66) are the ML estimates of the change points.

2. For each rectangle where at least one candidate is not the ML estimate of the change point, the boundaries of the rectangle have to be investigated, i.e. fit model (2.67) respectively (2.69) to the data.
3. For each rectangle where pattern (d) - (f) occur, a subsequent analysis of the corners of the rectangle is necessary, i.e. fit (2.70) to the data.

Thus, in the worst case, each rectangle has to be investigated in nine different ways to get the ML estimates of the change points restricted to the considered rectangle. For a global ML estimate of the change point, this has to be done for all feasible rectangles, thus the computational complexity increases rapidly with the number of observations.

As mentioned in Küchenhoff (1997), this exact method can be extended to the case of more than two change points. For instance, in the case of three change points, all feasible cubes  $[x_s; x_{s+1}] \times [x_r; x_{r+1}] \times [x_u; x_{u+1}]$  have to be investigated.

### Example 2.5 (Continuation of Example 1.4)

Recall, that a GLM with two continuous change points is considered for the data of recorded discharge of a river after a heavy rainstorm. Again, for the call of the function `ccp` a GLM without a continuous change point has to be fitted to the data. As the implemented `power` link function does not support any negative power, a new family `hyd.const` was implemented with Gaussian distribution for the response variable and a power link function with three different negative values, i.e.  $-0.5$ ,  $-0.4$  and  $-0.3$  noted as `w05`, `w04` and `w03` respectively. The fitted GLM is stored as object `hyd.glm`. This object is the main argument of `ccp`. The number of change points (`nOfCp`) is set to 2 and the exact method is used to estimate the change points. The estimated coefficients are the intercept and the slope of the leftmost segment and the slopes of the remaining segments.

```
> hyd.glm <- glm(discharge~x, data=sulm.dat,
+               family=hyd.const(link=w05),
+               subset=which(sulm.dat[,1]=="rez8"))
> hyd.ccp <- ccp(hyd.glm, nOfCp=2, method="exact")
> hyd.ccp
```

Call:

```
ccp(object = hyd.glm, nOfCp = 2, method = "exact")
```

Coefficients:

(Intercept)	x.seg1	x.seg2	x.seg3
0.125999	0.136961	0.047631	0.008015

Continuous Change Points at :

```

cp.1    cp.2
0.3259  1.7128

```

```

Degrees of Freedom: 28 Total (i.e. Null);  23 Residual
Null Deviance:      5808
Residual Deviance:  13.22
AIC : 73.51
BIC : 81.71

```

*In the case of two continuous change points the profile log likelihood depending on these two change points is shown as a contour plot. The location of the estimated change points is marked as + and additional horizontal and vertical dashed lines. In addition, the values of the estimated change points are printed as subtitle (Figure 2.24).*

```
> plot(hyd.ccp, which=4)
```

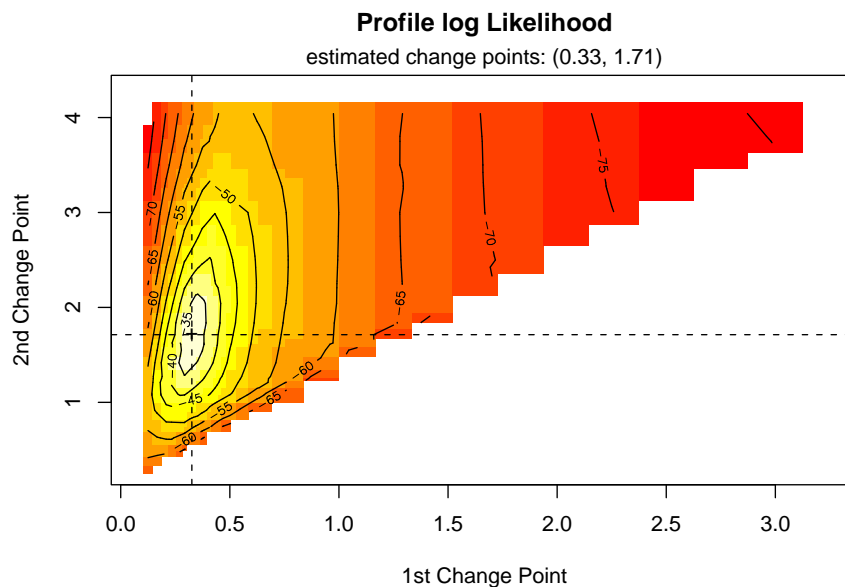


Figure 2.24: Contour plot of the profile log likelihood depending on two continuous change points.

## 2.4 Continuous Change Points for Panel Data

In this section, the analysis of continuous change point models is extended to panel data. First, the case of independent observations both between and within the panels is considered (Subsection 2.4.1). Second, the same kind of

model is considered but it is additionally assumed that the slope parameter for the rightmost segment is the same for all panels (Subsection 2.4.2). Then, a fitting algorithm to estimate the change points and the parameters of interest for such a model is presented (Subsection 2.4.3). Finally, an MC simulation study evaluates the performance of this new algorithm (Subsection 2.4.4).

### 2.4.1 The Independent Case

Let us consider panel data with  $j = 1, \dots, n$  panels (clusters), where each panel has  $i = 1, \dots, n_j$  observations. This notation allows either balanced panels,  $n_1 = n_2 = \dots = n_n$ , or unbalanced panels with  $n_{j_1} \neq n_{j_2}$  for at least one  $j_1 \neq j_2$  with  $j_1, j_2 \in \{1, \dots, n\}$ . In general, different models for each panel are possible (e.g. see Subsection 2.2.1). Nevertheless, in this work it is assumed, that the distribution of the response variable is the same for all panels and segments within each panel. Furthermore, it is assumed that the link and the variance function are the same for all segments and all panels. Even though it is not important for the remaining analysis if the number of change points is fixed to a certain value or may vary from panel to panel, in the remainder of this section and without loss of generality, GLMs with two continuous change points are considered for each panel.

Let  $\mathbf{y}_j = (y_{j1}, \dots, y_{jn_j})^T$  be the response vector of panel  $j$ , where  $y_{ji}$  follows a distribution from the exponential family. We assume that the corresponding explanatory variable  $x_{ji}$  has a natural ordering, i.e.  $x_{ji} \leq x_{j,i+1}$ . Moreover, two continuous change points  $\gamma_{j1}, \gamma_{j2} \in [x_{j1}; x_{jn_j}]$ , with  $\gamma_{j1} < \gamma_{j2}$ , partition the data into three segments. Besides the change points, the parameters of interest, are the linear parameters  $\beta_{jk} = (\beta_{jk0}, \beta_{jk1})^T$ ,  $k = 1, 2, 3$ , of these three segments. The continuous change point model for panel  $j$  can then be given by

$$g(\mu_{ji}) = \begin{cases} \beta_{j10} + \beta_{j11}x_{ji} & a_j \leq x_{ji} \leq \gamma_{j1} \\ \beta_{j20} + \beta_{j21}x_{ji} & \gamma_{j1} < x_{ji} \leq \gamma_{j2} \\ \beta_{j30} + \beta_{j31}x_{ji} & \gamma_{j2} < x_{ji} \leq b_j, \end{cases} \quad (2.71)$$

with  $a_j = \min x_{ji}$  and  $b_j = \max x_{ji}$ . As the link function is the same for all three segments, the continuity constraint at the change points can be simplified to

$$\beta_{jk0} + \beta_{jk1}\gamma_{jk} = \beta_{j(l)0} + \beta_{j(l)1}\gamma_{jk}, \quad k = 1, 2 \quad (2.72)$$

with  $l = k + 1$ .

First, consider a given pair of change points  $\boldsymbol{\gamma}_j = (\gamma_{j1}, \gamma_{j2})^T$  and, without loss of generality, consider  $x_{js} \leq \gamma_{j1} < x_{j,s+1}$  and  $x_{jr} \leq \gamma_{j2} < x_{j,r+1}$  for some  $s + 2 < r$  with  $s, r \in \{1, \dots, n_j\}$ . Then, using (2.72), the intercepts can be built up recursively and model (2.71) can be written as

$$g(\mu_{ji}) = \begin{cases} \beta_{j10} + \beta_{j11}x_{ji} & a_j \leq x_{ji} \leq \gamma_{j1} \\ \beta_{j10} + \beta_{j11}\gamma_{j1} + \beta_{j21}(x_{ji} - \gamma_{j1}) & \gamma_{j1} < x_{ji} \leq \gamma_{j2} \\ \beta_{j10} + \beta_{j11}\gamma_{j1} + \beta_{j21}(\gamma_{j2} - \gamma_{j1}) + \beta_{j31}(x_{ji} - \gamma_{j2}) & \gamma_{j2} < x_{ji} \leq b_j. \end{cases} \quad (2.73)$$

We denote the mean vector of panel  $j$  as  $\boldsymbol{\mu}_j = (\mu_{j1}, \dots, \mu_{jn_j})^T$  and the parameter vector of interest as  $\boldsymbol{\beta}_j^* = (\beta_{j1}^T, \beta_{j21}, \beta_{j31})^T$ . Then (2.73) can be written in matrix notation as

$$g(\boldsymbol{\mu}_j) = \mathbf{Z}_j \boldsymbol{\beta}_j^* \quad (2.74)$$

with design matrix

$$\mathbf{Z}_j = \begin{pmatrix} 1 & x_{j1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{js} & 0 & 0 \\ 1 & \gamma_{j1} & x_{j,s+1} - \gamma_{j1} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \gamma_{j1} & x_{jr} - \gamma_{j1} & 0 \\ 1 & \gamma_{j1} & \gamma_{j2} - \gamma_{j1} & x_{j,r+1} - \gamma_{j2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \gamma_{j1} & \gamma_{j2} - \gamma_{j1} & x_{j,n_j} - \gamma_{j2} \end{pmatrix}.$$

Remember, model (2.74) is a GLM with two change points, as discussed in Subsection 2.3.2. In addition, the log likelihood is similar to (2.54) but with an extra term for the third segment, i.e.

$$\begin{aligned} \ell(\boldsymbol{\beta}_j^*, \phi | \boldsymbol{\gamma}_j, \mathbf{y}_j) &= \ell(\boldsymbol{\beta}_{j1}, \phi | \boldsymbol{\gamma}_j, y_{j1}, \dots, y_{js}) \\ &+ \ell(\boldsymbol{\beta}_{j1}, \beta_{j21}, \phi | \boldsymbol{\gamma}_j, y_{j,s+1}, \dots, y_{jr}) \\ &+ \ell(\boldsymbol{\beta}_{j1}, \beta_{j21}, \beta_{j31}, \phi | \boldsymbol{\gamma}_j, y_{j,r+1}, \dots, y_{jn_j}). \end{aligned}$$

As for a GLM with one continuous change point, the ML estimate of  $\boldsymbol{\beta}_j^*$  is obtained by analyzing model (2.74). Again, the estimates of both intercepts  $\beta_{j20}$  and  $\beta_{j30}$ , are provided by solving restriction (2.72) and plugging in the ML estimate of  $\boldsymbol{\beta}_j^*$ .

Combining these  $n$  models for the entire panel data set leads to the global model

$$\begin{pmatrix} g(\boldsymbol{\mu}_1) \\ g(\boldsymbol{\mu}_2) \\ \vdots \\ g(\boldsymbol{\mu}_n) \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\beta}_2^* \\ \vdots \\ \boldsymbol{\beta}_n^* \end{pmatrix}. \quad (2.75)$$

Assuming that all observations are mutually independent, i.e. all observations within a panel are independent as well as all observations between all panels are independent, the log likelihood of model (2.75) is

$$\ell(\boldsymbol{\beta}, \phi | \boldsymbol{\gamma}, \mathbf{y}) = \sum_{j=1}^n \ell(\boldsymbol{\beta}_j^*, \phi | \boldsymbol{\gamma}_j, \mathbf{y}_j),$$

with  $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \dots, \boldsymbol{\gamma}_n^T)^T$ . Furthermore, the terms on the right hand side of this equation are autonomous, thus each panel can be analyzed separately. As this

model represents a GLM for given pairs of change points, it is the basis for the analysis if either the change points are known or are estimated by applying an ordinary grid search.

Next, consider the case where the change points are unknown and an exact method should be applied to estimate the change points. Again assume  $x_{js} \leq \gamma_{j1} < x_{j,s+1}$  and  $x_{jr} \leq \gamma_{j2} < x_{j,r+1}$ , then the model for panel  $j$  is

$$g(\boldsymbol{\mu}_j) = \mathbf{X}_j \boldsymbol{\beta}_j \quad (2.76)$$

with  $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{j1}^T, \boldsymbol{\beta}_{j2}^T, \boldsymbol{\beta}_{j3}^T)^T$  and design matrix

$$\mathbf{X}_j = \begin{pmatrix} 1 & x_{j1} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{js} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{j,s+1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{jr} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{j,r+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & x_{jn_j} \end{pmatrix}.$$

The log likelihood is similar to (2.56) but with an additional term for the third segment, i.e.

$$\begin{aligned} \ell(\boldsymbol{\beta}_j, \phi, \boldsymbol{\gamma}_j | \mathbf{y}_j) &= \ell(\boldsymbol{\beta}_{j1}, \phi, \boldsymbol{\gamma}_j | y_{j1}, \dots, y_{js}) \\ &+ \ell(\boldsymbol{\beta}_{j2}, \phi, \boldsymbol{\gamma}_j | y_{j,s+1}, \dots, y_{jr}) \\ &+ \ell(\boldsymbol{\beta}_{j3}, \phi, \boldsymbol{\gamma}_j | y_{j,r+1}, \dots, y_{jn_j}). \end{aligned}$$

To estimate the change points of model (2.76), the exact method introduced in Subsection 2.3.2 can be applied.

Combining these models gives to the global model for the entire panel data set

$$\begin{pmatrix} g(\boldsymbol{\mu}_1) \\ g(\boldsymbol{\mu}_2) \\ \vdots \\ g(\boldsymbol{\mu}_n) \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix}. \quad (2.77)$$

Again, assuming that all observations are mutually independent, the log likelihood of this global model is

$$\ell(\boldsymbol{\beta}, \phi, \boldsymbol{\gamma} | \mathbf{y}) = \sum_{j=1}^n \ell(\boldsymbol{\beta}_j, \phi, \boldsymbol{\gamma}_j | \mathbf{y}_j).$$

As in the case of known change points, the terms on the right hand side are autonomous and each panel can be analyzed separately.



## 2.4.2 Common Slope in the Last Segment

For this special model, we again consider panel data with a GLM with continuous change points for each panel. In addition, we assume that the slope parameter of the last segment is the same for all panels. The parameter of interest, besides the change points, are the panel specific parameters  $\boldsymbol{\beta}_j = (\boldsymbol{\beta}_{j1}^T, \boldsymbol{\beta}_{j2}^T, \beta_{j30})^T$ ,  $j = 1, \dots, n$ , and the common slope  $\delta$  for the last segment of all panels. Then the model for panel  $j$  can be written as

$$g(\boldsymbol{\mu}_{ji}) = \begin{cases} \beta_{j10} + \beta_{j11}x_{ji} & a_j \leq x_{ji} \leq \gamma_{j1} \\ \beta_{j20} + \beta_{j21}x_{ji} & \gamma_{j1} < x_{ji} \leq \gamma_{j2} \\ \beta_{j30} + \delta x_{ji} & \gamma_{j2} < x_{ji} \leq b_j \end{cases} \quad (2.78)$$

with continuity constraints

$$\begin{aligned} \beta_{j10} + \beta_{j11}\gamma_{j1} &= \beta_{j20} + \beta_{j21}\gamma_{j1} \\ \beta_{j20} + \beta_{j21}\gamma_{j2} &= \beta_{j30} + \delta\gamma_{j2}. \end{aligned} \quad (2.79)$$

First, consider the case where the change points are known. Then, model (2.75) is extended to

$$\begin{pmatrix} g(\boldsymbol{\mu}_1) \\ g(\boldsymbol{\mu}_2) \\ \vdots \\ g(\boldsymbol{\mu}_n) \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{T}_1^* \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} & \mathbf{T}_2^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_n & \mathbf{T}_n^* \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \vdots \\ \boldsymbol{\beta}_n^* \\ \delta \end{pmatrix}, \quad (2.80)$$

with  $\boldsymbol{\beta}_j^* = (\boldsymbol{\beta}_{j1}^T, \beta_{j21})^T$ ,  $\mathbf{T}_j^* = (0, \dots, 0, x_{j,r+1} - \gamma_{j2}, \dots, x_{j,n_j} - \gamma_{j2})^T$  and design matrix for the  $j$ th panel given by

$$\mathbf{Z}_j = \begin{pmatrix} 1 & x_{j1} & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_{js} & 0 \\ 1 & \gamma_{j1} & x_{j,s+1} - \gamma_{j1} \\ \vdots & \vdots & \vdots \\ 1 & \gamma_{j1} & x_{jr} - \gamma_{j1} \\ 1 & \gamma_{j1} & \gamma_{j2} - \gamma_{j1} \\ \vdots & \vdots & \vdots \\ 1 & \gamma_{j1} & \gamma_{j2} - \gamma_{j1} \end{pmatrix}.$$

In contrast to model (2.75), the design matrix for the entire panel data set is no longer block diagonal. Moreover, the estimation of  $\delta$  depends on the observations of the last segment of all panels. Furthermore, the observations belonging to the last segment are determined by the location of the right-most change point. Consequently, the entire panel data has to be analyzed to estimate the parameter of interest.

As the log likelihood of panel  $j$  is

$$\begin{aligned} \ell(\boldsymbol{\beta}_j^*, \delta, \phi | \boldsymbol{\gamma}_j, \mathbf{y}_j) &= \ell(\boldsymbol{\beta}_{j1}, \phi | \boldsymbol{\gamma}_j, y_{j1}, \dots, y_{js}) \\ &+ \ell(\boldsymbol{\beta}_j^*, \phi | \boldsymbol{\gamma}_j, y_{j,s+1}, \dots, y_{jr}) \\ &+ \ell(\boldsymbol{\beta}_j^*, \delta, \phi | \boldsymbol{\gamma}_j, y_{j,r+1}, \dots, y_{jn}) , \end{aligned}$$

the log likelihood of the global model (2.80) is

$$\ell(\boldsymbol{\beta}, \delta, \phi | \boldsymbol{\gamma}, \mathbf{y}) = \sum_{j=1}^n \ell(\boldsymbol{\beta}_j^*, \delta, \phi | \boldsymbol{\gamma}_j, \mathbf{y}_j) .$$

Due to the common slope  $\delta$ , the terms on the right hand side are no longer autonomous. Let  $A_j = \{i | x_{ji} > \gamma_{j2}\}$  denote the set of indices of  $x_{ji}$  which corresponds to the last segment of the  $j$ th panel. Then the parameter score function with respect to the common slope  $\delta$  is

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}, \delta, \phi | \boldsymbol{\gamma}, \mathbf{y})}{\partial \delta} &= \sum_{j=1}^n \frac{\partial \ell(\boldsymbol{\beta}_j^*, \delta, \phi | \boldsymbol{\gamma}_j, \mathbf{y}_j)}{\partial \delta} \\ &= \sum_{j=1}^n \sum_{i \in A_j} \frac{y_{ji} - \mu_{ji}}{\phi V(\mu_{ji})} \frac{\partial \mu_{ji}}{\partial \eta_{ji}} x_{ji} \end{aligned} \quad (2.81)$$

with

$$\eta_{ji} = \beta_{j30} + \delta x_{ji} , \quad \gamma_{j2} < x_{ji} \leq x_{jn_j} .$$

This score (2.81) contains parameters of each panel. Hence, the estimation of  $\delta$  depends on several observations of all panels. Thus, the panels can not be analyzed separately. In contrast, the global model (2.80) has to be analyzed to estimate the parameters of interest.

Next, consider the case of unknown change points. There are two ways to estimate these change points. We either apply a grid search over all feasible combinations of change points or an exact method. For the grid search model, (2.80) is used. In contrast, for the exact method, the global model for the panel data set is

$$\begin{pmatrix} g(\boldsymbol{\mu}_1) \\ g(\boldsymbol{\mu}_2) \\ \vdots \\ g(\boldsymbol{\mu}_n) \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{T}_1 \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} & \mathbf{T}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_n & \mathbf{T}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_n \\ \delta \end{pmatrix} , \quad (2.82)$$

with  $\mathbf{T}_j = (0, \dots, 0, x_{j,r+1}, \dots, x_{jn_j})^T$  and design matrix for panel  $j$  given by

$$\mathbf{X}_j = \begin{pmatrix} 1 & x_{j1} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{js} & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{j,s+1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{jr} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Recall, that the exact method consists of two steps. We first fit a GLM to each segment separately, calculate the solutions of the change points, and investigate if these solutions are the ML estimates of the change points. Second, if this is not the case, the boundaries of each rectangle have to be considered (see Subsection 2.3.2). In the case of independent observations (Subsection 2.4.1), this can be done for each panel separately. To emphasize that this is not possible in the case of a common slope for the last segment, consider the log likelihood of the  $j$ th panel, namely

$$\begin{aligned} \ell(\boldsymbol{\beta}_j, \delta, \phi, \boldsymbol{\gamma}_j | \mathbf{y}_j) &= \ell(\boldsymbol{\beta}_{j1}, \phi, \boldsymbol{\gamma}_j | y_{j1}, \dots, y_{js}) \\ &+ \ell(\boldsymbol{\beta}_{j2}, \phi, \boldsymbol{\gamma}_j | y_{j,s+1}, \dots, y_{jr}) \\ &+ \ell(\boldsymbol{\beta}_{j30}, \delta, \phi, \boldsymbol{\gamma}_j | y_{j,r+1}, \dots, y_{jn_j}). \end{aligned}$$

The log likelihood for the global model is then given by

$$\ell(\boldsymbol{\beta}, \delta, \phi, \boldsymbol{\gamma} | \mathbf{y}) = \sum_{j=1}^n \ell(\boldsymbol{\beta}_j, \delta, \phi, \boldsymbol{\gamma}_j | \mathbf{y}_j).$$

Due to the same reasons as in the case of known change points, the terms on the right hand side are no longer autonomous. Consequently, the entire panel data has to be analyzed simultaneously.

Regardless of using either a grid search or the exact method, the computational complexity for fitting this global model increases rapidly with the number of panels and the number of change points within each panel. Thus, analyzing the entire panel data simultaneously is not feasible in rational time and an appropriate algorithm has to be developed.

### 2.4.3 Fitting Procedure

In this section, a new algorithm for fitting such kind of models is introduced. As mentioned above, applying either a grid search or an exact method to estimate the unknown change points and the parameters of interest is not feasible in rational time. This is due to the fact that a common slope for the last segments

for all panels induces dependency of the estimation of the change points. Thus, the idea of this new algorithm is to divide the fitting procedure into two steps. In the first step we estimate the common slope for the last segments. In the second step, given this estimate of the common slope we estimate the change points. The advantage of this separation is, that for a given last slope  $\delta$ , the design matrix of the model (2.75) and (2.77) respectively, can be divided into two terms. That is, the linear terms corresponding to the common slope can be interpreted as an offset in these models. Thus, for given pairs of change points, model (2.75) can be written as

$$\begin{pmatrix} g(\boldsymbol{\mu}_1) \\ g(\boldsymbol{\mu}_2) \\ \vdots \\ g(\boldsymbol{\mu}_n) \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \vdots \\ \boldsymbol{\beta}_n^* \end{pmatrix} + \delta \begin{pmatrix} \mathbf{T}_1^* \\ \mathbf{T}_2^* \\ \vdots \\ \mathbf{T}_n^* \end{pmatrix}, \quad (2.83)$$

and for unknown change points model (2.77) can be written as

$$\begin{pmatrix} g(\boldsymbol{\mu}_1) \\ g(\boldsymbol{\mu}_2) \\ \vdots \\ g(\boldsymbol{\mu}_n) \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix} + \delta \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \vdots \\ \mathbf{T}_n \end{pmatrix}. \quad (2.84)$$

The consequence is that the remaining design matrices are block diagonal. Therefore, each panel can be analyzed separately. Regardless which method is used to estimate the change points and the parameters of interest, the model for panel  $j$  is either

$$g(\boldsymbol{\mu}_j) = \mathbf{Z}_j \boldsymbol{\beta}_j^* + \delta \mathbf{T}_j^* \quad (2.85)$$

using a grid search, or

$$g(\boldsymbol{\mu}_j) = \mathbf{X}_j \boldsymbol{\beta}_j + \delta \mathbf{T}_j \quad (2.86)$$

using the exact method. Moreover, the model for the  $j$ th panel is autonomous of the remaining  $j - 1$  panels. Thus, the parameter  $\boldsymbol{\beta}_j$  of each panel can be estimated separately.

Therefore, the two tasks of this algorithm are:

1. Estimate a common slope  $\delta$  for the last (right most) segments for all panels.
2. Given an estimate of  $\delta$ , we can estimate a GLM with continuous change points for each panel separately using (2.85) and (2.86), respectively.

For the first task it is necessary to know which observations belong to the last segment of each panel. This is determined by the location of the last change points usually unknown at the beginning of the algorithm. Hence, an appropriate method has to be developed to obtain a common slope, though no information about the rightmost change points are available. There are three

different methods considered here. As usually these methods provides only a first approximation of the last change points and a common slope, respectively, these two tasks are iterated.

The first method is to fit a GLM to the entire panel data. This GLM takes into account panel specific intercepts  $\alpha_j$  and one common slope  $\delta$  for the last segment of each panel. Hence, for the estimation of  $\delta$  only the observations of the last segment of each panel should be used. Usually at the beginning of the algorithm, that is, the first iteration, no information about which observations belong to the last segment is available. Thus, at the beginning a GLM based only on the last  $n_j - p$  observations of each panel is considered. Of course this is not a satisfactory situation. The idea is to augment these observations step by step with one additional observation until the right most change point of each panel is reached. As no information about the right most change point is available, we used the method of Brown to detect the rightmost continuous change point for each panel. In contrast to the description of this method in Subsection 2.1.4, we start the analysis on the right side of the data. That is, we first consider the last  $n_j - p$  observations of each panel. Let us denote the index of the left most observation of the last segment as  $a_j$ . Then, at the beginning of this method  $a_j = n_j - p$  and the GLM is stated as

$$\begin{pmatrix} g(\mu_{1a_1}) \\ \vdots \\ g(\mu_{1n_1}) \\ g(\mu_{2a_2}) \\ \vdots \\ g(\mu_{2n_2}) \\ \vdots \\ g(\mu_{na_n}) \\ \vdots \\ g(\mu_{nn_n}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & x_{1a_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & x_{1n_1} \\ 0 & 1 & \cdots & 0 & x_{2a_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & x_{2n_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x_{na_n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x_{nn_n} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \\ \delta \end{pmatrix}.$$

After fitting this model, recursive residuals are calculated for the  $a_j - 1$  observation of each panel. The method of Brown is applied to decide whether a change point between the  $(a_j - 1)$ th and  $a_j$ th observation is detected or not. If a change point is detected, then for the remaining iterations of this method the number of observations stays constant for this panel. Denote the index before a change point is detected as  $\tilde{a}_j$ . Then for the remaining iterations only the observations  $\tilde{a}_j, \dots, n_j$  are considered. For all panels, with no change point detected,  $a_j = a_j - 1$  for the next iteration of this method. Thus these segments are augmented with the  $(a_j - 1)$ th observation for the next iteration. This is repeated until a change point is detected for all panels. Consequently, the indices of observations which belong to the last segments are  $\tilde{a}_j, \dots, n_j$ ,

$j = 1, \dots, n$ , and we further concentrate on

$$\begin{pmatrix} g(\mu_{1\bar{a}_1}) \\ \vdots \\ g(\mu_{1n_1}) \\ g(\mu_{2\bar{a}_2}) \\ \vdots \\ g(\mu_{2n_2}) \\ \vdots \\ g(\mu_{n\bar{a}_n}) \\ \vdots \\ g(\mu_{nn_n}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & x_{1\bar{a}_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & x_{1n_1} \\ 0 & 1 & \cdots & 0 & x_{2\bar{a}_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & x_{2n_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x_{n\bar{a}_n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x_{nn_n} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \\ \delta \end{pmatrix}.$$

Fitting this model yields an estimate  $\hat{\delta}$  of the common slope for the last segments. In what follows, this method is denoted as *glm* method.

Consider the following example to illustrate this *glm* method in detail. Panel data with three panels are used. These panels are colored green, magenta, and blue in Figure 2.25. In the first step, a GLM based on the last three observations of each panel is considered, i.e.  $a_j = n_j - 2$ . These observations are shown as  $\times$ , and the corresponding GLM is plotted as three parallel lines in Figure 2.25 (a). Next, the recursive residuals for the  $a_j - 1$  observations are calculated in order to decide whether a change point between the  $(a_j - 1)$ th and  $a_j$ th observation is detected or not. We assume that this is the case for the green model which is illustrated as a vertical red line in Figure 2.25 (b). Thus, in the remaining iterations only the last three observations of the green model are used. In contrast, the last segment of the magenta and blue model is augmented with the  $(a_j - 1)$ th observation (see Figure 2.25 (c)). At this iteration, it is assumed that a change point in the blue model is detected. Consequently, for the remaining iterations, only the last four observations of the blue model are used. Finally, the last segment of the magenta model is augmented step by step with one observation until a change point between the  $(n_3 - 6)$ th and  $(n_3 - 5)$ th observation is detected. Therefore, the last segment consists of the last three observations of the green model, the last six observations of the magenta model, and the last four observations of the blue model. The result is a common slope for the last segment of these three models.

The second and third method to estimate a common slope is based on an individual change point model for each panel. This involves first estimating a continuous change point model without any restriction for the slope of the last segment to each panel separately using (2.74) or (2.76). Then, a rather informal estimate of the common slope is given by an appropriate statistic of the estimated slopes  $\hat{\beta}_{j31}$ . In this work, the mean and the median of these

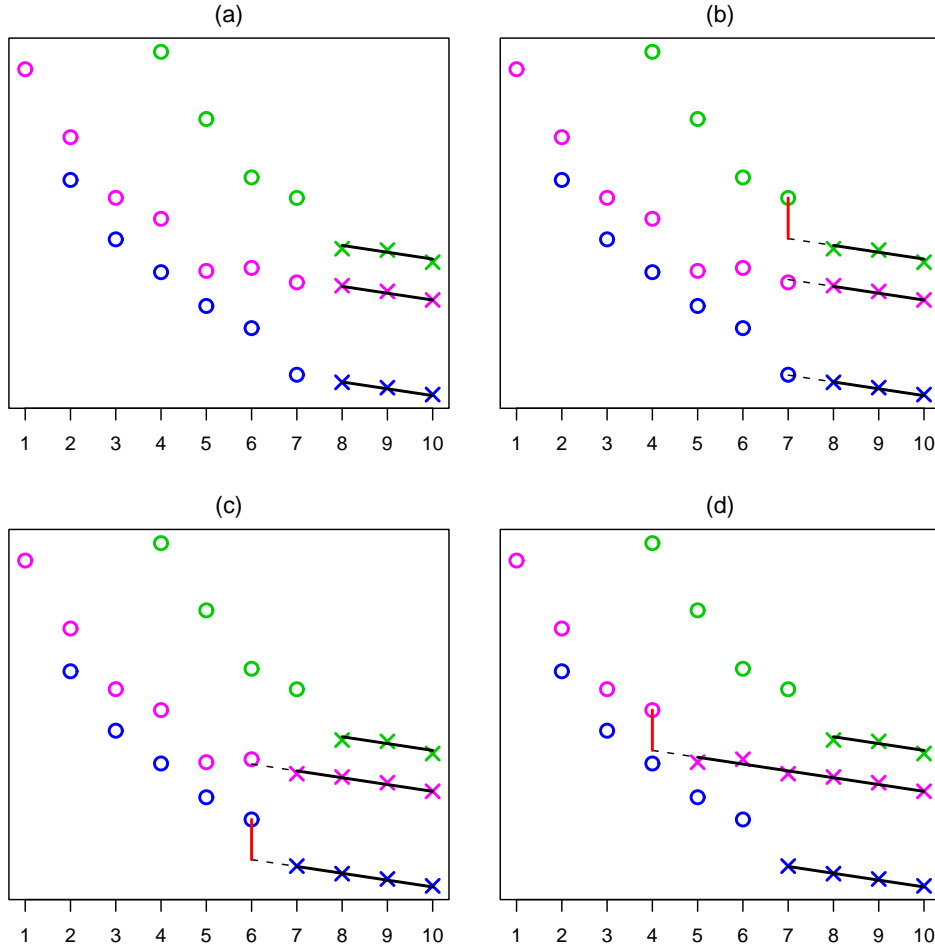


Figure 2.25: Illustration of the *glm* method to estimate a common slope for the last segments.

estimated slopes are considered, i.e.

$$\hat{\delta} = \frac{1}{n} \sum_{j=1}^n \hat{\beta}_{j31} \quad \text{and}$$

$$\hat{\delta} = \text{median} \left( \hat{\beta}_{131}, \dots, \hat{\beta}_{n31} \right).$$

In the remainder of this section, these two methods are denoted as the *mean* and *median* method, respectively.

Given an estimate  $\hat{\delta}$  of the common slope, the second task of the algorithm can be applied. This involves fitting a GLM with continuous change points to each panel separately but with the slope in the last segment as a given quantity  $\hat{\delta}$ . This can be done by applying either a grid search or the exact method (see Subsection 2.3.2) on model (2.85) and (2.86), respectively. If  $\hat{\delta}$  is close to the true value we stop. However, these two tasks should be iterated

until the algorithm has converged. Three different criteria to decide whether the algorithm is converged or nor converged are described later.

For the second iteration of this algorithm, again the first task is utilized but now estimates of the last change points exist. These determine which observations belong to the last segments and the common slope for these last segments can be easily estimated. The three methods from above are considered. The first one again uses a GLM with panel specific intercepts but a common slope for all panels. In contrast to the beginning of the algorithm, estimates of the last change points are available. Thus, the estimated slope of this special GLM provides an estimate of  $\hat{\delta}$ . For the second and third method we fit a GLM with continuous change points to each panel separately, but with a given last change point and unrestricted slope for the last segment. The estimate for the common slope is then an appropriate statistic of the  $n$  estimates  $\hat{\beta}_{j31}$ . Given this new  $\hat{\delta}$ , the second task is applied.

As we are interested in maximizing the log likelihood, an obvious stopping rule might be the difference between the values of the maximized log likelihoods of the fitted model of two consecutive iterations. If this difference is smaller than a given quantity  $\xi$ , the algorithm is said to be converged. The other two stopping rules are based on the parameter estimates. Thus, the algorithm can be ruled as converged if either the relative difference between the common slope of two consecutive iterations,  $t$  and  $t + 1$ , is less than a given constant  $\xi$ , i.e.

$$\left| \frac{\hat{\delta}^{(t)} - \hat{\delta}^{(t+1)}}{\hat{\delta}^{(t+1)}} \right| < \xi$$

or the maximum of the relative difference of the last change points is less than a given constant  $\xi$ , i.e.

$$\max_j \left| \frac{\hat{\gamma}_{j2}^{(t)} - \hat{\gamma}_{j2}^{(t+1)}}{\hat{\gamma}_{j2}^{(t+1)}} \right| < \xi. \quad (2.87)$$

Briefly, this new algorithm consists of the following steps.

1. Calculate a starting value  $\hat{\delta}^{(0)}$  for the common slope of the last segments using either the *glm*, *mean*, or *median* method.
2. Given an estimate of the common slope  $\delta$ , fit a GLM with continuous change points and fixed slope for the last segment for each panel separately.
3. Given the estimated last change point of each panel from step 2, estimate the common slope  $\hat{\delta}$  again using either the *glm*, *mean*, or *median* method.
4. Given  $\hat{\delta}$  from step 3, fit a GLM with continuous change points and fixed slope for the last segment for each panel separately.
5. If the algorithm has converged then stop, otherwise, repeat from step 3 above.



We investigate the performance of this new algorithm in the following MC Simulation.

### 2.4.4 Simulation Study

The following MC simulation is performed to answer two questions about the performance of this new algorithm.

- Does this new algorithm provide appropriate estimates of the change points as well as the common slope of the rightmost segments?
- Which method for estimating the common slope in the first task should be preferred, i.e. the *glm*, *mean*, or *median* method?

The considered panel data consists of three different panels with a common slope  $\delta = -0.1$  for the last segment of all three panels. For each panel, a GLM with one continuous change point is considered. The response variable of all six segments follows a Poisson distribution with the canonical link function,  $g(\mu) = \log(\mu)$ , for its mean  $\mu$ . Furthermore, the parameters of the left segments are different as well as the domain of the explanatory variable. Moreover, the location of the continuous change points are different for each panel as well.

To account for such a setup let the explanatory variable of the first model  $x_i = i/2$ ,  $i = 0, \dots, 40$ , with change point  $\gamma_1 = 7.75$ , i.e.

$$g(\mu_i) = \begin{cases} 6.55 - 0.3x_i & 0 \leq x_i \leq 7.75 \\ 5 - 0.1x_i & 7.75 < x_i \leq 20. \end{cases}$$

The explanatory variable of the second model is  $x_i = 2i/5$ ,  $i = 0, \dots, 30$ , with change point  $\gamma_2 = 8.2$  such that

$$g(\mu_i) = \begin{cases} 6.23 - 0.25x_i & 0 \leq x_i \leq 8.2 \\ 5 - 0.1x_i & 8.2 < x_i \leq 12. \end{cases}$$

Finally, for the third model the explanatory variable is  $x_i = i/2$ ,  $i = 0, \dots, 44$ , with change point  $\gamma_3 = 5.75$ ,

$$g(\mu_i) = \begin{cases} 6.4375 - 0.35x_i & 0 \leq x_i \leq 5.75 \\ 5 - 0.1x_i & 5.75 < x_i \leq 22. \end{cases}$$

Here, without loss of generality, the intercept of the last segment is the same for all panels.

For the following MC simulation, 1000 panel data sets were generated and the new algorithm was applied to all these replications to estimate the common slope of the last segments and the continuous change points of all three panels. To compare the performance of the three different methods, *glm*, *mean*, or *median*, of course the same 1000 replications were used. As stopping rule, the difference of the maximized log likelihood of two consecutive iterations with  $\xi = 10^{-5}$  is used.

The main result of this MC simulation is that this new algorithm worked well. The common slope as well as the continuous change point, both were estimated with negligible biases. The performance of all three methods are nearly the same, whereas the *mean* method yields the smallest bias and the estimates based on the *glm* have the smallest variance.

In particular, all three methods estimate the common slope  $\delta$  with a negligible bias and small standard deviation (Table 2.3). Frequency histogram of the 1000  $\hat{\delta}$  estimates are plotted in Figure 2.26, where the smallest variability of these estimates can be observed for the *glm* method and the largest one for the *mean* method.

parameter	method	mean	median	std.dev.	true.value
$\hat{\delta}$	mean	-0.101	-0.100	0.009	-0.10
$\hat{\delta}$	median	-0.100	-0.100	0.006	-0.10
$\hat{\delta}$	glm	-0.100	-0.100	0.005	-0.10
$\hat{\gamma}_1$	mean	7.749	7.757	0.363	7.75
$\hat{\gamma}_1$	median	7.731	7.734	0.319	7.75
$\hat{\gamma}_1$	glm	7.726	7.730	0.293	7.75
$\hat{\gamma}_2$	mean	8.200	8.186	0.445	8.20
$\hat{\gamma}_2$	median	8.217	8.193	0.435	8.20
$\hat{\gamma}_2$	glm	8.220	8.192	0.429	8.20
$\hat{\gamma}_3$	mean	5.741	5.711	0.329	5.75
$\hat{\gamma}_3$	median	5.723	5.703	0.273	5.75
$\hat{\gamma}_3$	glm	5.710	5.710	0.250	5.75

Table 2.3: Comparison of the mean, median and standard deviation of the last slope estimates and last change point estimates, with respect to the three different methods.

Comparing the estimates of the change points,  $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$  establish that the *mean* method yields the smallest bias, whereas the largest bias can be observed when applying the *glm* method (Table 2.3). In contrast, for the variance of the estimates it is the other way around. The smallest variance of the estimates is obtained through the *glm* method, and the largest variance is observed for the *mean* method. The frequency histograms of the change point estimates with respect to the three different methods are plotted in Figure 2.27 for the first panel, in Figure 2.28 for the second panel, and in Figure 2.29 for the third panel.

Furthermore, we compare the number of iterations until the algorithm converged. The *glm* method converges very quickly and needs, at most, seven iterations, where in more than 55% the algorithm converges after four iteration (Figure 2.30). The *mean* method and the *median* method needs up to 25 and 28 iterations until convergence, respectively. To compare these two methods, it seems that the *mean* performs slightly better than the *median* method.

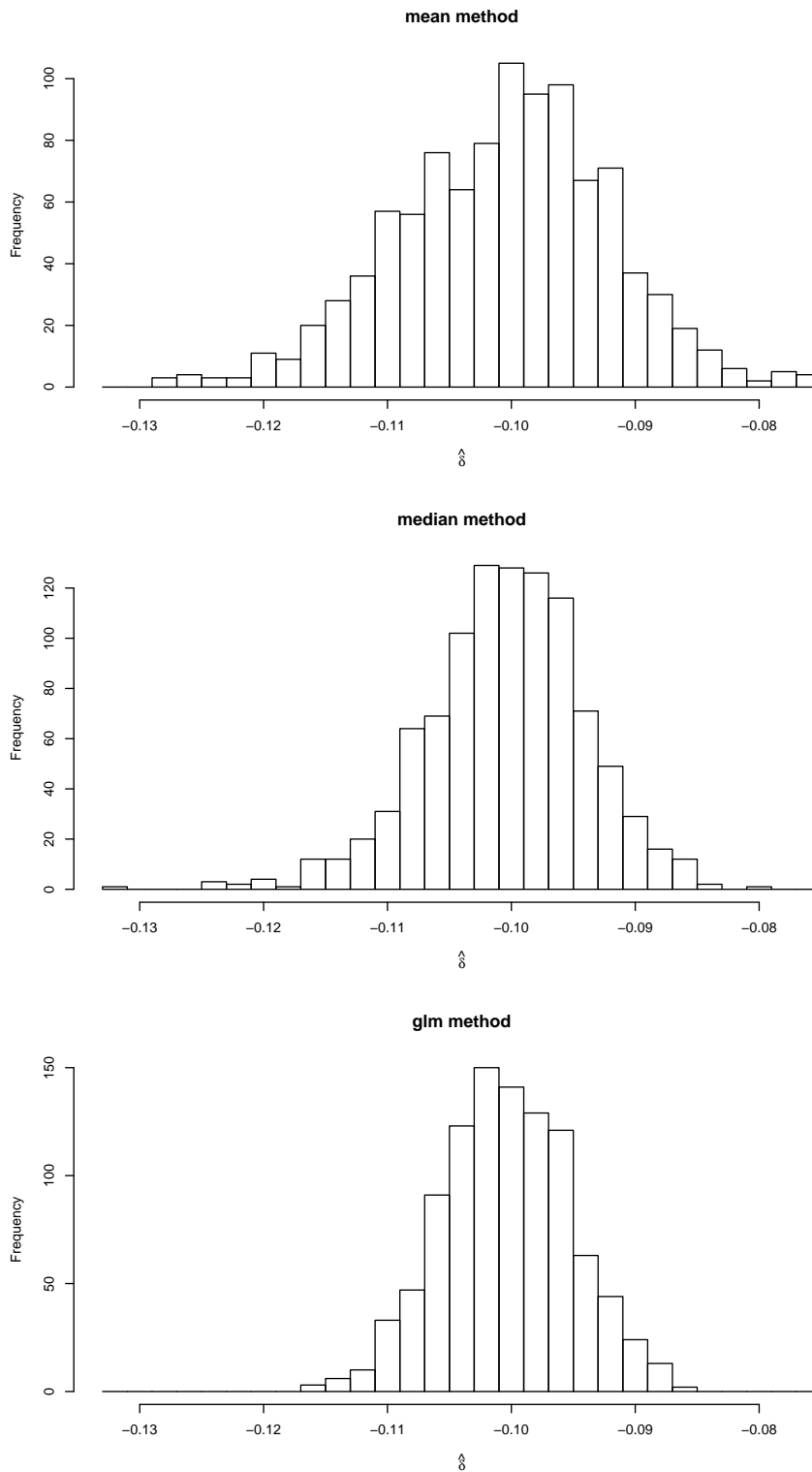


Figure 2.26: Comparison of the MC frequencies of the common last slope estimates with respect to the three different methods.

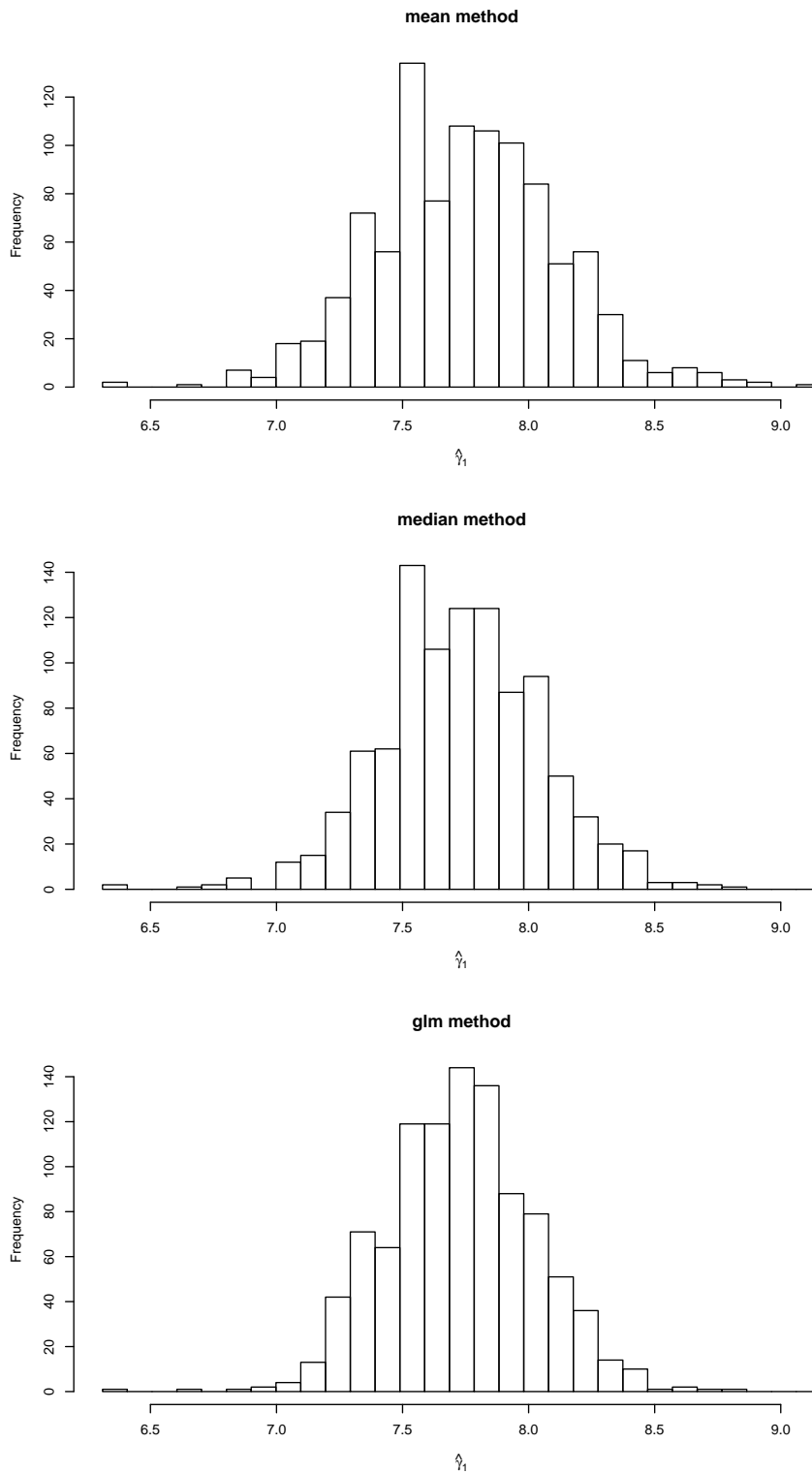


Figure 2.27: Comparison of the MC frequencies of the change point estimates of the first model, with respect to the three different methods.

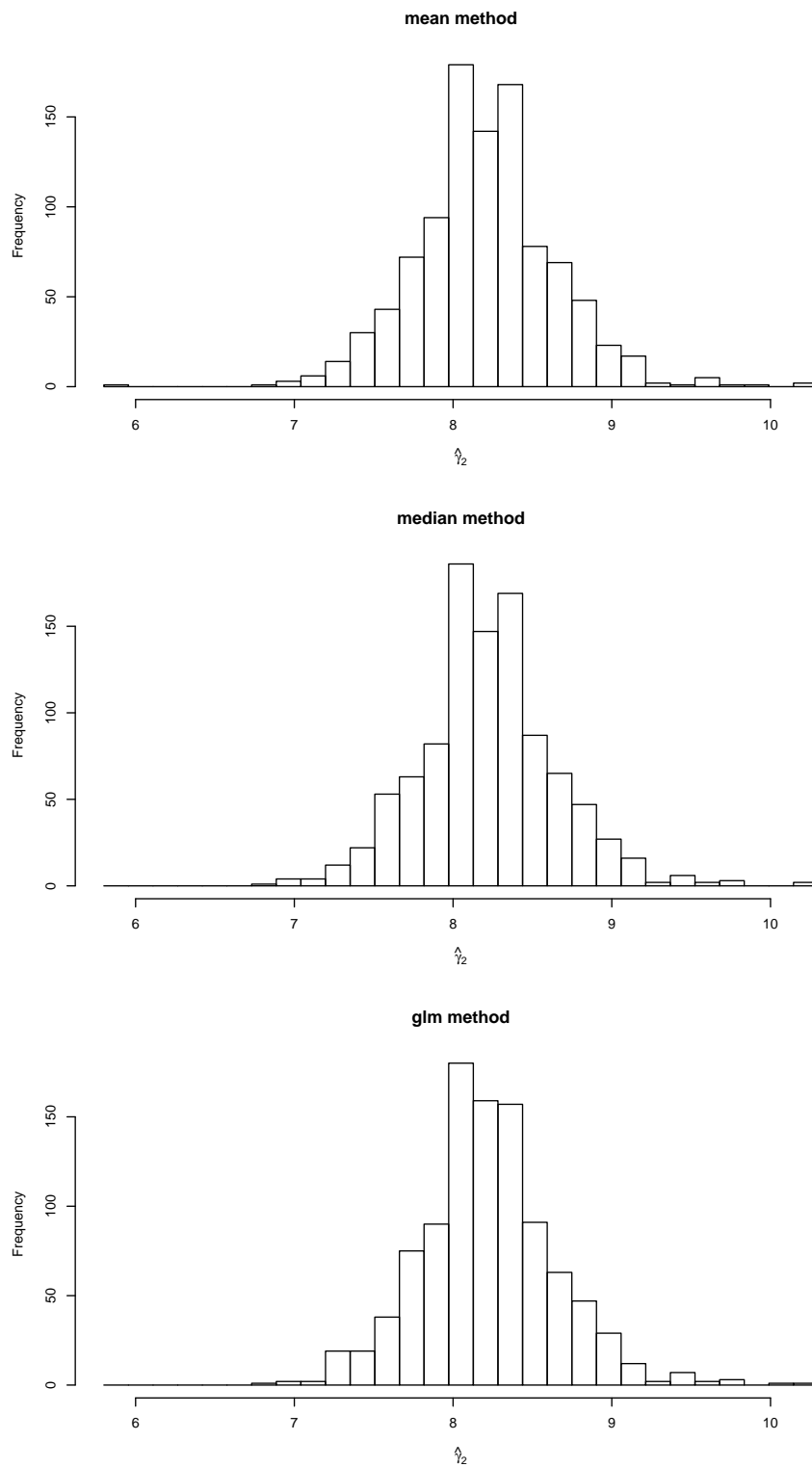


Figure 2.28: Comparison of the MC frequencies of the change point estimates of the second model, with respect to the three different methods.

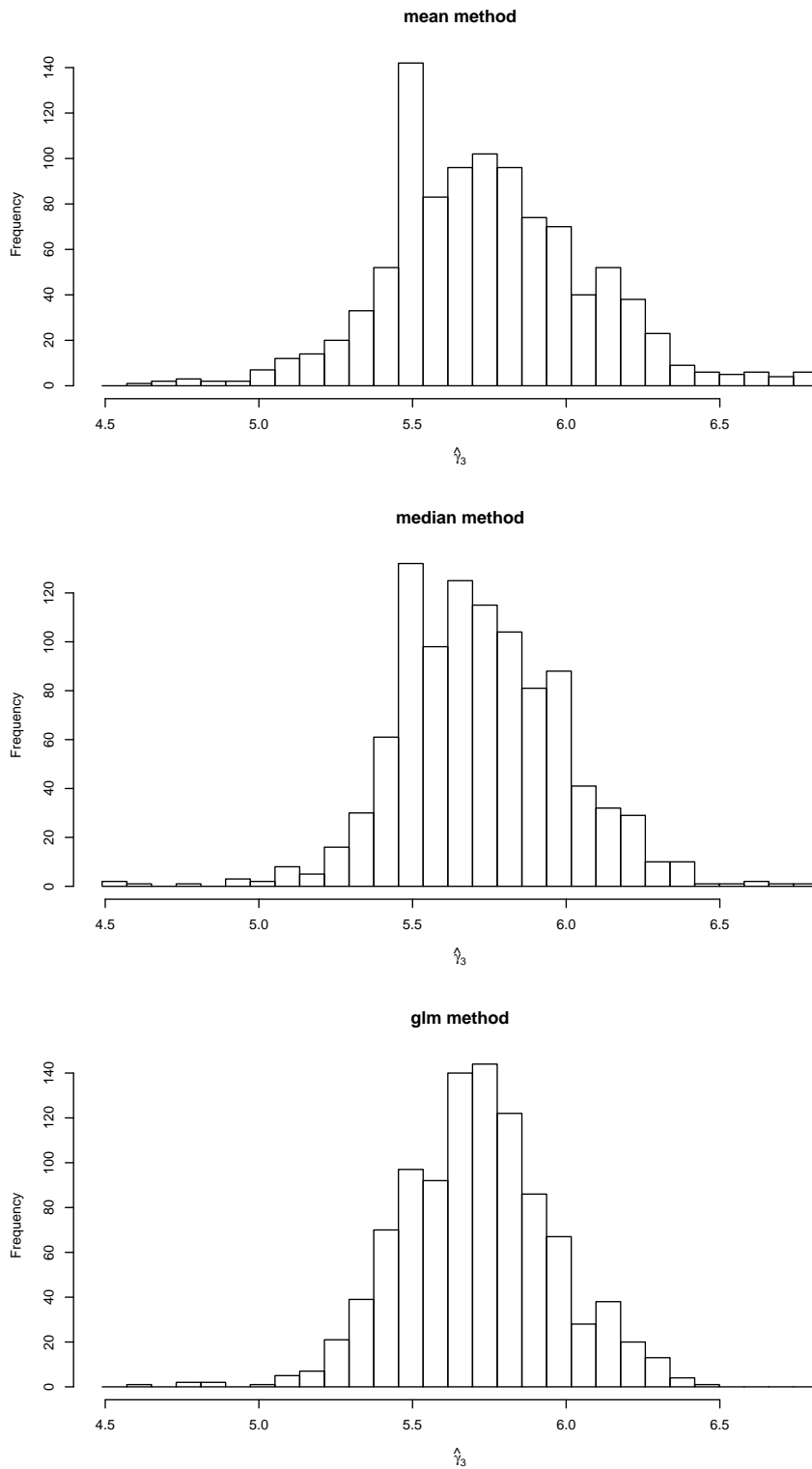


Figure 2.29: Comparison of the MC frequencies of the change point estimates of the third model, with respect to the three different methods.

The cost for this speed of convergence of the *glm* method is the larger bias compared with the other two methods.

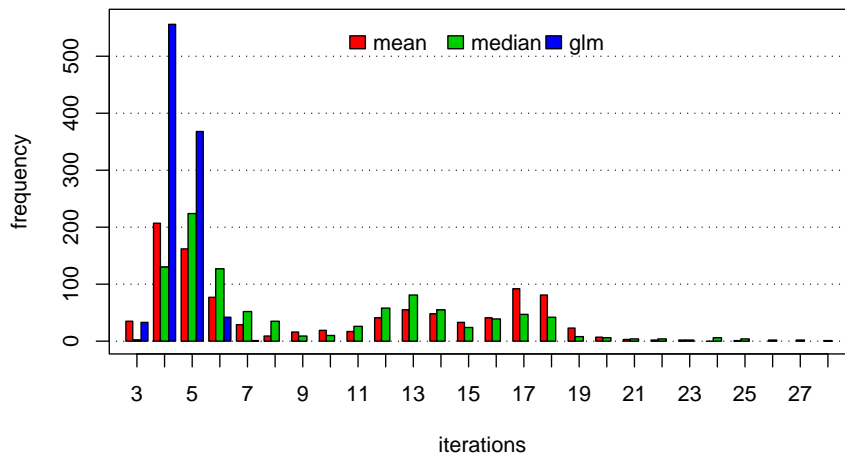


Figure 2.30: Frequencies of the number of iterations until the algorithm converged, with respect to the three considered methods.

Summing up these results, this new method can be applied to estimate the parameters of such kind of models for panel data. In the case where the accuracy of the estimates is important, the *mean* method should be preferred to the other ones. For large data sets and when computational time plays an important role the *glm* method should be used.

### Example 2.6 Panel data Set

We use one of the simulated data generated in the MC simulation before. This panel data is plotted in Figure 2.31 where the first and third panel is shifted on the *x*-axis such that the true change points of all panels are the same, i.e.  $x = 8.2$ .

To fit such a model to panel data, the function `ccpComSlope` is used. The call of this function is very similar to the call of the well known `glm` function. The first argument is a formula describing the relationship between the response and the explanatory variable. The arguments `data` and `family` have the same meaning as those in the function `glm`. In addition, there are some more arguments, i.e. `iPanel` gives the index of the column which determines the panel, `comSlopeMethod` determines which method should be used to estimate the common slope, and `nOfCp` fixes the number of change points for the model in each panel. There are some more arguments, mainly to control the algorithm. These are described in detail in the help of the package `CpInGLM` (see Appendix B).

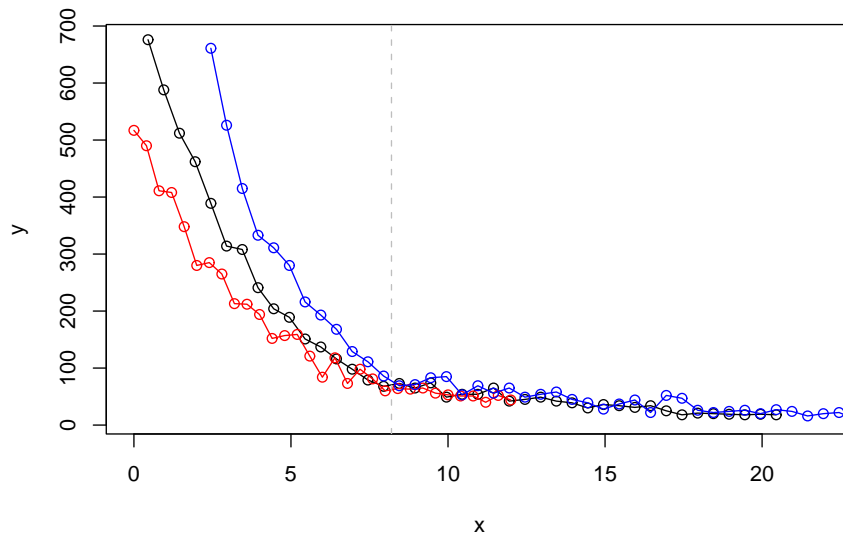


Figure 2.31: Simulated panel data, where the explanatory variables of each panel are centered at  $x = 8.2$ , the change point of the second panel.

The data set has to be loaded into the workspace. As this data set is already included in the package `CpInGLM`, this can be done using the function `data`. Next, the function `ccpComSlope` is applied to this panel data set where the median method is used to estimate the common slope for the last segment, i.e. `comSlopeMethod='median'`. The output of this function is stored in the object `m1` which is of the new class `ccpComSlope`. For this class, the generic functions `print` and `plot` exist. The `print` function gives a short overview about the fitted model, summarizing the call of the function, the estimated common slope and last change points of each panel, and indicating if the algorithm has converged or not.

```
> data(panel.dat)
>
> m1 <- ccpComSlope(y~x, data=panel.dat, iPanel=3,
+                 family=poisson(link=log), nOfCp=1,
+                 comSlopeMethod="median", verbose=FALSE)
> m1
```

Call:

```
ccpComSlope(formula = y ~ x, data = panel.dat, iPanel = 3,
            family = poisson(link = log), nOfCp = 1,
            comSlopeMethod = "median", verbose = FALSE)
```

Common slope for the last segment: -0.1036



Last change point for each panel at:  
7.655 8.000 5.777

Algorithm converged after: 7 iterations

*The algorithm converged after the 7th iteration. The evolution of the estimated common slope, the last change points and the index of the explanatory variable after which the last change point was detected can be generated using the `plot` function and are plotted in Figure 2.32.*

```
> plot(m1)
```

*The estimated common slope is  $\hat{\delta} = -0.1036$ , which is very close to the true value of  $-0.1$ . The estimated change points for the three panels are  $\hat{\gamma}_1 = 7.655$ ,  $\hat{\gamma}_2 = 8$  and  $\hat{\gamma}_3 = 5.777$ . The fitted model for all three panels is plotted in Figure 2.33 and shows clearly, that a common slope for this model is appropriate. To enable some diagnostic tools for the fitted model of each panel, one element of `m1` is a list of these fitted models, named `ccpList`. Each element of this list has class `ccp`, thus the generic function for this class, mentioned above, can be applied. For instance, to take a closer look at the model of the first panel, one can use*

```
> plot(m1$ccpList[[1]])
```

*to obtain some diagnostic plots for the fitted continuous change point model of the first panel (Figure 2.34).*

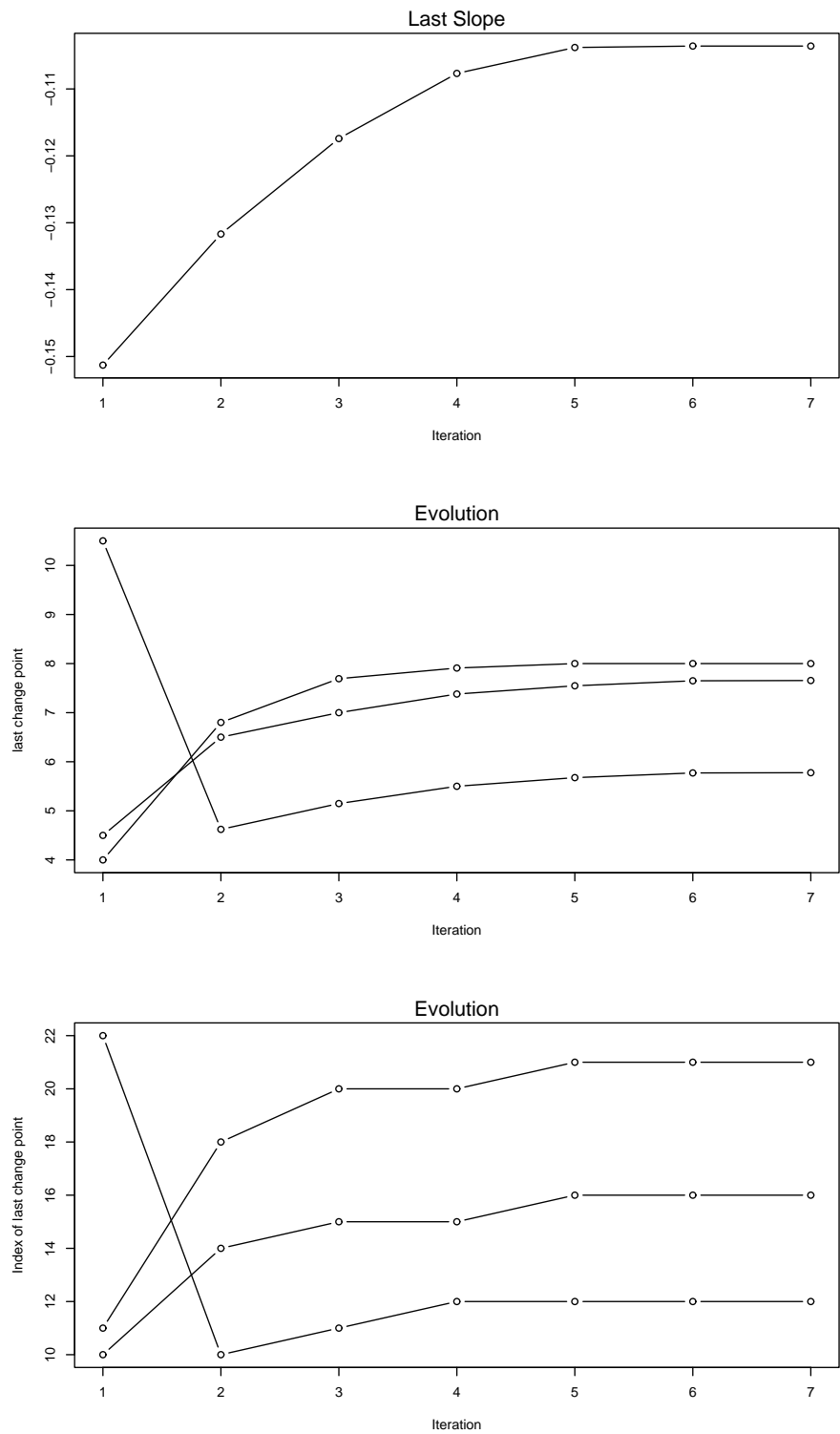


Figure 2.32: Evolution of the estimated last slope, last change points and index of the last change points.

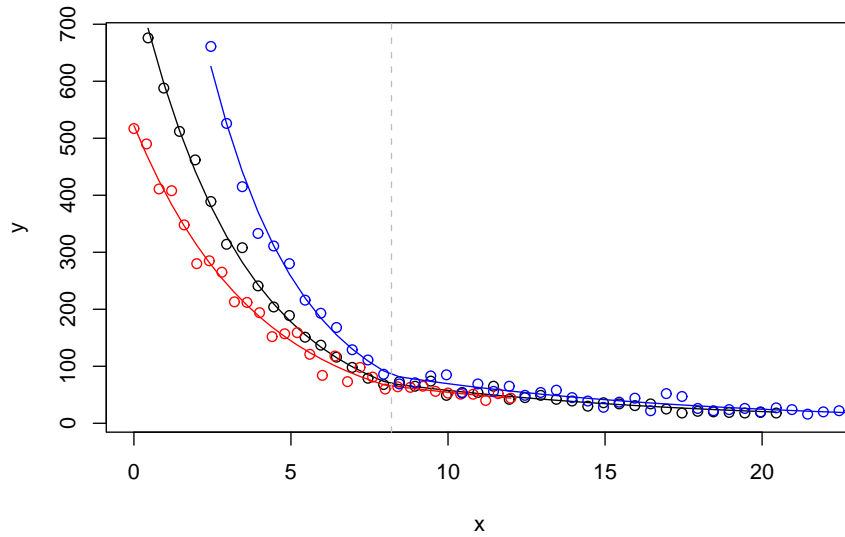


Figure 2.33: Fitted model for the panel data set, with a common slope for the last segment.

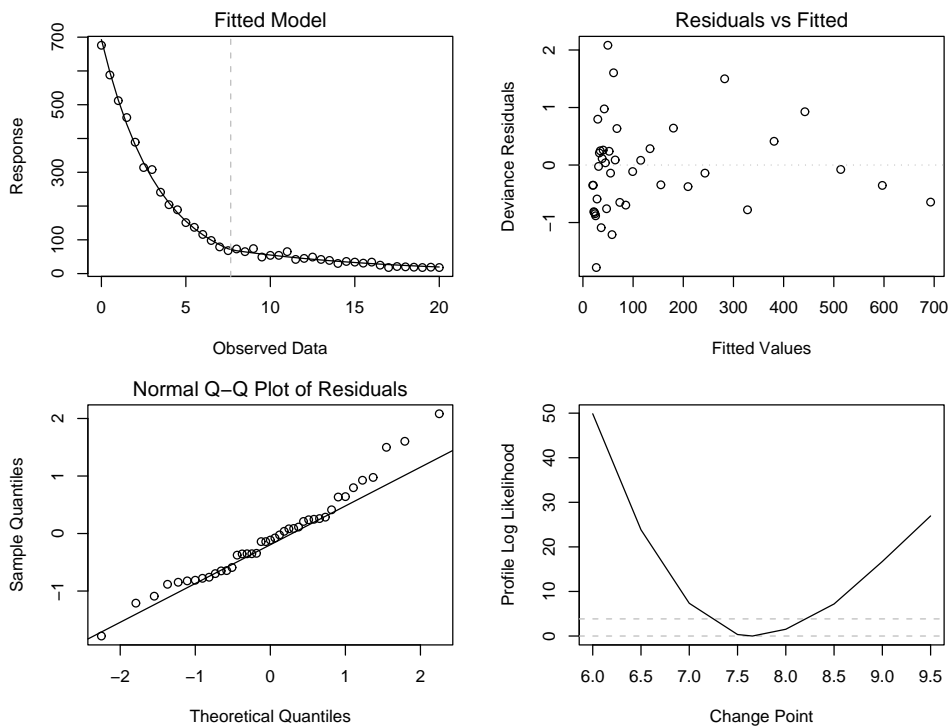


Figure 2.34: Some diagnostic plots for the fitted model of the first panel.

# Chapter 3

## Application

In this chapter the methods discussed in Chapter 2 to estimate multiple change points are applied to hydrological data. These hydrological data describes the discharge of a creek or a river. The main interest is to detect change points in the hydrograph of the discharge under certain circumstances and to estimate the parameters of the model of the obtained segments. This is done to get some important hydrological information about the creek. In particular, the groundwater storage of the catchment of this creek is one information, which is an important quantity for water resources management.

### 3.1 Hydrological Runoff Model

As mentioned above, one way to obtain information about the groundwater storage of a catchment is to analyze a model of the discharge of the creek of this catchment. The catchment of a creek is defined as the area of the landscape, where all rain falling on the area recharges into this creek. The plot of the discharge is called *hydrograph*. Usually the discharge is measured by an automatic gauging station, recording the discharge in  $m^3/sec$  at its time  $t$ . At modern gauging stations the recording interval is about 5 to 15 minutes. A central interest in the analysis of the discharge is the recession curve. The recession curve is defined as the discharge after a rainstorm. It starts at the peak of the hydrograph and ends at the next marked increase of the hydrograph. Thus, the recession curve decreases between the starting and end point, provided that there is no marked precipitation in between. In other words, the recession curve starts at the end of a rainstorm and ends at the following marked precipitation.

The discharge of a recession curve has three types of sources, the direct flow, the surface flow, and the base flow (see Figure 3.1). To observe all these three types of flows, the recession curve has to hold on for several days. The number of these days depends mainly on the properties of the catchment. The direct flow and the surface flow are composed by the water falling directly into the creek or recharging the creek through the hill slope. The base flow is the outflow of the groundwater storage. Thus when the recharge ceases, the re-

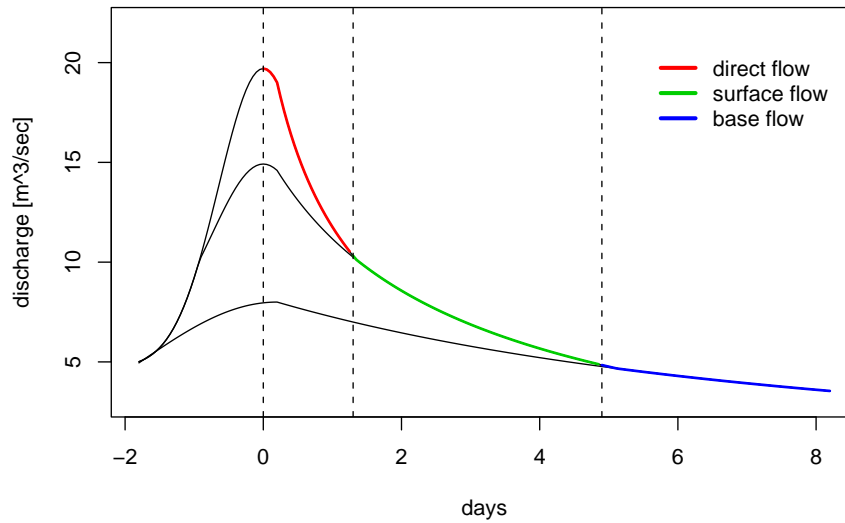


Figure 3.1: Direct, surface and base flow of the theoretical runoff model

cession curve reflects only the base flow. Moreover, it reflects the relationship between the groundwater storage and the base flow of the catchment, provided that outer influences like surface storage, groundwater abstraction or evapotranspiration are negligible. The time points, when the direct flow and then the surface flow ends can be recognized at two, more or less marked, changes in the recession curves. Thus to get information about the groundwater storage from the knowledge of the base flow, the base flow has to be separated from the total flow. This separation consists of two tasks. First, the time when the recharge ceases have to be determined. This is equivalent with determining the second change in the recession curve. Second, a theoretical model for the relationship between the groundwater storage and the base flow has to be supposed. For this relationship several theoretical models were considered in the last century. A very popular model is the model from Maillet (1905), which considered a linear relationship. Though it is very old it is still widely used nowadays. But a number of numerical analyses of flow recession curves in the last decade reveals a nonlinear relationship between base flow  $Q$  and groundwater storage  $S$  for which the equation

$$S = aQ^b \quad \text{with} \quad 0 < b \leq 1 \quad \text{and} \quad 0 < a \quad (3.1)$$

was adopted (Wittenberg, 1999). The coefficient  $a$  is related to catchment properties, mainly the area and the shape of the basin, pore volume and transmissivity. The exponent  $b$  describes the relationship between base flow and groundwater storage. The relationship is linear for  $b = 1$  and quadratic if  $b = 1/2$ , which are two particular cases with theoretical hydrological back-

ground. Empirical studies (Wittenberg, 1999) showed that  $b$  is around  $1/2$ . Therefore in this thesis a nonlinear relationship (3.1) with  $b = 1/2$  will be considered. Moreover, it is assumed that there is no inflow in the hydrological system. Then the continuity equation of a groundwater reservoir is given by

$$\frac{dS}{dt} = -Q. \quad (3.2)$$

Applying the chain rule, i.e.

$$\frac{dS}{dt} = \frac{dS}{dQ} \frac{dQ}{dt} = -Q$$

and combining this with equation (3.1) yields

$$abQ^{b-1} \frac{dQ}{dt} = -Q,$$

respectively

$$\frac{dQ}{dt} = -\frac{1}{abQ^{b-2}},$$

which can be written as

$$abQ^{(b-2)} dQ = -dt.$$

Now the base flow  $Q_t$  at some time  $t$  can be derived by solving this differential equation as follows. First integrate the terms on both sides,

$$\int_{Q_0}^{Q_t} abQ^{(b-2)} dQ = - \int_0^t dt,$$

where  $Q_0$  is the discharge at time  $t = 0$ . Solving these two integrals yields

$$\begin{aligned} \left. \frac{ab}{b-1} Q^{(b-1)} \right|_{Q_0}^{Q_t} &= -t \Big|_0^t \\ Q_t^{(b-1)} - Q_0^{(b-1)} &= -\frac{b-1}{ab} t \\ Q_t^{(b-1)} &= Q_0^{(b-1)} - \frac{b-1}{ab} t. \end{aligned}$$

This leads to the recession curve function

$$Q_t = \left[ Q_0^{b-1} + \frac{1-b}{ab} t \right]^{\frac{1}{b-1}} \quad (3.3)$$

for the base flow  $Q_t$  at time  $t$ , where  $Q_0$  is any initial discharge at time  $t = 0$ . In the special case of  $b = 1/2$ , i.e. a quadratic relationship between the groundwater storage and the base flow, (3.3) simplifies to

$$Q_t = \left( Q_0^{-\frac{1}{2}} + a^{-1} t \right)^{\frac{1}{2}}. \quad (3.4)$$

With the function  $g(u) = u^{b-1}$  and a reparametrization  $\alpha = Q_0^{b-1}$  and  $\beta = (1 - b)/ab$ , (3.3) can be written as

$$g(Q_t) = \alpha + \beta t. \quad (3.5)$$

This model is similar to the mean structure used in a GLM. Thus, assuming that at least the first and second moments of the observed discharge  $Q_t$  are known, or that  $Q_t$  follows a distribution included in the exponential family, this model can be analyzed within the GLM framework. Therefore, finding the two time points, where consecutively the direct flow and the surface flow ends, corresponds to finding two change points in a GLM. Hence, to separate the base flow from the total flow, the second change point has to be estimated.

In general not only one recession curve is analyzed to obtain some information about the groundwater storage, but usually several recession curves observed at one gauging station are investigated. As it is assumed, that outer influence to the hydrological system is negligible, it is still present in reality. Thus the base flow of one recession curve does not exactly reflect the relationship between the groundwater storage and the base flow. But it varies depending on some unobserved outer influences. Therefore, all observed recession curves should be analyzed simultaneously in order to determine a common function of the base flow. This makes sense, because the properties of the catchment are usually constant over time. Therefore, the parameters  $a$  and  $b$  in (3.1) are the same for all recession curves. In contrast, the initial discharge  $Q_0$  may differ, that is the parameter  $\beta$  in (3.5) is the same for all recession curves and the parameter  $\alpha$  may differ for each recession curve. Moreover, the parameters of the first and second model differ between each recession curve, due to the fact, that the parameter of these two segments depends on the amount of rain falling during the rainstorm. Thus the corresponding statistical model is a model for a panel data set, where each panel corresponds to an observed recession curve after different rainstorms. Each panel can be described by a GLM with two continuous change points, but a common slope for the last segment for all panels. As a consequence, the new algorithm introduced in Section 2.4 is applied to estimate the common slope, which describes the common decrease of the base flow and hence the relationship between the groundwater storage and the base flow.

## 3.2 Data Description

The discharge of the river Sulm at Leibnitz is analyzed in this thesis. The Sulm is a river in the south of Styria (Austria). The discharge was measured at an automatic gauging station recording the discharge every 15 minutes. Here, hourly data are the basis of the investigations. The available time period of measurements starts at 01/01/1999 00:00 and continues until 01/01/2000 00:00 (Figure 3.2). In this time period several rainstorms have been observed. These rainstorm are prerequisites for analyzing recession curves. But not all recession

curves after a rainstorm are suitable for a recession curve analysis. They have to last for a minimal time period to enable the observation of all three types of flow. This minimal time period was determined to be at least three days. For analyzing all suitable recession curves the data has to be separated in the individual recession curves.

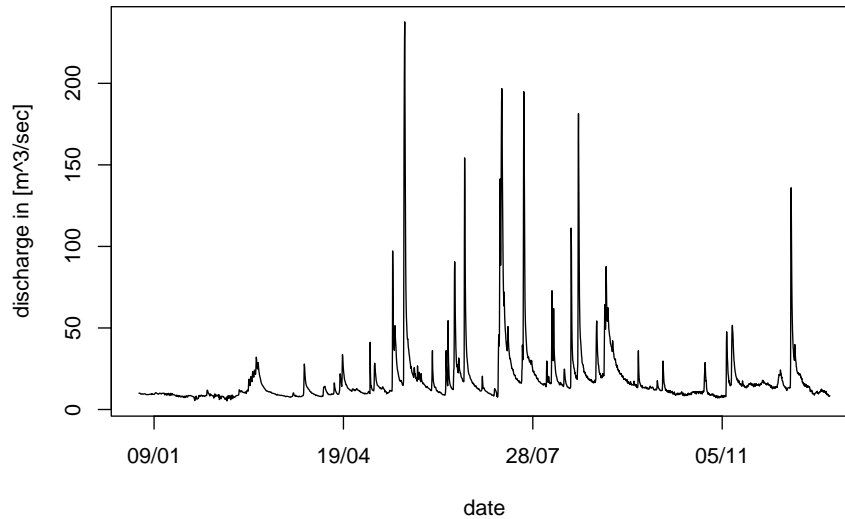


Figure 3.2: Discharge of Sulm at Leibnitz from 01/01/1999 until 01/01/2000.

### 3.2.1 Extracting of Recession Curves

To extract suitable recession curves out of the whole data set, several restrictions were specified, which describe a recession curve. First, the discharge after a rainstorm has to exceed a predetermined level of  $25 \text{ m}^3/\text{sec}$ . Second, the decrease of the discharge has to hold up for more than three days. Third, a short and small increase of the discharge was accepted. This insignificant increase may be the result of a measurement error or of a short rainfall and should be included in the analysis. Finally, the endpoint of the recession curve was determined at a significant increase of the discharge. An increase was ruled as significant, if the relative increase was larger than 10% and/or hold up for more than twelve hours.

These restrictions result in 13 suitable recession curves. The dates of the start and end points and the duration of each recession curve are listed in Table 3.1.

To illustrate these restrictions the discharge from 02/06/1999 to 01/07/1999 is considered (Figure 3.3). This figure shows five recession curves from consecutive rainstorm. The recession curves after the first and fifth rainstorm satisfy



number	start point	end point	duration in days
1	29/03/99 09:00	07/04/99 04:00	8.79
2	18/04/99 12:00	23/04/99 10:00	4.92
3	21/05/99 09:00	26/05/99 00:00	4.62
4	04/06/99 22:00	11/06/99 09:00	6.46
5	22/06/99 01:00	30/06/99 20:00	8.79
6	14/07/99 21:00	21/07/99 16:00	6.79
7	23/07/99 08:00	02/08/99 15:00	10.29
8	08/08/99 00:00	12/08/99 22:00	4.92
9	17/08/99 07:00	20/08/99 07:00	3.00
10	21/08/99 02:00	28/08/99 21:00	7.79
11	08/09/99 04:00	18/09/99 20:00	10.67
12	04/10/99 18:00	12/10/99 21:00	8.12
13	10/11/99 09:00	15/11/99 07:00	4.92

Table 3.1: Start- and endpoints and duration of each of the 13 recessions.

the restrictions defined above. The vertical lines represents the starting and end points of each of the two recession curves. The recession curves after the second and third rainstorm do not last for at least three days. The recession curve after the fourth rainstorm does hold up for more than three days, but there is a significant increase of about more than 10% in between. Therefore, the discharges after the second, third, and fourth rainstorm were not included in the analysis.

### 3.2.2 Data Reduction

The average duration of a recession curve is about seven days, giving almost 170 observations per recession curve. This is a lot of data for a change point analysis, especially when analyzing several recession curves together. Therefore it would be useful to reduce the data if possible.

To illustrate the utilized method of data reduction, a closer look at the recession curve five is taken (Figure 3.4). This recession curve starts at 22/06/99 01:00 and ends at 30/06/99 20:00. Thus the duration is 8 days and 19 hours, which yields  $n = 211$  observations. One way to reduce the amount of data is to use only every say fourth data point. But this strategy would result in only a few data at the beginning of the recession curve, where most information about the curvature of the model is included, and where also the first change point is probably located. As it can be seen in Figure 3.4, the decrease of the discharge at the end of the recession curve is very small. So, there is huge amount of data, which could be reduced without losing to much information about the functional form of this part of the data. Accordingly to that, the data has to be reduced such, that at the beginning of the recession curve nearly all data are available and at the end only a few data points are used.

Let  $i$  be the index of an observation in the entire data set, then the indices

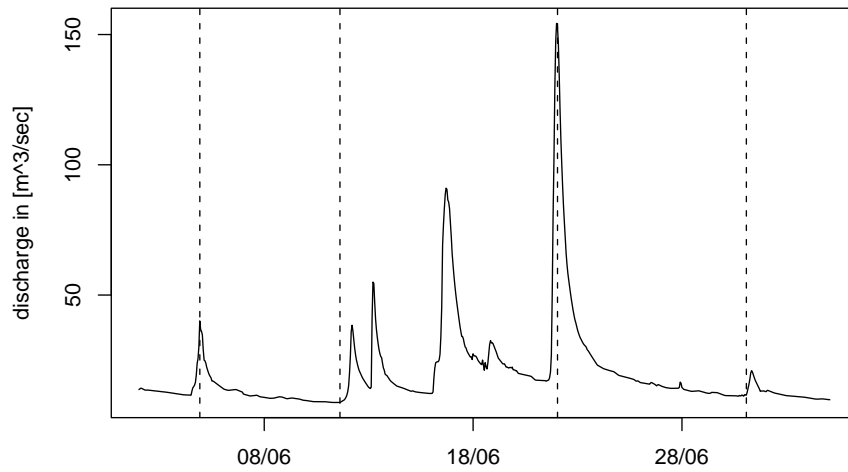


Figure 3.3: Discharge of the Sulm at Leibnitz from 02/06/1999 until 01/07/1999, showing five recession curves, where only the first and the fifth recession curve satisfy the restrictions for the subsequent analysis.

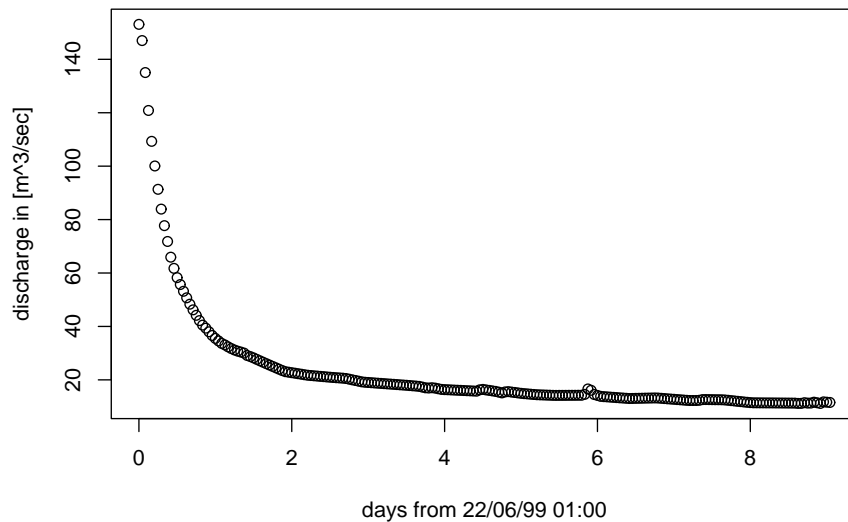
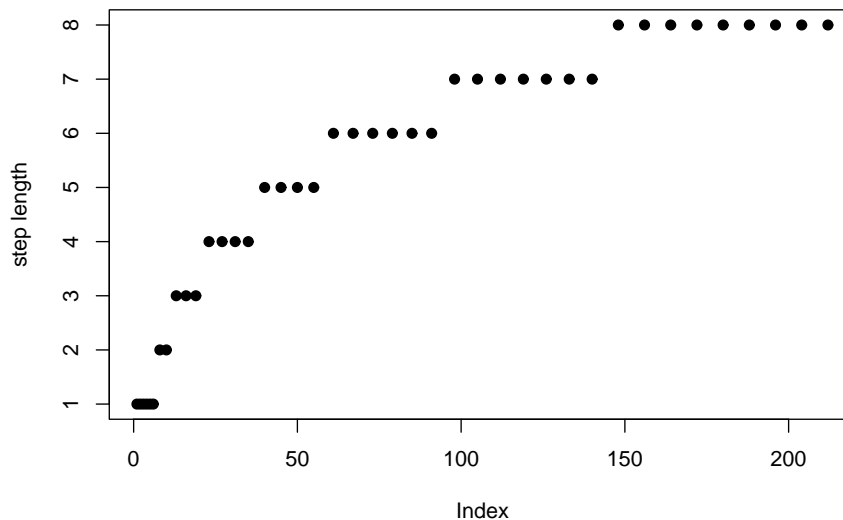


Figure 3.4: Hourly recorded discharge of the fifth recession curve of Sulm at Leibnitz.

of the reduced observations are determined by

$$j(k) = \sum_{i=1}^k \left\lfloor \log \left[ \left( \frac{i}{5} \right)^4 + 2 \right] \right\rfloor,$$

where  $k$  is increased by step one ( $k = 1, 2, \dots$ ) until  $j(k) > n$ . Using this function to calculate the remaining indices  $j(k)$  reduces the data as mentioned above. At the beginning every data point is taken, but after a few data points only every second data point is selected, and so on. Thus, the step length between the selected data points increases. The step length in dependence of the index for the fifth recession curve is plotted in Figure 3.5. It can be seen that for the fifth recession curve at the beginning every data point is selected and at the end only every eighth. Therefore, the number of observations is reduced to  $n = 41$ , which is about 20% of the original size. The reduced data of the fifth recession curve is plotted in Figure 3.6. The amount of data of all recession curves were reduced by applying this method.



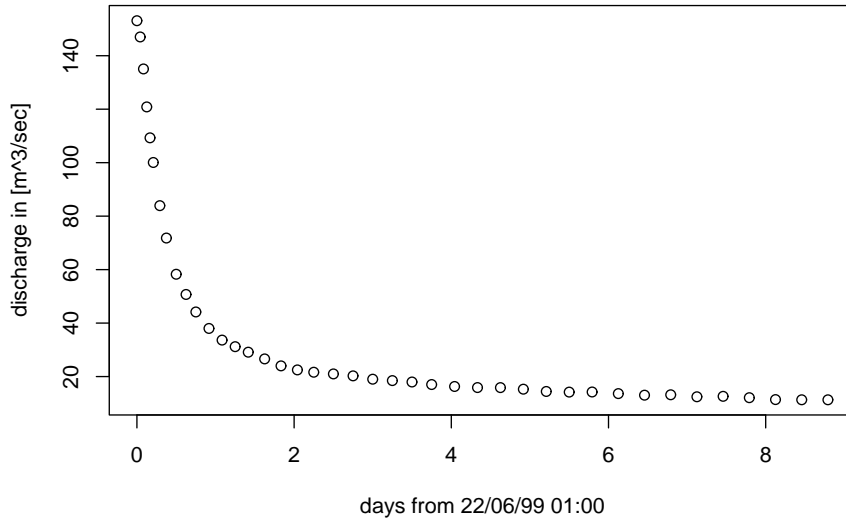


Figure 3.6: Reduced data of the fifth recession curve of the Sulm.

slope for the base flow for all curves should be the same. In addition, a nonlinear relationship (3.1) between base flow and groundwater storage is assumed. Moreover, it is assumed that this model also holds for the period where a direct and/or surface flow exists. Thus, the mean model for all segments has the same structure, i.e.

$$g(Q_t) = \alpha + \beta t,$$

but of course with different parameters for each segment. In addition it is assumed that for each recession curve all three kinds of flow are observed. In Figure 3.7 the discharges of all 13 recession curves are plotted. The discharge at the starting points of each recession curve varies a lot, but after the third or fourth day the slopes are all very similar. This supports the assumption of a common slope for the base flow of all curves.

A first investigation of the recession curves indicates that there is a constant variance of all the observations around their means. Thus, for the remaining analysis we assume a normal distribution for the responses.

### 3.2.4 Methods for Recession curve analysis

Five different methods are applied to analyze the data of the 13 recession curves. For the first two methods, no continuity constraints at the change points are assumed. The parameters of a GLM with two discontinuous change points are estimated in each panel separately. This is done by using the HS algorithm as well as the DP approach. For the third and fourth method additional continuity constraints at the change points are considered. Thus, a GLM

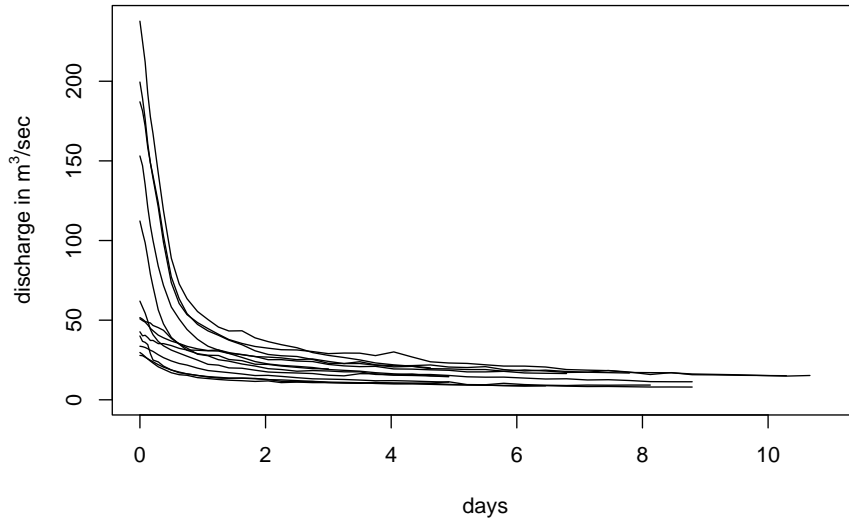


Figure 3.7: 13 recessions of Sulm at Leibnitz after various rainfall events.

with two continuous change points is fitted to each panel separately, where a grid search and the exact method are used to estimate the change points. For all four methods, the common slope for the last segment is obtained by an appropriate summary statistic, i.e. the mean and the median. Finally in the fifth method, all panels are analyzed simultaneously. This is done using the new algorithm introduced in Section 2.4.

The analyses are done by utilizing the functions `dcp`, `ccp`, and `ccpComSlop` of the R- package `CpInGLM`. The results of these analysis are summarized in the next Section.

### 3.3 Results

As mentioned above, five different methods were used to analyze the 13 recession curves. For the first four methods each recession curve was analyzed separately under two different assumptions each with two different algorithms. The results are of course different estimates for the slope of the last segment. These estimated slopes are listed in Table 3.2. The means and the medians as well as the standard deviations of these slopes are also listed in this table. A comparison of the different algorithms, i.e. HS with DP and grid search with exact method, indicates, that the differences in the estimated slopes between HS and DP are negligible. The difference between the grid search and the exact method for continuous change points is larger. This indicates that in most cases the ML estimates of the change points lies between two observations.

Moreover, from this table it can be seen that only in a few cases an additional continuity constraint for the model at the change points leads to different estimates for the last slope (recession curves number 2, 7, and 10). This results in a slightly larger estimate of a common slope, i.e. the mean as well as the median of these estimated slopes are larger in the case of continuous change points. The empirical standard deviations of these slopes are nearly the same for all methods.

recession	discontinuous		continuous	
	DP	HS	grid search	exact
1	0.0084	0.0084	0.0086	0.0085
2	0.0154	0.0154	0.0117	0.0114
3	0.0239	0.0239	0.0259	0.0247
4	0.0137	0.0137	0.0138	0.0137
5	0.0134	0.0137	0.0135	0.0134
6	0.0075	0.0075	0.0089	0.0088
7	0.0108	0.0101	0.0123	0.0126
8	0.0096	0.0106	0.0108	0.0080
9	0.0232	0.0232	0.0233	0.0257
10	0.0080	0.0080	0.0157	0.0165
11	0.0084	0.0084	0.0074	0.0074
12	0.0063	0.0074	0.0066	0.0063
13	0.0111	0.0111	0.0127	0.0137
mean	0.0123	0.0124	0.0132	0.0131
median	0.0108	0.0106	0.0123	0.0126
std.dev	0.0057	0.0056	0.0057	0.0061

Table 3.2: Estimated slopes of the third segments using four different methods.

The fifth considered method is a common analysis of the recession curves, i.e. analyzing the whole panel data set with the assumption of a common slope for the last segment. This is done utilizing the algorithm described in Section 2.4 by using the *median* method for estimating the common slope. This estimated common slope of the last segments for all panels is  $\hat{\delta} = 0.0126$ , which equals the median of the estimated slopes using the exact method for estimating continuous change points for each recession curve separately. As the fitting algorithm is an iterative one, the number of iterations until the algorithm converged as well as the evolution of the estimated common slope and the estimated last change points are of interest. For this data, the algorithm was ruled to converge after the fourth iteration, where the stopping rule (2.87), the maximum of the difference between the change points of two consecutive iterations, was applied. The evolution of the estimated common slope and the last change points are plotted in Figure 3.8 and Figure 3.9, respectively.

To illustrate the fact that the difference of the fitted models either analyzing the whole panel data set or utilize a separate analysis using the exact method

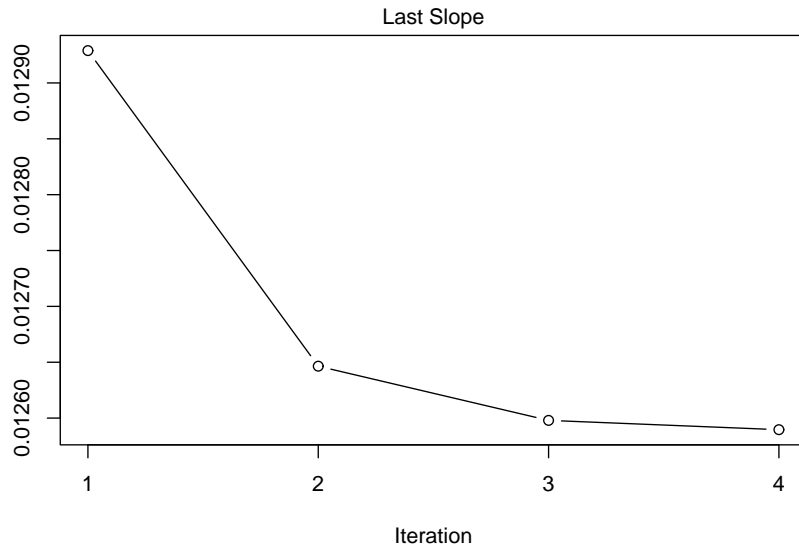


Figure 3.8: Evolution of the common slope  $\hat{\delta}^{(t)}$  for all recession curves.

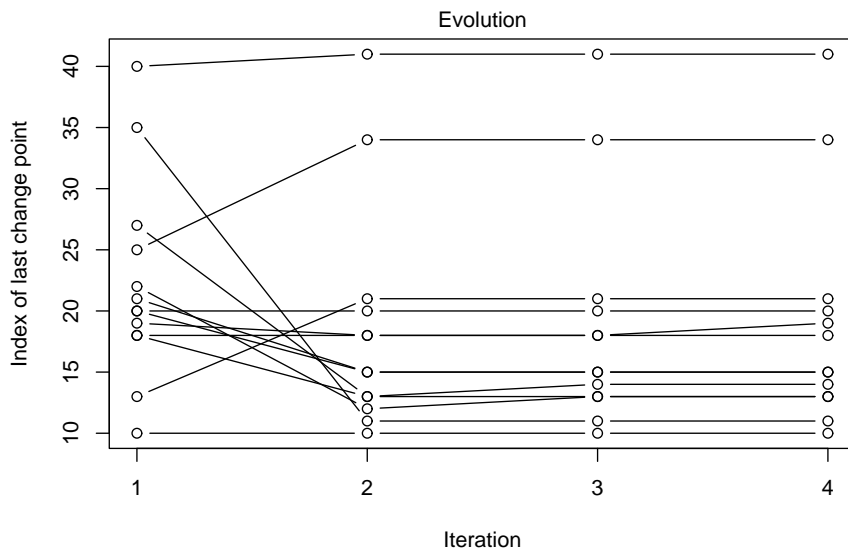


Figure 3.9: Evolution of the second change points for all recession curves.

is not that much, the fitted models for the eighth recession curve are compared (see Figure 3.10). The estimates of the first change point are nearly the same, whereas the estimates of the second change point differ slightly. Of course this is due to the fact, that the estimated slopes for the last segments are different. But overall, such a difference between these two models can be hardly recognized in the figure.

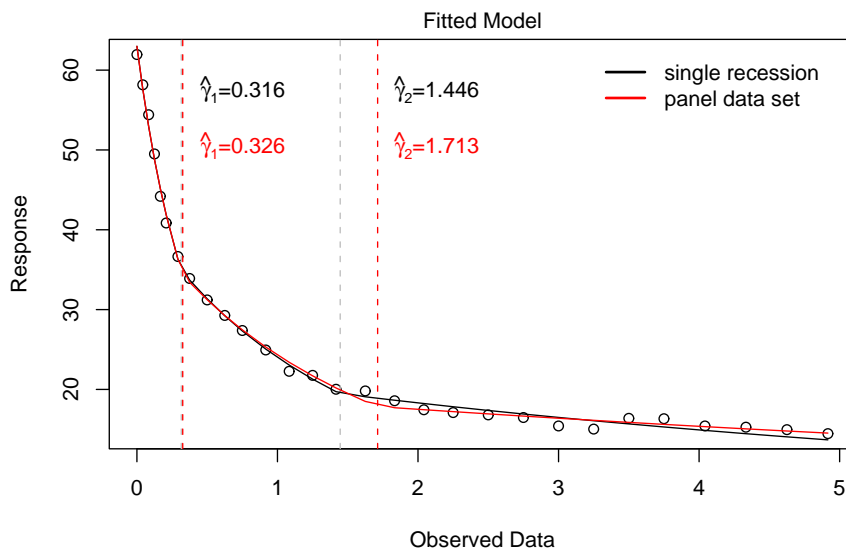


Figure 3.10: Comparison between the fitted models of a single analysis of the eighth recession (red) and an analysis within the panel data set assuming a common slope for the last segment (black).

Summing up, the new method applied on hydrological data leads to an appropriate fitted model for each recession curve with a common slope for the last segment. The advantage of this analysis is, that the number of parameters which have to be estimated reduces by  $n - 1$ , where  $n$  is the number of investigated recession curves.



# Appendix A

## Generalized Linear Models

The generalized linear model (GLM) is a generalization of an OLM. A comprehensive introduction in GLMs is given in McCullagh and Nelder (1989). First, the distribution of the response  $y$  is no longer restricted to the normal distribution. In GLMs  $y$  follows a distribution within the exponential family with probability density function

$$f(y|\theta, \phi) = \exp \left[ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right].$$

The two unknown parameters are the canonical parameter  $\theta$  and the dispersion  $\phi$ .

Second, a link function  $g(\cdot)$  connects the mean  $E(y) = \mu$  to the linear predictor  $\eta = \mathbf{x}^T \boldsymbol{\beta}$ . Therefore, the GLM can be written as

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

where  $\mathbf{X}$  is again the  $n \times p$  design matrix and  $\boldsymbol{\beta}$  a  $p$ -dimensional column vector of unknown parameters. The link function can be any monotonic differentiable function. A special link function is  $g(\cdot) = \theta$ , the so called canonical link.

Standard results of the exponential family are

$$E(y) = \mu = b'(\theta)$$

and

$$Var(y) = a(\phi)b''(\theta) = a(\phi)V(\mu),$$

where  $V(\cdot)$  is called the variance function. Thus,  $Var(y)$  is proportional to a function in  $\mu$ . The assumed distribution of  $y$  determines the relationship between the mean and the variance. In particular,  $V(\mu) = 1$  for the normal distribution or  $V(\mu) = \mu$  for Poisson variates.

Consider a sample of independent responses  $\mathbf{y} = (y_1, \dots, y_n)^T$ , then the log likelihood function is given by

$$\ell(\boldsymbol{\theta}, \phi | \mathbf{y}) = \sum_{i=1}^n \log f(y_i | \theta_i, \phi) = \sum_{i=1}^n \left[ \frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right]. \quad (\text{A.1})$$

Since  $\boldsymbol{\beta}$  is the parameter of interest, usually the log likelihood is written as

$$\ell(\boldsymbol{\beta}, \phi|\mathbf{y}) = \sum_{i=1}^n \log f(y_i|\boldsymbol{\beta}, \phi).$$

To estimate  $\boldsymbol{\beta}$  with the ML method, the score function

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi|\mathbf{y})}{\partial \boldsymbol{\beta}} = 0$$

has to be solved. The chain rule yields to

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi|\mathbf{y})}{\partial \boldsymbol{\beta}} = \frac{\partial \ell}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}}, \quad (\text{A.2})$$

where  $\partial \boldsymbol{\mu}/\partial \boldsymbol{\eta}$  depends on the specification of the link function. Thus

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi|\mathbf{y})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left[ \frac{y_i - \mu_i}{\phi V(\mu_i)} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} x_i \right].$$

This function has no analytical solution, therefore it is solved by applying the Iterative Weighted Least Square (IWLS) method. This method provides a ML estimate for  $\boldsymbol{\beta}$  with large sample distribution given by

$$\sqrt{n} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \sim N_p \left( 0, n (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \right),$$

where  $N_p$  denotes the  $p$ -dimensional Normal distribution and  $\mathbf{W}$  is a diagonal matrix with diagonal elements  $w_i$  defined as

$$1/w_i = \phi V(\mu_i) [g'(\mu_i)]^2.$$

Note, that  $W$  depends on the dispersion parameter  $\phi$ , which may be unknown and thus has to be estimated. For this case we will use the mean Pearson statistic

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \frac{1}{n-p} X^2 \quad (\text{A.3})$$

as estimate of  $\phi$

A common goodness-of-fit statistic is the deviance

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}, \phi) = 2\phi \{ \ell(\mathbf{y}, \phi|\mathbf{y}) - \ell(\hat{\boldsymbol{\mu}}, \phi|\mathbf{y}) \},$$

where  $\ell(\mathbf{y}, \phi|\mathbf{y})$  is the log likelihood function of the saturated model. In this case the fitted data are represented by the observed data, i.e.  $\hat{\mu}_i = y_i$ . Therefore, for given responses  $\mathbf{y}$  and dispersion  $\phi$ ,  $\ell(\mathbf{y}, \phi|\mathbf{y})$  is constant and the maximization of  $\ell(\hat{\boldsymbol{\mu}}, \phi|\mathbf{y})$  implies a minimization of the deviance. For a normal model the deviance is

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}, \phi) = (\mathbf{y} - \hat{\boldsymbol{\mu}})^T (\mathbf{y} - \hat{\boldsymbol{\mu}})$$

and thus equals the residual sum of squares and follows a  $\chi^2$ -distribution. One property of the deviance is additivity for nested models. Therefore, it can be used similar to the analysis of variance table for nested models. As for nested models the deviance is  $-2$  times the log likelihood ratio test statistic, it follows asymptotically a  $\chi^2$ -distribution with  $q$  degrees of freedom, where  $q$  is the difference in the number of parameters between these two models.

Common residuals in GLMs are the standardized Pearson residuals

$$r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{\phi V(\hat{\mu}_i)} \sqrt{1 - h_i}},$$

where  $h_i$  is the  $i$ th diagonal element of the generalized hat matrix

$$\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}.$$

A well known property of  $r_{Pi}$  is that they inherit the skewness of the response variable. Thus, for a skewed distribution a normal approximation for the  $r_{Pi}$  is insufficient. Another type of residuals is based on the contribution of each observation to the deviance. This contribution  $d_i$  is given by

$$d_i = 2(\ell(y_i, \phi|y_i) - \ell(\hat{\mu}_i, \phi|y_i))$$

and yields the definition of the standardized deviance residuals

$$r_{Di} = \frac{\text{sign}(y_i - \hat{\mu}_i) \sqrt{|d_i|}}{\sqrt{1 - h_i}}.$$

These standardized deviance residuals are closer to the standard normal distribution than the standardized Pearson residuals (Davison & Snell, 1991). For a more detailed discussion of GLMs see McCullagh and Nelder (1989).

# Appendix B

## Help R-Package CpInGLM

### Package ‘CpInGLM’

January 2007

**Version** 0.9-28

**Date** 2007-01-04

**Title** Change Point Detection in GLMs

**Author** Johannes Hofrichter

**Maintainer** Johannes Hofrichter <johannes.hofrichter@joanneum.at>

**Description** Fitting GLMs with one or multiple change points. These change points can be continuous or discontinuous change points. In the case of continuous change points only one explanatory variable is allowed.

**Depends** boot, akima, R ( $\geq 2.3.1$ )

**License** GPL

### R topics documented:

---

`CpInGLM.internal` *Internal CpInGLM objects*

---

### Description

These are not to be called by the user.

---

 HSvsDP

*Artificial Data with three Discontinuous Change Points*


---

### Description

Artificial data based on a simple linear regression with three discontinuous change points

### Usage

`data(HSvsDP)`

### Format

The data was generated based on a simple linear regression with three discontinuous change points. The explanatory variable  $x$  is given by  $x_i = i/10$ , with  $i = 1, \dots, 60$ . The change points are located at the indices  $\tau_1 = 15$ ,  $\tau_2 = 30$  and  $\tau_3 = 45$ . Then, the linear model is given by

$$\begin{aligned}
 y_i &= x_i + \epsilon_i & 0 < i \leq 15 \\
 y_i &= 5 + x_i + \epsilon_i & 15 < i \leq 30 \\
 y_i &= 18 - 0.8x_i + \epsilon_i & 30 < i \leq 45 \\
 y_i &= 1 + 2x_i + \epsilon_i & 45 < i \leq 60,
 \end{aligned}$$

with  $\epsilon_i \sim N(0, 1)$ .

HSvsDP is a matrix of dimension (60 x 2):

[,1]	y	numeric	response
[,2]	x	numeric	explanatory variable

---

 ccp

*Continuous Change Points in GLMs*


---

### Description

ccp is used to estimate multiple change points in GLMs, where a continuity constraint at the change points is assumed.

### Usage

```
## Default S3 method:
ccp(object, ...)
## S3 method for class 'glm':
ccp(object, nOfCp, dps, method=c("grid","exact"),
      nameCp, lastCp, lastSlope, gr, ...)
```

### Arguments

<code>object</code>	an object of class "glm", usually the result of a call to <code>glm</code>
<code>nOfCp</code>	integer, specifying the number of change points
<code>dps</code>	integer, specifying the minimum number of observations in each segment
<code>method</code>	method for estimating the change points (see detail)
<code>nameCp</code>	character: column name of the explanatory variable which describes the change point
<code>lastCp</code>	an optional vector (or single value), determining all possible x-values of the rightmost change point
<code>lastSlope</code>	an optional value, determining the value of a fixed slope for the rightmost segment
<code>gr</code>	definition of the grid for the grid search, if <code>method = "grid"</code>
<code>...</code>	further arguments passed to or from other methods.

### Details

`ccp` is used to estimate continuous change point in GLMs. The considered GLM with  $k - 1$  change points is given by

$$\begin{aligned}
 g(\mu_i) &= \beta_{10} + \beta_{11}x_i, & a \leq x_i \leq \gamma_1 \\
 g(\mu_i) &= \beta_{20} + \beta_{21}x_i, & \gamma_1 \leq x_i \leq \gamma_2 \\
 & \vdots \\
 g(\mu_i) &= \beta_{k0} + \beta_{k1}x_i, & \gamma_{k-1} \leq x_i \leq b,
 \end{aligned}$$

Here it is assumed, that the link function  $g(\cdot)$  and the variance function  $V(\cdot)$  is the same for all segments. Moreover, it is assumed that the dispersion parameter  $\phi$  is constant for all observations. To ensure continuity of  $g(\cdot)$  at  $\gamma_i$ , the continuity constraint

$$\beta_{r0} + \beta_{r1}\gamma_r = \beta_{s0} + \beta_{s1}\gamma_r,$$

at each change point is assumed, where  $s = r + 1$ .

The estimation of the change points is either done by an ordinary grid search (`method="grid"`), or by an exact method (`method="exact"`). In the case of a grid search and missing argument `gr`, the grid is determined by the observed values of the explanatory variable.

The optional argument `lastCp` can be a vector or a single value of a guess of the last (rightmost) change point. In that case, the estimated rightmost change point is restricted to an element of `lastCp`.

The optional argument `lastSlope` can be used to fix the slope of the last (rightmost) segment to a certain value.

## Value

`ccp` returns an object of class `"ccp"`. The general generic accessor functions `coefficients`, `fitted.values`, `residuals`, and `logLik` can be used to extract various useful features of the value returned by `ccp`.

An object of class `"ccp"` is a list containing at least the following components:

<code>coefficients</code>	a named vector of coefficients
<code>residuals</code>	the working residuals, that is the residuals in the final iteration of the IWLS fit of each segment.
<code>fitted.values</code>	the fitted mean values, obtained by transforming the linear predictors by the inverse of the link function.
<code>weights</code>	the working weights, that is the weights in the final iteration of the IWLS fit.
<code>change.points</code>	values of the estimated continuous change points
<code>CpIndex</code>	indices of the explanatory variable after which the estimated continuous change points are detected.
<code>logLikelihood</code>	maximized log likelihood
<code>deviance</code>	the deviance of the fitted model, see <code>glm</code>
<code>null.deviance</code>	the deviance of the null model, which corresponds to a GLM with an intercept but without any change point.
<code>aic</code>	Akaike's Information Criterion, minus twice the maximized log-likelihood plus twice the number of the estimated parameters (coefficients and change points).
<code>bic</code>	Bayesian Information Criterion, minus twice the maximized log-likelihood plus the number of the estimated parameters (coefficients and change points) times the logarithm of the number of observations.

`df.residual` the residual degrees of freedom.  
`df.null` the residual degrees of freedom for the null model.  
`y` vector of the response variable.  
`orgDesign` design matrix of the original GLM  
`possibleCp` matrix of all possible combination of change points  
`pLL` Profile log-likelihood for `possibleCp`  
`family` the `family` object used.  
`model` the model frame used.  
`na.action` the `na.action` initially supplied.  
`prior.weights`  
the case weights initially supplied.  
`call` the matched call.

## Note

It is assumed that there is only one explanatory variable and this variable is given in ascending order!!

## Author(s)

Johannes Hofrichter

## See Also

For a change point model with a discontinuous change at the change point see `dcp`

## Examples

```
data(exmpl1)

plot(exmpl1$x,exmpl1$y)

m1 <- glm(y~x, data=exmpl1, family=poisson(link=log))
m1.ccp <- ccp(m1,nOfCp=2, dps=5)

summary(m1.ccp)
# plot(m1.ccp)

plot(m1.ccp, which=1)
lines(exmpl1$x, fitted(m1), col=2)
```



---

ccpComSlope                    *Change Point Model for Panel Data*

---

## Description

ccpComSlope is used to fit a change point model for panel data with a common slope for the last (rightmost) segment in all panels.

## Usage

```
ccpComSlope(formula, data, iPanel,
             family=gaussian(link=identity),
             nOfCp=2, dps=5,
             comSlopeMethod = c("mean", "median", "glm"),
             control = ccpComSlope.control(...),
             verbose=FALSE, ...)
```

## Arguments

<code>formula</code>	a symbolic description of the model to be fit. (See <code>glm</code> for details of formula specifications.)
<code>data</code>	a <code>data.frame</code> of the panel data
<code>iPanel</code>	integer, specifying the column of the factor determining the panels
<code>family</code>	a description of the error distribution and link function to be used in the model. (See <code>family</code> for details of family functions.)
<code>nOfCp</code>	integer, specifying the number of change points
<code>dps</code>	integer, specifying the minimum number of observations in each segment
<code>comSlopeMethod</code>	method for calculating the common slope for the last segment
<code>control</code>	list of control properties of the algorithm
<code>verbose</code>	an optional logical value. If <code>TRUE</code> information on the evolution of the iterative algorithm is printed. Default is <code>FALSE</code> .
<code>...</code>	further arguments passed to or from other methods.

## Details

`ccpComSlope` is used to fit a change point model for panel data sets, assuming that the slope of the model in the last segment is the same for all panels. The fitting procedure is an iterative algorithm, which consists of two steps. The first step is to estimate the common slope for the last segment given the rightmost change point of each panel. The second step is to estimate the change points given the new coefficient of the slope of the last segment from the first step. These two steps are iterated until the maximum number of iterations are reached or the algorithm is ruled to be converged (see `ccpComSlope.control`).

## Value

`ccpComSlope` returns an object of class `"ccpComSlope"`.

An object of class `"ccpComSlope"` is a list containing at least the following components:

<code>lastSlope</code>	the vector of the evolution of the common last slope estimate.
<code>lastCp</code>	the vector or matrix of the evolution of the estimates of the last change points.
<code>lastCpIndex</code>	the vector or matrix of the evolution of the indices after which the the last change point is detected.
<code>mll</code>	the vector of the evolution of the maximized log likelihood.
<code>model</code>	the model frame used.
<code>iterations</code>	the number of iterations
<code>converged</code>	logical. Was the algorithm judged to have converged.
<code>call</code>	the matched call
<code>ccpList</code>	a list of the estimated continuous change point model for each panel. Each element has class <code>"ccp"</code> .

## Author(s)

Johannes Hofrichter

## See Also

For a continuous change point model for one data set see `ccp`.

## Examples

```

data(panel.dat) #included in CpInGLM

contr <- ccpComSlope.control(epsilon = 1e-02, maxit = 40,
                             method = "MLL")

m1 <- ccpComSlope(y~x, data=panel.dat,
                 iPanel=3, family=poisson(link=log), nOfCp=1,
                 comSlopeMethod="mean", control=contr)

m1

# plot of the evolution of the common slope and
# the indices of the change points
plot(m1)

# summary and plot of the continuous change point model
# of the first panel

m1$ccpList[[1]]
plot(m1$ccpList[[1]])

```

---

```
ccpComSlope.control
```

*Control the Iterations in ccpComSlope*

---

## Description

Various parameters to control aspects of the ccpComSlope fit.

## Usage

```
ccpComSlope.control(epsilon = 1e-05, maxit = 40,
                    method = c("MLL", "slope", "cp"))
```

## Arguments

epsilon	A positive numeric value specifying the tolerance of the convergence criterion.
maxit	A positive integer specifying the maximum number of iterations allowed.
method	Method that determines which convergence criterion is used.

## Details

`method="MLL"` The algorithm is ruled to be converged, if the difference between the maximized log likelihood of two consecutive iterations is smaller than `epsilon`.

`method="slope"` The algorithm is ruled to be converged, if the relative difference between the estimate of the common slope of two consecutive iterations is smaller than `epsilon`.

`method="cp"` The algorithm is ruled to be converged, if the maximum of the relative differences between the estimates of the last change points of two consecutive iterations is smaller than `epsilon`.

## Value

A list with exactly three components: `epsilon`, `maxit`, and `method` with meanings as explained under Arguments.

## Author(s)

Johannes Hofrichter

## See Also

`ccpComSlope`

---

dcp

*Discontinuous Change Points in GLMs*

---

## Description

`dcp` is used to estimate multiple discontinuous change points in GLMs.

## Usage

```
## Default S3 method:  
dcp(object, ...)  
## S3 method for class 'glm':  
dcp(object, nOfCp=NULL, dps=NULL, method=c("dp","hs"), ...)
```

## Arguments

<code>object</code>	an object of class "glm", usually the result of a call to <code>glm</code> .
<code>nOfCp</code>	integer, specifying the number of change points
<code>dps</code>	integer, specifying the minimum number of observations in each segment
<code>method</code>	string, specifying the method for change point estimation, see below
<code>...</code>	further arguments passed to or from other methods.

## Details

`dcp` is used to estimate discontinuous change points in GLMs. The considered GLM with  $k - 1$  change points is given by

$$\begin{aligned}
 g(\mu_i) &= x_i\beta_1, & 0 < i \leq \tau_1 \\
 g(\mu_i) &= x_i\beta_2, & \tau_1 < i \leq \tau_2 \\
 & \vdots \\
 g(\mu_i) &= x_i\beta_k, & \tau_{k-1} < i \leq n
 \end{aligned}$$

It is assumed, that the link function  $g(\cdot)$  and the variance function  $V(\cdot)$  is the same for all segments. In contrast, the dispersion parameter changes at the change points, as well. As there are not any constraints at the change points, the GLMs across the segments are autonomous and therefore the parameters can be estimated separately. Thus, once the change points are determined the model of each segment can be estimated with the usual IWLS algorithm. There are two different methods implemented for estimating the change points, the hierarchic binary splitting and a dynamic programming algorithm. The method used in `dcp` can be selected via the argument `method`.

The hierarchic binary splitting algorithm is chosen by `method="hs"` and splits the model in the same way as regression trees do. It is fast, but can lead to a suboptimal solution. For details see `dcpBinarySplit`.

The dynamic programming algorithm is chosen by `method="dp"`. The computational complexity is greater than that one of the hierarchic binary splitting, but leads to an optimal solution. For details see `dcpDynamicProg`

## Value

`dcp` returns an object of class "dcp". The general generic accessor functions `coefficients`, `fitted.values`, `residuals`, and `logLik` can be used to extract various useful features of the value returned by `dcp`.

An object of class "dcp" is a list containing at least the following components:

<code>coefficients</code>	a named vector of coefficients
<code>residuals</code>	the working residuals, that is the residuals in the final iteration of the IWLS fit of each segment.
<code>fitted.values</code>	the fitted mean values, obtained by transforming the linear predictors by the inverse of the link function.
<code>change.points</code>	values of the estimated discontinuous change points
<code>CpIndex</code>	indices of the observations after which the discontinuous change points are detected.
<code>logLikelihood</code>	maximized log likelihood
<code>y</code>	vector of the response variable.
<code>deviance</code>	the deviance of the fitted model, see <code>glm</code>
<code>null.deviance</code>	the deviance of the null model, which corresponds to a GLM with an intercept but without any change point.
<code>aic</code>	Akaike's Information Criterion, minus twice the maximized log-likelihood plus twice the number of the estimated parameters (coefficients and change points).
<code>bic</code>	Bayesian Information Criterion, minus twice the maximized log-likelihood plus the number of the estimated parameters (coefficients and change points) times the logarithm of the number of observations.
<code>df.residual</code>	the residual degrees of freedom.
<code>df.null</code>	the residual degrees of freedom for the null model.
<code>family</code>	the <code>family</code> object used.
<code>model</code>	the model frame used.
<code>na.action</code>	the <code>na.action</code> initially supplied.
<code>prior.weights</code>	the case weights initially supplied.
<code>call</code>	the matched call.
<code>method</code>	the name of the method used for detecting the change points.

### Author(s)

Johannes Hofrichter

### See Also

For a change point model with a continuous change see `ccp`.

## Examples

```

data(HSvsDP)

plot(HSvsDP$x,HSvsDP$y)

m1 <- glm(y~x, data=HSvsDP)

# via dynamic programming algorithm
m1.dcp1 <- dcp(m1, nOfCp=3, dps=5)
plot(m1.dcp1)

# via hierarchic splitting
m1.dcp2 <- dcp(m1, nOfCp=3, dps=5, method="hs")
plot(m1.dcp2)

```

---

dcpBinarySplit	<i>Discontinuous Change Points via Hierarchic Splitting</i>
----------------	---

---

## Description

Estimation of a discontinuous change points in GLMs via a hierarchic binary splitting algorithm

## Usage

```
dcpBinarySplit(x, nOfCp, dps)
```

## Arguments

x	an object of class "glm", usually the result of a call to <code>glm</code> .
nOfCp	integer, specifying the number of change points
dps	integer, specifying the minimum number of observations in each segment

## Details

`dcpBinarySplit` splits the data in a hierarchical way. The first step is to split the data into two segments. These two segments are partitioned by that change point which maximizes the sum of the log likelihood of both segments. The second step is to find a potential change point in each of these two segments. From these new potential change points, choose that one which increases the sum of the log likelihood of all three segments the most. Thus, one step of the HS algorithm consists of two tasks,

1. find a potential change point in each segment,
2. given these potential change points in each segment, choose that one which increases the log likelihood the most.

This procedure is repeated until the final number of change points, given by the argument `nOfCp`, is achieved. For more details see Hawkins (2001).

Note that the change points of a model with  $r$  change points is always a subset of the change points of a model with  $s > r$  change points.

## Value

Returns the indices of the estimated change points

## Author(s)

Johannes Hofrichter

## References

Hawkins D.M. (2001), Fitting multiple change-point models to data, *Computational Statistics & Data Analysis*, **37**, 323–341.

## See Also

For estimating discontinuous change point models based on GLMs use `dcp`. A second method for estimating discontinuous change points is `dcpDynamicProg`, which yields to an optimal solution, but with greater computational complexity.

## Examples

```
data(HSvsDP)
plot(HSvsDP$x,HSvsDP$y)
m1 <- glm(y~x, data=HSvsDP)

# assuming at least 5 observations in each segments
cp1 <- dcpBinarySplit(m1, nOfCp=3, dps=5)
```



---

dcpDynamicProg      *Discontinuous Change Points via Dynamic Programming*

---

## Description

dcpDynamicProg is used to estimate discontinuous change points in GLMs via a dynamic programming algorithm

## Usage

```
dcpDynamicProg(x, nOfCp, dps)
```

## Arguments

x	an object of class "glm", usually the result of a call to glm.
nOfCp	integer, specifying the number of change points
dps	integer, specifying the minimum number of observations in each segment

## Details

dcpDynamicProg is used to estimate discontinuous change point in GLMs. The algorithm is based on Bellman's principal of optimality. Consider a GLM and a data set of size  $n$  with  $k - 1$  change points ( $k$  segments). Then knowing the rightmost change point, the problem can be reduced to split the remaining data on the left hand side of this change point into  $k - 1$  segments. This is repeated until all  $k$  segments are obtained. For more details see Hawkins (2001). This algorithm is based on the values of the maximized log likelihood of the models of all possible connected subsets. Thus the computational complexity is greater, than that one of the hierarchic binary splitting, but the algorithm leads to an optimal solution.

## Value

Returns the indices of the estimated change points

## Author(s)

Johannes Hofrichter

## References

Hawkins D.M. (2001), Fitting multiple change-point models to data, *Computational Statistics & Data Analysis*, **37**, 323–341.

**See Also**

For estimating discontinuous change point models based on GLMs use `dcp`. A second method for estimating discontinuous change points in GLMs is `dcpBinarySplit`, which is faster, but may lead only to a suboptimal solution.

**Examples**

```
data(HSvsDP)
plot(HSvsDP$x,HSvsDP$y)
m1 <- glm(y~x, data=HSvsDP)

cp1 <- dcpDynamicProg(m1, nOfCp=3, dps=3)

# assuming at least 10 observations in each segments
cp2 <- dcpDynamicProg(m1, nOfCp=3, dps=10)
```

---

exmpl1

*Artificial Data with two Continuous Change Points*


---

**Description**

Artificial data based on a GLM with two continuous change points

**Usage**

```
data(exmpl1)
```

**Format**

This data is based on a GLM with one continuous change point. The explanatory variable  $x$  is given by  $x_i = 1/50$ ,  $i = 1, \dots, 50$ . The response variable follows a Poisson distribution and the mean value satisfies a GLM with two continuous change points at  $x_i = 0.36$  and  $x_i = 0.7$ . The considered model is given by

$$\begin{aligned} g(\mu_i) &= 2 + 1x_i & 0 \leq x_i \leq 0.36 \\ g(\mu_i) &= 0.92 + 4x_i & 0.36 < x_i \leq 0.7 \\ g(\mu_i) &= 2.67 + 1.5x_i & 0.7 < x_i \leq 1 \end{aligned}$$

with canonical link function  $g(\mu) = \log(\mu)$  for all segments. Based on this model 50 Poisson distributed variates  $y_i$  were generated.

`exmpl1` is a matrix of dimension (50 x 2):

```
[,1] y numeric response
[,2] x numeric explanatory variable
```

---

exmpl2

*Artificial Data with one Continuous Change Point*

---

## Description

Artificial data based on a GLM with one continuous change point

## Usage

```
data(exmpl2)
```

## Format

This data is based on a GLM with one continuous change point. The explanatory variable  $x$  is given by  $x_i = 1/4$ ,  $i = 1, \dots, 40$ . The response variable follows a Poisson distribution and the mean value satisfies a GLM with one continuous change point at  $x_i = 5$ . The considered model is given by

$$\begin{aligned} g(\mu_i) &= 2 + 0.2x_i & 0 \leq x_i \leq 5 \\ g(\mu_i) &= 0.5 + 0.5x_i & 5 < x_i \leq 10 \end{aligned}$$

with canonical link function  $g(\mu) = \log(\mu)$ . Based on this model 40 Poisson distributed variates  $y_i$  were generated.

exmpl2 is a matrix of dimension (40 x 2):

```
[,1] y numeric response
[,2] x numeric explanatory variable
```

---

panel.dat

*Panel data*

---

## Description

Panel data with three panels. For each panel a GLM with one continuous change point is assumed, where the slope of the rightmost segment is the same for all panels.

**Usage**

```
data(panel.dat)
```

**Format**

panel.dat is a `data.frame` with 117 observations of three variables

[,2]	y	numeric	response
[,3]	x	numeric	explanatory variable
[,1]	panel	factor	number of the panel

---

```
plot.ccpComSlope Plot a ccpComSlope object
```

---

**Description**

Several useful plots for an object of class `ccpComSlope`

**Usage**

```
plot.ccpComSlope(x, which = 1:3,
  caption = c("Last Slope", "Evolution", "Evolution"),
  ask = prod(par("mfcol")) < length(which) && dev.interactive(),
  ...)
```

**Arguments**

<code>x</code>	an object of class "ccpComSlope", usually the result of a call to <code>ccpComSlope</code> .
<code>which</code>	If a subset of the plots is required, specify a subset of the numbers 1:3
<code>caption</code>	Captions to appear above the plots
<code>ask</code>	logical; if TRUE, the user is asked before each plot, see <code>par(ask=)</code> .
<code>...</code>	other parameters to be passed through to plotting functions.

**Details**

The first plot illustrates the evolution of the estimated common slope.

The second plot illustrates the evolution of the estimates of the last change points.

The third plot illustrates the evolution of the indices after which the of the last change points are detected.

**Author(s)**

Johannes Hofrichter

**See Also**

ccpComSlope

**Examples**

```
data(panel.dat)

m1 <- ccpComSlope(y~x, data=panel.dat, iPanel=3,
                 family=poisson(link=log), nOfCp=1,
                 comSlopeMethod="mean")

# evolution of the estimate of the common slope
plot(m1, which=1)

# evolution of the estimates of the last change points
plot(m1, which=2)

# evolution of the indices
plot(m1, which=3)
```

---

plot.ccp

*Plot a ccp Object*

---

**Description**

Several usefull plots for an object of class `ccp`

**Usage**

```
plot.ccp(x, which=1:4,
         caption=c("Fitted Model", "Normal Plot of Residuals",
                  "Residuals vs Fitted"),
         ask = prod(par("mfcol")) < length(which) && dev.interactive(),
         ...)
```

**Arguments**

`x` an object of class "ccp", usually the result of a call to `ccp`.

`which` If a subset of the plots is required, specify a subset of the numbers 1:4

caption	captions to appear above the plots
ask	logical; if TRUE, the user is asked before each plot, see <code>par(ask=)</code> .
...	other parameters to be passed through to plotting functions.

## Details

The first plot illustrates the data and the fitted model. Additionally the location of the change points are plotted.

The second plot illustrates a normal plot of the deviance residuals.

The third plot illustrates the deviance residuals vs the fitted values.

The fourth plot illustrates the profile log-likelihood with respect to the change points. If the number of change points is one, then the profile log-likelihood is plotted as a single line. If the number of change points is two, then the profile log-likelihood is plotted as a contour plot. At a number of change points greater than two, the third plot will be neglected.

## Author(s)

Johannes Hofrichter

## See Also

`ccp`

## Examples

```
data(exmpl1)

plot(exmpl1$x,exmpl1$y)

m1 <- glm(y~x, data=exmpl1, family=poisson(link=log))
m1.ccp <- ccp(m1,nOfCp=2, dps=5)

#plot(m1.ccp)

plot(m1.ccp, which=1)
lines(exmpl1$x, fitted(m1), col=2)
```

---

plot.dcp                      *Plot a dcp Object*

---

## Description

Several usefull plots for an object of class dcp

## Usage

```
plot.dcp(x, which=1:3,
         caption=c("Fitted Model", "Normal Plot of Residuals",
                  "Residuals vs Fitted"),
         ask = prod(par("mfcol")) < length(which) && dev.interactive(),
         ...)
```

## Arguments

x	an object of class "dcp", usually the result of a call to dcp.
which	If a subset of the plots is required, specify a subset of the numbers 1:3
caption	captions to appear above the plots
ask	logical; if TRUE, the user is asked before each plot, see par(ask=).
...	other parameters to be passed through to plotting functions.

## Details

The first plot illustrates the data and the fitted model and the location of the change points are plotted. In the case of more than one explanatory variable this plot is neglected.

The second plot illustrates a normal plot of the deviance residuals.

The third plot illustrates the deviance residuals vs the fitted values.

## Author(s)

Johannes Hofrichter

## See Also

dcp

## Examples

```

data(HSvsDP)
m1 <- glm(y~x, data=HSvsDP)
m1.dcp <- dcp(m1, nOfCp=3, dps=5)

# plot of the fitted model
plot.dcp(m1.dcp, which = 1)

# Normal Q-Q plot of the deviance residuals
plot.dcp(m1.dcp, which = 2)

# plot of the deviance residuals vs fitted values
plot.dcp(m1.dcp, which = 3)

```

---

quandt

*Quandt's Data*

---

## Description

Quandt's Data:

## Usage

```
data(quandt)
```

## Format

This data was first introduced by Quandt (1959). A simple linear regression model with one discontinuous change point is considered. The artificial data was generated in the following way. For the explanatory variable  $x$ , a single independent variable was generated by a random permutation of the first 20 integers. The errors  $\varepsilon_i$  were generated as 20 iid normal variates with zero mean and variance equal to one. These errors were added to a mean model that changed after the first 12 observations, giving

$$y_i = 2.5 + 0.7x_i + \varepsilon_i \quad i = 1, \dots, 12$$

$$y_i = 5.0 + 0.5x_i + \varepsilon_i \quad i = 13, \dots, 20.$$

quandt is a matrix with two columns:

```

[ ,1] x  numeric  explanatory variable
[ ,2] y  numeric  response

```



## Source

Quandt, (1958)

## References

Quandt R. E. (1958), The estimation of the Parameters of a Linear Regression system Obeying Two Separate Regimes. *JASA*, **53**, 873–880

---

recres

*Recursive Residuals for GLMs*

---

## Description

recres is used to calculate standardized recursive residuals for GLMs.

## Usage

```
## Default S3 method:
recres(object, ...)
## S3 method for class 'glm':
recres(object, method=c("response", "delta", "deletion"),
       forward=TRUE, ...)
```

## Arguments

<code>object</code>	an object of class "glm", usually the result of a call to <code>glm</code>
<code>method</code>	string, specifying the method for calculating recursive residuals, for more information see <code>Details</code>
<code>forward</code>	logical, if <code>TRUE</code> , the recursive residuals are based on the fitted model on the left hand side, for more information see <code>Details</code> .
<code>...</code>	further arguments passed to or from other methods.

## Details

recres is used to calculate standardized recursive residuals for GLMs.

The argument `forward` specifies the direction for calculating the recursive residuals. If `forward=TRUE`, then the first  $r - 1$  observations are used to fit the model and based on this model the recursive residual for the  $r$ th observation is derived. In the case of `forward=FALSE` it's the other way around.

The argument `method` specifies the method for calculating the standardized recursive residuals. If `method="delta"`, then the variance of the recursive residuals is calculated with the Delta Method. In the case of `method="deletion"`, the standardized recursive residuals are based on the one step approximation of the deletion residuals (Pregibon, 1979).

## Value

`recres` returns a vector of standardized recursive residuals.

## Author(s)

Johannes Hofrichter

## References

Pregibon D. (1979), *Data Analytic Methods for Generalized Linear Models, Unpublished Ph.D Thesis: Univ. of Toronto.*

## See Also

`residuals`

---

<code>residuals.ccp</code>	<i>Residuals for Continuous Change Point Model</i>
----------------------------	--

---

## Description

Residuals for Continuous Change Point Model

## Usage

```
## S3 method for class 'ccp':
residuals(object,
          type = c("deviance", "pearson", "working", "response"),
          ...)
```

## Arguments

<code>object</code>	an object of class <code>"ccp"</code> , usually the result of a call to <code>ccp</code> .
<code>type</code>	the type of residuals which should be returned. The alternatives are: <code>"deviance"</code> (default), <code>"pearson"</code> , <code>"working"</code> and <code>"response"</code>
<code>...</code>	further arguments passed to or from other methods.

## Details

The references define the types of residuals: Davison & Snell is a good reference for the usages of each.

## Value

Named vector of residuals

## Author(s)

Johannes Hofrichter

## References

Davison, A. C. and Snell, E. J. (1991) *Residuals and diagnostics*. In: Statistical Theory and Modelling. In Honor of Sir David Cox, FRS, eds. Hinkley, D. V., Reid, N. and Snell, E. J., Chapman & Hall.

## See Also

ccp, residuals

---

residuals.dcp

*Residuals for Discontinuous Change Point Model*

---

## Description

Residuals for Discontinuous Change Point Model

## Usage

```
## S3 method for class 'dcp':
residuals(object,
           type = c("deviance", "pearson", "working", "response"),
           ...)
```

## Arguments

**object** an object of class "dcp", usually the result of a call to dcp.

**type** the type of residuals which should be returned. The alternatives are: "deviance" (default), "pearson", "working" and "response"

... further arguments passed to or from other methods.

## Details

The references define the types of residuals: Davison & Snell is a good reference for the usages of each.

## Value

Named vector of residuals

## Author(s)

Johannes Hofrichter

## References

Davison, A. C. and Snell, E. J. (1991) *Residuals and diagnostics*. In: Statistical Theory and Modelling. In Honor of Sir David Cox, FRS, eds. Hinkley, D. V., Reid, N. and Snell, E. J., Chapman & Hall.

## See Also

dcp, residuals

---

uniqueMatrix	<i>Extract unique rows of a matrix</i>
--------------	--

---

## Description

uniqueMatrix extracts unique rows of a matrix as unique does, but returns additional information

## Usage

```
uniqueMatrix(x, ...)
```

## Arguments

x	a matrix
...	not used yet

**Value**

Returns a list with four elements

<code>x</code>	original <code>x</code> of the function call
<code>x.unique</code>	matrix of unique rows of <code>x</code>
<code>index</code>	vector of row indices of <code>x.unique</code> , where <code>x.unique[index,]</code> results <code>x</code>
<code>index.unique</code>	vector of indices of that rows which are unique in <code>x</code>

**Author(s)**

Johannes Hofrichter

**See Also**

`unique`

**Examples**

```
set.seed(1234)
x <- matrix(sample(2,18,replace=TRUE), ncol=3)
x.u <- uniqueMatrix(x)
y <- x.u[x.unique[x.u$index,]]
all.equal(x,y)
```

# Index

- \*Topic **datasets**
  - exmpl1, 121
  - exmpl2, 122
  - HSvsDP, 108
  - panel.dat, 122
  - quandt, 127
- \*Topic **hplot**
  - plot.ccp, 124
  - plot.ccpComSlope, 123
  - plot.dcp, 126
- \*Topic **internal**
  - CpInGLM.internal, 107
- \*Topic **misc**
  - ccpComSlope.control, 114
  - uniqueMatrix, 131
- \*Topic **models**
  - ccp, 108
  - ccpComSlope, 112
  - dcp, 115
  - dcpBinarySplit, 118
  - dcpDynamicProg, 120
  - recres, 128
  - residuals.ccp, 129
  - residuals.dcp, 130
- calcCoef (*CpInGLM.internal*),  
107
- calcCpEdge (*CpInGLM.internal*),  
107
- calcEdgeSolution  
(*CpInGLM.internal*), 107
- calcPanelSize  
(*CpInGLM.internal*), 107
- ccp, **108**, 113, 117, 124, 125, 129,  
130
- ccpComSlope, **112**, 123, 124
- ccpComSlope.control, 113, **114**
- ccpExact (*CpInGLM.internal*),  
107
- ccpGammaExact  
(*CpInGLM.internal*), 107
- ccpKomb (*CpInGLM.internal*), 107
- ccpLogLikExact  
(*CpInGLM.internal*), 107
- ccpPossibleCp  
(*CpInGLM.internal*), 107
- ccpRefineGrid  
(*CpInGLM.internal*), 107
- commonSlope  
(*CpInGLM.internal*), 107
- cp2index (*CpInGLM.internal*),  
107
- CpInGLM.internal, **107**
- cpmat2list (*CpInGLM.internal*),  
107
- dcp, 111, **115**, 119, 121, 126, 130,  
131
- dcpBinarySplit, 116, **118**, 121
- dcpDynamicProg, 116, 119, **120**
- dpLogLik (*CpInGLM.internal*),  
107
- exmpl1, **121**
- exmpl2, **122**
- family, 111, 112, 117
- generateBlockDiag  
(*CpInGLM.internal*), 107
- generateCpDesign  
(*CpInGLM.internal*), 107
- generateCpGrid  
(*CpInGLM.internal*), 107
- generateCpNames  
(*CpInGLM.internal*), 107

glm, 109, 110, 112, 116–118, 120,  
128

hsLogLik (*CpInGLM.internal*),  
107

HSvsDP, 108

isCpSolution  
(*CpInGLM.internal*), 107

logLik.ccp (*CpInGLM.internal*),  
107

logLik.ccpComSlope  
(*CpInGLM.internal*), 107

logLik.dcp (*CpInGLM.internal*),  
107

na.action, 111, 117

panel.dat, 122

par, 123, 125, 126

plot.ccp, 124

plot.ccpComSlope, 123

plot.dcp, 126

print.ccp (*CpInGLM.internal*),  
107

print.ccpComSlope  
(*CpInGLM.internal*), 107

print.dcp (*CpInGLM.internal*),  
107

print.summary.ccp  
(*CpInGLM.internal*), 107

print.summary.dcp  
(*CpInGLM.internal*), 107

quandt, 127

recres, 128

recResGlmDelta  
(*CpInGLM.internal*), 107

residuals, 129–131

residuals.ccp, 129

residuals.dcp, 130

splitByFixedCp  
(*CpInGLM.internal*), 107

summary.ccp  
(*CpInGLM.internal*), 107

summary.dcp  
(*CpInGLM.internal*), 107

unique, 131, 132

uniqueMatrix, 131

validLastCp  
(*CpInGLM.internal*), 107

## References

- Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1–22.
- Beckman, R. and Cook, R. (1979). Testing for two-phase regression. *Technometrics*, 21(1), 65–69.
- Bellman, R. and Dreyfus, S. (1962). *Applied dynamic programming* (Tech. Rep.). Princeton University Press.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth.
- Brown, R., Durbin, J., and Evans, J. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B*, 37, 149–192.
- Burnham, K. P. and Anderson, D. R. (1998). *Model Selection and Inference: A practical Information-Theoretic Approach*. New York: Springer.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference* (Second ed.). Pacific Grove, California: Duxbury advanced Series.
- Chen, J. and Gupta, A. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association*, 92(438), 739–747.
- Chen, J. and Gupta, A. (2000). *Parametric Statistical Change Point Analysis*. Boston: Birkhäuser.
- Chow, G. (1960). Test of equality between sets of coefficients in two linear regressions. *Econometrica*, 28, 591–605.
- Davison, A. and Snell, E. (1991). Residuals and diagnostics. In D. Hinkley, N. Ried, and E. Snell (Eds.), *Statistical Theory and Modelling: In Honor of Sir David Cox* (pp. 81–106). London: Chapman and Hall.
- Durbin, J. (1971). Boundary-crossing probabilities for the Brownian motion and Poisson process and techniques for computing the power of the Kolmogorov-Smirnov test. *Journal of Applied Probability*, 8, 431–453.
- Esterby, S. and El-Shaarawi, A. (1981). Inference about the point of change in a regression model. *Applied Statistics*, 30(3), 227–285.
- Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics*, 13, 342–368.
- Fank, J., Harum, T., and Stadler, H. (1993). *Erfassung von Abflußvorgängen in kleinen Einzugsgebieten; Speicherverhalten kleiner Einzugsgebiete* (Tech. Rep.). Institut f. Geothermie und Hydrologie; Joanneum Research, Graz, Austria.
- Farley, J. and Hinich, M. (1970). A test for a shifting slope coefficient in a linear model. *Journal of the American Statistical Association*, 65(331), 1320–1329.
- Farley, J., Hinich, M., and McGuire, T. (1975). Some comparisons of tests for a shift in the slope of multivariate linear time series models. *Journal of Econometrics*, 3, 297–318.



- Friedl, H. (1998). *Computer statistics*. Lecture Notes, Graz University of Technology.
- Hawkins, D. (1980). A note on continuous and discontinuous segmented regressions. *Technometrics*, 22, 443–444.
- Hawkins, D. (2001). Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37, 323–341.
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
- Krämer, W., Plomberger, W., and Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica*, 56, 1355–1369.
- Küchenhoff, H. (1997). An exact algorithm for estimating breakpoints in segmented generalized linear models. *Computational Statistics*, 12(2), 235–247.
- Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 11(8), 1–18.
- Maillet, E. (1905). *Essai d'hydraulique souterraine et fluviale*. Librairie scientifique. Hermann, Paris.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.
- Pregibon, D. (1979). *Data analytic methods for generalized linear models*. PhD thesis, Univ. of Toronto.
- Quandt, R. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53, 873–880.
- Quandt, R. and Ramsey, J. (1978). Estimating mixtures of normal distribution and switching regression. *Journal of the American Statistical Association*, 73(364), 730–738.
- Seber, G. and Wild, C. (1989). *Nonlinear Regression*. New York: John Wiley & Sons.
- Stasinopoulos, D. and Rigby, R. (1992). Detecting break points in generalized linear models. *Computational Statistics & Data Analysis*, 13, 461–471.
- Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S-plus*. New York: Springer.
- Wittenberg, H. (1999). Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes*, 13, 715–726.
- Worsley, K. (1982). An improved Bonferroni inequality and applications. *Biometrika*, 69, 297–302.
- Worsley, K. (1983). Testing for a two-phase multiple regression. *Technometrics*, 25(1), 35–42.
- Yao, Y.-C. (1988). Estimating the number of change-points via Schwartz' criterion. *Statistics and Probability Letters*, 6, 181–189.
- Zeileis, A. (2004). Alternative boundaries for CUSUM tests. *Statistical Papers*, 45, 123–131.