Technische Universität Dortmund

Fakultät Statistik

Robust Modelling of Count Data:
Applications in Medicine

Doctoral Thesis

Submitted by

Hanan Abdel kariem Abdel latif Elsaied

From Egypt

Supervisor:

Prof. Dr. Roland Fried

Dortmund, 09.01.2012

To my family
and
to all members, who took part
in the January 25 Revolution in Egypt

ACKNOWLEDGEMENT

First of all I would like to express my deepest gratitude to my supervisor, Professor Dr. Roland Fried for his continuous encouragement, patience, guidance and amiability throughout this work, without which this thesis would not have been accomplished.
In Addition to my supervisor, I am deeply grateful to PD Dr. Sonja Kuhnt, who accepted to be the second referee. Also to Professor Dr. Katja Ickstadt, who accepted to be the heading of my commission.

I would especially like to thank Dr. Christian H. Weiss who gave me the references, which I asked. I am also grateful to Professor Dr. Konstantinos Fokianos, who gave me some comments on my dissertation before the discussion.

I would especially like to thank every one at the statistics department at the university in Dortmund, in particular, Dipl.-Stat. Tobias Liboschik who gave me the functions, which he implemented in R to remove outliers in Chapter 5. And my colleagues at work Arsene Ntiwa, Oliver Morell, Anita Thieler, Katrin Hainke and Dr. Daniel Vogel for their help. My thanks go to Sebastian Krey, who helped me in R, and Sabine Bell, who helped me from the first day I came to Germany. Also I am grateful to Omniah Abdulazim and Nadja Bauer, who I knew lately in our department for their help.

Further, I thank all of my friends who encouraged me all during the years I was working on my thesis. I would especially like to thank Doaa Elaidi, Haiam Elkatry from Egypt and Marina Umar Muchtar, Rohmatul Fajriyah from Andonisa. Also my best friends far away Seham Nabil, Hanaa Shehata, Mona Nazieh, and Professor Dr. Safaa Abdeldayem.

And of course, I am also deeply grateful to all members of my family, my parents Abdel kariem Elsaied, Samia Abou zyed and my brothers Khaled, Sameh who are far away, for their encouragement and support.

Last but not least, I would definitely like to thank my husband Mohamed Elenany and my sons Ziad and Moaz, who stay with me through the difficult moments during this work and for their encouragement and support.

I apologize beforehand to every one that I forgot.

Abstract

M-estimators as modified versions of maximum likelihood estimators and their asymptotic properties play an important role in the development of modern robust statistics since the 1960s. In our thesis, we construct new M-estimators based on Tukey's bisquare function to fit count data robustly. The Poisson distribution provides a standard framework for the analysis of this type of data.

In case of independent identically distributed Poisson data, M-estimators based on the Huber and Tukey's bisquare function are compared to already existing estimators implemented in R via simulations in case of clean data and of additive outliers. It turns out that it is difficult to combine high robustness against outliers and high efficiency under ideal conditions if the Poisson parameter is small, because such Poisson distributions are highly skewed. We suggest an alternative estimator based on adaptively trimmed means as a possible solution to this problem. Our simulation results indicate that a modified version of the R-function glmrob with external weights gives the best robustness properties among all estimation procedures based on the Huber function. A new modified Tukey M-estimator provides improvements over the other procedures which depend on the Tukey function and also those which depend on the Huber function, particularly in case of moderately large and very large outliers. The estimator based on adaptive trimming provides even better results at small Poisson means.

Furthermore, our work constitutes a first treatment of robust M-estimation of INGARCH models for count time series. These models assume the observation at each point in time to follow a Poisson distribution conditionally on the past, with the conditional mean being a linear function of previous observations and past conditional means. We focus on the INGARCH(1,0) model as the simplest interesting variant. Our approach based on Tukey's bisquare function with bias correction and initialization from a robust AR(1) fit provides good efficiencies in case of clean data. In the presence of outliers, the bias-corrected Tukey M-estimators perform better than the uncorrected ones and the conditional maximum likelihood estimator. The construction of adequate Tukey M-estimators or the development of other robust estimators for INGARCH models of higher orders remains an open problem, albeit some preliminary investigations for the INGARCH(1,1) model are presented here.

Some applications to real data from the medical field and artificial data examples indicate that the INGARCH(1,0) model is a promising candidate for such data, and that the issue of robust estimation tackled here is important.

Keywords: Count data; Poisson model; INGARCH models; GLM models; Huber M-estimator; Tukey M-estimator; Robustness; Asymptotic properties; Medical applications.

Contents

List of Figures

List of Tables

List of abbreviations

Abbreviation	Meaning	Page
ABP	Asymptotic breakdown point	12
AO	Additive outlier	110
ARE	Asymptotic relative efficiency	23
BP	Breakdown point	12
condlikest	Conditional maximum likelihood function for INARCH(1) model	72
condlikest11	Conditional maximum likelihood function for INGARCH(1,1) model	72
FBP	Finite breakdown point	12
GES	Gross-error sensitivity	14
glmrob with external weights	glmrob in the R-package robustbase using external weight function	28
huberpois	Huber estimator for i.i.d. Poisson data	24
Hubercorr	Huber estimator with bias correction for INARCH(1) model	134
Huberuncorr	Huber estimator without bias correction for INARCH(1) model	75
tukeypois	Tukey estimator for i.i.d. Poisson data	26
IF	Influence function	13
INGARCH	Integer-valued generalized autoregressive conditional heteroscedasticity model	59
LS	Level shift outlier	110
poissmall	Initial estimator for i.i.d. Poisson data	24
RMSE	Root mean square error	29
trimmeanfit	Adaptive trimming estimator	54
robPoisTShuber1	Huber function without bias correction and with independence initialization for INARCH(1)	72
robPoisTShuber2	Huber function without bias correction and with initialization from the AR(1) for INARCH(1)	72
robPoisTShuber3	Huber function with bias correction and with initialization from the AR(1) for INARCH(1)	72
robPoisTStukey1	Tukey function without bias correction and with independence initialization for INARCH(1)	72
robPoisTStukey2	Tukey function without bias correction and with initialization from the AR(1) for INARCH(1)	72
robPoisTStukey3	Tukey function with bias correction and with initialization from the AR(1) for INARCH(1)	72
robPoisTStukey11	Tukey estimator without bias correction for INGARCH(1,1) model	99
robPoisTStukey22	Tukey estimator without bias correction for INGARCH(1,1) using independent initialization	99
robPoisTStukey33	Tukey estimator with bias correction for INGARCH(1,1) using ARMA(1,1) initialization	100
SC	Sensitivity curve	13
TS	Transient shift outlier	110
Tukeyuncorr	Tukey estimator without bias correction for INARCH(1) model	80
Tukeycorr	Tukey estimator with bias correction for INARCH(1) model	80

Chapter 1

Introduction

Robust statistics provides inference methods, which are not sensitive to unusual observations or other small deviations from ideal models. Finding the best fit to the majority of the data is one of the most important aims of robust methods.
M-estimators are a general class of robust estimators and their asymptotic properties played an important role in the development of modern robust statistics since the 1960s, where the most important property is that for any asymptotically normal estimator exists an asymptotically equivalent M-estimator, see Staudte and Sheather (1990, Page 116).
In general there are two main approaches to finding robust M-estimators described in Peracchi (1990). The first one is Huber's minimax approach, see Huber (1964, 1981). The second one is Hampel's infinitesimal approach, see Hampel (1968) and Hampel et al. (1986). Both approaches assume a parametric model for the observations and try to construct estimators that perform well over a neighborhood of the assumed model. Huber's approach is to consider a neighborhood of the assumed parametric model and then to safeguard within that neighborhood in a minimax sense. This approach also described in Kordzakhia et al. (2001) is based on the minimization of some functional of the likelihood process, namely so-called Huber's ρ functions, see Section 2.1 and Section 2.2. Hampel's approach focuses on the asymptotic behavior of an estimator in an infinitesimal neighborhood of a given model. In our work, we follow the first approach to finding robust estimators, but we use modified versions of likelihood functions, which are suitable for our model and we use estimators computed using the second approach for the purpose of comparison, as we will see later in Chapter 3.
This thesis considers the problem of robust modelling for count data, where the Poisson model provides a standard framework for the analysis of this type of data. It consists of six chapters and the relationship between them is illustrated in Figure 1.1. The first chapter is this introduction and the outline of the other chapters is as follows:

- In Chapter 2, we review some concepts for location M-estimators in robust estimation theory such as their definition, types, properties and computation.

- In Chapter 3, we construct new Tukey M-estimators with bias correction of the Poisson mean in case of i.i.d. data. We propose a new algorithm for estimating the mean of the Poisson distribution, which is based on the Tukey function. We modify the R-function glmrob by adding a bias correction term and external weights. Then we compare modified bias-corrected M-estimators based on the Huber and the Tukey functions to already existing estimators implemented in R via simulation in case of clean and additive outliers data. We will finish this chapter by considering alternative estimators as a solution to the problem of combining high robustness against outliers and high efficiency relatively to the sample mean when the true mean is small.

- In Chapter 4, we introduce robust M-estimation for so called INGARCH models for count time series data in the presence of outliers, where we focus on robust estimation for the INGARCH(1,0), or more briefly INARCH(1), model. We start with the definition and the properties of these models. We apply conditional maximum likelihood as a classical approach to estimate the parameters of these models. We discuss robust estimation of the marginal mean in case of time series data from INGARCH models using our best functions given in Chapter 3 for i.i.d. Poisson data. Then we modify the classical estimation approach by giving robust estimators for the parameters of the INARCH(1) model. We investigate some of the basic properties of these estimators. Afterwards we compute the estimates using some functions, which we have implemented in R, and compare them via simulations in case of clean and contaminated Poisson time series data. We will finish this chapter by trying to extend robust estimation to more general INGARCH models.

- In Chapter 5, we apply our methods proposed in Chapter 4 to two real data examples in the medical field. The first example is the poliomyelitis data. The second example is the campylobacterosis data. We start with an analysis of the poliomyelitis data. We give a description of these data, then we fit an INGARCH(1,0) model to them using conditional maximum likelihood as a non robust method and Tukey M-estimation as a robust method for parameter estimation. After that, we fit an INGARCH(1,0) model using the same methods but after having cleaned the data from outliers. To verify the reliability of our proposed methods, we analyse an artificial data example generated to resemble the poliomyelitis data. We fit an INGARCH(1,0) model to the artificial data using the same methods as for the poliomyelitis data, then we fit an INGARCH(1,0) model again but after having cleaned the artificial data from outliers. For the campylobacterosis data, we repeat what we did for the poliomyelitis data.

- In Chapter 6, we provide a summary, conclusions and an outlook.

Each chapter starts with a short description of its contents. Additionally, we begin Chapter 3 and Chapter 4 with a brief review of the previous treatments in the literature for the topics treated in these chapters.
We use R software version 2.11.1 (2010-05-31). Under R, we use packages "MASS", "robustbase", "dplR" and "ROptEst". Along with this dissertation comes a CD, which contains the .pdf of this document and the codes in R, which we wrote to calculate our estimates, to run our simulations and to plot our figures.

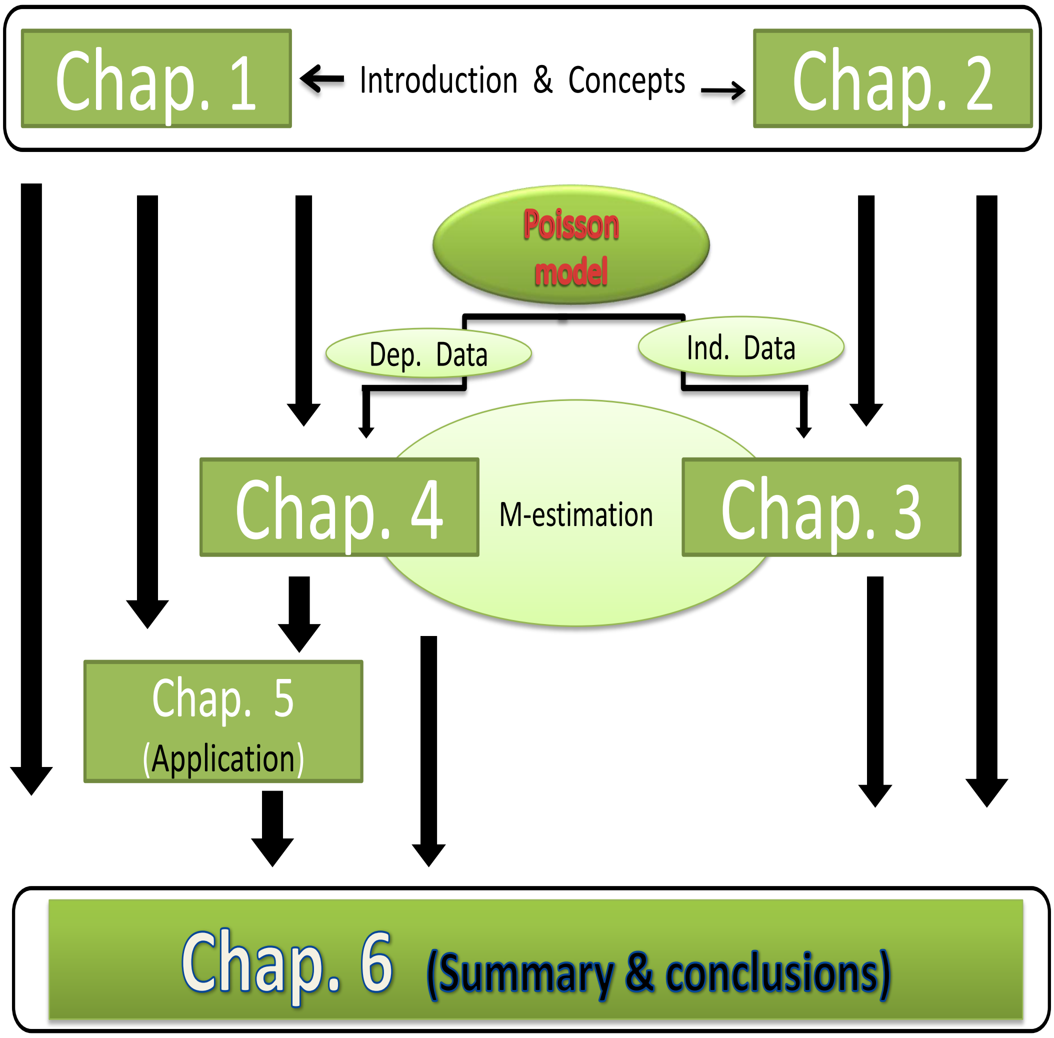


Figure 1.1: Thesis outline

Chapter 2

Basic concepts of location M-estimators

The purpose of this chapter is to present some basic concepts of location M-estimators in robust estimation theory, which we will need afterwards. We start by giving a definition of location M-estimators. Then we present Huber's M-estimators and Tukey's biweight or bisquare M-estimators as different types of M-estimators. Afterwards we review criteria used for studying whether robust estimators have good properties: qualitative robustness, quantitative robustness, and infinitesimal robustness. We finish this chapter by the computation of location M-estimators with previously computed dispersion.

2.1 Definition of location M-Estimators

M-estimators naturally estimate M-measures of location, so we can define them through the following three parts:

- Location model
- Measures of location
- M-Estimators

We first consider the simple *location model* as described in Maronna et al. (2006, Page 17),

$$X_i = \mu + u_i \quad (i = 1, ..., n), \tag{2.1}$$

where the outcome X_i of each observation depends on the true value of μ and on some random error u_i, with $u_1, ..., u_n$ being assumed to be independent and identically distributed random variables with the same symmetric distribution F_0, which is symmetric to 0. It follows that $X_1, ..., X_n$ are independent with common distribution function F, where

$$F(x) = F_0(x - \mu), \quad x \in \mathcal{R} \tag{2.2}$$

A measure μ maps a class $\mathcal{F}$ of distribution functions F onto the real line ($\mathcal{R}$) by constructing $\mu(F)$. According to Staudte and Sheather (1990, Page 101), a measure μ is called *a measure of location* if for any constants a and b and random variable X with distribution F holds:

- $\mu(X + b) = \mu(X) + b$ (location equivariance).
- $\mu(-X) = -\mu(X)$ (symmetry).
- $X \geq 0$ implies $\mu(X) \geq 0$.
- $\mu(aX) = a\mu(X)$ for all $a > 0$.

Bickel and Lehmann (1975) require measures of location to be stochastic order preserving, so they added

- If X is stochastically larger than Y, then $\mu(X) \geq \mu(Y)$.

We can use the classical estimator of μ, e.g., least squares or maximum likelihood methods, assuming the data to come from the same normal distribution, $F_0 \sim N(0, \sigma^2)$, which implies that $F \sim N(\mu, \sigma^2)$ if there are no outliers.
Applying the maximum likelihood method, the likelihood function for a realization $x_1, ..., x_n$ of $X_1, ..., X_n$ is

$$L(x_1, ..., x_n; \mu) = \prod_{i=1}^{n} f_0(x_i - \mu), \tag{2.3}$$

with f_0 being a density of F_0. The maximum likelihood estimate (MLE) of μ is the value $\hat{\mu}$ - depending on $(x_1, ..., x_n)$ - that maximizes $L(x_1, ..., x_n; \mu)$,

$$\hat{\mu} = \hat{\mu}(x_1, ..., x_n) = argmax_{\mu} L(x_1, ..., x_n; \mu), \tag{2.4}$$

where "$argmax$" stands for "the value maximizing the function".
If F was exactly a normal distribution, the MLE which is the sample mean would be an optimal estimator. But if F is only approximately normal, then our goal is an estimator that is almost as good as the mean when F is exactly normal. To achieve this goal, we use modified versions of maximum likelihood methods for estimation, such that $\hat{\mu}$ is close to μ with high probability, e.g., M-estimators.
M-estimators are a broad class of estimators which are obtained as the solution of the problem of minimizing certain objective functions of the data or as the root of a system of equations equating certain functions of the data to 0. Given observations $x_1, ..., x_n$, an M-estimator of a location parameter μ is the minimizer of the following objective function

$$\sum \rho(x_i, \mu) \tag{2.5}$$

where the ρ function measures the agreement between an observation x_i and any possible value of μ. Using $\rho(x, \mu) = -log f_0(x, \mu)$, i.e. the negative logarithm of the model density, gives the maximum likelihood estimator.
If we assume that ρ has a derivative ψ with respect to its second argument, $\psi(x_i, \mu) = \frac{\partial}{\partial\mu}\rho(x_i, \mu)$, then an M-estimator can be defined as the solution of the following equation for $\hat{\mu}$

$$\sum_{i=1}^{n} \psi(x_i, \hat{\mu}) = 0. \tag{2.6}$$

According to Maronna et al. (2006, Page 23), we can calculate $\hat{\mu}$ using the maximum likelihood estimation as a special case of M-estimation for model (2.1) assuming the scale parameter σ to be known, e.g., $\sigma = 1$, as follows:
Since the logarithm is an increasing function if f_0 is everywhere positive, (2.4) can be written as

$$\hat{\mu} = \hat{\mu}(x_1, ..., x_n) = argmin_{\mu} \sum_{i=1}^{n} \rho(x_i - \mu), \tag{2.7}$$

If ρ is differentiable, differentiating (2.7) with respect to μ yields

$$\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0 \tag{2.8}$$

Note that if f_0 is symmetric, then ρ is even and hence ψ is odd.
Generally, we have the following possible solutions of (2.8) according to the type of ψ:

- if ψ is monotone nondecreasing with $\psi(-\infty) < 0 < \psi(\infty)$, a solution to (2.8) exists and all solutions form an interval since ρ is convex.
- If ψ is continuous and strictly increasing, the solution is unique.
- If ψ is discontinuous, a solution to (2.8) might not exist and in this case we shall interpret (2.8) to mean that the left-hand side changes its sign at $\hat{\mu}$.
- If ψ is not a strictly monotone function, then there can be more than one solution of (2.8).

Using different structures of ρ and ψ functions gives different types of estimates of μ, for example the mean and the median:
If $F_0 = N(0, 1)$, then $f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, and apart from a constant we have for the MLE $\rho(x) = \frac{x^2}{2}$ and $\psi(x) = x$, and equation (2.8) becomes

$$\sum_{i=1}^{n} (x_i - \hat{\mu}) = 0 \tag{2.9}$$

which has $\hat{\mu} = \bar{x}$, the sample mean, as its unique solution.
If F_0 is the double exponential distribution, $f_0(x) = \frac{1}{2}e^{-|x|}$, then for the MLE $\rho(x, \mu) = |x - \mu|$, $\psi(x, \mu) = sign(x - \mu)$, and any sample median of $x_1, ..., x_n$ will be a solution of (2.8), what is equivalent to

$$\sum_{i=1}^{n} sign(x_i - \hat{\mu}) = 0 \tag{2.10}$$

Under normality the sample mean is the most efficient estimator for μ, while the median has asymptotic efficiency $\frac{2}{\pi} \approx 64\%$, since the asymptotic variance of the sample median is $\pi/2$, see Maronna et al. (2006, Page 26). On the other hand, the median is a robust measure of central tendency while the mean is not.
Now our problem is, how can we choose appropriate ρ or ψ functions, which give us the best compromise between efficiency and robustness? One of the most popular ψ functions is the Huber function

$$\psi(r_i) = \begin{cases} r_i, & |r_i| \leq k \\ k \cdot sign(r_i), & |r_i| > k \end{cases} \tag{2.11}$$

where for $i = 1, ..., n$, $r_i = x_i - \mu$ is the residual of x_i and k is a tuning constant to be determined suitably.
The calculation of the Huber location estimate defined in (2.11) is commonly done by iteratively reweighted least squares (IRWLS), derived from writing the solution of equation (2.8) as a weighted mean with weights depending on the distances between the data points and the current solution as follows:
Define the weight function $W(r_i)$

$$W(r_i) = \psi(r_i)/r_i = \begin{cases} 1, & |r_i| \leq k \\ k/|r_i|, & |r_i| > k \end{cases} \tag{2.12}$$

Rewrite (2.8) as

$$\sum_{i=1}^{n} W(r_i)(x_i - \hat{\mu}) = 0 \tag{2.13}$$

so that

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \omega_i x_i}{\sum_{i=1}^{n} \omega_i} \tag{2.14}$$

with $\omega_i = W(r_i)$. If $W(r_i)$ is bounded and nonincreasing for $r_i > 0$, IRWLS converges to a solution of (2.8).
For Huber's ψ function, choosing a larger value of k increases the efficiency but reduces the robustness to outliers. Now our problem becomes how can we choose a reasonable value of k? If the model distribution F is a normal distribution with a unit scale, it is reasonable to choose the tuning constant k of the Huber function within the interval $[1, 3]$, since such distributions rarely generate values with distances from the mean larger than 3 (standard deviations), whereas all values within the range $[-1, 1]$ are typical.
We will give later on more details about ψ functions and the results of applying these functions for different choices of the tuning constants.

If F depends on an unknown scale parameter σ, we can derive M-Estimators in two ways:

1. with previous estimation of dispersion,
2. simultaneous M-Estimates of location and dispersion.

In these cases, k can be chosen as a corresponding multiple of an estimate $\hat{\sigma}$, which can be calculated a-priori or simultaneously, see Maronna et al. (2006, Page 36). We will discuss these options with some details in Section 2.4.

2.2 Types of M-Estimators

Several types of M-estimators have been developed depending on the choice of ρ or ψ functions. One usually tries to obtain ρ or ψ functions, which lead to some desirable properties. If ρ is differentiable, the computation of the estimate $\hat{\mu}$ is usually easier. If the derivative ψ of ρ is continuous and strictly monotone increasing, there is a unique solution like in case of Huber's ψ function. When ψ is not monotone a solution of (2.8) is called a redescending M-estimator. According to Staudte and Sheather (1990, Page 118), redescending M-estimators are popular since they have some additional desirable properties. Their ψ functions are non decreasing near the origin but decrease for large arguments. Many of them satisfy $\psi(x) = 0$ for all x with $|x| \geq k$, where k is a finite number which is called the minimum rejection point.
Another property is that they can be quite efficient and have a high breakdown point if we find the solution by iteration, beginning with an initial estimator with a high breakdown point, and unlike other outlier rejection techniques they do not suffer from masking effects. Their efficiency is due to the fact that they completely reject large outliers, but use the exact values of all reasonable observations, as opposed to the median. This is because their ψ function is chosen to redescend smoothly to 0.
Examples for this type of estimators:

- Hampel's three part M-estimators
- Tukey's biweight or bisquare M-estimators

Table 2.1 and Figure 2.1 show the different ρ, ψ and ω (weights) functions, for the Huber, Hampel and Tukey proposals, which are the most popular ones.

From Table 2.1, we find that Hampel's ψ function is more complicated than the Huber and Tukey functions since it needs fixing three tuning constants a, b and c instead of only one constant. So in our study, we will concentrate on the Huber function and the Tukey function, which have been successfully used in a wide variety of applications. Under the normal model with a unit scale, Tukey's ψ function needs larger values of k between 3 and 5, because k does not limit the range of typical, but the range of plausible observations generated from the normal model.

Table 2.1: ρ, ψ **and** ω **functions for Huber, Tukey and Hampel proposals**

Criterion	$\rho(r_i)$	$\psi(r_i) = \frac{\partial}{\partial\mu}\rho(r_i)$	$\omega(r_i) = \psi(r_i)/r_i$	range
Huber	$\frac{1}{2}r_i^2$	r_i	1	$\|r_i\| \leq k$
	$k\,\|r_i\| - \frac{1}{2}k^2$	$k \cdot sign(r_i)$	$k/\,\|r_i\|$	$\|r_i\| > k$
Tukey's biweight	$1 - [1 - (\frac{r_i}{k})^2]^3$	$r_i[1 - (\frac{r_i}{k})^2]^2$	$[1 - (\frac{r_i}{k})^2]^2$	$\|r_i\| \leq k$
	1	0	0	$\|r_i\| > k$
Hampel	r_i^2	r_i	1	$\|r_i\| \leq a$
	$2a\,\|r_i\| - a^2$	$a \cdot sign(r_i)$	$\frac{a}{\|r_i\|}$	$a < \|r_i\| \leq b$
	$\dfrac{a(2c\,\|r_i\| - r_i^2)}{c - b} - \dfrac{7}{6}a^2$	$\dfrac{a \cdot sign(r_i)(c - \|r_i\|)}{c - b}$	$\dfrac{a}{\|r_i\|} \cdot (c\,\|r_i\|)/(c - b)$	$b < \|r_i\| \leq c$

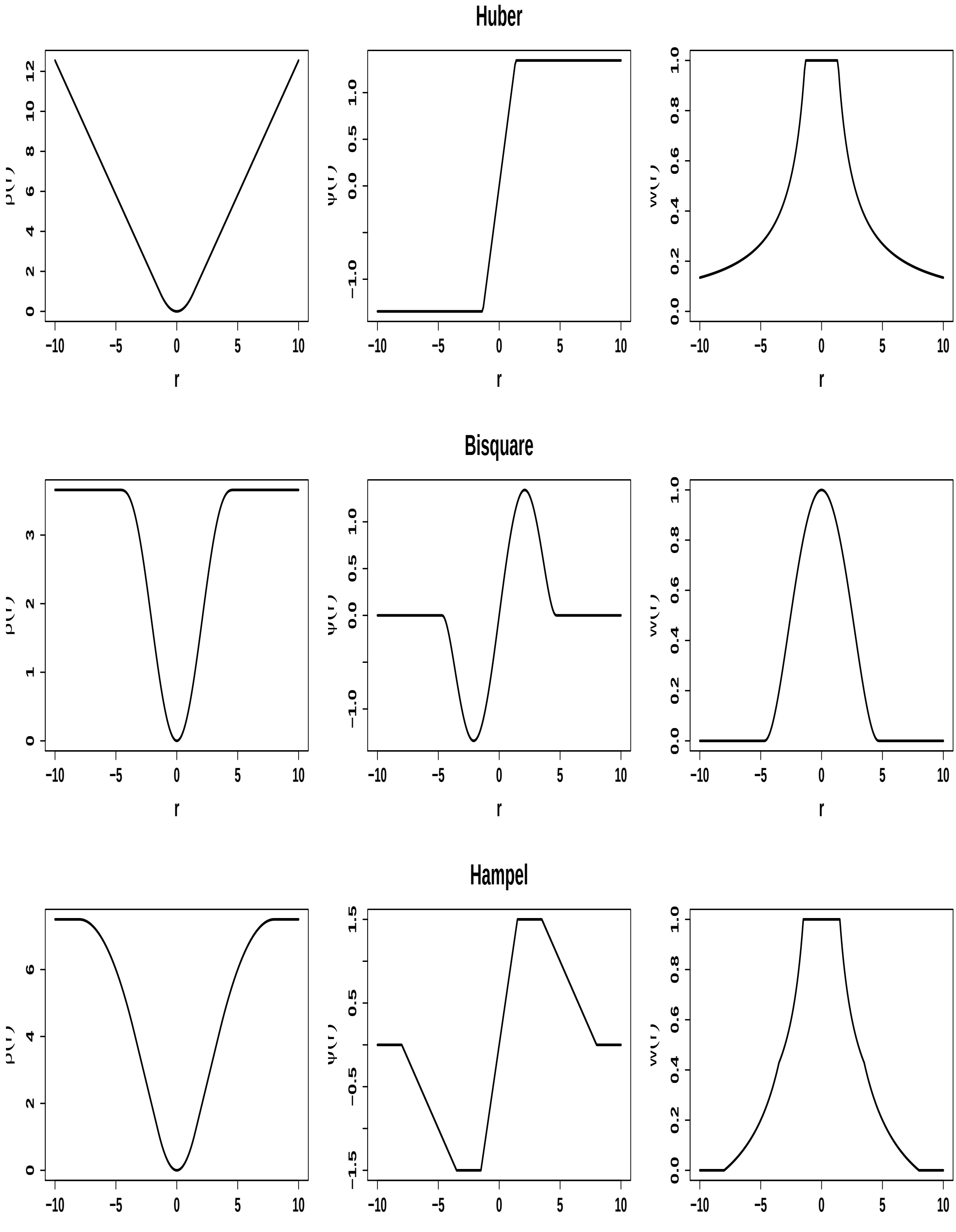


Figure 2.1: ρ , ψ and ω functions (from left to right) for Huber, Tukey and Hampel proposals

2.3 Properties of M-Estimators

There are three basic concepts used to establish whether robust estimators have good properties:

- Qualitative robustness
- Quantitative robustness
- Infinitesimal robustness

Qualitative robustness:
The definition of qualitative robustness is very closely related to continuity of the statistic in the weak topology viewed as a functional. Hampel et al. (1986, Page 99) relate continuity and qualitative robustness with each other, where they note that qualitative robustness is closely related, but not identical with, a nonzero breakdown point. Hampel (1968) calls an estimator qualitatively robust if its sampling distribution is equicontinous. That is, roughly stated, a small change in distribution of the observations should cause only a small change in the distribution of the estimator.

Quantitative robustness (global reliability):
The general idea here is to measure quantitatively how the effect of a small change in the underlying distribution F changes the distribution of an estimator or statistic $\hat{\theta}$. The breakdown point (BP) addresses this property. The BP concepts have been presented by Maronna et al. (2006, Page 58) as follows: Let $\mathbf{X} = (x_1, ..., x_n)$ be a data set and let $\hat{\theta}_n$ be an estimator of the parameter θ, which belongs to a given parameter space Θ. The finite-sample breakdown point (FBP) is the largest fraction of contamination m/n, such that $\hat{\theta}_n$ is bounded away from the boundary of Θ, if at most m data points are changed arbitrarily. More formally: Let $d : \Theta \times \Theta \to \mathcal{R}_0^+$ be a distance measure and call $\mathcal{N}(\mathbf{X}, m) = \{\mathbf{Z} = (z_1, ..., z_n) : \#\{i : z_i \neq x_i\} = m\}$. Then the FBP of $\hat{\theta}$ at the sample $\mathbf{X}$ is

$$\epsilon_n(\hat{\theta}, \mathbf{X}) = \frac{1}{n} \max\{m : \exists K, d(\hat{\theta}_n(\mathbf{Z}), \hat{\theta}_n(\mathbf{X})) < K \quad \forall \mathbf{Z} \in \mathcal{N}(\mathbf{X}, m)\}. \tag{2.15}$$

The asymptotic contamination breakdown point (ABP) of an estimate $\hat{\theta}$ at F, denoted by $\epsilon^*(\hat{\theta}, F)$, is the largest fraction $\epsilon \in (0, 1)$ of contamination, such that $\hat{\theta}$ is bounded away from the boundary of Θ for all distributions in the corresponding contamination neighborhood. More formally: Let $d : \Theta \times \Theta \to \mathcal{R}_0^+$ be a distance measure and $\mathcal{N}(F, \epsilon) := \{(1 - \epsilon)F + \epsilon G, G \in \mathcal{G}\}$ be a contamination neighborhood of F, where $\mathcal{G}$ is a family of contamination distributions. Then

$$\epsilon^*(\hat{\theta}, F) = \sup\{\epsilon > 0 : \exists K \in \mathcal{R}, \ \ d\Big(\hat{\theta}(F), \hat{\theta}((1 - \epsilon)F + \epsilon G)\Big) < K \quad \forall G \in \mathcal{G}\} \tag{2.16}$$

In most cases of interest, the FBP does not depend on $\mathbf{X}$ and tends to the ABP when $n \to \infty$.

Note that both the sample mean and the sample variance have an FBP and ABP of 0 if the parameter space is $\mathcal{R}$. On the other hand, the median has a breakdown point of 50% asymptotically if the parameter space is $\mathcal{R}$, because at least half of the observations need to be moved arbitrarily far away before the median becomes completely wrong.
For location M-estimators with monotonic but not necessarily odd ψ function, the breakdown point is

$$\epsilon^*(\hat{\theta}; F) = \frac{\min(k_1, k_2)}{k_1 + k_2} \tag{2.17}$$

where $k_1 = -\psi(-\infty)$ and $k_2 = \psi(\infty)$ are the limits of ψ as its argument goes to $-\infty$ or ∞. If ψ is odd as for the Huber M-estimator, then $k_1 = k_2$ and $\epsilon^* = 0.5$. For all practical purposes the bisquare M-estimator with previous scale (median absolute deviation, MAD) has $\epsilon^* = 0.5$, while for simultaneous estimation, ϵ^* is usually lower than 0.5, see Maronna et al. (2006, Page 60).

Infinitesimal robustness (local stability)
Infinitesimal robustness shows us what happens if we add one more observation with value x_0 to a large sample. The Influence Function (IF) measures the effect of an infinitesimal perturbation, where IF in its finite sample version is known as the Sensitivity Curve (SC). The SC is defined in Maronna et al. (2006, Page 55) as follows: Let $x_1, ..., x_n$ be an i.i.d. sample from a distribution F and $\hat{\theta}_n$ an estimator of a parameter θ, then

$$SC_{\hat{\theta}_n}(x_1, ..., x_n, x_0) = (n+1)[\hat{\theta}_{n+1}(x_1, ..., x_n, x_0) - \hat{\theta}_n(x_1, ..., x_n)]$$

$$= \frac{\hat{\theta}\Big((1 - \frac{1}{n+1})F_n + \frac{1}{n+1}\delta_{x_0}\Big) - \hat{\theta}(F_n)}{\frac{1}{n+1}}$$

The SC is computed by calculating the estimator $\hat{\theta}_n$ as a function of the empirical distribution ($F_n = \Sigma_{i=1}^n \delta_{x_i}/n$) with and without an observation, where δ_{x_0} is the point-mass at the point x_0, and is proportional to the size of the sample. This means the SC describes the effect of an individual observation on the estimator for a specific data set. When we let the sample size n tend to infinity (asymptotic behavior), we usually have $\hat{\theta}_n \xrightarrow{p} \hat{\theta}_\infty(F)$, where $\hat{\theta}_\infty(F)$ is the asymptotic value of the estimate at F, and the resulting limit is the IF. According to Hampel (1974), the IF of an estimator $\hat{\theta}$ at a distribution F is

$$IF_{\hat{\theta}}(x_0, F) = lim_{\epsilon \to 0} \frac{\hat{\theta}_\infty\Big((1-\epsilon)F + \epsilon\delta_{x_0}\Big) - \hat{\theta}_\infty(F)}{\epsilon}, \tag{2.18}$$

where $\hat{\theta}_\infty$ is a functional on a set of reasonable distributions (a contamination neighborhood of F).
Note that the SC is obtained if F is replaced by the empirical distribution F_n, which assigns mass $\frac{1}{n+1}$ to each of $(x_0, x_1, ..., x_n)$, and setting ϵ to $\frac{1}{n+1}$.
For location M-estimators with bounded and continuous ψ function, Croux (1998) shows that for each x_0

$$SC_{\hat{\theta}_n}(x_0) \xrightarrow{a.s.} IF_{\hat{\theta}}(x_0, F)$$

The IF of an M-estimator is

$$IF_{\hat{\theta}}(x_0, F) = -\frac{\psi(x_0, \hat{\theta}_\infty)}{B(\theta, \psi)} \tag{2.19}$$

with $B(\theta, \psi) = \frac{\partial}{\partial\theta} E\psi(X, \theta)$. Equation (2.19) shows an important connection between the IF and the ψ function of an M-estimator. The IF provides a lot of information on the estimators. According to Maronna et al. (2006, Page 62), the most important one being that the IF gives us a picture of the asymptotic bias caused by a small contamination of size ϵ in the data (robustness stability). To show this: Consider a contamination neighborhood of F_θ with parameter θ, and $\hat{\theta}$ an estimator of θ, such that

$$\mathcal{N}(F_\theta, \epsilon) := \{(1-\epsilon)F_\theta + \epsilon G, G \in \mathcal{G}\},$$

where $\mathcal{G}$ is a family of contamination distributions. Then the asymptotic bias of $\hat{\theta}$ at any $F \in \mathcal{N}(F_\theta, \epsilon)$ is

$$\mathbf{B}(F, \theta) = \hat{\theta}_\infty(F) - \theta$$

and the maximum asymptotic bias (MB) is

$$MB_{\hat{\theta}}(\epsilon, \theta) = \sup\{|\mathbf{B}(F, \theta)| : F \in \mathcal{N}(F_\theta, \epsilon)\}$$

where the maximum bias is related to the IF via

$$MB_{\hat{\theta}}(\epsilon, \theta) \approx \epsilon \cdot \gamma(\hat{\theta}, \theta)$$

for small values of ϵ, where $\gamma(\hat{\theta}, \theta)$ is the gross-error sensitivity (GES) of $\hat{\theta}$ at θ, which is equal to

$$\gamma(\hat{\theta}, \theta) = \max_{x_0} |IF_{\hat{\theta}}(x_0, F_\theta)|. \tag{2.20}$$

If the parameter space is the whole set of real numbers, the relationship between MB and BP is

$$\epsilon^*(\hat{\theta}, F_\theta) = \max\{\epsilon \geq 0 : MB_{\hat{\theta}}(\epsilon, \theta) < \infty\}$$

Note that two estimators may have the same BP but different MBs.
For M-estimators with nondecreasing and bounded ψ function, let $F_\mu(x) = F_0(x-\mu)$, where F_0 is symmetric about 0, $k = \psi(\infty)$ and $\epsilon < 0.5$, then the maximum bias is the solution b_ϵ of the following equation

$$E_{F_0}\psi(x, b_\epsilon) = \frac{k\epsilon}{1-\epsilon}, \tag{2.21}$$

see Maronna et al. (2006, Page 79).
Secondly, Martin (1978) shows that the IF provides an intuitively appealing representation for a robust estimator from which the asymptotic distribution of the estimate may be formally deduced. Under some regularity conditions, $\hat{\theta}_n$ may be represented as

$$\hat{\theta}_n = \theta + \frac{1}{n}\sum_{i=1}^{n} IF_{\hat{\theta}}(x_i, F) + R_n \tag{2.22}$$

and often the remainder term satisfies $n^{\frac{1}{2}}R_n \xrightarrow{d} 0$. Then the difference

$$\sqrt{n}\{[\hat{\theta}_n - \theta] - \tfrac{1}{n}\textstyle\sum_{i=1}^{n} IF_{\hat{\theta}}(x_i, F)\}$$

converges to zero in probability, so that we have the approximation

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{a}{\sim} \mathcal{N}(0, V_{\hat{\theta}}(\theta))$$

i.e. $\hat{\theta}_n - \theta$ is asymptotically normal with parameters

$$E_F[IF_{\hat{\theta}}(X, F)] = 0 \text{ (i.e. } E(\hat{\theta}_n) \to \theta)$$

and

$$V_{\hat{\theta}}(\theta) = E_F(IF_{\hat{\theta}}(X, F))^2 \tag{2.23}$$

This formula shows that a bounded IF implies a bounded asymptotic variance.

Maronna et al. (2006, Page 64) define the asymptotic relative efficiency (ARE) of $\hat{\theta}$ at θ as the ratio of variances as follows:

$$ARE(\theta) = V_{\min}(\theta)/V_{\hat{\theta}}(\theta) \tag{2.24}$$

where $V_{\min}(\theta)$ is the smallest possible asymptotic variance within a reasonable class of estimators (e.g. equivariant ones) or $V_{\min}(\theta) = \min\{V_{\hat{\theta}}(\theta) : \hat{\theta}$ is a "reasonable" estimator of $\theta\}$. Under reasonable regularity conditions $V_{\min}(\theta)$ is usually achieved by the MLE.
At the normal model, M-estimators with the Huber function (2.11) have an ARE larger than redescending M-estimators if we choose the constants to obtain the same maximal bias. For several symmetric, wider tailed distributions, suitable redescending M-estimators are slightly more efficient than M-estimators with the Huber function, and for the Cauchy distribution redescending M-estimators are much more efficient (about 20% more) than the Huber estimator. This is because they completely reject large aberrant observations, while the Huber estimator effectively treats them like moderate outliers, see Staudte and Sheather (1990, Page 119).

If F does not belong to the family F_θ but is in a neighborhood of F_θ, $F \in \mathcal{N}(F_\theta, \epsilon)$, the bias will dominate the variance component of MSE for n large since $(\hat{\theta}_n - \theta) \overset{a}{\sim} \mathcal{N}(b, w/n)$ with $b = \hat{\theta}_\infty(F) - \theta$ being the asymptotic bias and w the asymptotical variance of $\hat{\theta}_n$ at F, i.e. the variance of $\hat{\theta}_n - \theta$ goes to zero while the bias does not. Thus we must balance the asymptotic efficiency and asymptotic bias using different approaches, which are given in Maronna et al. (2006, Page 64). One approach to achieve this is to minimize the maximum bias for a fixed efficiency. For example among Huber and Tukey bisquare M-estimators with previous MAD scale which have BP=0.5, we choose k to fix a certain efficiency and compare the maximum biases thereafter. Hampel (1974) states another approach to balance the problem between bias and efficiency as minimizing the asymptotic variance under the constraint that the gross-error sensitivity (GES) is bounded. A criterion for the construction of optimally robust M-estimators is to use the ψ function that minimizes $tr(V_{\hat{\theta}}(\theta))$ under the constraint that $\gamma(\hat{\theta}, \theta)$ is bounded by a finite constant k. An optimal choice of ψ is

$$\psi^{opt}(x, \mu) = \Big(s(x, \mu) - a\Big) min\Big[1, \frac{k}{\parallel s(x, \mu) - a \parallel}\Big] \tag{2.25}$$

where $s(x, \mu)$ denotes the score function for μ or the objective function and a satisfies $E(\psi^{opt}(X, \mu)) = 0$. The term $min[\quad]$ in (2.25) is often called the robust weight, see Simpson et al. (1987).
When working with parametric models, Hampel (1974) shows that in the class of M-estimators with bounded influence functions, a type of modified log-likelihood function (Huber function) offers highest asymptotic efficiency, and is therefore asymptotically optimal in this sense.

Huber (1972) links the three concepts (qualitative robustness, influence function and breakdown point) to the stability aspects of, say, a bridge: (1) qualitative robustness - a small perturbation should have small effects; (2) the influence function measures the effects of infinitesimal perturbations; and (3) the breakdown point tells us how big the perturbation can be before the bridge breaks down, see Hampel et al. (1986, Page 42).

Staudte and Sheather (1990, Page 115) summarize some of the desirable properties of M-estimators as follows:

- M-estimators can be tuned to be robust against large proportions of outliers.
- For every asymptotically normal estimator $\hat{\theta}$ there is an equivalent M-estimator (so from the point of view of asymptotic normality only M-estimators need to be studied).
- M-estimators can be chosen to completely reject large outliers maintaining a large breakdown point and high efficiency at the model.
- The IF of an M-estimator is proportional to ψ (from 2.19), hence this function may be chosen to bound the influence of outliers and achieve high efficiency for a particular model.

As an example to show the latter property of M-estimators, returning to (2.9) and (2.10), the ψ functions of the mean and the median are $x_i - \hat{\mu}$ and $sign(x_i - \hat{\mu})$, respectively and their influence functions are proportional to their score functions, so the mean is not B-robust (unbounded influence function), while the median is B-robust (bounded influence function). So one criterion for a robust measure of location is that its influence function is bounded.
However, we note that M-estimators also have some drawbacks, such as they are not in general scale equivariant and algorithms for their computation possibly do not converge if we do not have a good initialization.

2.4 Computation

For many choices of ψ, no closed form solution for the corresponding M-estimator exists and an iterative approach to computation is required, such as a Newton Raphson algorithm. Alternatively, in many cases an iteratively re-weighted least squares fitting algorithm can be applied. According to Maronna et al. (2006, Page 39), we will use the latter algorithm to compute M-estimates with a previously computed dispersion, since in general estimation with a previously computed dispersion is more robust than simultaneous estimation of location and scale as follows:
Location with previously computed dispersion estimation
The weighted average expression (2.14) suggests an iterative procedure, starting with a robust estimate $\hat{\sigma}_0$ of σ and some initial estimate $\hat{\mu}_0$ of μ. For $j = 0, 1, ...,$ given $\hat{\mu}_j$ compute

$$\omega_{j,i} = W\left(\frac{x_i - \hat{\mu}_j}{\hat{\sigma}_0}\right) \quad (i = 1, ..., n), \tag{2.26}$$

where W is the function defined in (2.12). Let

$$\hat{\mu}_{j+1} = \frac{\sum_{i=1}^{n} \omega_{j,i} x_i}{\sum_{i=1}^{n} \omega_{j,i}}. \tag{2.27}$$

If $W(r_i)$ is bounded and nonincreasing for $r_i > 0$, then the sequence $\hat{\mu}_j$ converges to a solution of

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}_0}\right) = 0 \tag{2.28}$$

The algorithm, which requires a stopping rule based on a tolerance parameter ϵ, is thus:

1. Compute $\hat{\sigma}_0$ (for instance, the normalized median absolute deviation, MADN) and $\hat{\mu}_0$ (for instance the sample median, Med(x)).

2. For $j = 0, 1, 2, ...,$ compute the weights (2.26) and then $\hat{\mu}_{j+1}$ in (2.27).

3. Stop when $|\hat{\mu}_{j+1} - \hat{\mu}_j| < \epsilon\hat{\sigma}_0$.

We note that we can use the same algorithm, if we want to estimate location and dispersion simultaneously, by adding another iterative procedure for the scale estimate $\hat{\sigma}_j$.

To combine high breakdown point and high efficiency at the normal distribution, we add a variant of M-estimators called MM-estimators. MM estimators have been introduced by Yohai (1987). These estimators combine an M-estimator μ of location with an S estimator S_n of scale. Here our interest is the location parameter μ, so we treat the scale parameter σ as an unknown nuisance parameter.
The following steps to compute MM-estimators are given in Maronna et al. (2006, Page 124). We start by finding an M-estimate as the solution which minimizes

$$L(x_1, ..., x_n) = \sum_{i=1}^{n} \rho\left(\frac{r_i(\mu)}{\hat{\sigma}}\right) \tag{2.29}$$

where $r_i(\mu) = x_i - \mu$ and $\hat{\sigma}$ is a preliminary scale M-estimator.
Then we can apply the following steps:

1. We compute a consistent initial estimate $\hat{\mu}_0$ with high breakdown point but possibly low normal efficiency (this initial point will also be used to compute the robust scale $\hat{\sigma}$ required to define the M-estimate).

2. Compute a robust scale $\hat{\sigma}$ from the residuals $r_i(\hat{\mu})$.

3. Find a solution $\hat{\mu}$ of $\sum_{i=1}^{n} \psi(\frac{r_i}{\hat{\sigma}}) = 0$ by using an iterative procedure starting at $\hat{\mu}_0$.

For the purpose of comparison we will include the results of MM-estimators in Chapter 3.

There are several functions available for computation of M-estimators with the Huber or the Tukey function implemented in R software version 2.11.1 (2010):
Firstly, for Huber's ψ function we have the following functions available:

1. Function huberM in the R-package "robustbase"
   huberM gives a modified "safe" (and more general) Huber estimator, which is a function of y, $k = 1.5$, weights = NULL, tol = 1e-06, mu , s ,..., where mu is the initial location estimator and s is the scale estimator held constant through the iterations.
   This function is a compatible improvement of huber() in MASS because it returns median() if the median absolute deviation, mad (), equals 0.

2. Function glmrob in the R-package "robustbase"
   glmrob is a function in formula, family, data, weights, ..., method = "Mqle",.... This function is used to fit generalized linear models by robust methods, where formula is a symbolic description of the model to be fitted. Here it is $y \sim 1$ for estimation of the mean of identically distributed data. Weights is an optional vector to be used in the fitting process. The method is maximum quasi likelihood "Mqle", which is used to fit a generalized linear model using Huber's ψ function as described in Cantoni and Ronchetti (2001).

3. Function rlm with method M-estimation in the R-package "MASS"
   rlm is a function in formula, data, weights, ..., psi =c("psi.huber", "psi.hampel", "psi.bisquare", d = 1.345, method = c("M", "MM"), ...).
   This function is used to fit a linear model by robust regression using an M estimator. The formula $y \sim 1$ fits a constant mean. Weights is a vector of prior weights for each case. Fitting is done by iterated re-weighted least squares (IWLS). Selecting method = "MM" ensures the estimator to be highly robust and highly efficient.

Secondly, for Tukey's biweight ψ function the following functions exist:

1. Function tbrm in the R-package "dplR"
   tbrm calculates Tukey's biweight robust mean, a robust average that is unaffected by outliers. It is a function in a numeric vector y and a constant k, where k determines the point at which outliers are given a weight of 0.

2. Function rlm with the methods M-estimation and MM-estimation in the R-package "MASS"
   rlm is described above for Huber's ψ function, but we use psi = "psi.bisquare" instead of psi = "psi.huber".

We will compare the different resulting estimators in Chapter 3 in the context of estimation of the Poisson parameter.

Chapter 3

M-estimation of the Poisson parameter

We start this chapter with some of the previous treatments in the literature in chronological order as follows:
Huber (1964) proposes M-estimation to estimate a location parameter robustly.
Hampel (1968) develops a useful optimality theory for robust M-estimation of a univariate parameter and he conjectures that the optimal M-estimate for the Poisson parameter is asymptotically normal provided that the truncation points of the score function are not integers.
Huber (1981) provides a solid foundation in robustness to both theoretical and applied statisticians. In Chapter 3, he discusses three basic types of estimates: Maximum likelihood type estimates (M-Estimates), linear combinations of order statistics (L-Estimates) and estimates derived from rank tests (R-Estimates), and their qualitative and quantitative robustness properties. He emphasizes M-estimates because of their flexibility and their possibility for generalization. He observes that M-estimators with score functions which are not everywhere differentiable have a non normal limit at certain distributions.
Hampel et al. (1986, Page 92) show that robustness theory is not only for location parameters and symmetric distributions. They use the Poisson model as an example for an asymmetric model, with the Poisson distribution getting closer to symmetry as the mean of the distribution increases. The sample mean is fully efficient at the Poisson model, but its influence function is unbounded. Hence the gross error sensitivity measured by the supremum of the absolute value of the influence function is unbounded, so the estimator is not B-robust (B from "bias").
Simpson et al. (1987) note that the score function for Hampel's optimal M-estimator is not smooth, that is, it is not everywhere differentiable and this can lead to complications in the asymptotic theory when the data are discrete. They show asymptotic non-normality over neighborhoods of Hampel's optimal M-estimators when the underlying distribution is discrete and they propose smooth score functions for the Poisson distribution to retain asymptotic normality.
Cantoni and Ronchetti (2001) propose a robust approach to inference for generalized linear models based on robust deviances, which are natural generalizations of quasi-

likelihood functions. They focus in particular on the estimation for binomial and Poisson models.
Cadigan and Chen (2001) show that the Huber M-estimator for the Poisson mean is reasonably efficient and has negligible bias even for small sample sizes.
Kohl (2005) uses Hampel's approach to finding robust estimators. He proposes optimally robust influence curves as solutions to certain optimization problems based on the maximum mean square error (MSE) criterion. He proofs asymptotic normality of his estimators in the framework of infinitesimal (shrinking at a ratio of $\sqrt{n}$) neighborhood by the smoothness of the underlying L2 differentiable parametric models such as Normal, Gamma, Binomial, Poisson and a suitable estimator construction, see Kohl (2005, Page 64).
Maronna et al. (2006) concentrate on M-estimators which play a major role throughout their book, because of their desirable properties such as the asymptotic normality under the assumptions of Theorem 10.7 on Page 339 in their book.

The remainder of this chapter is organized as follows: We introduce new Tukey M-estimators modified with bias correction. We propose a new algorithm for estimating the mean of the Poisson distribution, which is based on the Tukey function. We modify the function glmrob, implemented in the R package "robustbase", by adding a bias correction term and external weights. Then we compare the results of these estimators to already existing estimators implemented in R software by simulation in case of clean data and data with additive outliers. After that we compare the best procedures obtained for the Huber function and the Tukey function. Finally, we finish this chapter by comparing our best estimators with some alternative estimators suggested as solutions for Huber's and Tukey's problem at small means caused by the strong asymmetry of such Poisson distributions.

3.1 M-estimation using Huber's ψ function

Consider a Poisson random variable Y with independent realizations $y_1, ..., y_n$, with mean θ and probability density function $f_\theta(y) = \frac{e^{-\theta}\theta^y}{y!}$, $y \in \mathbb{N}_0$. An M-estimator $\hat{\theta}$ is the solution of

$$\sum_{i=1}^{n} \psi(y_i, \hat{\theta}) = 0 \tag{3.1}$$

The ψ functions introduced in Chapter 2 are symmetric and thus implicitly rely on a symmetric distribution of the observations. Following Cadigan and Chen (2001), a modified version of (2.25) for Huber's ψ function is

$$\psi_{k,a}(y_i, \theta) = \Big(\frac{y_i - \theta}{\theta^{1/2}} - a\Big) min\Big[1, \frac{k\theta^{1/2}}{|y_i - \theta - a\theta^{1/2}|}\Big] \tag{3.2}$$

where the tuning constant k is chosen to ensure a given asymptotic efficiency and $a = a(\theta, k)$ is a correction term to achieve asymptotical unbiasedness. To estimate θ we need

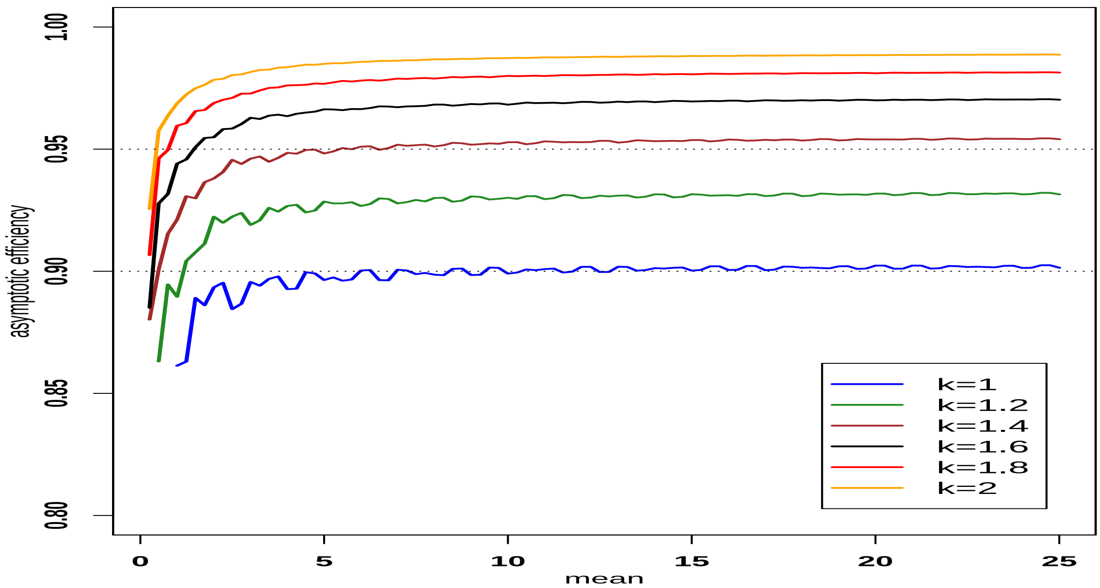


Figure 3.1: Relative asymptotic efficiency of the Huber M-estimator relatively to the sample mean as a function of underlying true mean θ for several tuning constants k.

to solve simultaneously both

$$E\psi_{k,a}(Y,\theta) = 0 \tag{3.3}$$

and

$$\sum_{i=1}^{n} \psi_{k,a}(y_i,\theta) = 0 \tag{3.4}$$

for θ and a. Let $\hat{\theta}_{poi}$ denote this solution. We can calculate the asymptotic efficiency of $\hat{\theta}_{poi}$ relatively to the maximum likelihood estimator of θ (ARE), which is the sample mean, as the ratio of variances as follows:

$$ARE = \theta / V_{\hat{\theta}}(\theta) \tag{3.5}$$

where $V_{\hat{\theta}}(\theta)$ is computed using (2.23), see Cadigan and Chen (2001).
Figure 3.1 illustrates the asymptotic efficiency (3.5) of the Huber M-estimator with different tuning constants $k \in \{1, 1.2, ..., 2\}$ for different true means of the Poisson distribution, $\theta \in \{0.25, 0.5, ..., 25\}$. Obviously, achieving a high relative asymptotic efficiency needs larger values of k if the true mean θ is small. This can be explained by the increasing asymmetry of the Poisson distribution for decreasing θ.

Based on the above formulas (3.2)-(3.4), we implement Cadigan and Chen's estimation algorithm, called "huberpois", in the following to compute Huber M-estimators:

huberpois algorithm
This algorithm can be described by its implementation in R using the following steps:

1. We define "psi.Huber" function as a function in the mean θ, tuning constant k, correction term a, values of the response variable Y. This function calculates the right hand side of (3.2).

2. We define the "optimHuber" function as the sum of the squares of the output of the "psi.Huber" function for all observations $y_1, ..., y_n$.

3. We calculate correction terms $a = a(\theta, k)$ as a function of the mean θ and the tuning constant choosing a certain range of values for the correction term a and equating the expected value of the left hand side of (3.3) under the Poisson distribution to 0, using the function "nullstelleH". For different values of $\theta \in \{0.1, 0.2, ..., 1, 1.2, ..., 5, 5.5, ..., 10, 11, ..., 50\}$ and $k \in \{2, 2.25, ..., 12.5\}$, we store the corrections a in the matrix aHuber.

4. We initialize our estimator using the function "poismall", which is a function in the data sample Y and in $l \in \{0, 1\}$. It calculates an alternative initial estimator based on the relative frequencies of small values or the sample median. We compute the mean θ from the estimate corresponding to the relative frequency either of zeros ($\hat{f}_0$): $\hat{\theta} = -\ln \hat{f}_0$, or of zeros and ones ($\hat{f}_0 + \hat{f}_1$): $\exp^{-\hat{\theta}}(1+\hat{\theta}) = \hat{f}_0 + \hat{f}_1$. For this, we use the uniroot function, which searches the interval from a lower (0) to an upper value (max(2*median, 5)) for a root with respect to the argument θ. Whenever the uniroot function fails to find a root, we use the sample median.

5. We define the "huberpois" function as a function in Y, k, init=poismall(1,0), tolerance. It iterates between estimating the value of θ for a given value of a and the specification of a based on the current estimate of θ.
   First we choose a value of a. Second we estimate θ. Then we derive the value of a which is suitable for this estimate of θ. Then we re-estimate θ by using the optimize function, and iterate this process until the absolute differences for two consecutive iterations of a do not exceed our desired accuracy of 0.0001, which we use to determine convergence.

The new estimates of θ are obtained using the optimize function, which searches the interval from a lower (0.5$*$ estimate) to an upper value (2$*$estimate+1) for a minimum of the function f = optimHuber with respect to the argument θ under the desired accuracy.

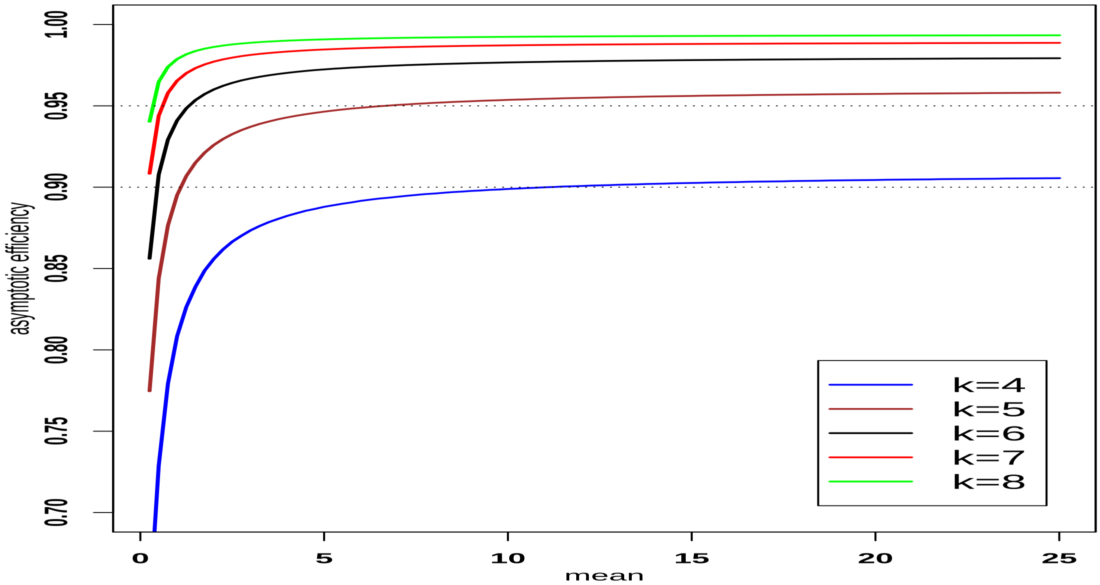


Figure 3.2: Relative asymptotic efficiency of the Tukey M-estimator relatively to the sample mean as a function of underlying true mean θ for several tuning constants k.

3.2 M-estimation using Tukey's ψ function

Analogously to (3.2), Tukey's biweight function also treats positive and negative deviations from the mean symmetrically. Therefore we modify Tukey's ψ function by introducing auto-scaling by the standard deviation $\sqrt{\theta}$ and a bias correction a. The modified Tukey ψ function is

$$\psi_{k,a}(y,\theta) = \left(\frac{y-\theta}{\sqrt{\theta}} - a\right)\left(k^2 - \left(\frac{y-\theta}{\sqrt{\theta}} - a\right)^2\right)^2 I_{[-k,k]}\left(\frac{y-\theta}{\sqrt{\theta}} - a\right), \qquad (3.6)$$

where I_A is the indicator function for a real set A, and $a = a(\theta, k)$ needs to fulfill (3.3). Similarly to Figure 3.1, Figure 3.2 illustrates the asymptotic efficiency (3.5) of the Tukey M-estimator with different tuning constants $k \in \{4, 5, ..., 8\}$. Figure 3.1 and Figure 3.2 illustrate that larger tuning constants lead to higher asymptotic efficiencies. But note that we again have difficulties to achieve a high efficiency at small means θ; we cannot choose tuning constants which guarantee the same desired level of efficiency for all considered means. We will check the validity of these asymptotic results in finite samples of size 100 within our simulation study for both the Huber and the Tukey functions and we will investigate this problem further later on in Section 3.6.

To compute an M-estimator based on (3.6), we define our a new estimation algorithm, called "tukeypois":

tukeypois algorithm
We can describe this algorithm by the following steps using Tukey's ψ function:

1. We define "psiTukey" as a function in the mean θ, tuning constant k, correction term a, values of the response variable Y. This function calculates the right hand side of (3.6).

2. We define the "optimTukey" function as the sum of the squares of the output of the "psiTukey" function for all observations $y_1, ..., y_n$.

3. We calculate a correction term a as a function of the mean θ and the tuning constant, choosing a certain range of values for the correction term a and equating the expected value of the left hand side of (3.3) under the Poisson distribution to 0, using the function "nullstelle". For different values of $\theta \in \{0.1, 0.2, ..., 1, 1.2, ..., 5, 5.5, ..., 10, 11, ..., 50\}$ and $k \in \{2, 2.25, ..., 12.5\}$, we store the values of a in the matrix aTukey.

4. We define "tukeypois" as a function in Y, k, init=poismall(1,0), tolerance, where "poismall" is defined in step 4 of the huberpois function and "tukeypois" iterates between estimating the value of θ for a given value of a and the specification of a based on the current estimate of θ. First we choose a value of a. Secondly we estimate θ. Then we derive the value of a which is suitable for this estimate of θ. Then we re-estimate θ by using the optimize function, and iterate this process until the absolute differences for two consecutive iterations of a do not exceed our desired accuracy of 0.0001, which we use to determine convergence.

In the following Sections 3.3 and 3.4, we perform some simulation experiments to compare the performance of the modified M-estimators based on the Huber function and the Tukey function. We generate contaminated data sets with sample size n=100 considering:

- Poisson distributions with means $\theta \in \{2, 5\}$.
- Different sizes of additive outliers (small = 5, medium = 10, high =30).
- Different percentages of outliers from 1% to 20%.

The results are based on 10000 independent and identically distributed random samples each.

3.3 Comparison of the Huber M-estimators

In this section, we compare several implementations based on Huber's ψ function in terms of bias and root of the mean square error (RMSE), namely:

- huberpois introduced in Section 3.1.
- huberM from the package "robustbase".
- rlm with method "M" and psi=psi.huber from the package "MASS".
- glmrob from the package"robustbase".

Additionally, we modify glmrob by using the following external weight function $Wt(r_i)$

$$Wt(r_i) = \psi_{k,a}(r_i)/r_i = \begin{cases} 1, & |r_i| \leq k \\ k/|r_i|, & |r_i| > k \end{cases} \tag{3.7}$$

where $\psi_{k,a}$ is defined in (3.2) and $r_i = ((y_i - \theta)/\sqrt{\theta} - a)$ is the residual. We compute (3.7) at the same values for the tuning constant k and the bias correction term a, which is suitable for k. We evaluate the bias correction for glmrob using the true mean because we want to know whether a bias correction is effective or not. Afterwards, if we have better results for the new version of glmrob with bias correction, then we will define another new version of glmrob where we iterate between estimating the value of θ and choosing a in the same way as in step 5 of the huberpois algorithm. This is because in practice the true mean is unknown. So we will add to our comparison under the Huber function

- glmrob with external weights and without bias correction term a, and
- glmrob with external weights and with bias correction term a.

For a first bias comparison of the different functions based on Huber's ψ function, we consider a Poisson distribution with mean 2, up to 20% additive outliers of size equal to 5 and tuning constants $k \in \{1, 1.5, 2, 2.5, 3\}$.
Figure 3.3 shows that the smaller the tuning constant the more robust is the estimator, except for the huberM and rlm estimators with small tuning constants in case of small percentages of additive outliers. Additionally, we find noteworthy differences between the different functions.
Figure 3.4 compares the efficiencies of these functions as measured by the percentage mean square error relatively to the sample mean as a function of the true mean of the Poisson, which is varied in $\{1, 2, 3, 4, 5, 6, 7, 10, 15, 20\}$ with the same tuning constants in the case of clean data. We find that for all means larger than 4 bigger tuning constants lead to higher efficiencies, except for the huberpois estimator. Again we note considerable differences between the different functions.

Our question once again is, which one of these functions should we choose?
Criteria for robustness are a small bias and a small root of the mean square error (RMSE) under contaminated data. However, efficiency and robustness are contradicting objectives. So we look for a method which gives a good compromise between robustness and efficiency. A common approach is to fix a certain good efficiency and then look for a method obtaining as much robustness as possible among those methods achieving the fixed level of efficiency. We realize this as follows:

1. We fix the efficiency relatively to the maximum likelihood estimator at a satisfactory level of 90% or 95% for uncontaminated data irrespective of the true mean of the Poisson. Therefore, for each method we try to find tuning constants which achieve for all values of θ at least approximately these levels of efficiencies, see Section 3.3.1.

2. We compare the biases and the root of the mean square error (RMSEs) for our tuning constants chosen in step 1 at means $\theta \in \{2, 5\}$ and sizes of additive outliers being 5, 10, and 30, see Section 3.3.2.

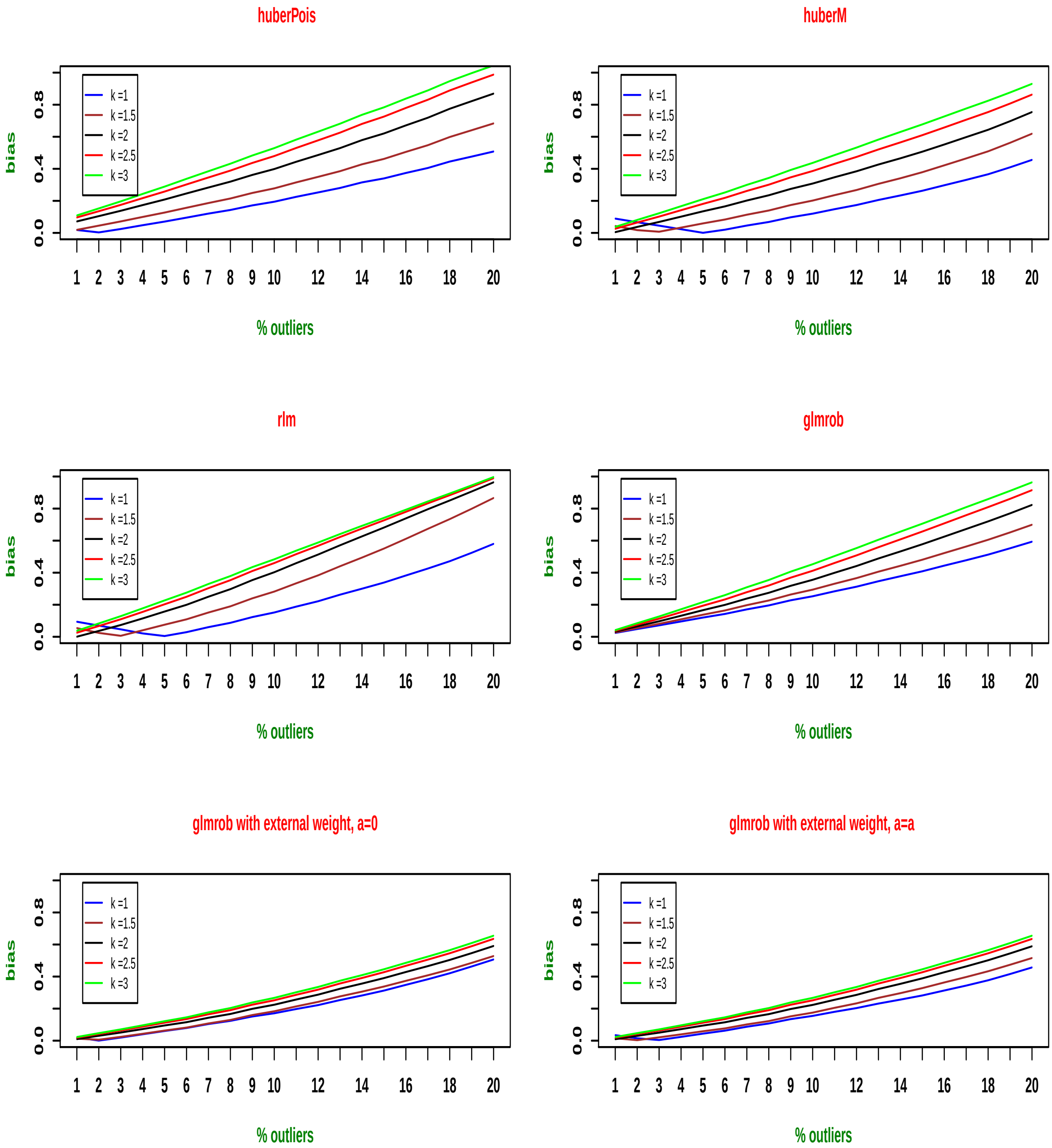


Figure 3.3: Comparison of the sample biases of all Huber procedures with several tuning constants k in case of an increasing percentage of additive outliers of size 5 and a Poisson distribution with mean 2, sample size n=100.

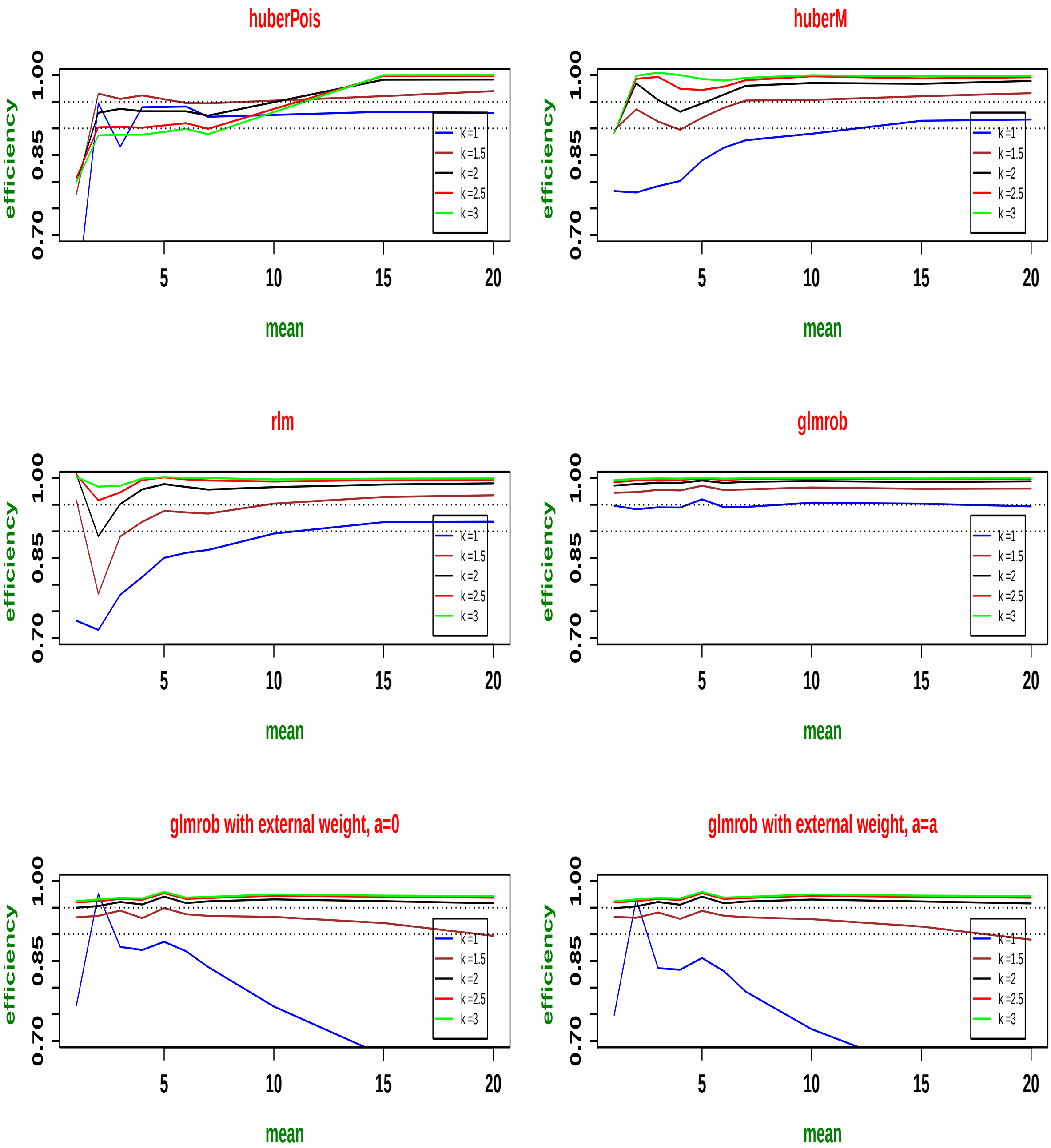


Figure 3.4: Comparison of the relative efficiencies measured by the mean square error of all Huber procedures with several tuning constants k relatively to the sample mean as a function of the underlying true mean, sample size n=100.

3.3.1 Choice of the tuning constant

In this subsection, we search for tuning constants which guarantee levels of efficiencies 90% or 95% over the whole range of values for θ considered here. For this, we generate 10000 samples of size n=100 each for several values of θ and calculate the percentage of the MSE of the sample mean relatively to the sample MSE of the considered estimator.
From Figures 3.4 and Figures 3.5, we find that the huberM and rlm functions have difficulties with reaching a high efficiency in case of a Poisson distribution with a small mean $\theta \leq 4$. huberpois has this difficulty for $\theta \leq 1$ and glmrob with external weights and with or without correction term only for $k = 1$. Therefore we try to choose tuning constants which guarantee the desired level of efficiency at least for most of the values of θ considered here.
Reasonable tuning constants for a 90% level of efficiency are:

1. For huberpois, we choose $k = 1$.
2. For huberM and rlm, we choose $k = 1.1$.
3. For glmrob, we choose $k = 0.8$.
4. For glmrob with external weights with bias correction term ($a \neq 0$) or without bias correction term ($a = 0$), we set $k = 1.5$.

Reasonable tuning constants for a 95% level of efficiency are:

1. For huberpois, we choose $k = 1.8$.
2. For huberM and rlm, we choose $k = 1.6$.
3. For glmrob, we choose $k = 1.1$.
4. For glmrob with external weights with bias correction term ($a \neq 0$) or without bias correction term ($a = 0$), we set $k = 2$.

Note that glmrob in all three versions considered here is the only procedure which achieves the desired level of efficiency over the whole range of values of θ.
If we look again at Figure 3.1, we find that a reasonable tuning constant for an asymptotic efficiency of 90% is 1 for the Huber M-estimator, which is the same as for the huberpois function, and a reasonable tuning constant for an asymptotic efficiency of 95% is 1.4 for the Huber M-estimator, which is less than the chosen tuning constants for all our functions at the same level of efficiency, except glmrob.
The results above are summarized in Figure 3.6, which gives the relative efficiencies with the tuning constants chosen to achieve approximately 90% and 95% levels of efficiencies, respectively.
Further simulation results indicate that the above tuning constants are also appropriate in case of samples of size n=50.

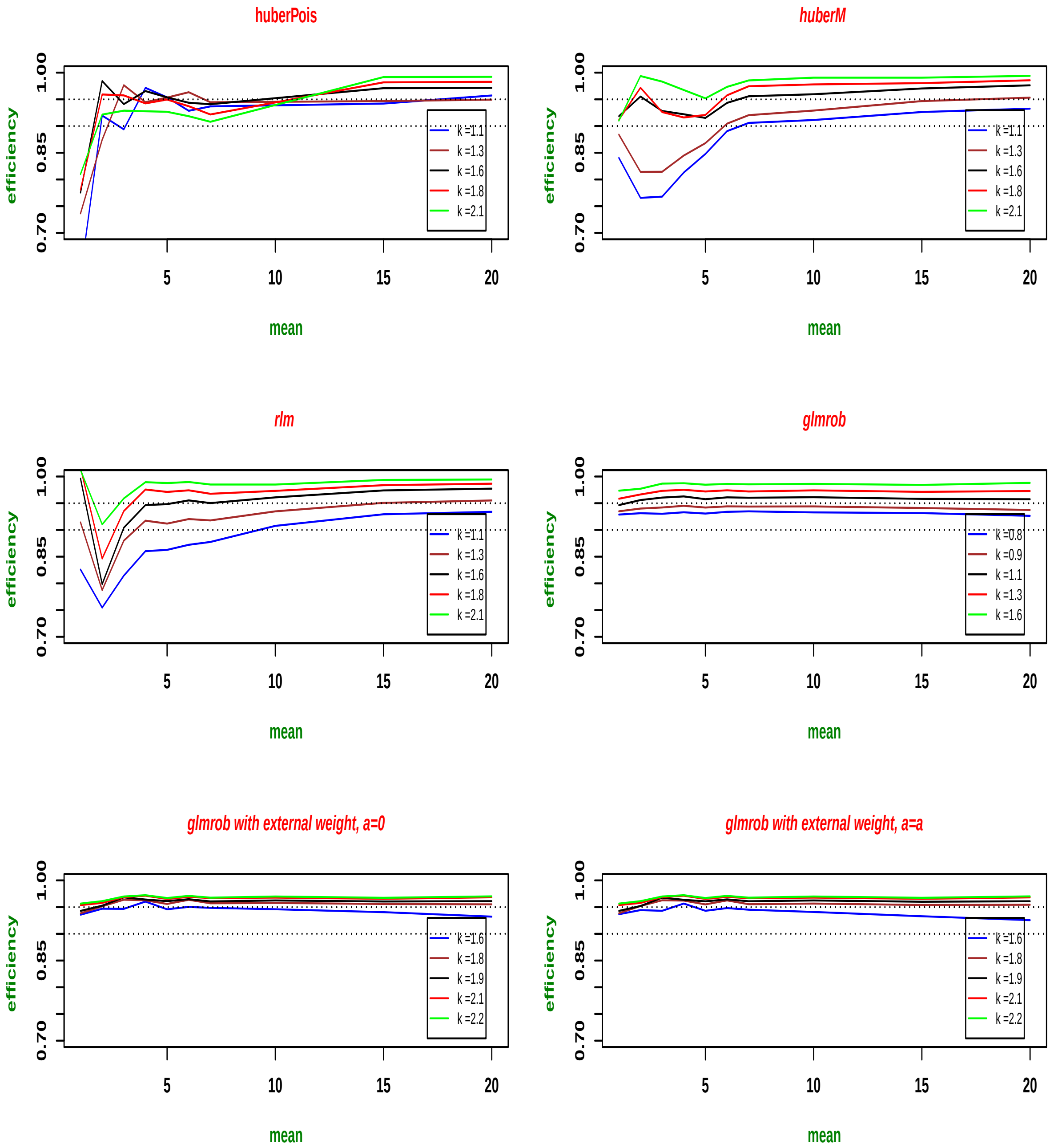
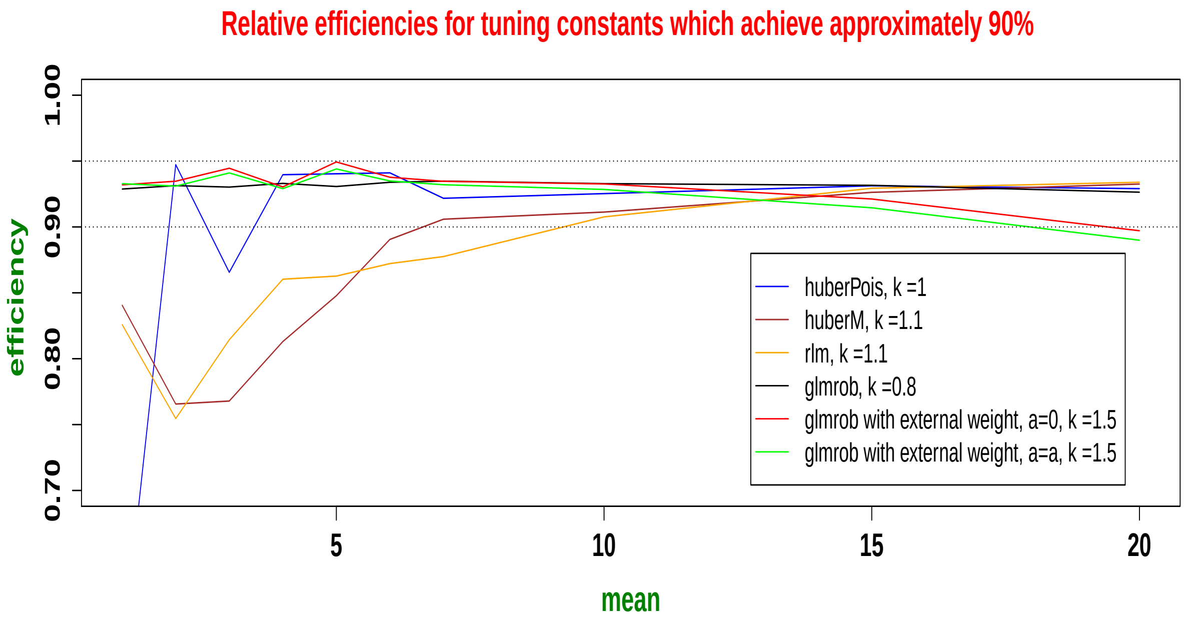


Figure 3.5: Comparison of the relative efficiencies measured by the mean square error of all Huber procedures relatively to the sample mean as a function of the underlying true mean for refined values of the tuning constants k, sample size n=100.



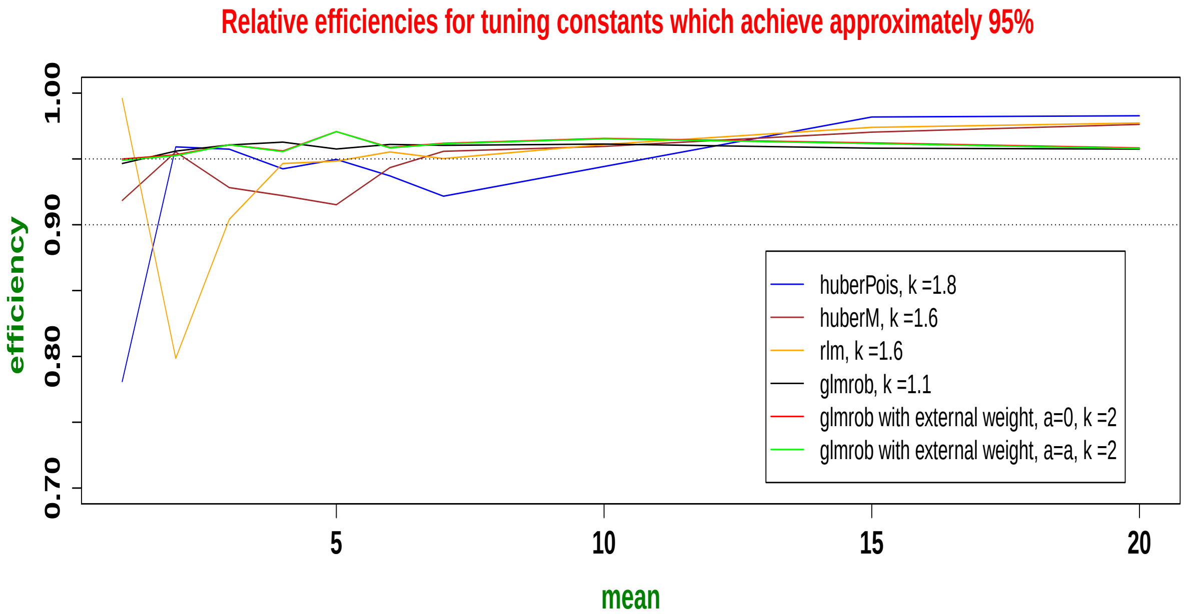


Figure 3.6: Comparison of the relatively efficiencies for the Huber M-estimators at tuning constants, which achieve approximately 90% or 95% level of efficiency (from top to bottom).

3.3.2 Robustness comparison

In this subsection, we compare the biases and the RMSEs under our choices of the tuning constant in Subsection 3.3.1. Figure 3.7 compares the biases of the methods with the tuning constants chosen to achieve a 90% level of efficiency. The Poisson mean is set to 2 and the sizes of the additive outliers to 5, 10 or 30. We find that:

1. At all sizes of the additive outliers the differences between the biases of the several methods are small, but the differences are noticeable at larger percentages of outliers for lager sizes of outliers 10 and 30.

2. rlm has the largest bias after 17% outliers, whereas glmrob with external weights with or without bias correction term (the green line and the red line are identical) has the smallest bias at larger percentages of outliers.

We obtain the same conclusions for the RMSEs since they are dominated by the biases.

Figure 3.8 compares the biases of the methods with the tuning constants chosen to achieve a 90% level of efficiency in case of a Poisson mean equal to 5 and the sizes of the additive outliers being 5, 10 and 30. We get similar results as in Figure 3.7. The main differences are:

1. In case of additive outliers of sizes 10 and 30, glmrob with external weights with or without bias correction term has the smallest bias. The differences between this function and the other functions become larger as the outliers increase.

2. The differences between the biases at the size of the additive outliers 30 are larger than the differences in the biases at the size 10.

For the RMSEs, we get again the same conclusions as for the bias.

Figure 3.9 compares the biases of the methods with the tuning constants chosen to achieve a 95% level of efficiency in case of a Poisson mean equal to 2 and the sizes of the additive outliers being 5, 10 and 30. We find that:

1. In case of additive outliers of size 5, glmrob, glmrob with external weights and huberM have smaller bias than the functions huberpois and rlm.

2. In case of additive outliers of sizes 10 and 30 the differences between the biases of the functions become larger than in case of additive outliers of size 5. For larger outliers, we can order the functions in ascending order with respect to the bias as follows:

   (a) glmrob with or without bias correction term ($a \neq 0, a = 0$),
   (b) followed by default glmrob and huberM,
   (c) then rlm for up to 10% outliers, followed by huberpois,

(d) and finally, after 10% outliers, rlm shows the largest bias.

The RMSE closely matches these findings since again the RMSEs are dominated by the bias.

Figure 3.10 compares the biases of the methods with the tuning constants chosen to achieve 95% level of efficiency in case of a Poisson mean equal to 5 and the sizes of the additive outliers being 5, 10 and 30. We get similar results as for Figure 3.9. The main differences are:

1. The differences between the biases of glmrob and glmrob with external weights with or without bias correction term become larger in case of additive outliers of size 30 than for the smaller additive outliers sizes.

2. We can order the functions in ascending order with respect to the bias in case of additive outliers of sizes 10 and 30 as follows:

   (a) glmrob with or without bias correction term ($a \neq 0, a = 0$),

   (b) followed by glmrob,

   (c) then rlm and huberM for up to 17% outliers, followed by huberpois,

   (d) and finally, after 17% outliers, rlm shows the largest bias.

Again these findings are confirmed with respect to the RMSEs since they are dominated by the bias.
According to the results of this section, glmrob with external weights gives the best results, which are similar irrespective whether the bias correction is used or not. So we do not need to iterate between estimating the value of our parameter and derive the corresponding bias correction in the same way as we did before in step 5 of the huberpois algorithm. From now, we call this function glmrob with external weights.

3.3.3 General conclusions

Altogether, glmrob with the external weights function gives the best robustness among all estimation procedures based on the Huber function for all sizes and percentages of outliers, both at 90% and 95% efficiency, and these results are better than for the ordinary glmrob function. The differences become larger for larger outliers and higher percentages of outliers.

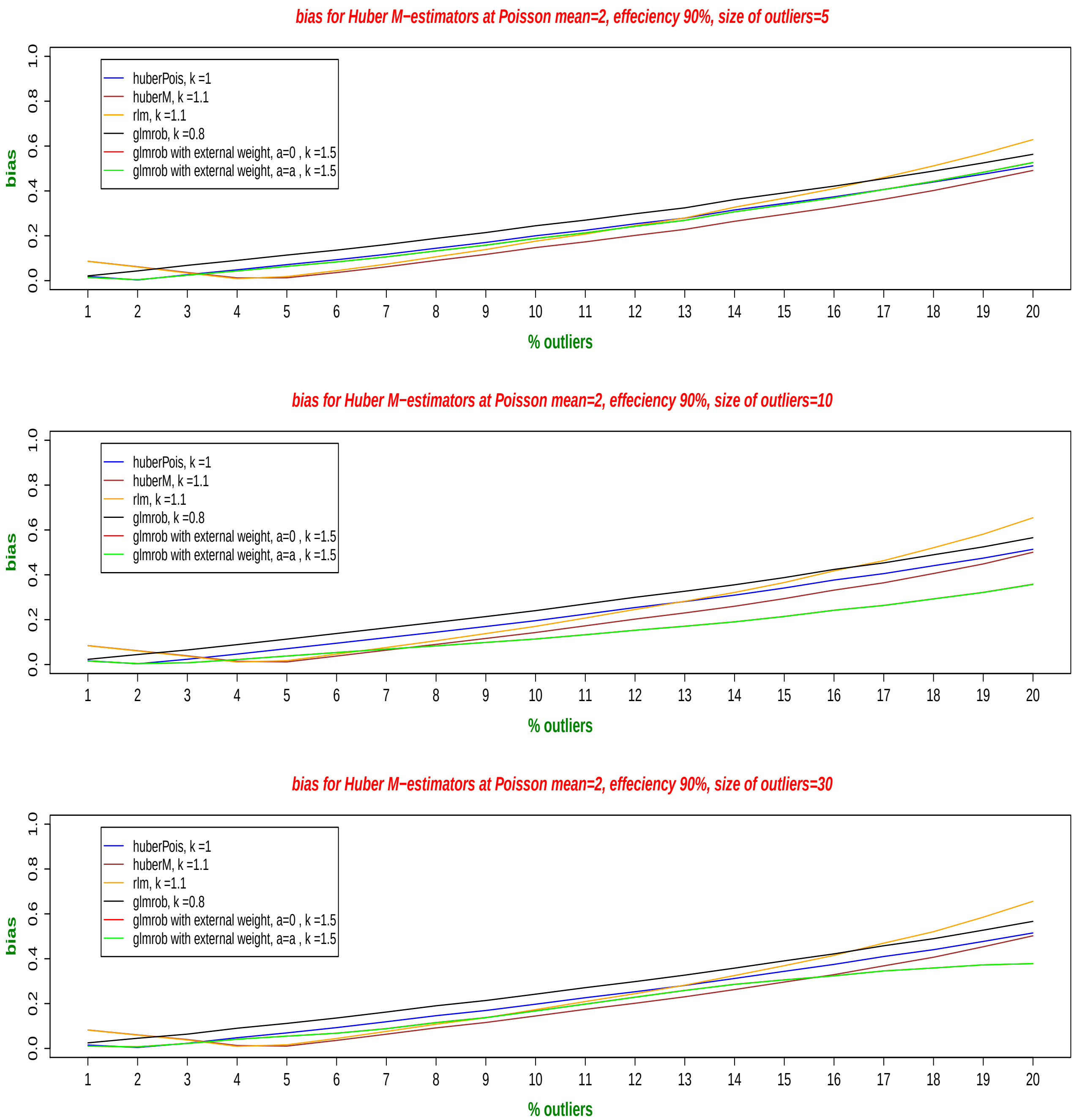


Figure 3.7: Comparison of the biases of the Huber procedures tuned to achieve 90% level of efficiency in case of a Poisson with mean 2 and the sizes of the additive outliers being 5, 10 and 30 (from top to bottom).

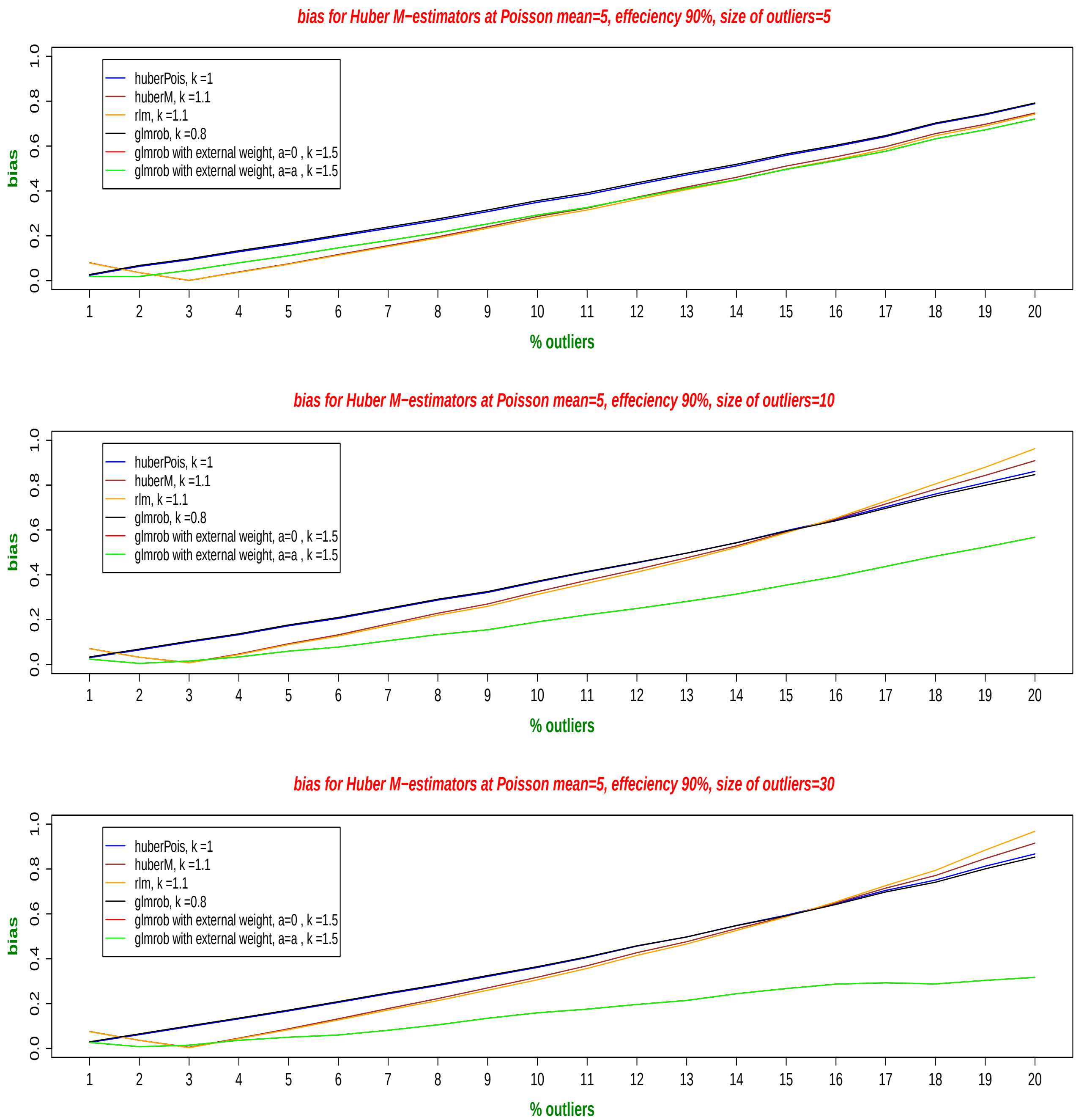


Figure 3.8: Comparison of the biases of the Huber procedures tuned to achieve 90% level of efficiency in case of a Poisson with mean 5 and the sizes of the additive outliers being 5, 10 and 30 (from top to bottom).

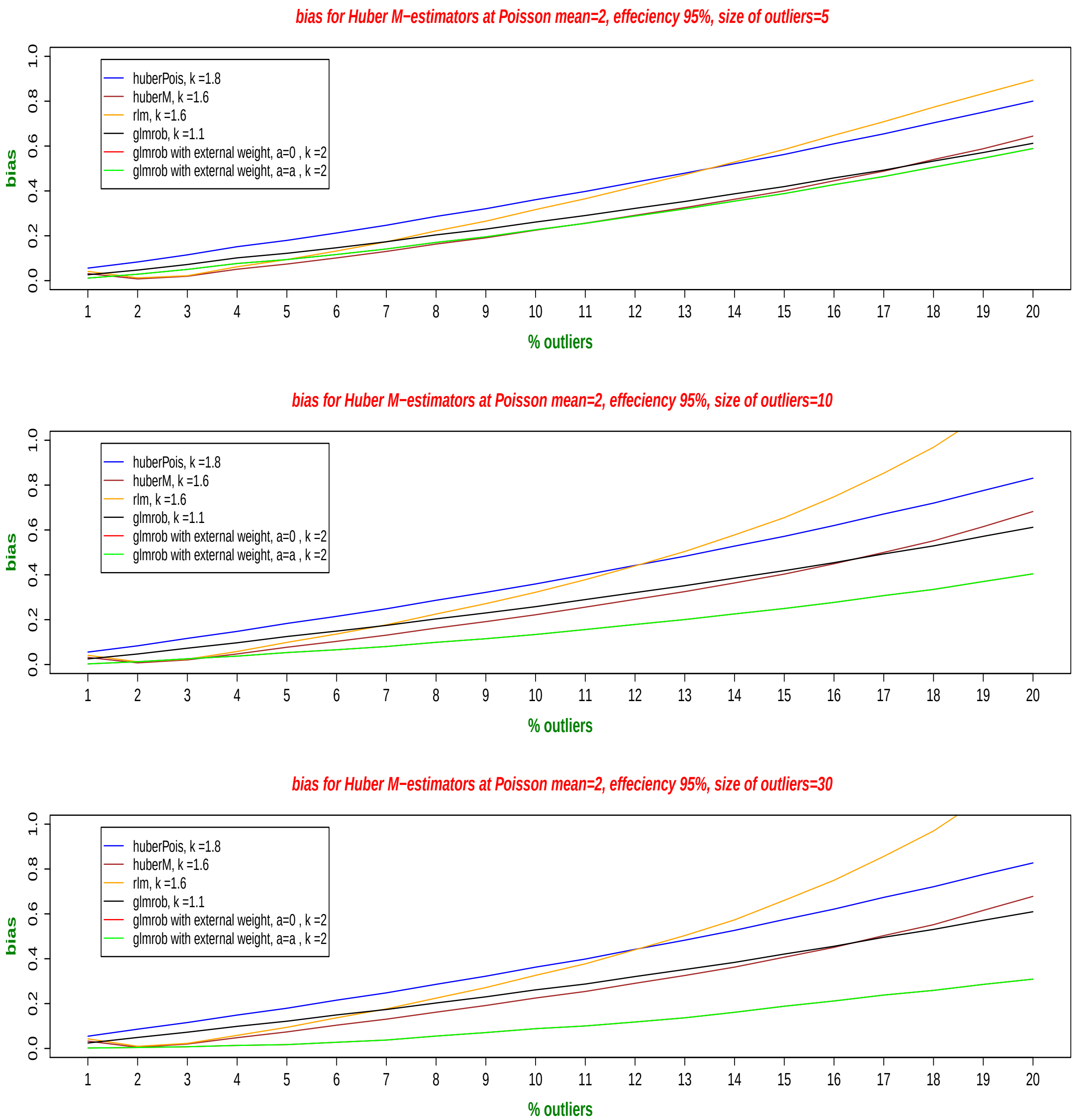


Figure 3.9: Comparison of the biases of the Huber procedures tuned to achieve 95% level of efficiency in case of a Poisson with mean 2 and the sizes of the additive outliers being 5, 10 and 30 (from top to bottom).

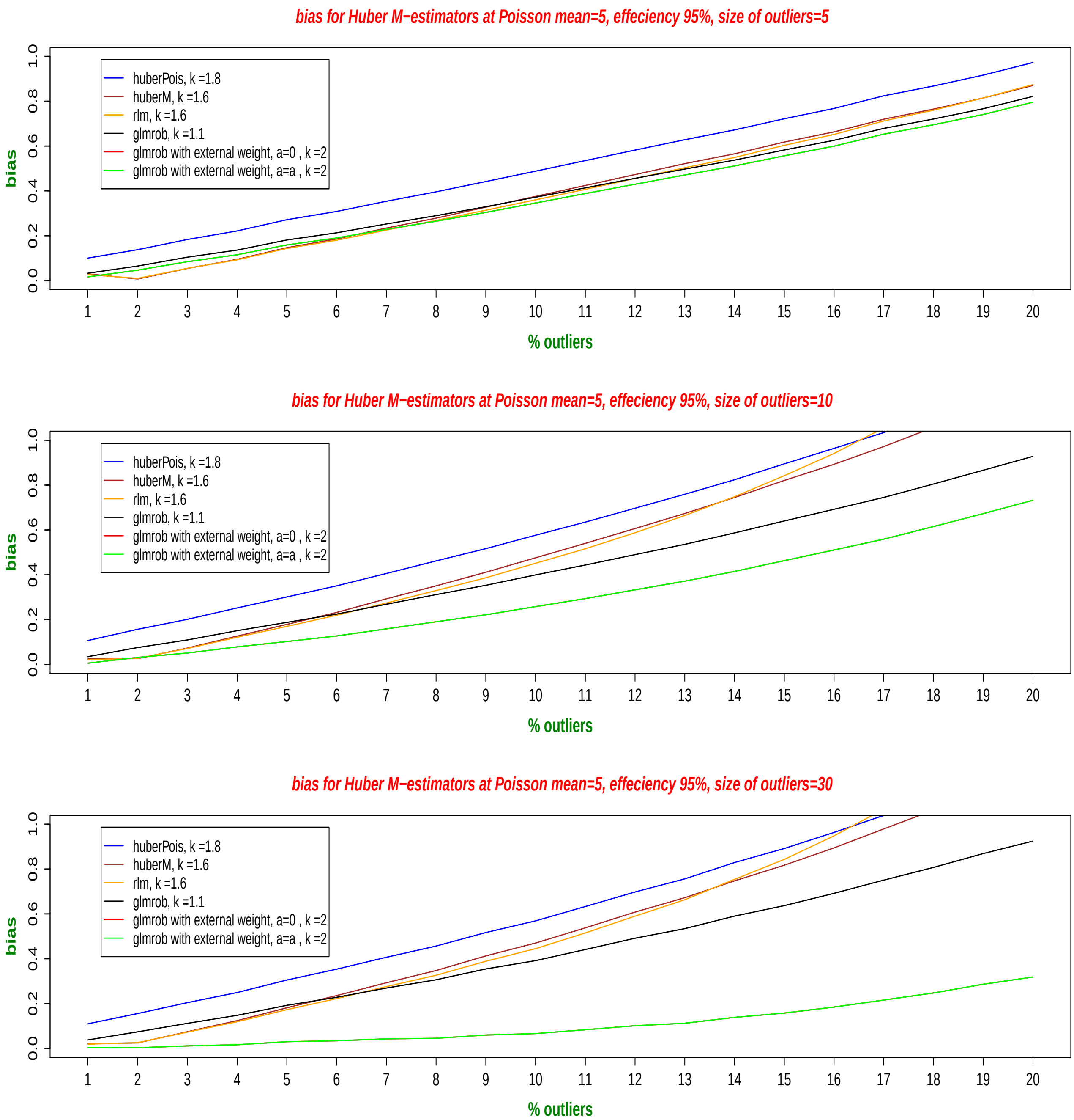


Figure 3.10: Comparison of the biases of the Huber procedures tuned to achieve achieve 95% level of efficiency in case of a Poisson with mean 5 and the sizes of the additive outliers being 5,10 and 30 (from top to bottom).

3.4 Comparison of the Tukey M-estimators

In this section, we compare the performance of the modified M-estimators based on Tukey's biweight function in terms of bias and root of the mean square error (RMSE), namely:

- tukeypois introduced in Section 3.2.
- tbrm from the package "dplR".
- rlm with methods "M","MM" and psi=psi.bisqure from the package "MASS".

Note that we add method "MM" because Tukey's bisquare function works well using this method, see Maronna et al. (2006, page 130). In the R Package "MASS", method "MM" is computed by using Tukey's bisquare function.
As in the case of the Huber function, we start by a first bias comparison of the different procedures based on Tukey's ψ function. We consider a Poisson distribution with mean 2, size of additive outliers equal to 5 and tuning constants $k \in \{5, 6, 7, 8, 9\}$.
Figure 3.11 shows that the smaller the tuning constant, the more robust is the estimator, except tbrm and rlm in both variants at small percentages of outliers. Additionally, we find noteworthy differences between the different procedures, which will be analysed in more details later on.
Figure 3.12 compares the efficiencies of the procedures with the same tuning constants in the case of clean data as a function of the true mean of the Poisson, which is varied in $\{1, 2, 3, 4, 5, 6, 7, 8, 10, 15, 20\}$ for a sample size n=100. We find that for all means bigger tuning constants lead to higher efficiencies, except for the tukeypois function, where a downward bias of tukeypois with small tuning constants explains this exception. Again we note considerable differences between the different functions.
We use the same approach as for the Huber function, by fixing a certain good level of efficiency of 90% or 95%, see Section 3.4.1. Then we search for a method which offers as much robustness as possible among those methods which guarantee this level of efficiency, see Section 3.4.2.

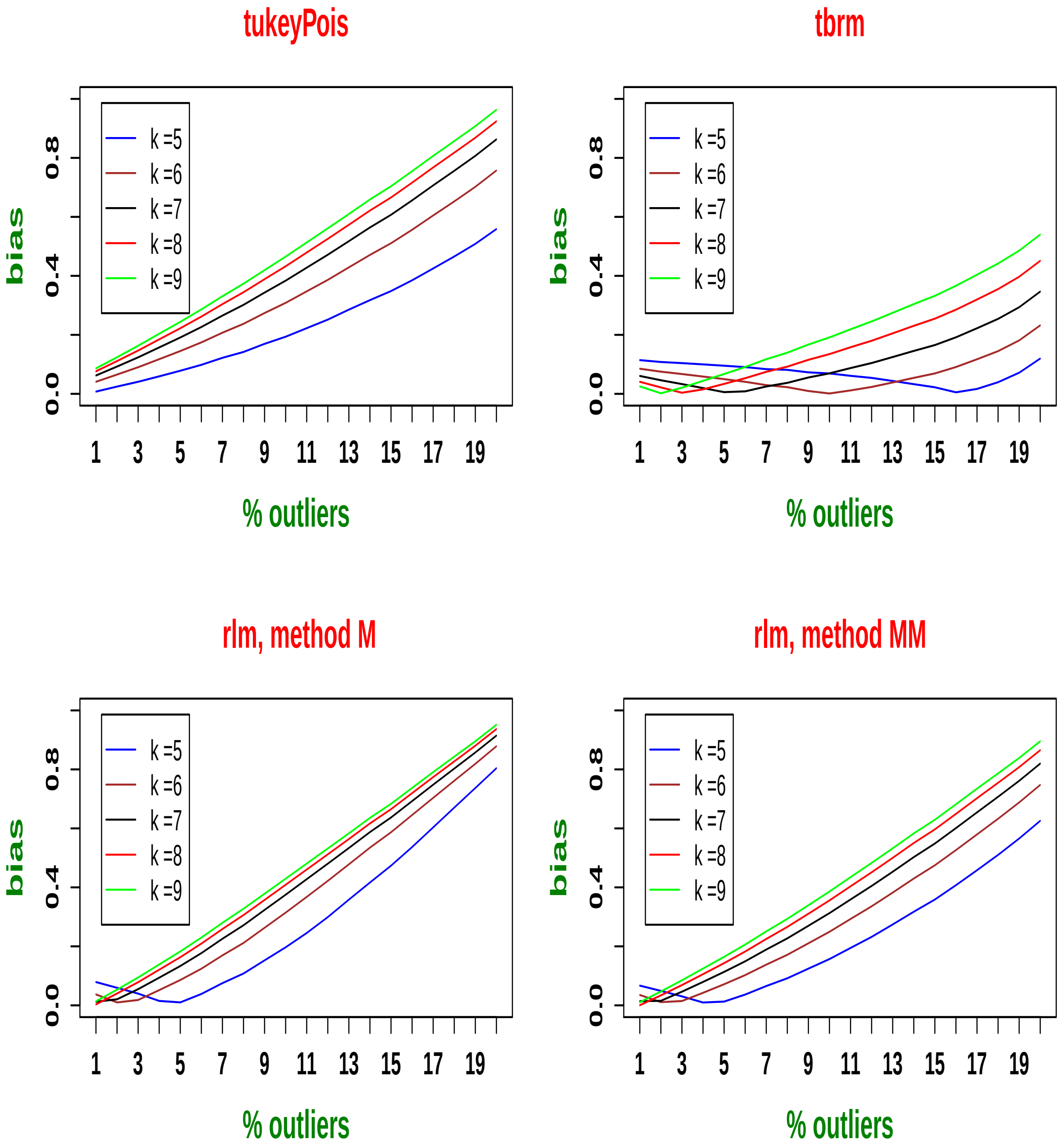


Figure 3.11: Comparison of the sample biases of all Tukey procedures in the case of additive outliers of size 5 and a Poisson distribution with mean 2.

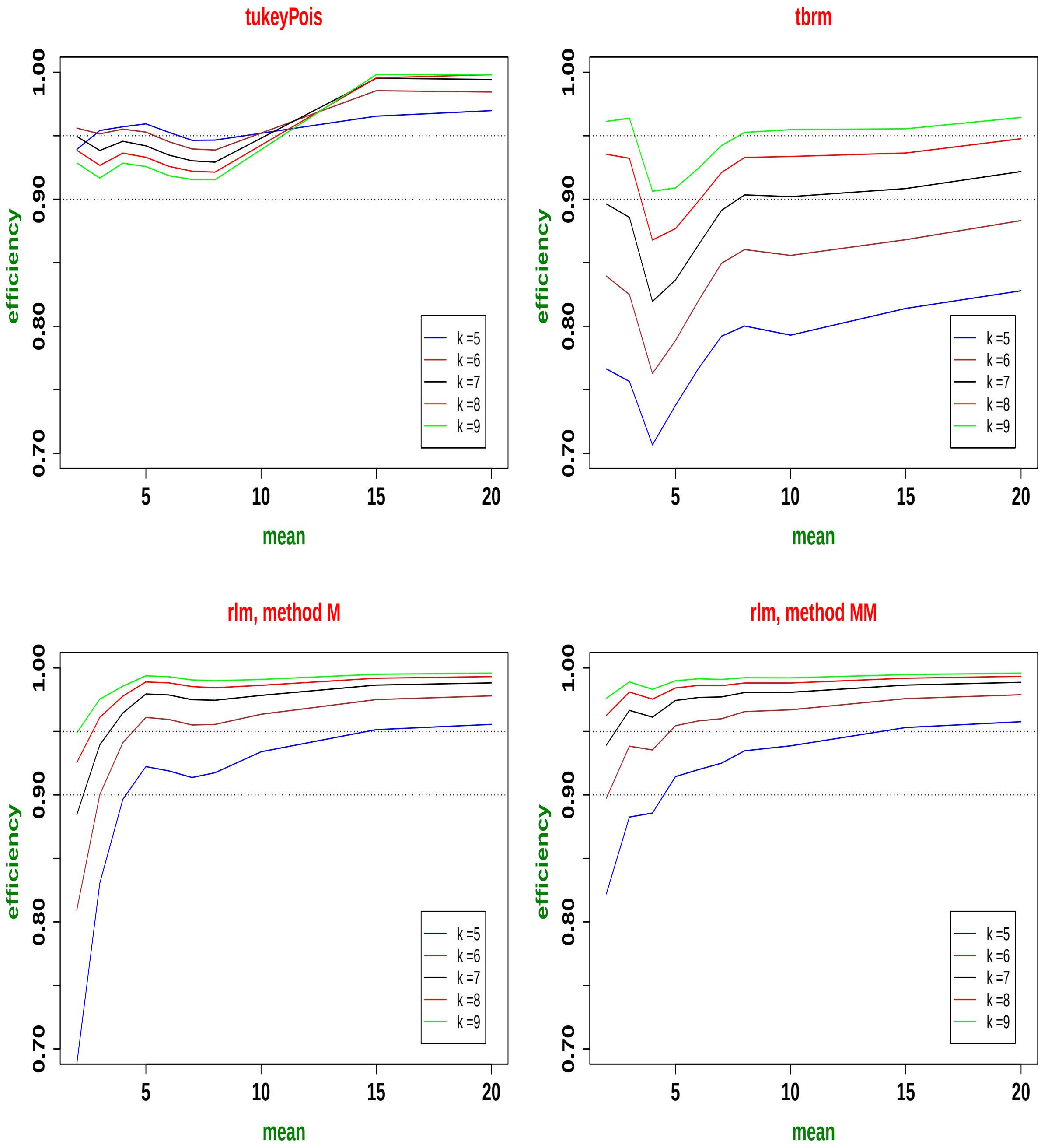


Figure 3.12: Comparison of the relative efficiencies of the Tukey procedures relatively to the sample mean as a function of the underlying true mean for several tuning constants k.

3.4.1 Choice of the tuning constant

In this subsection, we try to find tuning constants, which achieve levels of efficiencies of 90% or 95% as we did in Subsection 3.3.1. Figure 3.13 gives the efficiencies of our functions with further values of the tuning constants. From Figure 3.12 and Figure 3.13, we can choose reasonable tuning constants which achieve approximately levels of efficiencies of 90% or 95% as shown in Figure 3.14.
Reasonable tuning constants which achieve approximately a level of efficiency of 90% are:

1. For tukeypois, we choose $k = 4$.
2. For rlm with method "M" or "MM", we choose $k = 4.5$.
3. For tbrm, we choose $k = 7.5$.

Reasonable tuning constants which achieve approximately a level of efficiency of 95% are:

1. For tukeypois, we choose $k = 5.5$.
2. For rlm with method "M" or "MM", we choose $k = 5.5$.
3. For tbrm, we choose $k = 11$.

If we look again at Figure 3.2, we find that a reasonable tuning constant for an asymptotic level of efficiency of 90% is 4 for the Tukey M-estimator and a reasonable tuning constant for a level of efficiency of 95% is 5, which are less than the tuning constants for all procedures defined above, except tukeypois at a 90% level of efficiency.
Further simulation results indicate that the above tuning constants are also reasonable even for samples of size n=50.

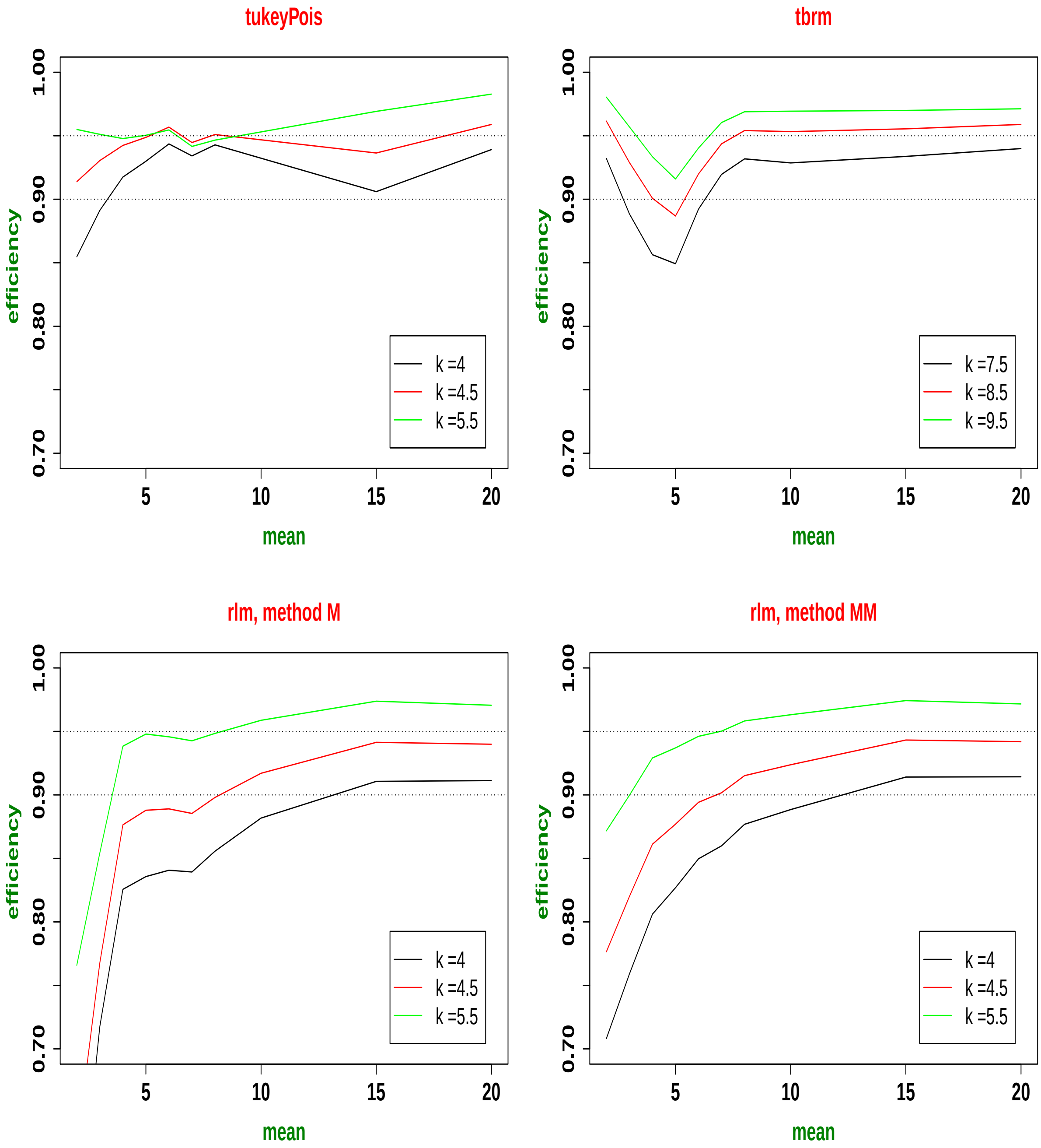
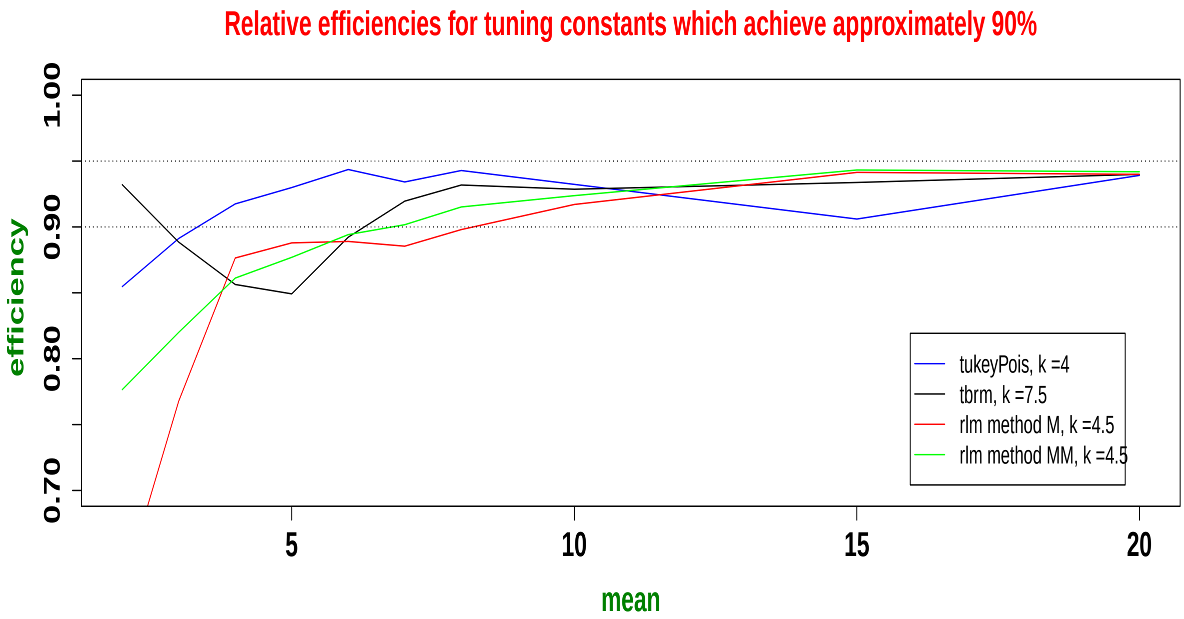


Figure 3.13: Comparison of the relative efficiencies of the Tukey procedures relatively to the sample mean as a function of the underlying true mean for refined values of the tuning constants k.



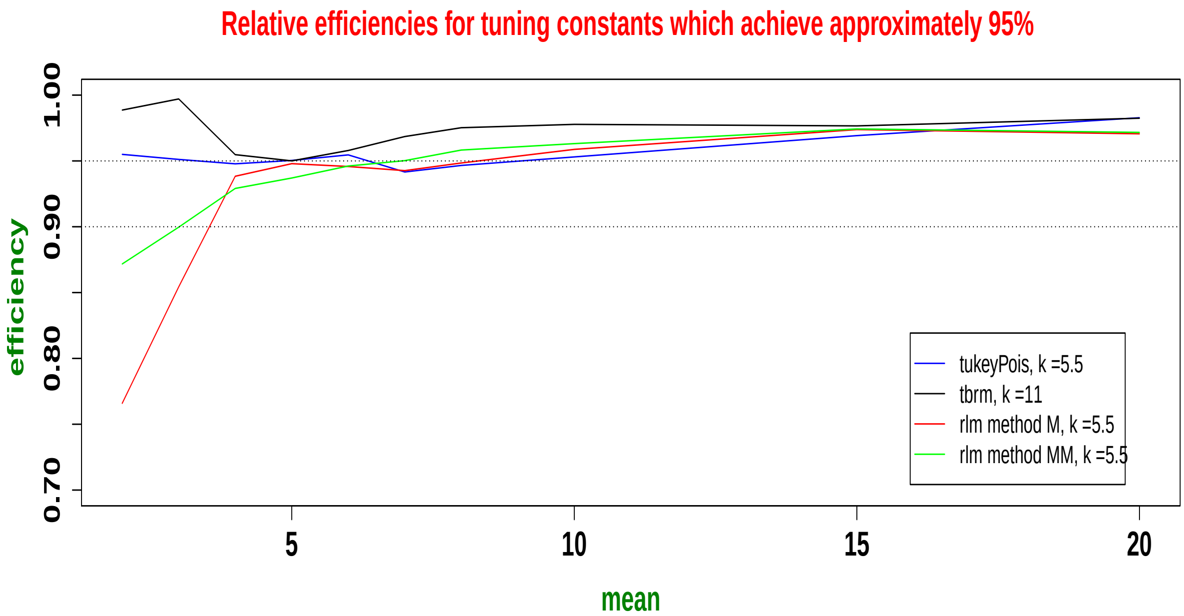


Figure 3.14: Comparison of the relative efficiencies of the Tukey procedures at tuning constants which achieve approximately 90% or 95% level of efficiency (from top to bottom).

3.4.2 Robustness comparison

In this subsection, we compare the biases and the RMSEs under our choices of the tuning constants from Subsection 3.4.1. Figure 3.15 compares the biases of the procedures with the tuning constants chosen to achieve a 90% level of efficiency. The Poisson mean is set to 2 and the sizes of the additive outliers are 5, 10 or 30. We find that in case of additive outliers of size 5, the differences between the biases of all procedures are small for up to 8% outliers. After 8% outliers, tukeypois shows the smallest bias, followed by tbrm, whereas rlm with method MM or method M show the largest bias. In case of additive outliers of sizes 10 and 30 all procedures perform very well, since the Tukey function is redescending and can thus ignore large outliers completely. We get the same results for the RMSEs since they are dominated by the bias.
Figure 3.16 compares the biases of the procedures with the tuning constants chosen to achieve a 90% level of efficiency in case of a Poisson mean equal to 5 and the sizes of the additive outliers being 5, 10 or 30. We find that, in case of additive outliers of sizes 5, the differences between the biases of the procedures are small for up to 20% outliers and the biases increase as the percentages of outliers increase. These differences are small up to 6% outliers in case of additive outliers of sizes 10. After 6% outliers tukeypois shows the smallest bias, followed by rlm with method MM, whereas tbrm and rlm with method M show the largest bias. In case of additive outliers of sizes 30, the differences between the biases are small for up to 20% outliers and appear as horizontal lines. For the RMSEs we have the same conclusions as for the biases.
Figures which compare the biases of the methods with tuning constants chosen to achieve 95% level of efficiency in case of a Poisson mean equal to 2 and 5 are omitted here, since they look similar to Figures 3.15 and 3.16. We only have the following remarks in case of additive outliers of size 10:

- In case of a Poisson distribution with mean equal to 2, tukeypois and rlm with method MM show the smallest bias, whereas tbrm amd rlm with method M show the largest bias after 16% outliers.
- In case of a Poisson distribution with mean equal to 5, tukeypois shows the smallest bias followed by rlm in both variants and tbrm after 10% outliers.

For the RMSEs, we have the same conclusions as for the biases.

3.4.3 General conclusions

The modified Tukey M-estimator seems to provide improvements over the other procedures, particulary in case of a Poisson distribution with mean $\theta = 2$ and small outliers, and at mean $\theta = 5$ in case of moderately large outliers. Very large outliers do not really affect Tukey's redescending bisquare function. It is well known that redescending M-estimators like Tukey can completely ignore very large outliers.

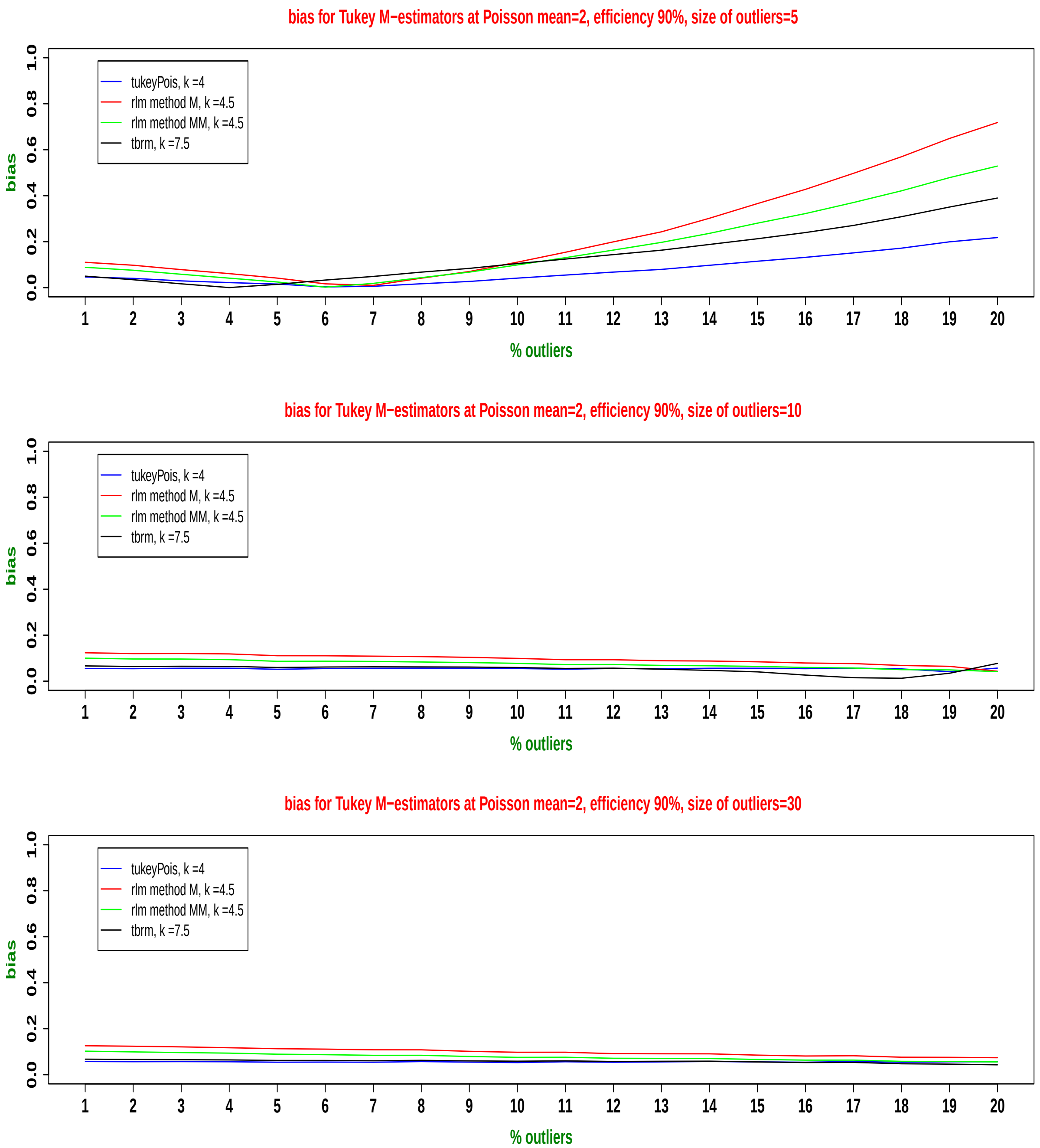


Figure 3.15: Comparison of the biases of the Tukey based procedures tuned to achieve 90% level of efficiency in case of a Poisson with mean 2 and the sizes of the additive outliers being 5, 10 and 30 (from top to bottom).

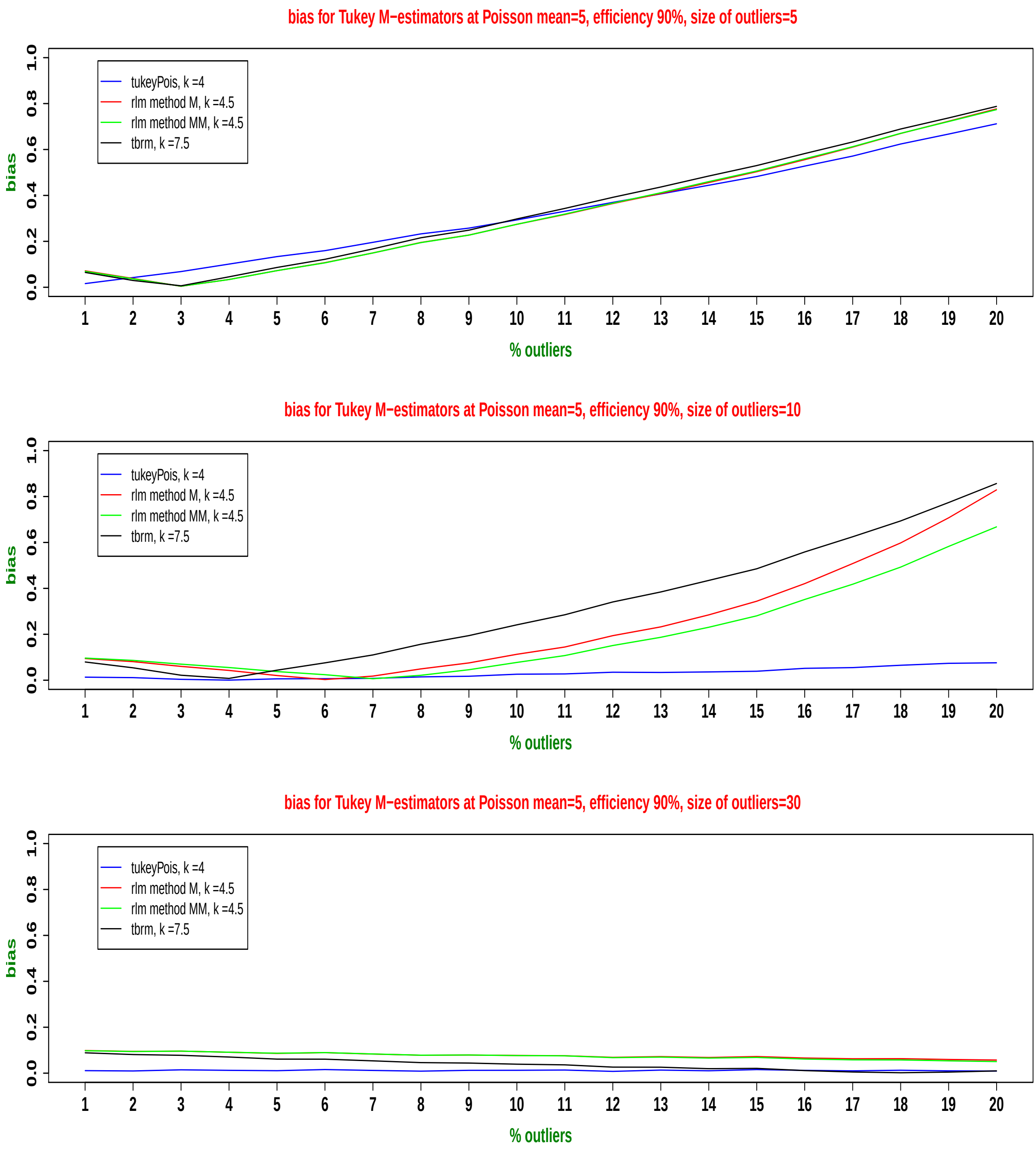


Figure 3.16: Comparison of the biases of the Tukey based procedures tuned to achieve 90% level of efficiency in case of a Poisson with mean 5 and the sizes of the additive outliers being 5, 10 and 30 (from top to bottom).

3.5 Comparison of Huber and Tukey M-estimators

In this section, we compare the best M-estimation procedures based on the Huber function and the Tukey function found before. In addition to the best procedures, we add glmrob and huberpois, and perform another robustness study with one additive outlier of increasing size $\{1, 2, ..., 10, 12, ..., 20\}$, since it is well known that Huber and Tukey M-estimators are most vulnerable to different outlier sizes. Finally, we compare the procedures in case of different percentages of outliers varying from 1% to 20%, fixing the outlier sizes to the most harmful ones identified in case of a single outlier.

Robustness comparison for an increasing number of additive outliers
Figure 3.17 compares the biases of the best M-estimation procedures based on the Huber function and the Tukey function in the same situations as before, namely, glmrob with external weights and tuning constant $k = 1.5$, and tukeypois with tuning constant k=4, chosen to achieve a 90% level of efficiency. The mean of the Poisson is set to 2 and the sizes of additive outliers are 5, 10 and 30. We find that the differences in biases between the two functions are small for up to 7% outliers of all sizes. After 7% outliers the differences in biases between the two functions increase with the percentages of outliers, with tukeypois having a smaller bias than glmrob with external weights. We get the same conclusions for the RMSEs.
We have similar results in case of a Poisson mean of 5, and also in case of Poisson means equal to 2 or 5 at a 95% level of efficiency, so we omit the corresponding figures here, but note that the differences between the two functions are small for up to 20% outliers at a 95% level of efficiency and in case of additive outliers of sizes 5.

Robustness comparison with one additive outlier
The left hand side of Figure 3.18 compares tukeypois, huberpois, glmrob and glmrob with external weights with the predefined tuning constants which achieve 95% efficiency, in case of a Poisson mean $\theta = 2$ and a single additive outlier of increasing size $\{1, 2, ..., 10, 12, ..., 20\}$. In case of small sizes less than 5, we find that glmrob with external weights leads to the smallest bias, followed by glmrob, then tukeypois and finally huberpois, which gives the largest bias. For lager outlier sizes, glmrob with external weights and tukeypois give the smallest biases, with the differences between both functions becoming smaller for an increasing size. glmrob gives larger bias then, followed by huberpois.
An outlier of size 3 causes the largest bias for tukeypois, whereas the procedures huberpois and glmrob, which are based on the Huber function, show an increasing bias which stabilizes for outlier sizes as the outlier size exceeds 5.
In case of a Poisson with mean $\theta = 5$, we get similar results as for $\theta = 2$, so we omit this figure here. But we note that under all outlier sizes, the two versions of glmrob give smaller biases than the modified Tukey M-estimator.

Robustness comparison for an increasing number of outliers of worst size
It would be interesting to compare the worst-case performances of the estimators for a varying percentage of outliers. However, the most harmful outlier size can depend on the percentage of outliers, and determination of the worst case for each percentage would be cumbersome. Instead, we consider an increasing percentage of outliers of a fixed size which was found to be particularly harmful in case of a single outlier. The right hand side of Figure 3.18 compares the predefined functions in case of different percentages of outliers from 1% to 20% with sizes of additive outliers being 3 for tukeypois, 5 for glmrob with external weights and sizes of additive outliers being 8 for the others. The Poisson mean is set to $\theta = 2$. We find that glmrob with external weights has the smallest bias for up to 17% outliers followed by tukeypois and glmrob. After 17% outliers, tukeypois gives smaller bias than glmrob with external weights and finally huberpois gives the largest bias.

General conclusions
According to the situations considered here, we conclude that for up to 4% additive outliers of different sizes glmrob with external weights provide better results than the modified Tukey M-estimator and glmrob, whereas in case of moderately large outliers and very large outliers, the modified Tukey M-estimator gives better results than glmrob with external weights. The differences between the two functions become larger for higher percentages of outliers.

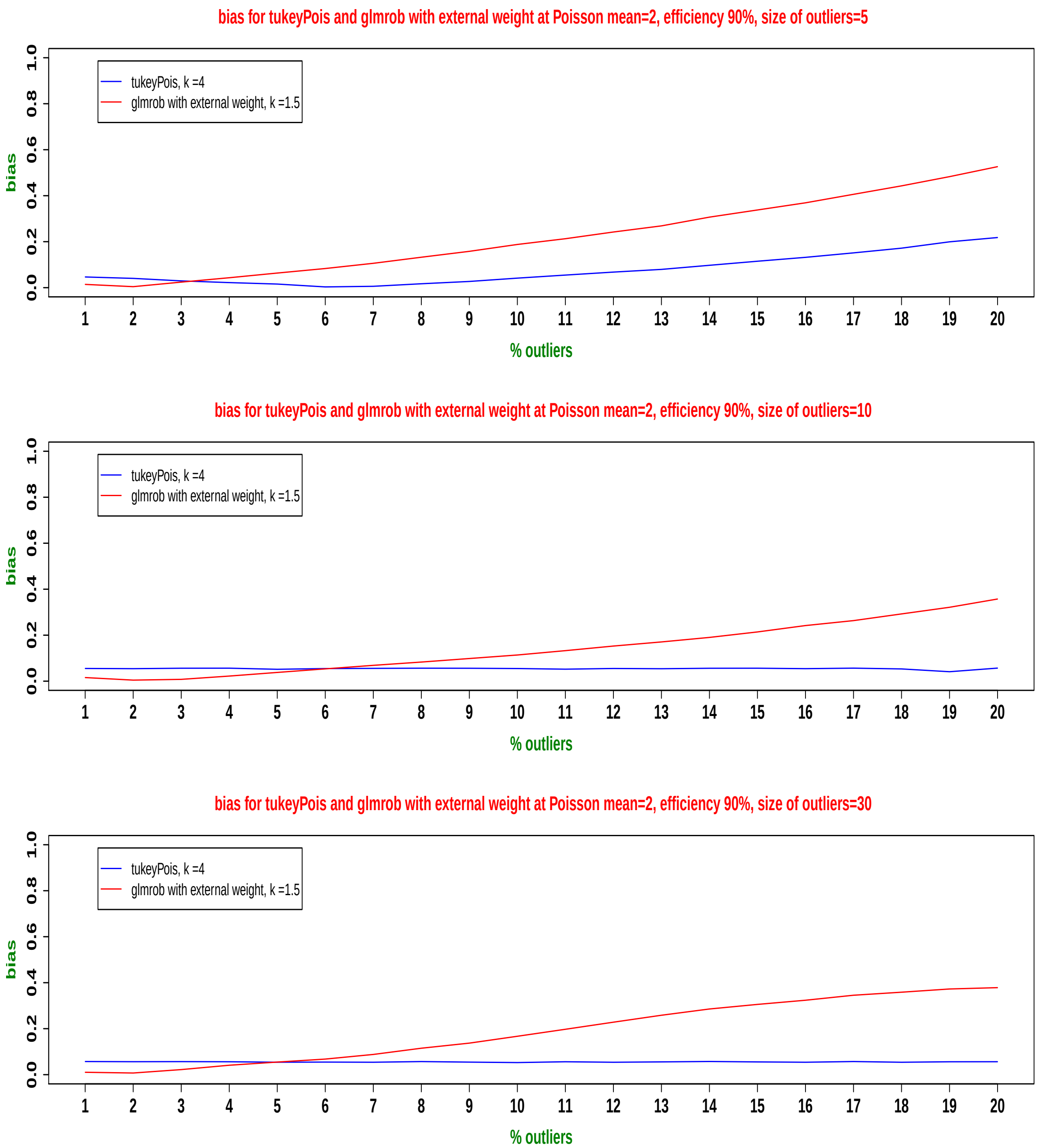


Figure 3.17: Comparison of the biases of glmrob with external weights and tukeypois, both tuned to achieve 90% level of efficiency in case of a Poisson with mean 2 and the sizes of the additive outliers being 5, 10 and 30 (from top to bottom).

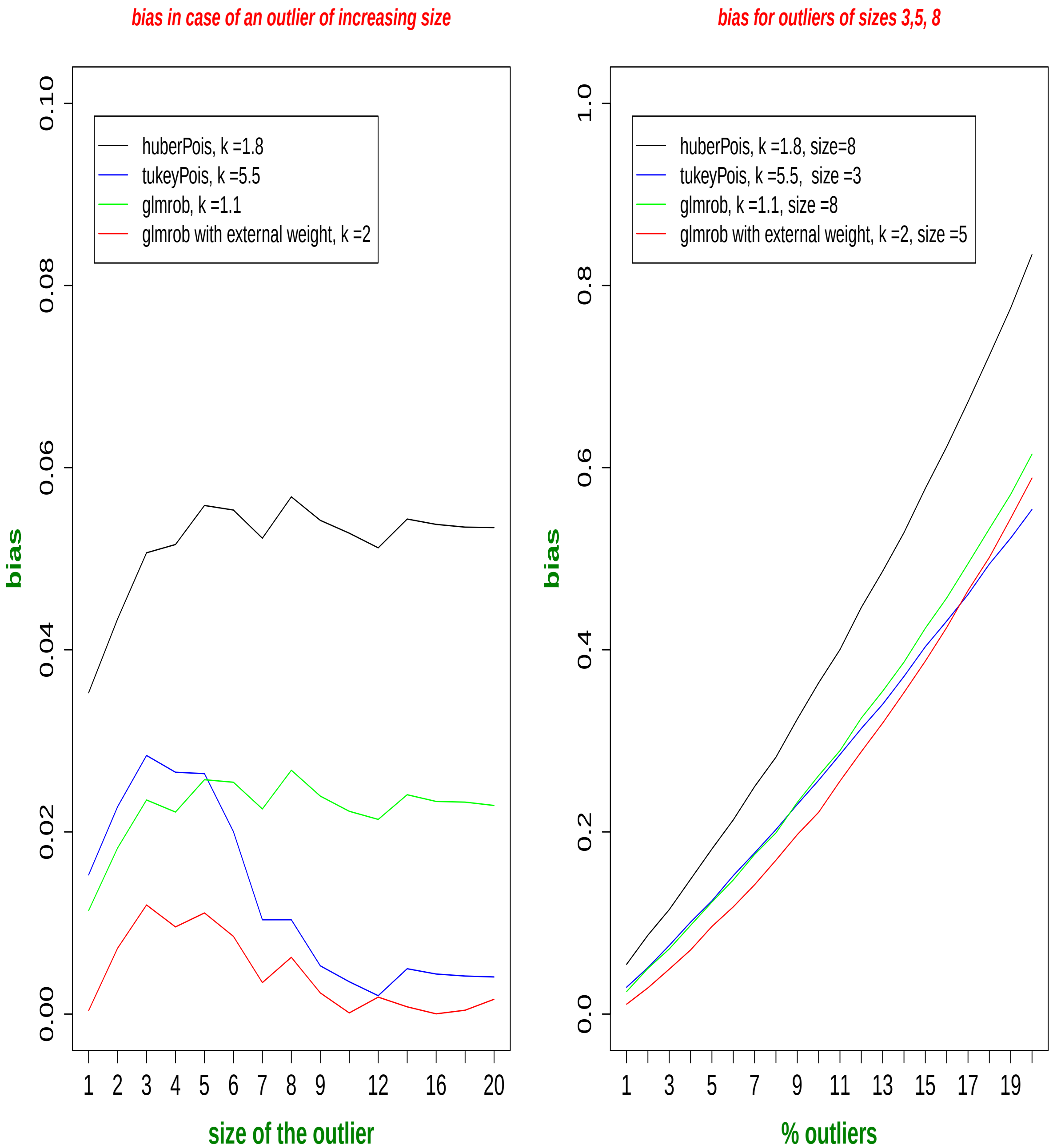


Figure 3.18: Comparison of the biases of tukeypois, huberpois, glmrob and glmrob with external weights, tuned to achieve 95% efficiency in case of an additive outlier of increasing size (left) and in case of the additive outliers of size 3 for tukeypois, size 5 for glmrob with external weights and of size 8 for the others (right), in case of a Poisson with mean 2, sample size n=100.

3.6 Alternative estimators suggested for small means

In this section, we focus on the problem of reaching large efficiencies in case of small Poisson means, say $\theta \leq 4$. In addition to our best Huber and Tukey procedures, namely glmrob with external weights and tukeypois, we will add the standard version of glmrob because we will treat small values of θ here and we want to see if this function works well under these values or not. We will also add the function roptest from the R-package "ROptEst", which can be described as follows:
roptest function
This is a function in the R package "ROptEst", which is written by Kohl and Ruckdeschel (2010). This function computes optimally robust estimates for L2-differentiable parametric families via k-step construction. It is a function in Y, L2Fam, eps, eps.lower, eps.upper, initial.est, steps, ... where Y is the data sample, L2Fam is the Poisson model as a member in this class, eps is the amount of gross errors considered, eps.lower is a lower bound for the amount of gross errors (contamination). eps.upper is an upper bound for the amount of gross errors (contamination), which is a rough estimate between 0 and 0.5. We choose eps.upper=0.05 as it is chosen in Kohl (2005) under the Poison model, initial.est is the initial estimate for the unknown parameter. If initial.est is missing, a minimum distance estimator is computed. steps is the integer number of steps used for k-steps construction. We choose the number of steps as 3. For more details, see the R package ROptEst. The main idea underlying this function is to minimize the asymptotic variance at the model subject to a bound on the supremum of the influence functions.

We construct an alternative new estimator, called trimmeanfit, based on the idea of trimmed means with adaptive trimming. We can describe it as follows:
trimmeanfit function
This is a function in Y, init, α, where Y is the data sample, the parameter init is either 2 in order to initialize by the median, or 0 to initialize by poismall with l=0, which is described before in the huberpois algorithm. Here, α is a trimming proportion chosen beforehand, which will be set to 0.01 in our simulations to control the efficiency of the procedure. The main idea underlying this function is to use an initial estimate (median or poismall), then trim all observations larger than the (1-$\alpha/2$) quantile or smaller than the ($\alpha/2$) quantile, and re-estimate the parameter as the mean of the non-trimmed observations. Iterate this procedure until convergence. Note that we might trim more or less than this fraction in each step depending on the position of the observations relatively to the quantiles of the current solution.

To choose a function which combines efficiency and robustness, we follow the same approach we used before to design Huber and Tukey functions, namely: By trial and error we fix the efficiency at a satisfactory level in case of uncontaminated data. For each function we try to find a tuning constant which achieves approximately the same level of efficiency. Then we compare the robustness of the procedures. To apply this approach, we compare the relative efficiencies for several values of $\theta \in \{0.1, 0.2, 0.3, 0.5, 0.8, 1.3, 2.1, 3.4, 5.5, 8.9, 14.4, 23.3\}$, and we compare biases at means $\theta \in \{0.5, 2\}$ for the following estimators.

1. glmrob with $k \in \{1.5, 1.8\}$.
2. glmrob with external weights and $k = 2.5$.
3. tukeypois with $k \in \{5, 6\}$.
4. trimmeanfit with init=2, $\alpha = 0.01$.
5. roptest with eps.upper = 0.05, steps = 3.

Figure 3.19 compares the efficiencies of our functions suggested above resulting from 1000 simulations runs at each value of θ for n=100, measured by the percentage mean square error relatively to the maximum likelihood estimator, which is the sample mean. If θ is very small, tukeypois and glmrob with external weights are not as efficient as glmrob, roptest and trimmeanfit. But we note that the efficiency for tukeypois with $k = 5$ is nearly 80% and nearly 85% with $k = 6$. For larger values of θ, all functions achieve much better efficiencies.
Figure 3.20 compares the biases of our functions with predefined tuning constants in case of a Poisson mean 0.5 and the sizes of additive outliers being 2 and 5. In case of additive outliers of size 2, the differences between the biases of the methods are small for up to 20% and the biases increase as the percentages of outliers increase. In case of additive outliers of size 5, we can order the functions in ascending order with respect to the bias as follows:

1. tukeypois with $k \in \{5, 6\}$, followed by trimmeanfit with $\alpha = 0.01$,
2. glmrob with external weights at $k = 2.5$,
3. then glmrob with $k = 1.5$ and roptest with eps.upper = 0.05, steps = 3,
4. and finally glmrob with $k = 1.8$.

We have the same order in case of additive outliers of size 10 (not shown here), but we note that the differences become larger at larger sizes and larger percentages of outliers.
Figure 3.21 compares the biases of our functions with predefined tuning constants in case of a Poisson mean 2 and sizes of additive outliers being 5 and 10. We find in case of additive outliers of size 5, the differences between the biases of all functions are small and tukeypois with $k = 5$ has the smallest bias. In case of additive outliers of size 10, we have the same order as in case of additive outliers of size 5 in Figure 3.20, and we have also the same order in case of additive outliers of size 30 (not shown here). But we note that in cases of additive outliers of sizes 10 and 30, the differences between the biases become larger.

General conclusions
We conclude that the estimator based on adaptive trimming is efficient and has similar robustness properties as the modified Tukey M-estimator. glmrob and roptest are also efficient at small means but they are less robust. In general, we can use the estimator based on adaptive trimming at small means, otherwise we should use the modified Tukey M-estimator.

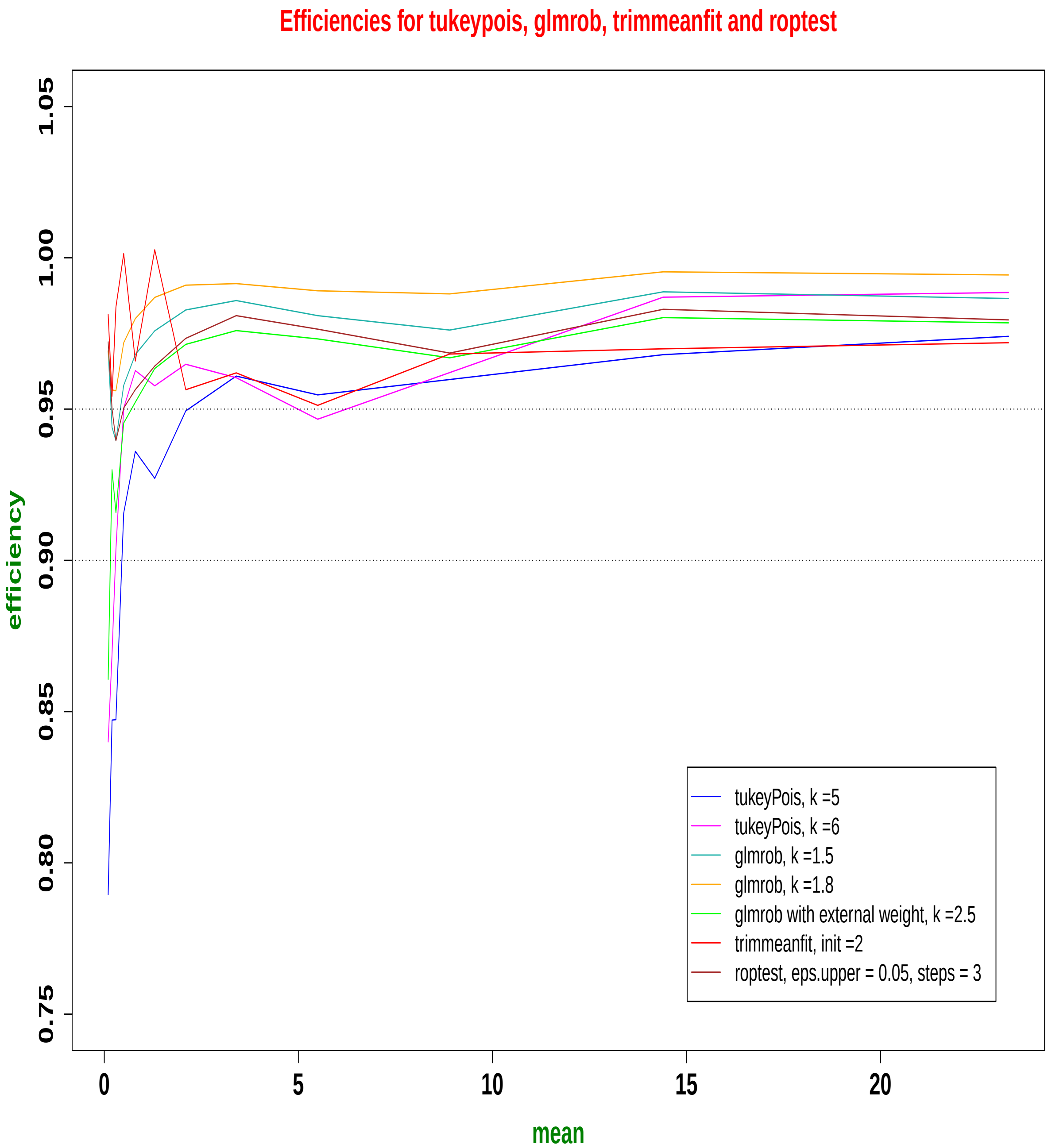


Figure 3.19: Relative efficiencies for tukeypois, glmrob, trimmeanfit and roptest measured by the percentage mean square error relatively to the sample mean with several tuning constants k, n=100

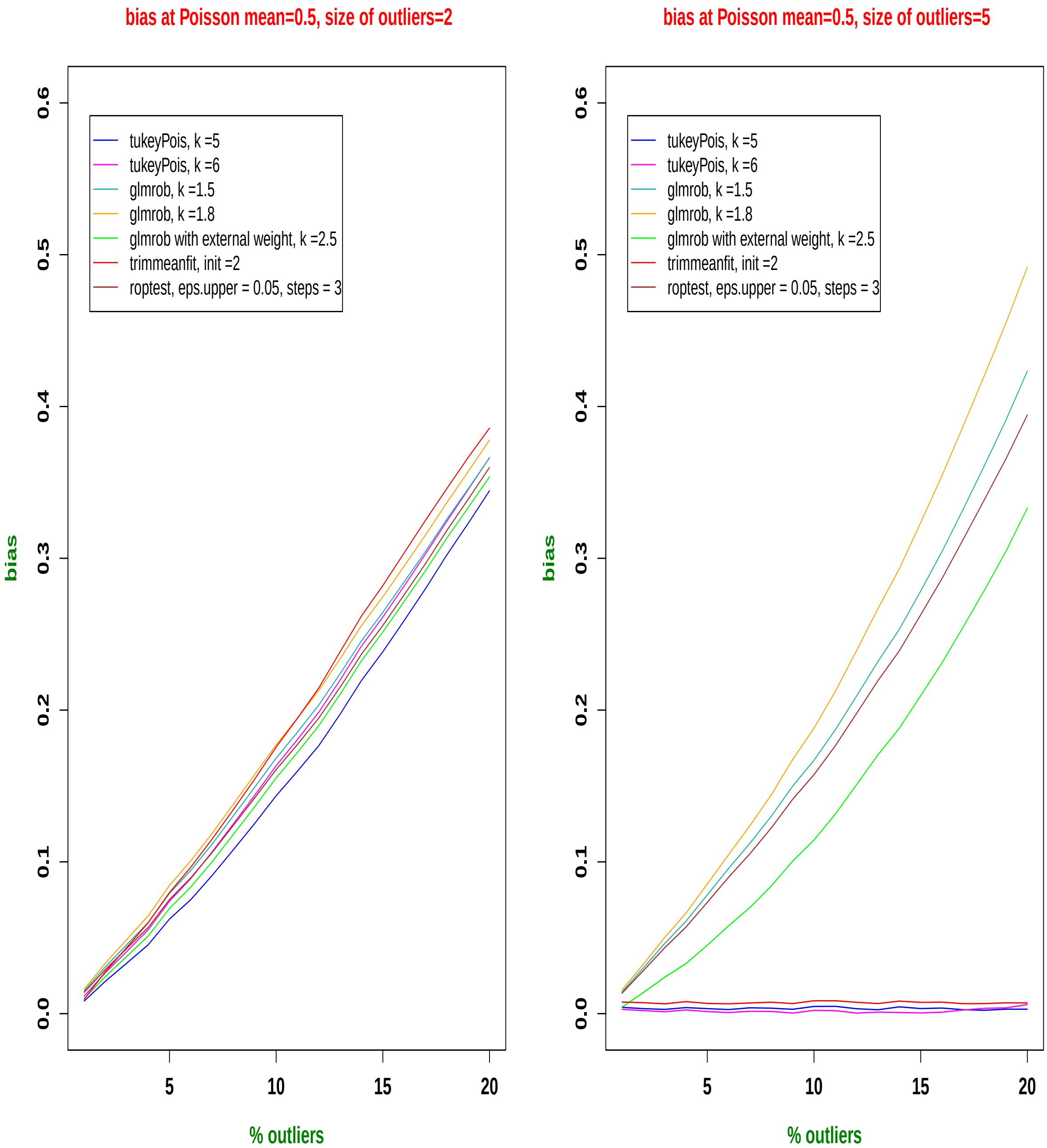


Figure 3.20: Comparison of the biases of tukeypois, glmrob, trimmeanfit and roptest with several tuning constants k, in case of a Poisson with mean 0.5 and the sizes of the additive outliers being 2 and 5, respectively.

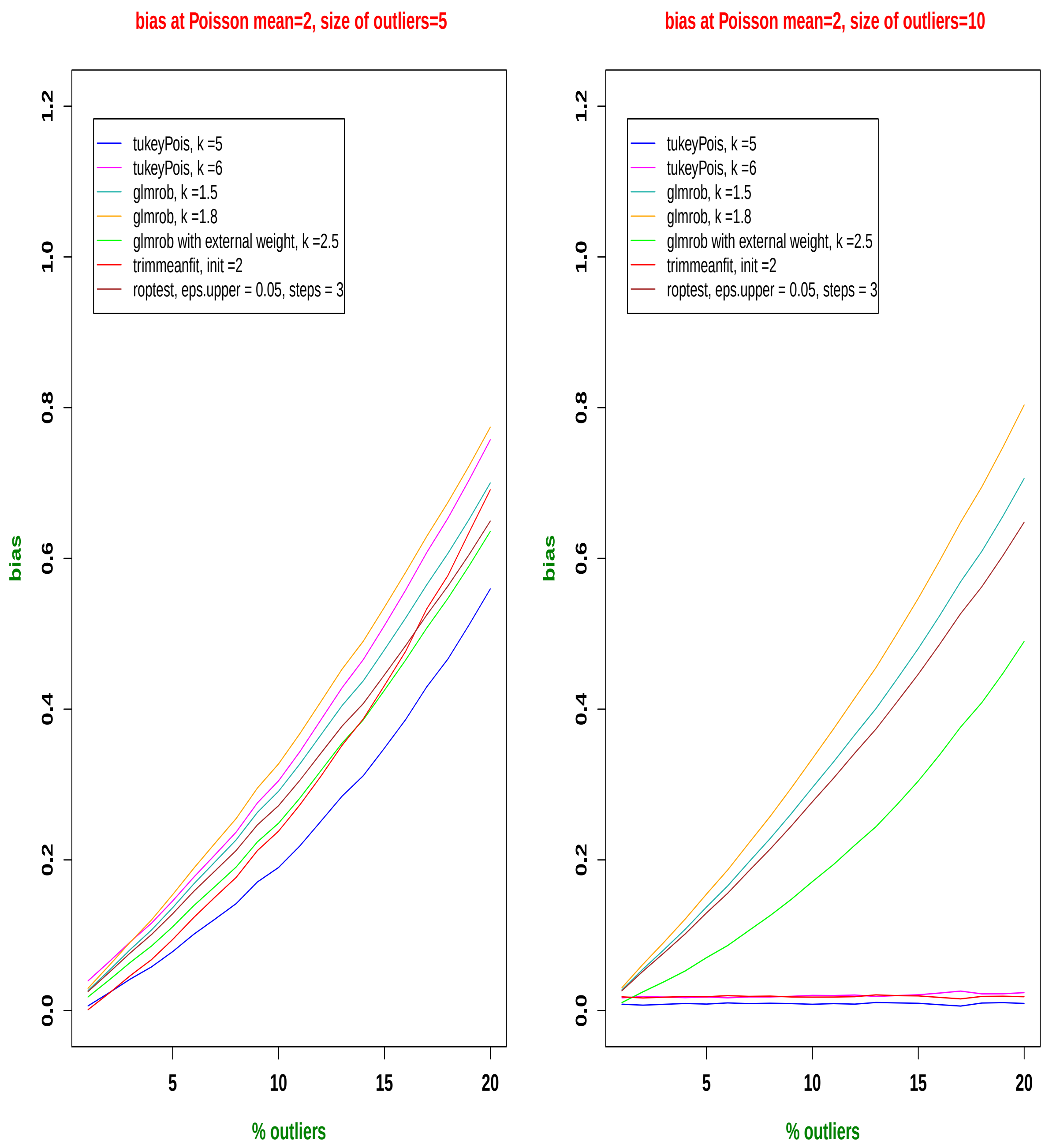


Figure 3.21: Comparison of the biases of tukeypois, glmrob, trimmeanfit and roptest with several tuning constants k, in case of a Poisson with mean 2 and the sizes of the additive outliers being 5 and 10, respectively.

Chapter 4

M-estimation for INGARCH Models

This chapter treats robust estimation for the parameters of INGARCH models for count time series data in the presence of outliers. We can start our previous treatments for this chapter from the 1980s, when the autoregressive conditional heteroscedastic (ARCH) model was proposed by Engle (1982), followed by the generalized autoregressive conditional heteroscedastic model (GARCH) introduced by Bollerslev (1986). These models are commonly used in the literature for estimation of the volatility of financial time series data.
MacDonald and Zucchini (1997) show the importance to use discrete valued time series models especially when the observations are categorical or quantitative but fairly small. Fokianos and Kedem (2002) propose in Chapter 4 of their book the regression model for count time series such as the Poisson model and the doubly truncated Poisson model. Fokianos and Kedem (2004) prove that these models fall within the broad class of generalized linear time series models and state that their analysis is based on partial likelihood inference.
Ferland et al. (2006) propose the integer-valued GARCH (INGARCH) model with Poisson deviates. They concentrate on a special case of this model, the INGARCH(1,1) model and its properties. They propose a conditional maximum likelihood (ML) approach, conditional on the pre-sample values, to fit this model. They conclude that the distribution of the ML estimator can be approximated by the normal distribution. The asymptotic covariance matrix of the ML estimator can be approximated from the observed Fisher information matrix.
Weiss (2009) shows that INGARCH models are able to describe integer-valued processes with over-dispersion. He investigates the purely autoregressive INGARCH(p,0) model, or more briefly INARCH(p), and shows that they are closely related to the standard AR(p) models. For p=1, he determines the stationary marginal distribution in terms of its cumulants. He also provides applications to real data for the INARCH(p) model.
The INGARCH model is studied further by Fokianos et al. (2009), who consider geometric ergodicity and likelihood based inference for linear and nonlinear Poisson autoregressions. In the linear case and under geometric ergodicity they prove that the maximum likelihood estimators of the INGARCH(1,1) model are asymptotically Gaussian.
Robust estimation and outlier detection for time series has been considered by several

authors since the 1970s, see Gastwirth and Rubin (1975), Denby and Martin (1979), Martin and Yohai (1985), Basawa et al. (1985) and Chen and Liu (1993).
Muler and Yohai (2002) present two robust estimates for ARCH processes: $\tau-$ and filtered $\tau-$estimates. These estimates are based on $\tau-$scale estimates, which have simultaneously a high breakdown point and high efficiency under normality.
Muler and Yohai (2008) introduce two classes of robust estimates for GARCH models: M-estimates and bounded M-estimates. The first class is an extension of M-estimates introduced by Huber (1964) for location and Huber (1973) for regression. They show that M-estimators are consistent and asymptotically normal. To improve robustness they propose bounded M-estimates, which are also consistent and asymptotically normal.
Mukherjee (2008) derives asymptotic normality of a class of M-estimators in the GARCH model. This class of estimators includes least absolute deviations and Huber's estimator in addition to quasi maximum likelihood estimator.
Fokianos and Fried (2010) develop techniques for estimation and detection of different types of outliers (intervention effects) within the framework of INGARCH models. They focus on the detection and estimation of sudden shifts and outliers and employ the maximum of score tests, whose critical values in finite samples are determined by parametric bootstrap to identify such unusual events successfully.

In the remainder of this chapter, we will focus on INARCH(1) and INGARCH(1,1) models as special variants of INGARCH models. We will start by the properties of these models. We will apply conditional maximum likelihood as a classical approach to estimate the parameters of these models. Then we will introduce the robust estimation of the marginal mean in case of time series data from INGARCH models using our best functions given in Chapter 3 for i.i.d. Poisson data to test if these functions are still suitable or not. Then we will modify the classical estimation approach by giving robust estimators for the parameters of INARCH(1) models. Thereafter we investigate some of the basic properties of these estimators. Afterwards we compute the estimates using some new functions, which we implemented in R, and compare between them via simulations in case of clean and contaminated Poisson time series data. We will finish this chapter by trying to extend robust estimation to general INGARCH models.

4.1 Properties of INGARCH(p,q) models

In this section, we will briefly give the definition and known basic properties of the INGARCH(p,q) models, concentrating on INARCH(1) and INGARCH(1,1) processes.
According to Weiss (2009), we can define the INGARCH(p,q) models as follows: Let (Y_t) be a process with range $\mathbb{N}_0 = \{0, 1, ...\}$. Let $\mathcal{Y}_t$ abbreviate the information on the process available at time t, $\mathcal{Y}_t = \{Y_t, Y_{t-1}, ...\}$.
The process $(Y_t : t \in \mathbb{N})$ follows an INGARCH(p,q) model, p $\geq$ 0, q $\geq$ 0 if

- Y_t conditioned on $\mathcal{Y}_{t-1}$ is Poisson distributed according to $Pois(\lambda_t)$, where
- the conditional mean $\lambda_t := E(Y_t|\mathcal{Y}_{t-1})$ fulfills the recursion

$$\lambda_t = \beta_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \tag{4.1}$$

with $\beta_0 > 0$ and $\alpha_1, ..., \alpha_p, \beta_1, ..., \quad \beta_q \geq 0$.
When $p = q = 0$, this leads to an i.i.d. process with marginal distribution $Pois(\beta_0)$.

Simple variants of INGARCH(p,q) models, defined in (4.1), are:

1. INGARCH(1,0) or INARCH(1), which has the following conditional mean form
$$\lambda_t = \beta_0 + \alpha_1 Y_{t-1} \tag{4.2}$$
2. INGARCH(1,1), which has the following conditional mean form
$$\lambda_t = \beta_0 + \alpha_1 Y_{t-1} + \beta_1 \lambda_{t-1}. \tag{4.3}$$

Weiss (2010) gives the following properties of the INARCH(1) model:

1. An INARCH(1) process with $\beta_0 > 0$ and $0 < \alpha_1 < 1$ is a stationary Markov chain with the transition probabilities
$$p_{i|j} := p(Y_t = i|Y_{t-1} = j) = exp(-\beta_0 - \alpha_1 \cdot j) \cdot \frac{(\beta_0 + \alpha_1 \cdot j)^i}{i!} > 0. \tag{4.4}$$
It is irreducible and aperiodic and hence ergodic.
2. All moments of the stationary marginal distribution exist and can be determined recursively. In particular, the marginal mean $\lambda = E(Y_t)$, marginal variance $V(Y_t)$, skewness (SK) and excess (EX) of Y_t are as follows:

$$\lambda = E(Y_t) = \frac{\beta_0}{1 - \alpha_1} \tag{4.5}$$

$$V(Y_t) = \frac{\beta_0}{(1-\alpha_1)(1-\alpha_1^2)} \tag{4.6}$$

$$SK = \frac{1+2\alpha_1^2}{1+\alpha_1+\alpha_1^2}\sqrt{\frac{1+\alpha_1}{\beta_0}} \tag{4.7}$$

$$EX = \frac{1+6\alpha_1^2+5\alpha_1^3+6\alpha_1^5}{\beta_0(1+\alpha_1+\alpha_1^2)(1+\alpha_1^2)} \tag{4.8}$$

3. The autocorrelation function $\rho_Y(h) := corr[Y_t, Y_{t-h}]$ simply equals α_1^h like in the standard AR(1) model. An INARCH(1) model can be described by an AR(1) structure or it has AR(1) like dependence structure, where the AR(1) satisfies the equation $Y_t = \alpha_1 Y_{t-1} + \epsilon_t$ with ϵ_t being a white noise error term.

4. An explicit expression for the marginal distribution of an INARCH(1) process is not known, but it can be approximated using the Poisson-Chalier expansion.

5. An INARCH(1) model is a generalized linear model with Poisson distribution as a random component and identity link as a systematic component, see Fokianos and Kedem (2002).

Ferland et al. (2006) and Fokianos et al. (2009) describe the following properties of the INGARCH(1,1) model:

1. The process λ_t is a stationary ergodic Markov chain provided that $0 < \alpha_1 + \beta_1 < 1$.

2. Under these conditions all moments of the stationary distribution exist. In particular, the expected value, $\lambda := E(Y_t)$, is given by

$$\lambda = \frac{\beta_0}{1-\alpha_1-\beta_1} \tag{4.9}$$

   and the variance, $V(Y_t)$, is given by

$$V(Y_t) = \frac{\lambda[1-(\alpha_1+\beta_1)^2+\alpha_1^2]}{1-(\alpha_1+\beta_1)^2}. \tag{4.10}$$

   Obviously, the mean and the variance being different, the marginal distribution of Y_t is not Poisson, except for $\alpha_1 = \beta_1 = 0$.
   The autocovariance function, $\gamma(r)$, is given by

$$\gamma(r) = \frac{\alpha_1[1-\beta_1(\alpha_1+\beta_1)](\alpha_1+\beta_1)^{r-1}\lambda}{1-(\alpha_1+\beta_1)^2} \quad \forall r \geq 1. \tag{4.11}$$

3. An INGARCH(1,1) process can be written as an ARMA(1,1) process: An ARMA(1,1) process (Y_t) satisfies the equation

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \tag{4.12}$$

where (ϵ_t) is a sequence of uncorrelated random variables with mean 0 and constant variance σ^2.
An INGARCH(1,1) process, given in (4.3), can be written as

$$Y_t = \beta_0 + (\alpha_1 + \beta_1) Y_{t-1} - \beta_1 e_{t-1} + e_t \tag{4.13}$$

where (e_t) is a white noise process with variance $\sigma^2 = \lambda = \frac{\beta_0}{1-\alpha_1-\beta_1}$.
Let $\phi_1 = \alpha_1 + \beta_1$, $\theta_1 = -\beta_1$ and $\sigma^2 = \lambda$, then the autocovariance function of the corresponding ARMA(1,1) process is the same as the autocovariance function of the INGARCH(1,1) process.
The autocorrelation function of an ARMA(1,1) process is given by

$$\rho_l = \begin{cases} 1 & l = 0 \\ \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{1+\theta_1^2+2\phi_1\theta_1}\ , & l = 1 \\ \phi_1 \rho_{l-1} & l \geq 2 \end{cases} \tag{4.14}$$

4. An INGARCH(1,1) process is related to the theory of generalized linear models for time series, see Kedem and Fokianos (2002).

4.2 Classical estimation in the INGARCH model

Let $y_1, ..., y_n$ be a time series from an INARCH(p) process, and $\theta = (\beta_0, \alpha_1, ..., \alpha_p)'$ denote the vector of its parameters. We can apply the conditional maximum likelihood approach to estimate θ as follows:
The conditional likelihood function of the n observations $y_1, ..., y_n$ conditionally on the first p values is given by

$$L(\theta) = \Pi_{t=p+1}^{n} \frac{e^{-\lambda_t} \lambda_t^{y_t}}{y_t!} \tag{4.15}$$

where λ_t is given in (4.1) with $q = 0$. The log-likelihood function is

$$\mathbf{l}(\theta) = lnL(\theta) = \sum_{t=p+1}^{n} (y_t ln\lambda_t - \lambda_t - lny_t!) = \sum_{t=p+1}^{n} l_t(\theta) \tag{4.16}$$

where $l_t(\theta) = y_t ln\lambda_t - \lambda_t - lny_t!$.
The score function is defined by

$$S_n(\theta) = \frac{\partial \mathbf{l}(\theta)}{\partial \theta} = \sum_{t=p+1}^{n} \frac{\partial l_t(\theta)}{\partial \theta} \tag{4.17}$$

The vector of the the first derivatives (gradient) of $l_t(\theta)$ with respect to θ is

$$\frac{\partial l_t(\theta)}{\partial\theta} = \left(\frac{y_t}{\lambda_t} - 1\right)\frac{\partial\lambda_t}{\partial\theta}, \tag{4.18}$$

where

$$\frac{\partial\lambda_t}{\partial\theta} = (1, y_{t-1}, ..., y_{t-p})'. \tag{4.19}$$

In case of the INARCH(1) model, this simplifies to

$$\frac{\partial\lambda_t}{\partial\theta} = (1, y_{t-1})' \tag{4.20}$$

Ferland et al. (2006) apply the conditional maximum likelihood approach for estimating the parameters of the INGARCH(p,q) model. Following their lines to estimate the parameters of INGARCH(1,1) models defined in (4.3), let $\theta = (\beta_0, \alpha_1, \beta_1)'$ denote the vector of its parameters. Then the first derivatives of $l_t(\theta)$ with respect to $\theta_i, \quad i = 0, 1, 2$, are

$$\frac{\partial l_t(\theta)}{\partial\theta_i} = \left(\frac{y_t}{\lambda_t} - 1\right)\frac{\partial\lambda_t}{\partial\theta_i} \tag{4.21}$$

where

$$\frac{\partial\lambda_t}{\partial\beta_0} = 1 + \beta_1\frac{\partial\lambda_{t-1}}{\partial\beta_0} \tag{4.22}$$

$$\frac{\partial\lambda_t}{\partial\alpha_1} = y_{t-1} + \beta_1\frac{\partial\lambda_{t-1}}{\partial\alpha_1} \tag{4.23}$$

$$\frac{\partial\lambda_t}{\partial\beta_1} = \lambda_{t-1} + \beta_1\frac{\partial\lambda_{t-1}}{\partial\beta_1} \tag{4.24}$$

Then the solution of the following set of estimation equations

$$S_n(\theta) = \sum_{t=2}^{n}\left(\frac{y_t}{\lambda_t} - 1\right)\frac{\partial\lambda_t}{\partial\theta_i} = \sum_{t=2}^{n}\left(\frac{y_t - \lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\frac{\partial\lambda_t}{\partial\theta_i} = 0 \tag{4.25}$$

gives the conditional maximum likelihood estimators with $\frac{\partial\lambda_t}{\partial\theta_i}$ defined in (4.20) for model (4.2) and with $\frac{\partial\lambda_t}{\partial\theta_i}$ defined in (4.22), (4.23) and (4.24) for model (4.3).

The conditional maximum likelihood estimators are calculated by numerical optimization of the function $S_n(\theta)$, see Section 4.5 in case of the INARCH(1) model and Section 4.8 in case of the INGARCH(1,1) model.

Zhu and Wang (2009, Theorem 3) show that the conditional maximum likelihood estimators of the INARCH(1) model are asymptotically normally distributed. Ferland et al. (2006) conclude that the distribution of the maximum likelihood estimator of the INGARCH(1,1) model can be approximated by the normal distribution where the asymptotic covariance matrix of the maximum likelihood estimator can be computed from the observed Fisher information matrix. Fokianos et al. (2009) prove that the maximum likelihood estimators of the INGARCH(1,1) model are asymptotically Gaussian under considering geometric ergodicity of this process.

4.3 Robust estimation of the marginal mean

We introduce this section as an entry of robust estimation for the parameters of INGARCH models in the next section. Here, we test if our best functions given in Chapter 3 for i.i.d. Poisson data are still suitable for estimation of the marginal mean in case of Poisson time series data from an INGARCH model or not. Namely, we compare biases and RMSEs for our best functions (glmrob with external weights, tukeypois, trimmeanfit, roptest) with the predefined tuning constants given in Chapter 3, which achieve 95% level of efficiency using Poisson time series data. We set the true marginal mean equal to 2 with $\beta_0 = 1$, $\beta_1 = 0.25$, and $\alpha_1 = 0.25$ and we consider one and two transient outliers of increasing size $\omega \in \{1, 2, ..., 20\}$. A transient outlier can be defined as an outlier whose effect on the time series decays exponentially. Following Fokianos and Fried (2010), we define the INARCH (1) model including this type of outliers as follows:

$$Z_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\kappa_t),$$

$$\kappa_t = \beta_0 + \alpha_1 Z_{t-1} + \omega X_t, \tag{4.26}$$

for $t \geq 1$, where $\mathcal{F}_t$ is the σ-field generated by $\{Z_0, ..., Z_t\}$ representing the whole information up to time t, κ_t is the new conditional mean after adding one transient outlier and Z_t is the contaminated process. ω is the size of the outlier and $\{X_t\}$ is a sequence of deterministic covariates, which has the following form

$$X_t = \delta^{t-\tau} I_t(\tau), \tag{4.27}$$

where $I_t(\tau)$ is an indicator function with $I_t(\tau) = 1$ if t equals a specific time τ and $I_t(\tau) = 0$ if $t \neq \tau$. In our simulations, we set $\delta = 0.8$ and $\tau = 30$.

1. **The results in case of one transient outlier**
   Figure 4.1 compares the biases between glmrob with external weights, tukeypois with $k \in \{5, 6\}$, trimmeanfit and roptest with the predefined tuning constants given in Chapter 3, which achieve 95% level of efficiency, in the presence of one transient outlier in Poisson time series data with mean equal to 2. We find that:

   (a) In case of outliers of sizes from 1 to 9, the differences in biases between all procedures are small.

   (b) In case of outliers of sizes larger than 9, tukeypois with $k = 5$ has the smallest bias, followed by trimmeanfit, then tukeypois with $k = 6$, and finally glmrob with external weights and roptest have the largest biases.

   We obtain the same conclusions with respect to the differences in the RMSEs.

2. **The results in case of two transient outliers**
   Figure 4.2 compares the biases between our predefined functions in Figure 4.1 if we have two transient outliers in Poisson time series data with mean equal to 2.

Figure 4.2 gives the same conclusions as in Figure 4.1, but we note that the biases become larger than before and also the differences in the biases between glmrob with external weights and roptest on one side, and tukeypois and trimmeanfit on the other side become larger than in case of one transient outlier.
Here we again have the same conclusions with respect to the differences in the RMSEs.

In general these results confirm our results in case of i.i.d Poisson data as a guideline for the suitability of our best functions. Next we search for suitable robust estimators of the INGARCH model parameters.

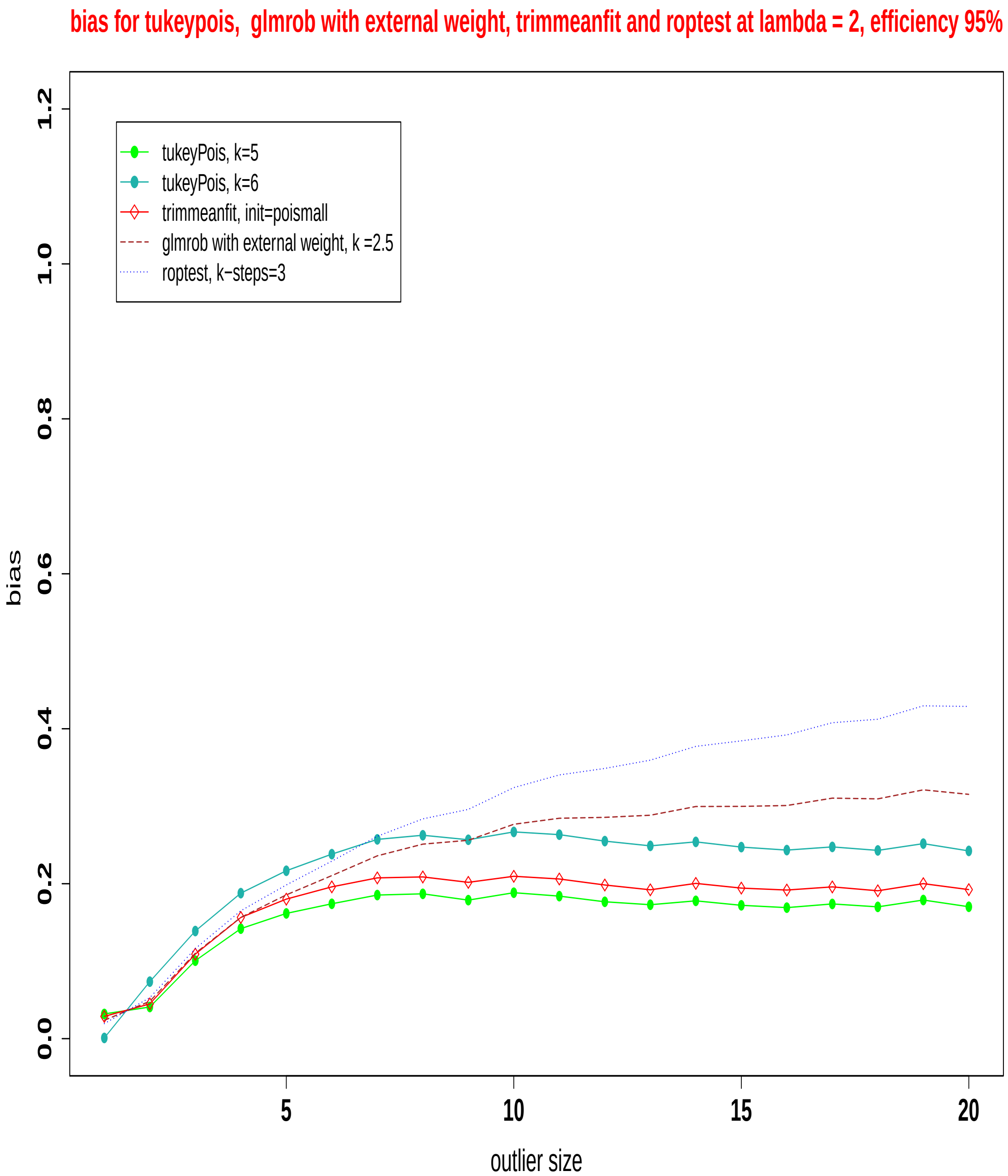


Figure 4.1: Simulated biases of tukeypois, glmrob with external weight, trimmeanfit and roptest tuned to achieve 95% level of efficiency in case of one transient outlier of increasing size from 1 to 20 in a Poisson time series with mean 2.

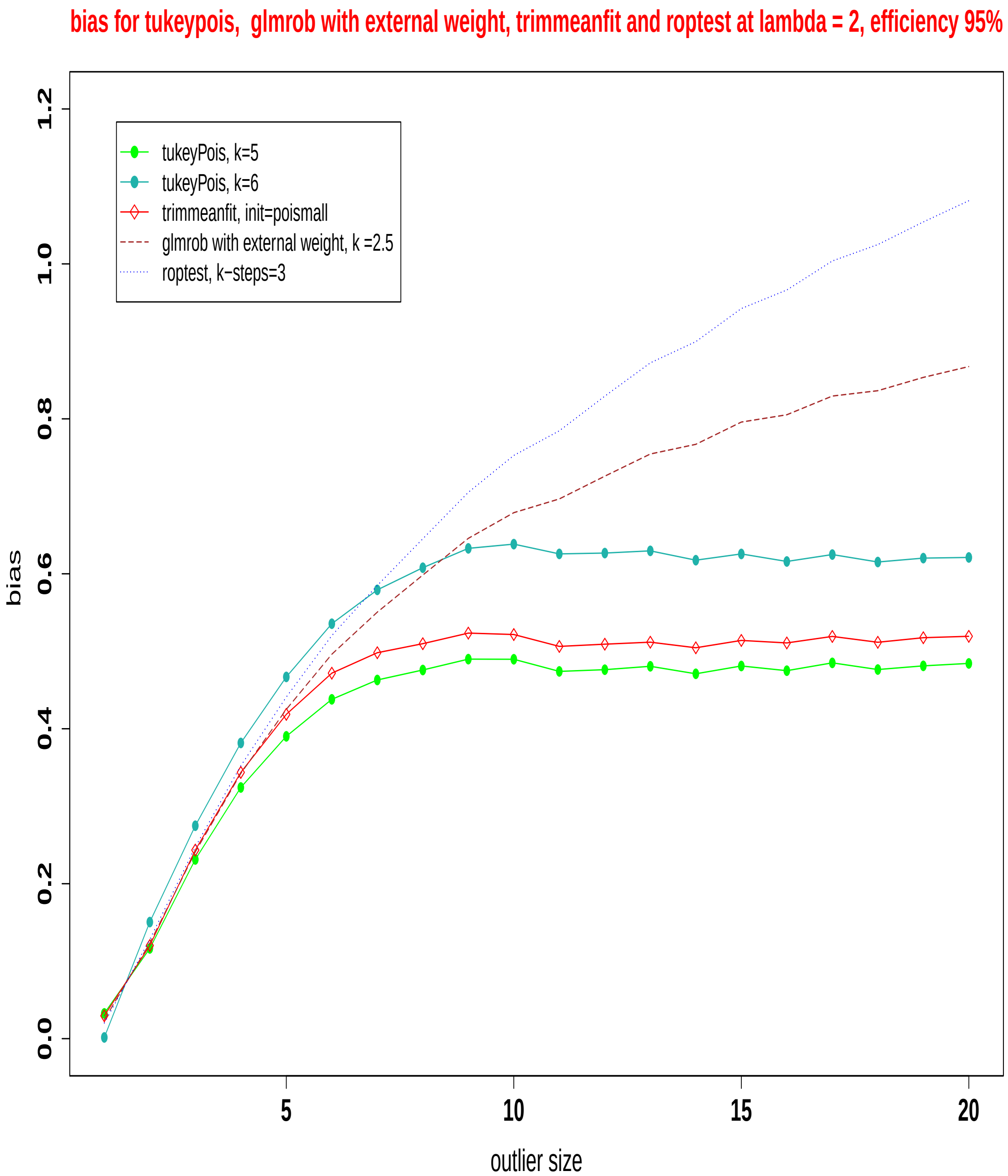


Figure 4.2: Simulated biases of tukeypois, glmrob with external weight, trimmeanfit and roptest tuned to achieve 95% level of efficiency in case of two transient outliers of increasing size from 1 to 20 in a Poisson time series with mean 2.

4.4 M-estimation in INARCH models

1. **M-estimation for the INARCH model without bias correction**
Returning to (4.25) and downweighting the influence of unusual observations in these equations leads to a straightforward robustification of the conditional likelihood estimators. For this, we truncate observations with large standardized residuals $(y_t - \lambda_t)/\sqrt{\lambda_t}$ using Huber's or Tukey's ψ function, and do the same with regressors $y_{t-1}, \ldots, y_{t-p}$ which are outlying w.r.t. the marginal distribution. This leads us to the following set of estimating equations:

$$\sum_{t=p+1}^{n} \psi\left(\frac{y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \frac{1}{\sqrt{\lambda_t}} \begin{pmatrix} 1 \\ \sigma\psi\left(\frac{y_{t-1}-\lambda}{\sigma}\right) + \lambda \\ \vdots \\ \sigma\psi\left(\frac{y_{t-p}-\lambda}{\sigma}\right) + \lambda \end{pmatrix} = 0, \tag{4.28}$$

where λ and σ^2 are the marginal mean and variance for the given set of parameters, respectively, see Elsaied and Fried (2010).
In case of the INARCH(1) model, (4.26) becomes simply

$$\sum_{t=2}^{n} \psi\left(\frac{y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \frac{1}{\sqrt{\lambda_t}} \begin{pmatrix} 1 \\ \sigma\psi\left(\frac{y_{t-1}-\lambda}{\sigma}\right) + \lambda \end{pmatrix} = 0, \tag{4.29}$$

where $\lambda = \beta_0/(1-\alpha_1)$ is the marginal mean and $\sigma^2 = \frac{\beta_0}{(1-\alpha_1)(1-\alpha_1^2)}$ is the marginal variance, see Fokianos et al. (2009).

2. **M-estimation for the INARCH model with bias correction**
After adding a bias correction term, (4.28) becomes

$$\sum_{t=p+1}^{n} \left(\psi\left(\frac{y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \frac{1}{\sqrt{\lambda_t}} \begin{pmatrix} 1 \\ \sigma\psi\left(\frac{y_{t-1}-\lambda}{\sigma}\right) + \lambda \\ \vdots \\ \sigma\psi\left(\frac{y_{t-p}-\lambda}{\sigma}\right) + \lambda \end{pmatrix} - \begin{pmatrix} a_0 \\ \vdots \\ \vdots \\ a_p \end{pmatrix} \right) = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \tag{4.30}$$

with bias correction $B = (a_0, \ldots, a_p)'$, such that the expectation of the term on the left hand side equals 0, see Elsaied et al. (2011).

In case of the INARCH(1) model the bias correction, $B=(a_0,a_1)'$, fulfills

$$E\left[\psi\left(\frac{y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\begin{array}{c}1\\ \sigma\psi\left(\frac{y_{t-1}-\lambda}{\sigma}\right)+\lambda\end{array}\right)-\left(\begin{array}{c}a_0\\ a_1\end{array}\right)\right]=0 \tag{4.31}$$

Following Cantoni and Ronchetti (2001), the bias correction, B, in (4.31) becomes

$$\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\psi\left(\frac{j-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\begin{array}{c}1\\ \sigma\psi\left(\frac{i-\lambda}{\sigma}\right)+\lambda\end{array}\right)\times P\left(Y_t=j|Y_{t-1}=i\right)\times P\left(Y_{t-1}=i\right) \tag{4.32}$$

where $P\left(Y_t=j|Y_{t-1}=i\right)$ is the conditional probability of $Y_t=j$ given $Y_{t-1}=i$, derived from a Poisson with parameter $\lambda_t=\beta_0+\alpha_1 i$, and $P\left(Y_{t-1}=i\right)$ is the marginal probability of $Y_{t-1}=i$, which can be estimated from the empirical relative frequencies in a Monte Carlo experiment, where we generate a long time series from this model.

The rest of this section is devoted to formulate a conjecture, which states some of the asymptotic properties of M-estimators for the INARCH(1) Parameters. Let:
$\theta=(\beta_0,\alpha_1)'$ denote the parameter vector of the INARCH(1) model,
θ_0 be the value of θ for which the expectation of the left hand side of (4.31) equals 0,
$\hat{\theta}=(\hat{\beta}_0,\hat{\alpha_1})'$ denote the vector of the estimators,
$G_n(\hat{\theta}_n)$ denote the left hand side of (4.30) with

$$G_n(\theta)=\left(\begin{array}{c}G_{n1}(\theta)\\ G_{n2}(\theta)\end{array}\right)=\sum_{t=2}^{n}X_t(\theta) \tag{4.33}$$

and let

$$G(\theta)=-E_{\theta_0}(X_t(\theta)) \tag{4.34}$$

where

$$X_t(\theta)=\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\begin{array}{c}1\\ \sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\end{array}\right)-\left(\begin{array}{c}a_0\\ a_1\end{array}\right)=\left(\begin{array}{c}0\\ 0\end{array}\right)$$

We need the following assumptions $(A1)$-$(A6)$, which are given with more details in Appendix A:

- $(A1)$: For model (4.2), the parametric space Θ is compact.
- $(A2)$: $G_n(\theta)$ is a continuous function (this is fulfilled whenever ψ function is continuous, as for the Tukey and the Huber ψ function).
- $(A3)$: $-\frac{1}{n-1}G_n(\theta)$ converges to $G(\theta)$ in probability uniformly in Θ for all $\theta\in\Theta$, and $G(\theta)$ has a unique root.
- $(A4)$: $U_n(\hat{\theta}_n)=\left.\frac{\partial G_n(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_n}$ exists and continuous.

- $(A5)$: $n^{-1}U_n(\theta_n^*)$ converges to a finite nonsingular matrix $A(\theta_0) = \lim En^{-1}U_n(\theta_0)$ in probability for any sequence θ_n^* such that plim $\theta_n^* = \theta_0$.

- $(A6)$: $n^{-\frac{1}{2}}G_n(\theta_0) \rightarrow N(0, B(\theta_0))$, where $B(\theta_0) = \lim_{n\longrightarrow\infty} En^{-1}G_n(\theta_0)G_n(\theta_0)'$

Conjecture 4.1:
Let (Y_t) be an INARCH(1) model such that $\alpha_1 < 1$. Under assumptions $(A1) - (A6)$, as $n \longrightarrow \infty$,

(a) $\hat{\theta}_n \xrightarrow{a.s} \theta_0$ (strong consistency) and

(b) $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\Big(0, A(\theta_0)^{-1}B(\theta_0)(A(\theta_0)^{-1})^t\Big)$ (normal convergence)

We investigate the background of this conjecture in Appendix A.

4.5 Computation

In this section we describe some new functions, which we implemented in R to compute

1. conditional maximum likelihood estimates for the parameters of the INARCH(1) model using function **condlikest**.

2. robust M-estimates for the parameters of the INARCH(1) model without bias correction and with robust initialization under the assumption of independence using functions **robPoisTShuber1** and **robPoisTStukey1**.

3. robust M-estimates for the parameters of the INARCH(1) model without bias correction and with robust initialization from the AR(1) model using functions **robPoisTShuber2** and **robPoisTStukey2**.

4. robust M-estimates for the parameters of the INARCH(1) model with bias correction and with robust initialization from the AR(1) model using functions **robPoisTShuber3** and **robPoisTStukey3**.

1. **condlikest**
   We can describe the function condlikest by the following steps:

   (a) We define the function "liklinear.poisson.un" as a function in theta and data, where theta is our parameter vector. This function calculates the value of the log-likelihood function (4.25).

   (b) We define the function "condlikest" as a function in y and theta. This function gives classical estimates for the parameters of the INARCH(1), β_0 and α_1. These estimates are obtained using the function constrOptim.

   (c) constrOptim is a function in theta, f, grad, ui, ci, mu = 1e-04, control = list(), method = if(is.null(grad)) "Nelder-Mead" else "BFGS", outer.iterations = 100, outer.eps = 1e-05, ...
   Here, theta is our parameter vector (β_0, α_1). We initialize β_0 with the mean of our data (y) and we let $\alpha_1 = 0.001$ or we initialize it using an AR(1) fit,
   f is our score function "liklinear.poisson.un", defined above in (a),
   grad is the function scorelinear.poisson.un, which we implemented to compute the derivatives of our score function with respect to its parameters,
   ui, ci are a matrix and a vector for our constraints β_0, $\alpha_1 \geq 0$ and $\alpha_1 < 1$,
   method is "BFGS", which is a quasi-Newton method. This method uses function values and gradients to build up a picture of the surface to be optimized, see the function optim in R for more details,
   outer.iterations is set to 100 and outer.eps to 1e-05.

2. **robPoisTStukey1**
   This function is based on Tukey's ψ function and we can describe it as follows:

   (a) We define the function "robPoistukeyscores1" as a function in param, y, k, which calculates the left hand side of (4.29) using Tukey's ψ function.

   (b) We define the function "robPoisTStukey1" as a function in y, k, iter=200. This function gives robust estimates for the parameters of the INARCH(1) model, β_0 and α_1. These estimates are obtained using the function constrOptim.

   (c) constrOptim is a function in theta, f, grad, ui, ci, mu = 1e-04, control = list (), method = if (is.null (grad)) "Nelder-Mead" else "BFGS", outer.iterations = 100, outer.eps = 1e-05, ...
   Here, theta is our parameter vector (β_0, α_1). We initialize theta under the assumption of independence with a robust estimate of β_0 using the function tukeypois and $\alpha_1 = 0.001$,
   f is our score function "robPoistukeyscores1", defined above in (a),
   grad is set to null because we do not make use of the derivatives,
   ui, ci are a matrix and a vector describing our constraints β_0, $\alpha_1 \geq 0$ and $\alpha_1 < 1$,
   method is "Nelder-Mead", which uses only function values and is robust but relatively slow. It will work reasonably well for non-differentiable functions, see the function optim in R for more details.
   outer.iterations is set to 200 and outer.eps to 1e-06.

3. **robPoisTStukey2**
   This function is a modified version of the function robPoisTStukey1, where we use a robust initialization for β_0 and α_1 to start the iterations in constrOptim using robust estimates in the AR(1) model. According to the properties of INARCH(1), we conclude that the first two moments of the INARCH(1) with parameters β_0 and α_1 are identical to those of an AR(1) model with the same dependence parameter α_1 and marginal mean $\lambda = \beta_0/(1-\alpha_1)$. Therefore, it suffices to estimate α_1 using a robust estimate of the lag-one dependence parameter $\alpha_1 = \rho(1)$ in the AR(1) model. Then we estimate the marginal mean using the function tukeypois and afterwards we use the formula of the marginal mean $\beta_0/(1-\alpha_1)$ to obtain an initial estimate of β_0. To estimate the lag-one parameter robustly, we apply the highly robust estimate of Ma and Genton (2000) as follows:

$$\hat{\rho}(1) = \frac{Q_{n-1}^2(y_2 + y_1, \ldots, y_n + y_{n-1}) - Q_{n-1}^2(y_2 - y_1, \ldots, y_n - y_{n-1})}{Q_{n-1}^2(y_2 + y_1, \ldots, y_n + y_{n-1}) + Q_{n-1}^2(y_2 - y_1, \ldots, y_n - y_{n-1})}, \qquad (4.35)$$

   using Rousseeuw and Croux (1993)'s Q_n for estimation of the unknown variances $Var(Y_t + Y_{t-1})$ and $Var(Y_t - Y_{t-1})$ because of its high robustness and considerable efficiency. The Q_n scale estimator of a sample $x_1, \ldots, x_m$ roughly corresponds to the 25% percentile of the sample of pairwise differences and is defined by

$$Q_m(x_1,\ldots,x_m) = c_m\{|x_i - x_j| : 1 \le i < j \le m\}_{(l)}, \quad l = \binom{\lfloor m/2 \rfloor + 1}{2}. \quad (4.36)$$

c_m is a finite sample correction factor to achieve unbiasedness at a sample of size m. It can be omitted in our context since it cancels out.

4. **robPoisTStukey3**.
This function is a modified version of the function robPoisTStukey2, where we use the same robust initialization for β_0 and α_1. However, here we correct our score function robPoistukeyscores3 by subtracting suitable bias correction terms. These correction terms have been computed based on (4.32). The estimates are obtained using the function constrOptim. constrOptim iterates between estimating β_0 and α_1 for a given value of bias corrections for β_0 and α_1 and the specification of bias corrections based on the current estimate of β_0 and α_1. First we choose a value of the correction terms. Second we estimate β_0 and α_1, then we derive the values of our correction terms which are suitable for these estimates. Then we re-estimate β_0 and α_1. We iterate this process until the absolute differences of our estimates for two consecutive iterations do not exceed our desired accuracy of 0.0001, which we use to determine convergence.

The functions **robPoisTShuber1**, **robPoisTShuber2** and **robPoisTShuber3** can be described in the same way as for the functions **robPoisTStukey1**, **robPoisTStukey2** and **robPoisTStukey3**, but we use Huber's ψ function instead of Tukey's ψ function.

4.6 Simulations

In this section, we perform some simulation experiments to compare the performance of the conditional maximum likelihood estimator, computed using the function condlikest, and the generalized M-estimators, computed using the different versions of the functions robPoisTStukey and robPoisTShuber. Namely, we consider the following situations of M-estimators:

- without bias correction and with robust initialization under the assumption of independence, see the results of the functions robPoisTStukey1 and robPoisTShuber1.
- with and without bias correction and with robust initialization using robust estimates in an AR(1) model, see the results of the functions robPoisTStukey2 and robPoisTStukey3.

Because the INARCH(1) model belongs to the class of generalized linear models, we include the function glmrob in our comparison using the identity link, family Poisson and using the lagged variables as regressors. We initialize our parameters estimates with $\beta_0 = 0.1$ and $\alpha_1 = 0.1$ or we use a robust initialization from robust AR(1) fit in the same way as in the function robPoisTStukey2.

These estimators are compared in terms of bias and root of mean square error (RMSE) in finite samples from an INARCH(1) model and with several tuning constants k.

4.6.1 Results for initialization from assuming independence

Here we compare

1. conditional maximum likelihood estimators for the parameters of the INARCH(1) model, with initialization of β_0 by the mean of our data (y) and $\alpha_1 = 0.001$, there were no differences between the two initialization used for this estmator, see function "condML".
2. robust M-estimators for the parameters of the INARCH(1) model using the function glmrob, with initialization $\beta_0 = 0.1$ and $\alpha_1 = 0.1$, see function "glmrob".
3. robust M-estimators for the parameters of the INARCH(1) model using the function robPoisTShuber1 with different tuning constants $k \in \{1.8, 2.5\}$, with initialization from robust estimate of β_0 using the function huperpois and $\alpha_1 = 0.001$, see function "Huberuncorr1".
4. robust M-estimators for the parameters of the INARCH(1) model using the function robPoisTStukey1 with initialization from robust estimate of β_0 using the function tukeypois and $\alpha_1 = 0.001$ and with different tuning constants $k \in \{5, 7\}$, and we add an adaptive choice of k between 5 and 7 depending on the estimate of α_1 (we use $k = 7$ if $0.7 < \hat{\alpha}_1$ or $\hat{\alpha}_1 < 0.3$ and $k = 5$ otherwise); we add this adaptive choice

of k to achieve more stable efficiency than using a fixed value of k over the whole parameter range, see function "Tukeyuncorr1".

These estimators are compared in several situations with and without outliers as follows:

Results in case of clean data

To compare the efficiencies and the biases of the predefined estimators, we generate data from INARCH(1) models with parameters $\beta_0 = 1$, $\alpha_1 \in \{0, 0.1, \ldots, 0.9\}$. The results are based on 2000 data sets of size 100 each.
Figure 4.3 illustrates relative efficiencies of the generalized M-estimators relative to the conditional maximum likelihood estimator as a function of α_1. Figure 4.3 shows that the estimators generally achieve good efficiencies, but have some problems as α_1 approaches 1, which is the non-stationary case. Only glmrob has some problems with the estimation of α_1 if it is small.
Figure 4.4 illustrates the resulting finite samples biases of the conditional maximum likelihood estimator, glmrob and the generalized M-estimators. Form Figure 4.4, the differences between the bias curves for β_0 are small except if α_1 approaches 1. For α_1, the estimators show a similar bias behavior except if α_1 approaches 1 and glmrob gives the smallest bias.

Results in case of contaminated data

To compare the robustness, bias curves are approximated using 1000 time series of length n = 100 from an INARCH(1) model with $\beta_0 = 1$ and $\alpha_1 = 0.4$. We consider a single transient outlier of increasing size $[j\sigma]$ (rounding to an increasing number j of multiples of the marginal standard deviation). Figure 4.5 shows the robustness of the estimators, where the conditional maximum likelihood estimator shows little bias for the intercept β_0 in this situation because α_1 absorbs almost all of the outlier effect. At sizes of the outlier ≥ 5, glmrob has the largest bias for the intercept β_0 and it has only slightly smaller bias than the conditional maximum likelihood estimator for α_1. robPoisTStukey1 with its different tuning constants shows better performance than robPoisTShuber1.

Next we will give the results for the other versions of the function robPoisTStukey (robPoisTStukey2 and robPoisTStukey3). The results of the functions robPoisTShuber2 and robPoisTShuber3 are illustrated in Appendix B.

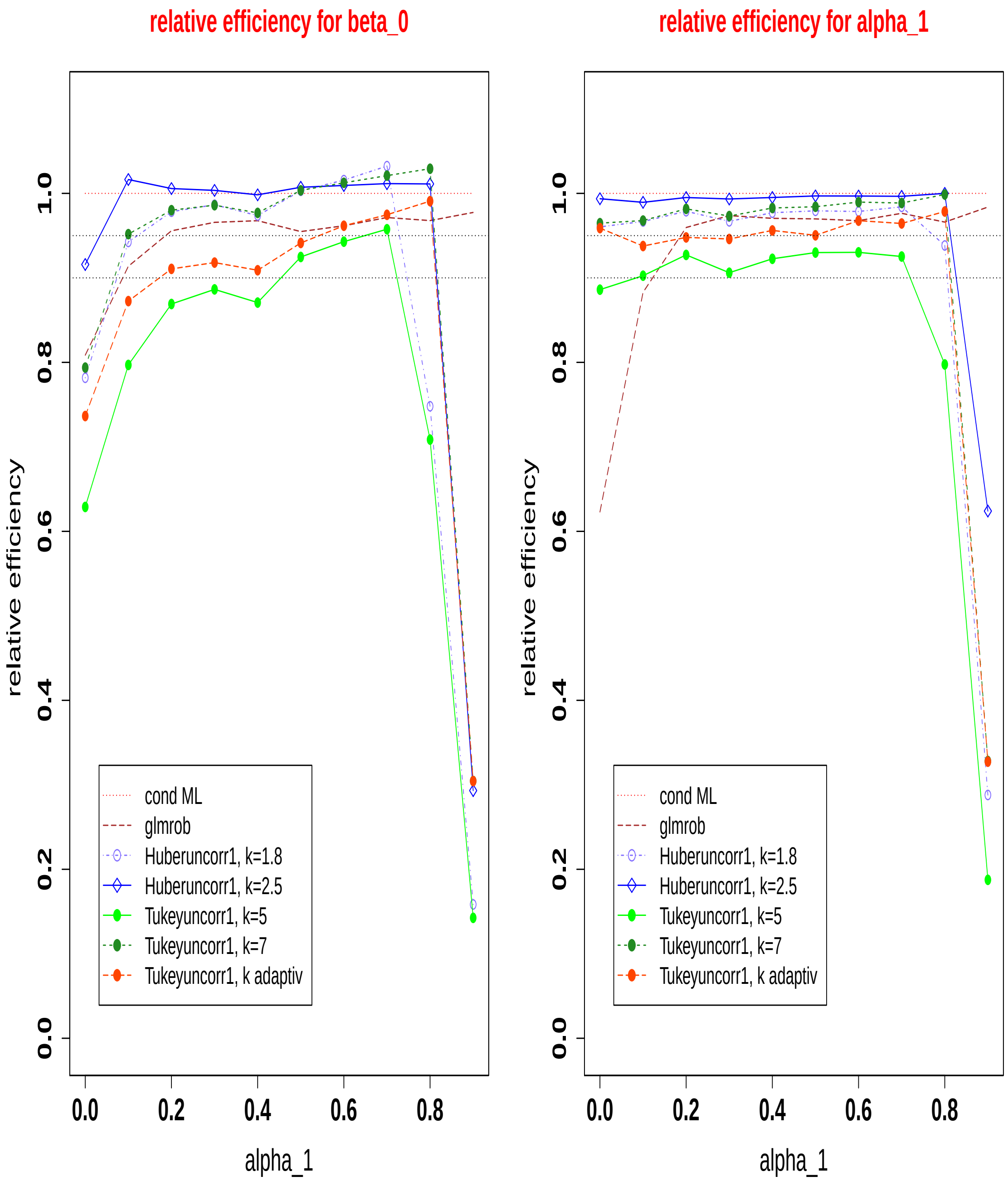


Figure 4.3: Simulated relative efficiencies of glmrob, Huber and Tukey M-estimators with different tuning constants k relative to the conditional maximum likelihood estimator for β_0 (left) and α_1 (right) as a function of the true α_1, n=100.

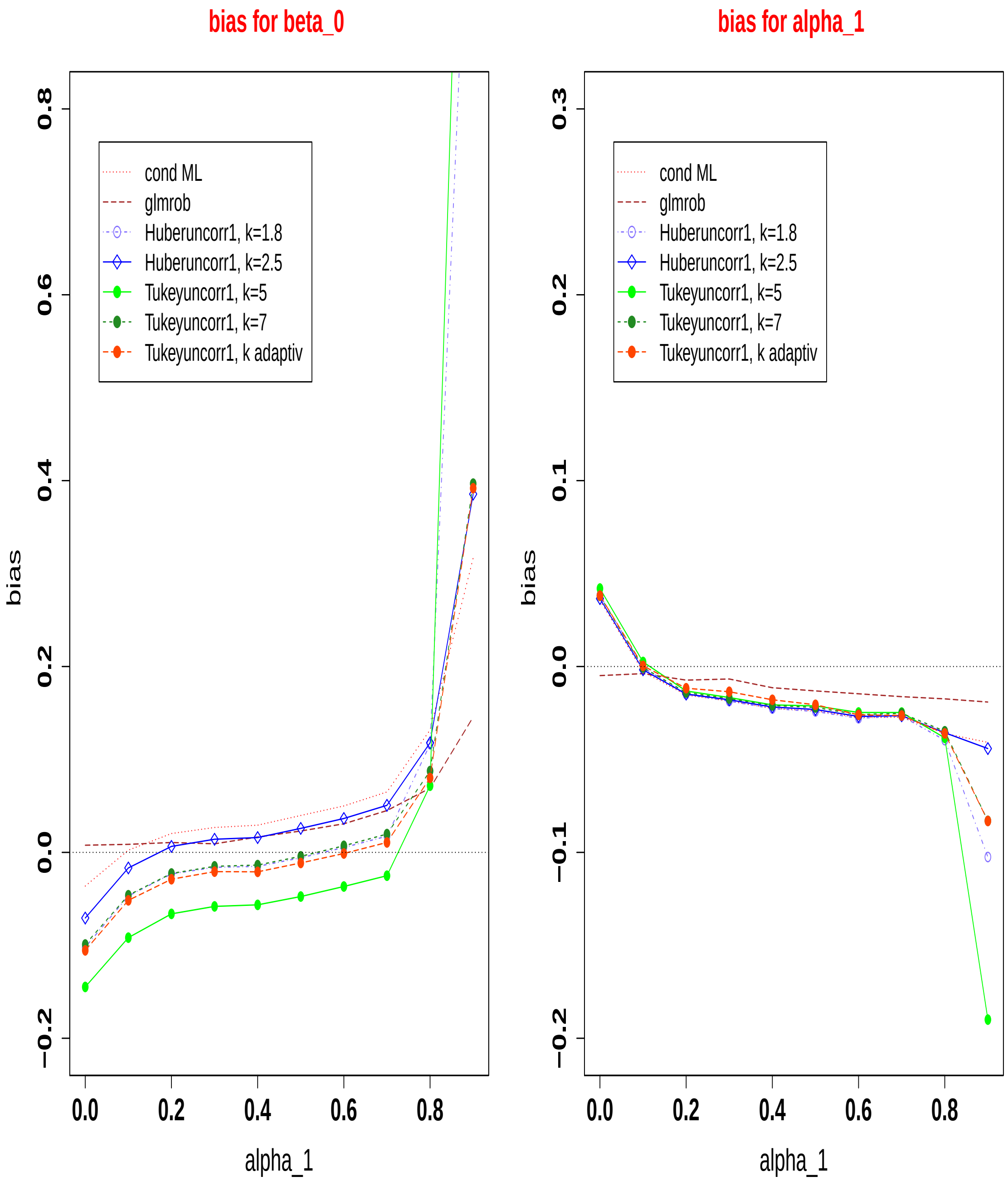


Figure 4.4: Simulated biases of the conditional maximum likelihood estimator, glmrob and of Huber and Tukey M-estimators with different tuning constants k for β_0 (left) and α_1 (right) as a function of the true α_1, n=100.

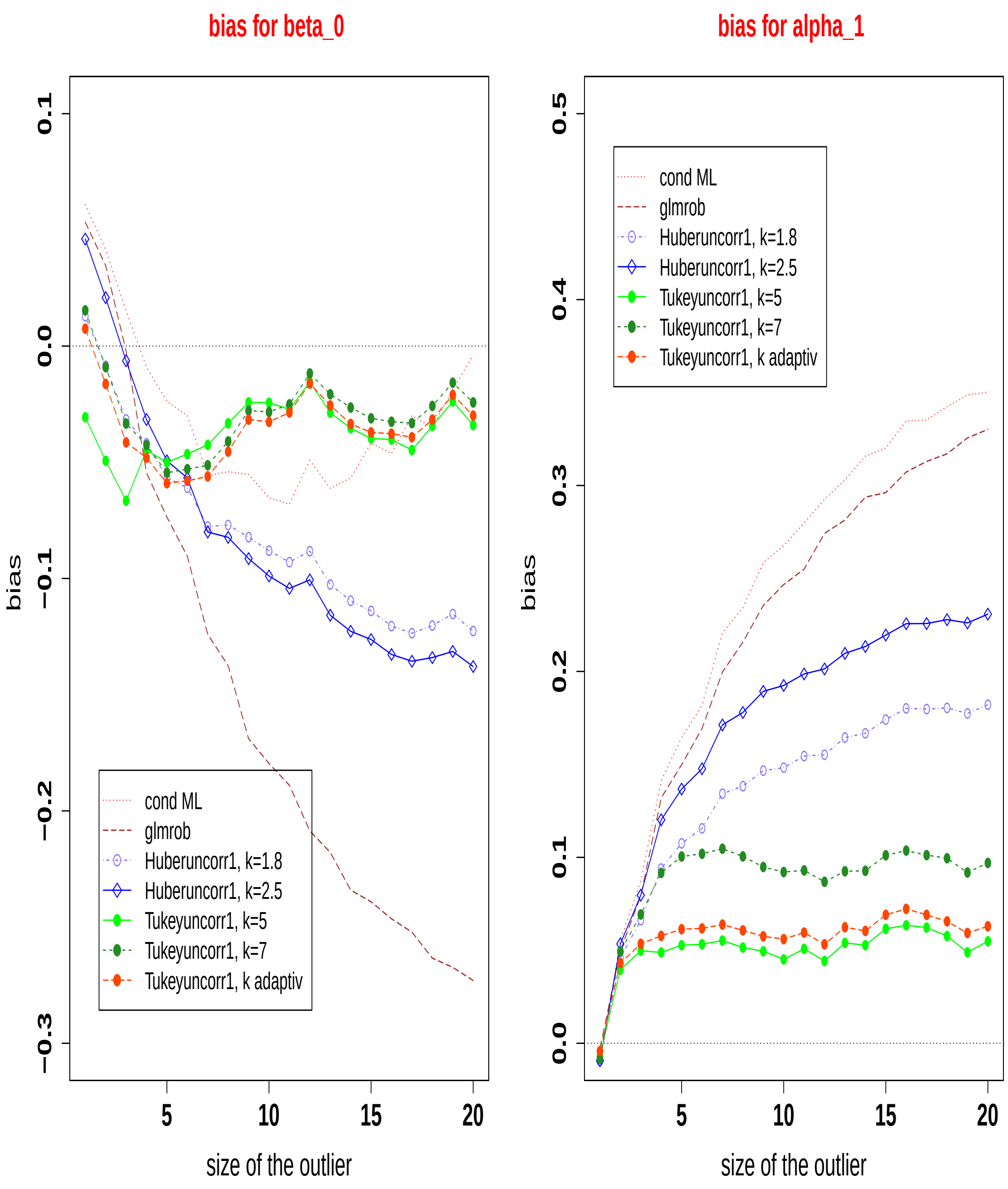


Figure 4.5: Simulated biases of the conditional maximum likelihood estimator, glmrob and of Huber and Tukey M-estimators with different tuning constants k for β_0 (left) and α_1 (right) in case of one transient outlier of increasing size with $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=100.

4.6.2 Results for initialization from robust AR(1) fit.

Here we compare

1. conditional maximum likelihood estimators for the parameters of the INARCH(1) model with initialization from an AR(1) model, see function "cond ML, initial AR(1)".

2. robust M-estimators for the parameters of the INARCH(1) model using the function glmrob with robust initialization from an AR(1) model, see function "glmrob, initial AR(1)".

3. robust M-estimators for the parameters of the INARCH(1) model with robust initialization from an AR(1) model and without bias correction using the function robPoisTStukey2, see function "Tukeyuncorr2".

4. robust M-estimators for the parameters of the INARCH(1) model with robust initialization from an AR(1) model and with bias correction using the function robPoisTStukey3, see function "Tukeycorr".

Robust M-estimators are compared with different tuning constants $k \in \{5, 7, 10\}$ and with an adaptive choice of k between 5, 7, and 10. The adaptive values depend on the initial estimate of α_1, which is computed using the function racf (we use $k = 7$ if $0.4 < \hat{\alpha}_1 < 0.8$, $k = 10$ if $0.8 < \hat{\alpha}_1$ and $k = 5$ otherwise). These estimates are compared in several situations with and without outliers as follows:

Results in case of clean data

We generate our data as follows:

- Data come from INARCH(1) models, with parameters $\beta_0 = 1$ and $\alpha_1 \in \{0, 0.1, \ldots, 0.9\}$.
- The results are based on 500 data sets for each of different sizes 100 and 200.

Figure 4.6 illustrates the efficiencies (left) and biases (right) for β_0 using the predefined functions with a sample size equal to 100. We find that:
The estimators generally achieve good efficiencies, except glmrob and Tukeyuncorr2, Tukeycorr with $k = 5$. We have small differences between the bias curves if $\alpha_1 \leq 0.8$, except the bias curve of Tukeyuncorr2 with $k = 7$, which has the largest bias. Tukeycorr with $k = 7$ and with adaptive k have the same efficiency and give the smallest biases (the orange line and the forest-green line are identical).
Visual comparison of Figures 4.3 (left) and 4.6 (left) indicates that the efficiencies for β_0 of Tukeyuncorr with $k \in \{5, 7\}$ in case of using robust initialization from an AR(1) fit is larger than for the initialization from independence. Comparison of Figures 4.4 (left) and 4.6 (right) indicates that this can be explained by the biases becoming smaller.

Figure 4.7 illustrates the efficiencies (left) and biases (right) for β_0 using the predefined functions with sample size equal to 200. We find that:
The bias-corrected estimators give good efficiencies, except glmrob and Tukeyuncorr2, Tukeycorr with $k = 5$. The differences in biases become smaller than in case of the sample size equal to 100. This means the bias correction works better in case of a larger sample size (the resulting estimators are asymptotically unbiased). Tukeycorr with $k = 7$ and with adaptive k again give the smallest biases.

Figure 4.8 illustrates the efficiencies (left) and biases (right) for α_1 using the predefined functions with a sample size equal to 100. The estimators generally achieve good efficiencies, except glmrob, Tukeyuncorr2 and Tukeycorr with $k = 5$. Among the estimators achieving high efficiency, Tukeycorr with $k = 7$ and with k adaptive give the smallest biases. The bias correction gives some improvements if α_1 is moderate to large.
By comparison between Figures 4.4 (right) and 4.8 (right), we find that the biases for α_1 of Tukeyuncorr with $k \in \{5, 7\}$ in case of using robust initialization from an AR(1) model become smaller than when initializing from independence, and the efficiencies become higher.

The results for a sample size equal to 200 are very similar, but with the bias again becoming substantially smaller, see Figure 4.9.

We run further simulation studies to check the consistency and the asymptotic normality of our estimators. For this, we draw boxplots and qq-plots for the parameters estimates obtained from conditional maximum likelihood, see function "cond ML", and the robust parameters estimates of the INARCH(1) model with robust initialization from an AR(1) model, without bias correction and with tuning constant $k = 5$, see function "Tukeyuncorr2".
Figure 4.10 depicts boxplots of the conditional maximum likelihood estimates and the uncorrected Tukey M-estimates with tuning constant $k = 5$ for β_0 (top) and α_1 (bottom) in case of data generated from an INARCH(1) with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, obtained from 5000 for each different sizes 100, 200, and 500 (from left to right). The boxplots indicate that the conditional maximum likelihood estimator and the uncorrected Tukey M-estimators with tuning constant $k = 5$ are consistent and the distributions become more symmetric and roughly normal when we increase the sample size. The qq-plots in Figures 4.11 and 4.12 for β_0 and α_1, respectively, confirm these results and support the conjecture of asymptotic normality.
Note that boxplots and qq-plots for the parameters estimators with bias correction are similar to the case without bias correction and thus not shown here. The results for the uncorrected Tukey M-estimators with tuning constants $k \in \{7, 10\}$ lie in between those for the conditional maximum likelihood estimator and the uncorrected Tukey M-estimators with tuning constant $k = 5$ (also not shown here).

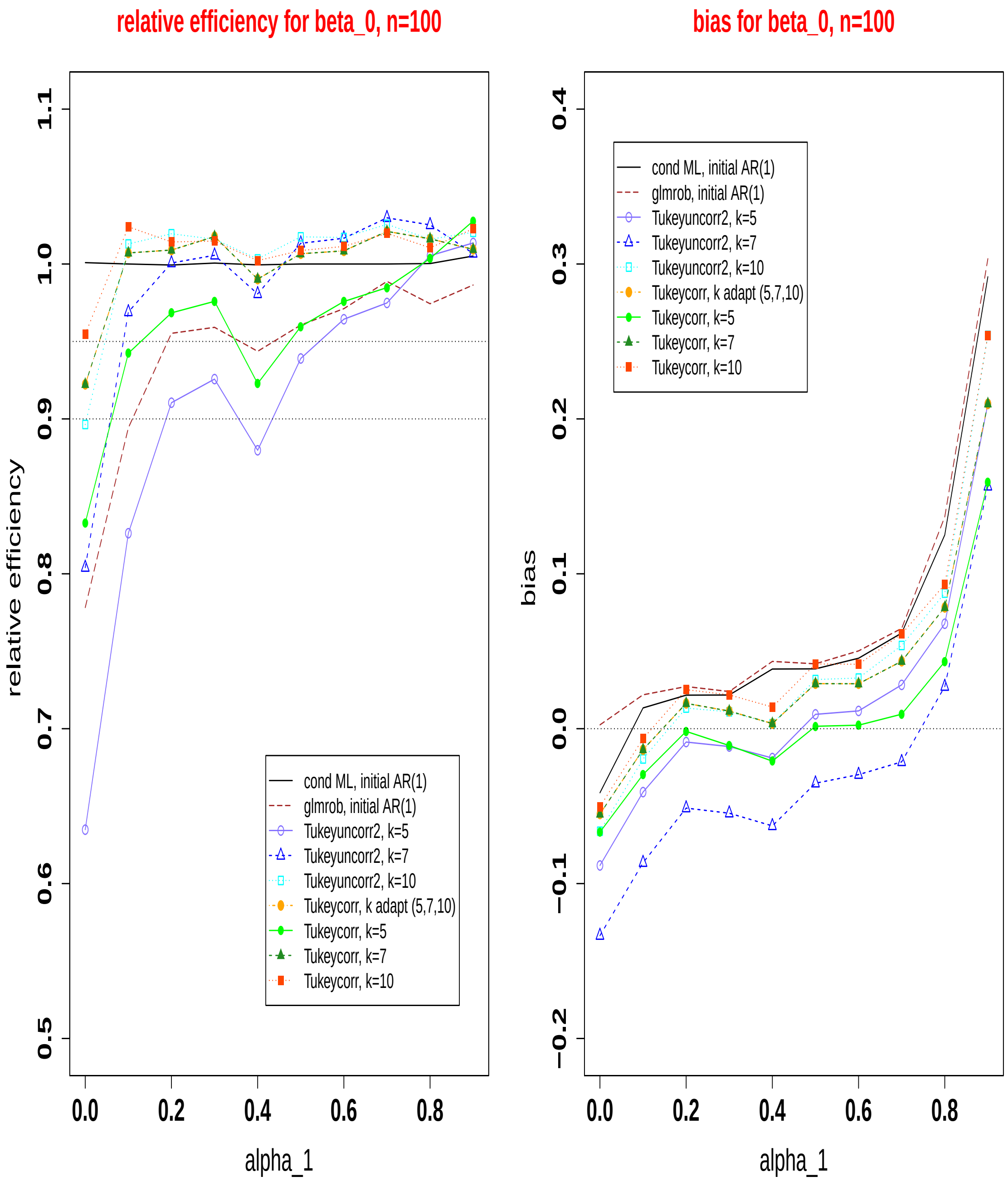


Figure 4.6: Simulated biases for β_0 (right) and relative efficiencies for β_0 (left) of glmrob, corrected and uncorrected Tukey M-estimators with different tuning constants k, relatively to the conditional maximum likelihood estimator, as a function of the true value of α_1, for $\beta_0 = 1$, n=100.

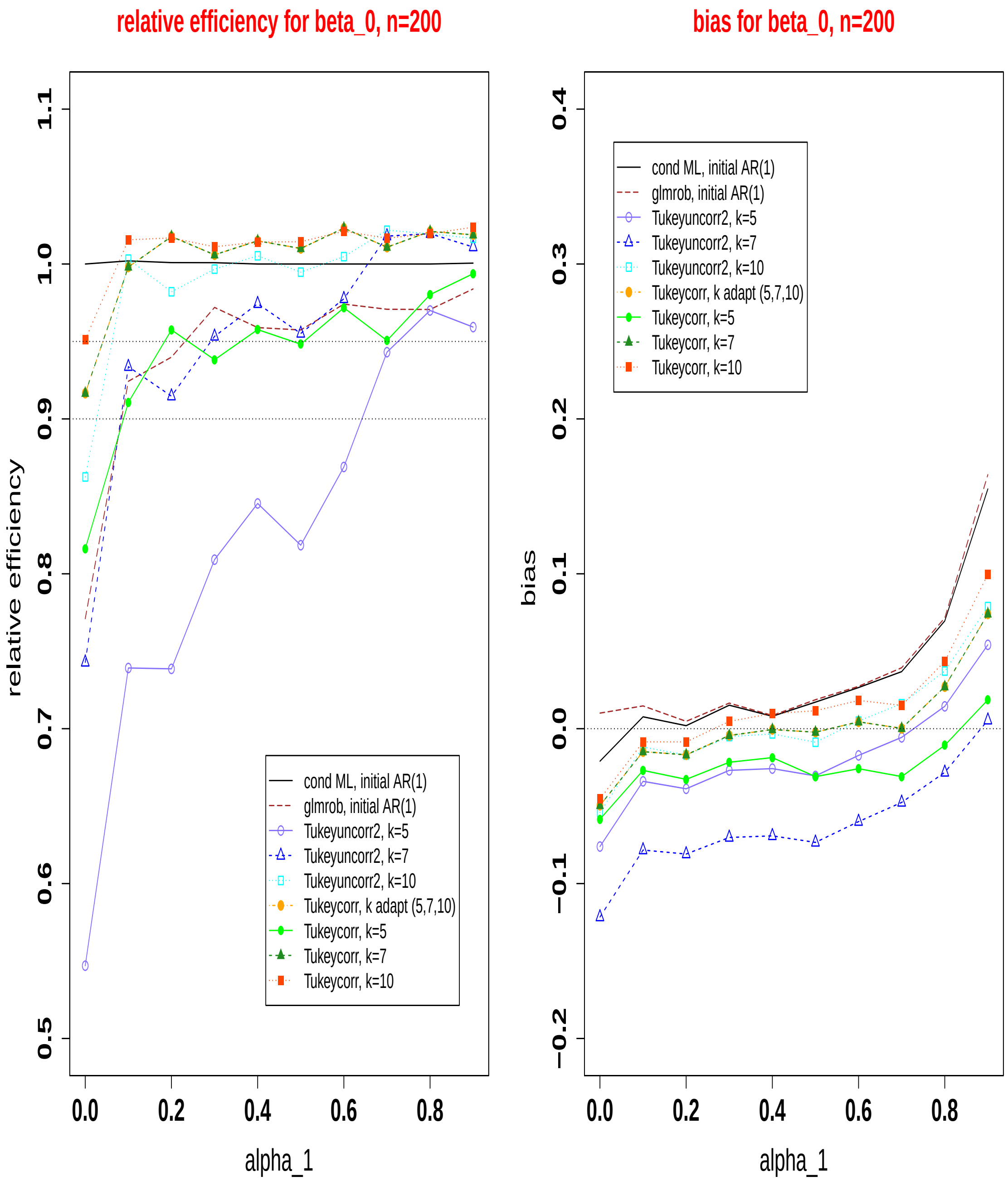


Figure 4.7: Simulated biases for β_0 (right) and relative efficiencies for β_0 (left) of glmrob, corrected and uncorrected Tukey M-estimators with different tuning constants k, relatively to the conditional maximum likelihood estimator, as a function of the true value of α_1, for $\beta_0 = 1$, n=200.

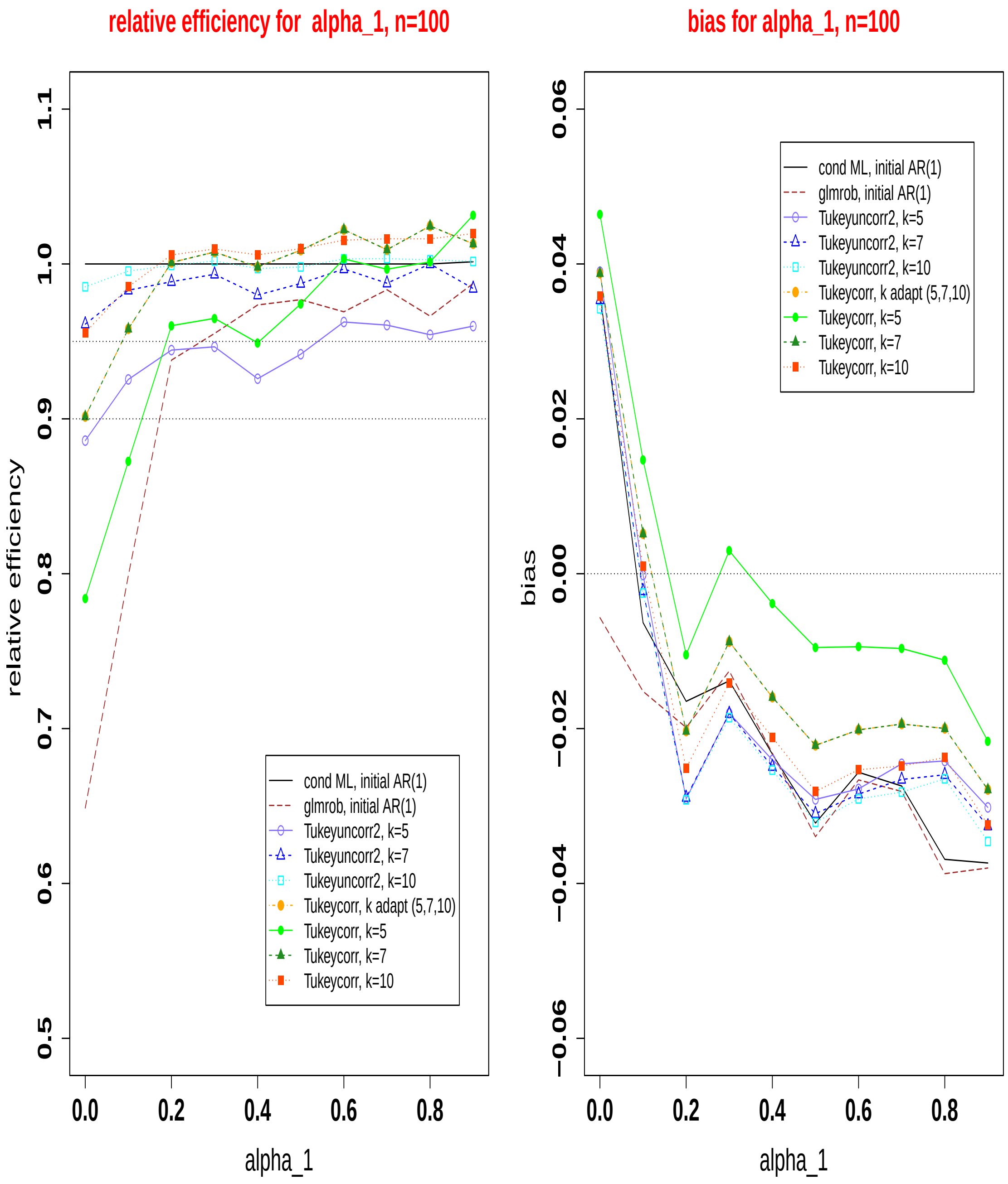


Figure 4.8: Simulated biases for α_1 (right) and relative efficiencies for α_1 (left) of glmob, corrected and uncorrected Tukey M-estimators with different tuning constants k, relatively to the conditional maximum likelihood estimator, as a function of the true value of α_1, for $\beta_0 = 1$, n=100.

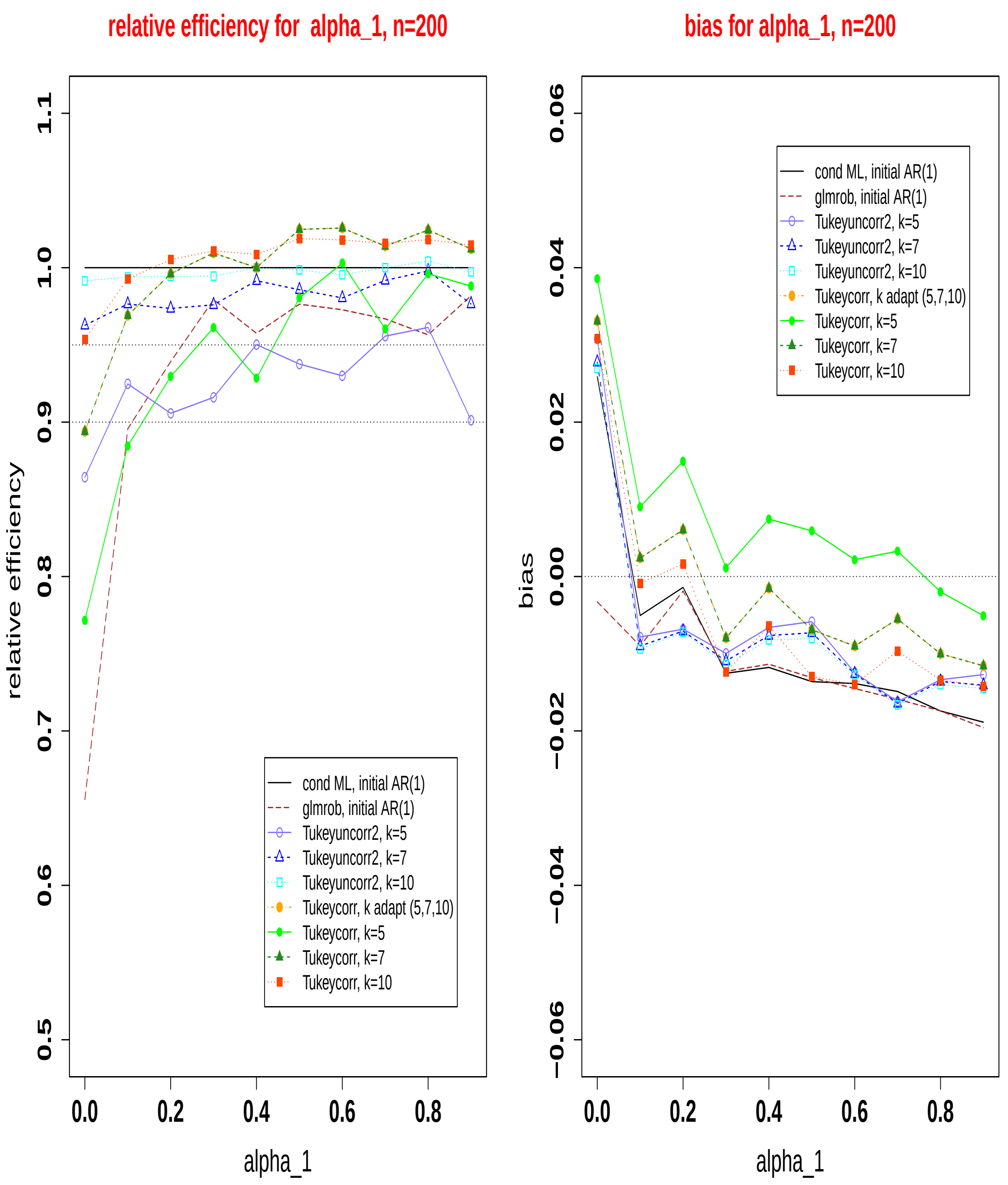


Figure 4.9: Simulated biases for α_1 (right) and relative efficiencies for α_1 (left) of glmrob, corrected and uncorrected Tukey M-estimators with different tuning constants k, relatively to the conditional maximum likelihood estimator, as a function of the true value of α_1, for $\beta_0 = 1$, n=200.

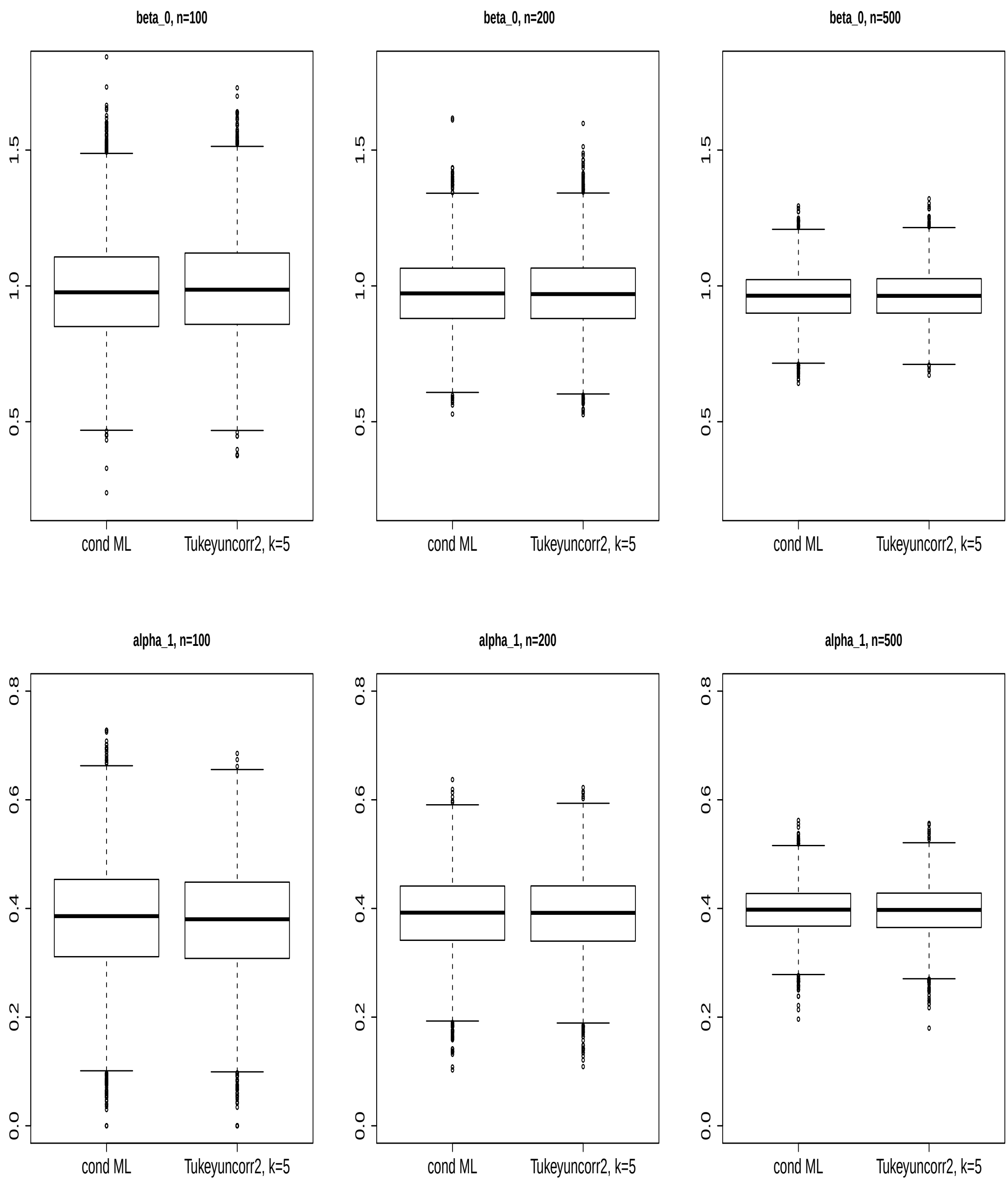


Figure 4.10: Boxplots of the conditional maximum likelihood estimator and uncorrected Tukey M-estimators with tuning constant $k = 5$ for β_0 (top) and α_1 (bottom) estimated from INARCH(1) with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, 5000 data sets of sizes 100, 200, and 500 (from left to right).

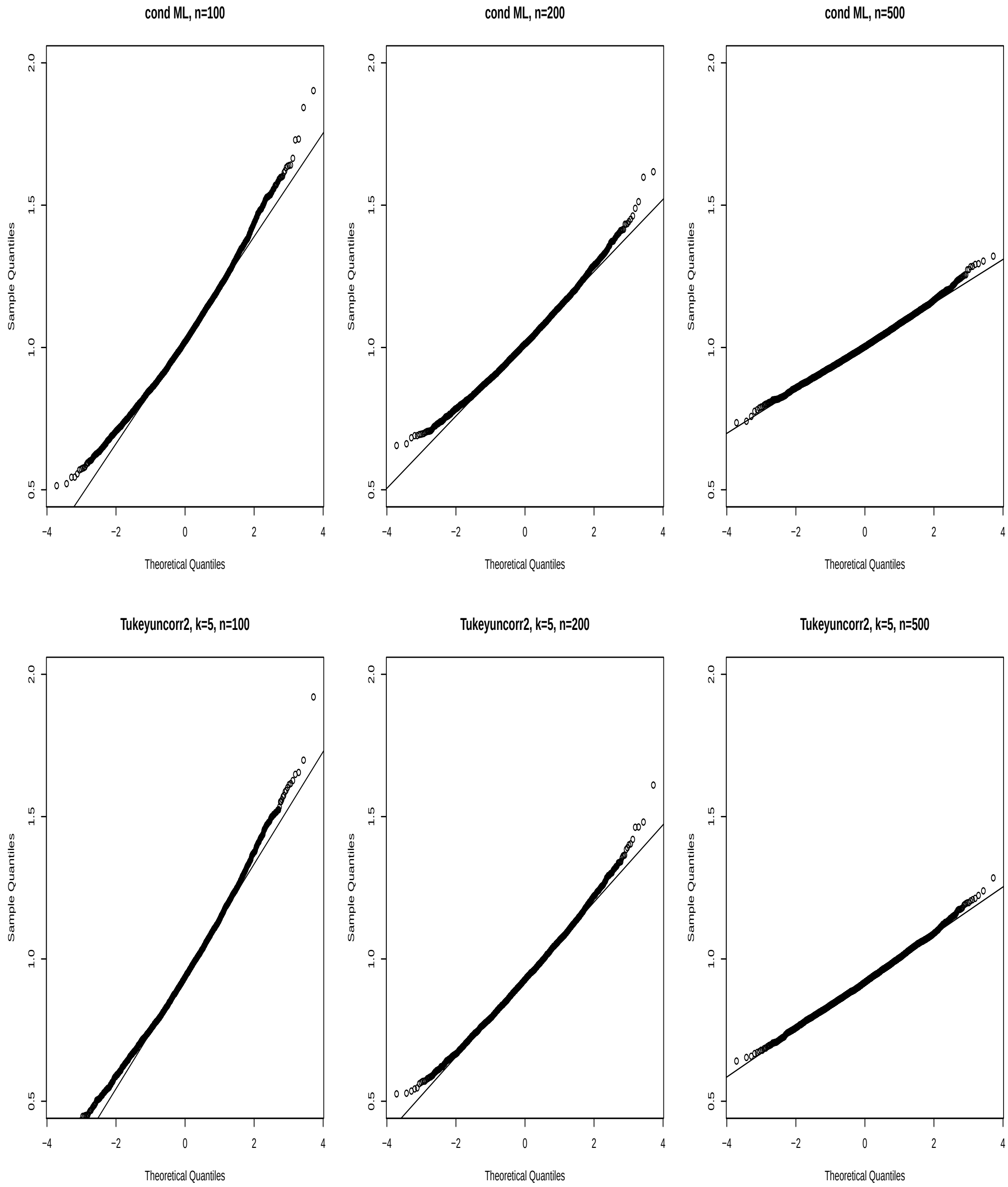


Figure 4.11: QQ-plots of the conditional maximum likelihood estimator and uncorrected Tukey M-estimators with tuning constant $k = 5$ for β_0 estimated from INARCH(1) with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, 5000 data sets of sizes 100, 200, and 500 (from left to right).

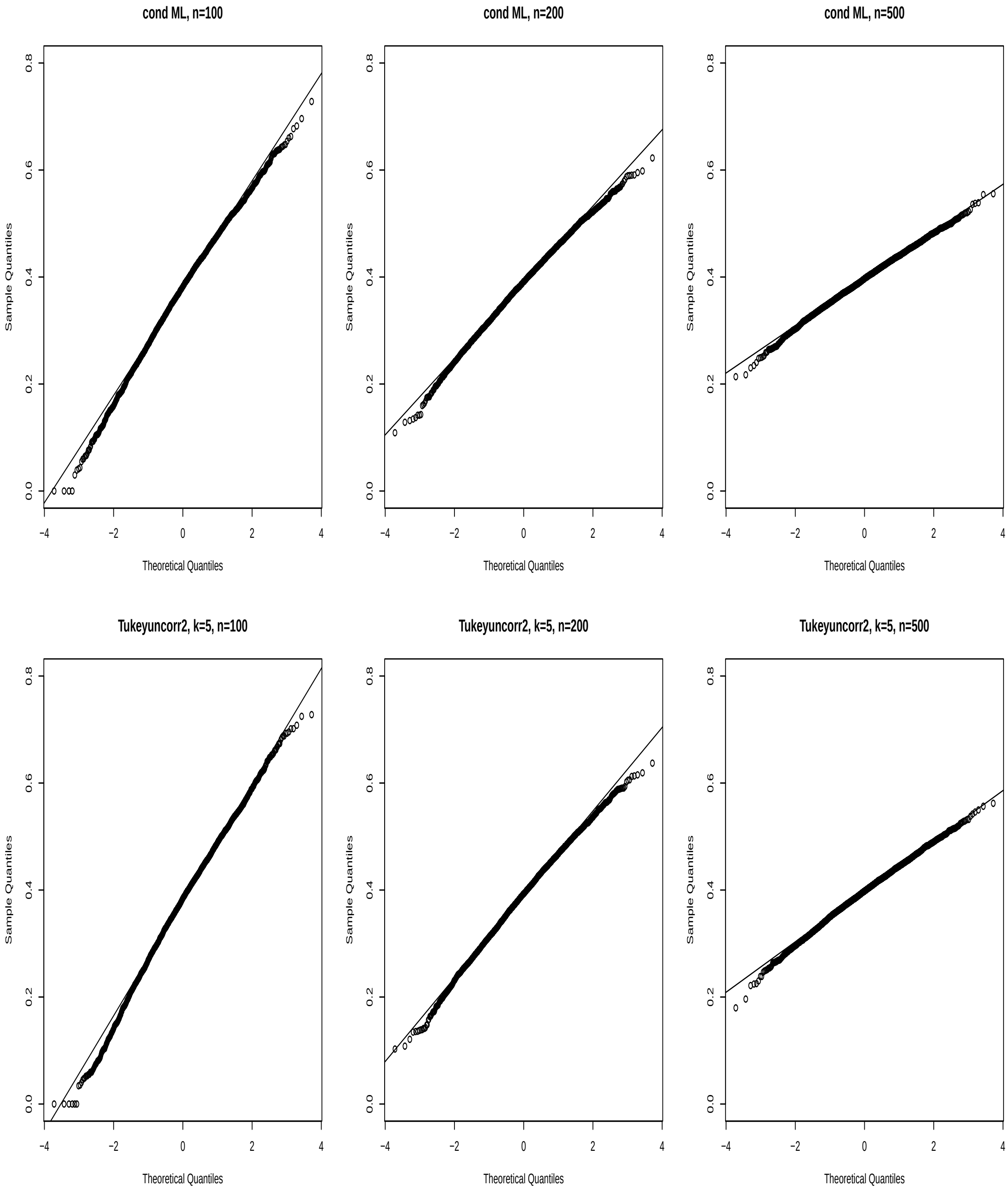


Figure 4.12: QQ-plots of the conditional maximum likelihood estimator and uncorrected Tukey M-estimators with tuning constant $k = 5$ for α_1 estimated from INARCH(1) with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, 5000 data sets of sizes 100, 200, and 500 (from left to right).

Results in case of contaminated data

Here we compare the predefined functions from Subsection 4.6.2 in the presence of outliers. We omit function Tukeycorr with adaptive k, because it gives the same results as for $k = 7$ in the situations considered here. Our data come from an INARCH(1) model with parameters $\beta_0 = 1$ and $\alpha_1 = 0.4$. The results are based on 200 data sets of size 200, where we consider the following four outlier scenarios:

1. **One transient outlier of increasing size**:
   We consider a single transient outlier of increasing size $[\omega = j\sigma]$ (rounding to an increasing number $j \in \{1, 2, ..., 20\}$ of multiples of the marginal standard deviation σ). The definition of the INARCH(1) including this type of outliers is given in Section 4.3. In our simulations, we set $\delta = 0.8$ and $\tau = 50$.

2. **One additive outlier of increasing size**:
   We consider a single additive outlier of increasing random size generated from Poisson distributions with means from 1 to 20 at time 50. The INARCH(1) with an additive outlier of increasing size can be defined by:

$$Z_\tau = Y_\tau + X_t, \tag{4.37}$$

   where Z_τ is the new contaminated observed value at a specific time τ and X_t is the size of the outlier such that $X_t \sim \text{Poisson}(\nu_t)$ with $\nu \in \{1, 2, ..., 20\}$. Here $Z_t = Y_t$ if $t \neq \tau$ with $\tau = 50$ and Y_t follows model (4.2).

3. **Increasing number of additive outliers of increasing size**:
   We consider an increasing number j of additive outliers of increasing size $[j\sigma]$ (rounding multiples of the marginal standard deviation). Note that the INARCH(1) model including this type of outliers is as in case of an additive outlier, however here we use an increasing number j of additive outliers, $X_t = j\sigma$ is the size of the outlier with $j \in \{1, 2, ..., 20\}$ and σ is the marginal standard deviation.

4. **Increasing number of additive outliers of fixed size**:
   We consider an increasing number j of additive outliers of fixed size $[4\sigma]$ (four marginal standard deviations after rounding). Note that the INARCH(1) model including this type of outliers is as in case of an additive outlier, however here we use an increasing number j of additive outliers of fixed size $X_t = 4\sigma$.

Figure 4.13 compares the biases for β_0 (left) and α_1 (right) using the estimators in case of one transient outlier of increasing size $[j\sigma]$. For β_0, Tukeyuncorr with $k = 5$ gives the largest bias at outlier sizes ≤ 5, whereas condML gives the largest bias at sizes of the outlier larger than 5 followed by glmrob. Tukeycorr with $k = 7$ gives the smallest bias. For α_1, we observe that the differences between the biases are small, except for glmrob and condML, which give the largest bias. Among the estimators achieving high efficiency, Tukeycorr with $k = 7$ gives the smallest bias.

Figure 4.14 compares the biases for β_0 (left) and α_1 (right) caused by an additive outlier of increasing size $[j\sigma]$. condML and glmrob overestimate β_0 and underestimate α_1 because of an additive outlier. The corrected Tukey M-estimators give better robustness than the uncorrected ones. Among the estimators achieving high efficiency, Tukeycorr with $k = 7$ gives the smallest bias.

Figure 4.15 compares the biases for β_0 (left) and α_1 (right) in case of an increasing number of additive outliers with increasing size $[j\sigma]$. For β_0, the differences between the bias curves are small, except with tuning constant $k = 10$ in case of a number of outliers less than 10, and condML, glmrob. condML and glmrob overestimate β_0 and give the largest bias. All estimators underestimate α_1, where the differences between the bias curves become smaller than for β_0, except the condML, glmrob and the bias curves for Tukeycorr and Tukeyuncorr with $k = 10$ in case of a number of outliers smaller than 10. Among the estimators achieving high efficiency, Tukeycorr with $k = 7$ gives the smallest bias.

Figure 4.16 compares the biases for β_0 (left) and α_1 (right) in case of an increasing number of outliers of fixed size $[4\sigma]$. The smaller values of the tuning constants give smaller biases; condML corresponds to $k = \infty$. For β_0 glmrob gives similar bias as our implementation of the Tukey estimator with tuning constant $k = 10$ in case of a number of outliers less than 10. For a number of outliers lager than 10, glmrob gives smaller bias than our estimator with tuning constant $k = 10$. The differences between the bias curves for α_1 are smaller than for β_0. We note that all estimators underestimate α_1 and overestimate β_0 except the bias curves of Tukeycorr and Tukeyuncorr with $k = 5$. Tukeycorr with $k = 5$ gives the smallest bias for β_0 and α_1. Among the estimators achieving high efficiency, Tukeycorr with $k = 7$ gives the smallest bias.

The results for the RMSE are the same as for the biases, because they are dominated by the bias and thus not shown here.

4.6.3 General conclusions

Generally, the uncorrected and the corrected versions of Tukey M-estimators show better results than the different versions of Huber M-estimators under all scenarios of outliers considered here, see Appendix B. Among the estimators achieving high efficiency, the corrected versions of Tukey M-estimators with robust initialization from AR(1) show good performance, where in our simulations Tukeycorr with $k = 7$ gives good robustness. Also our simulation results indicate that the consistency and the asymptotic normality of these estimators.

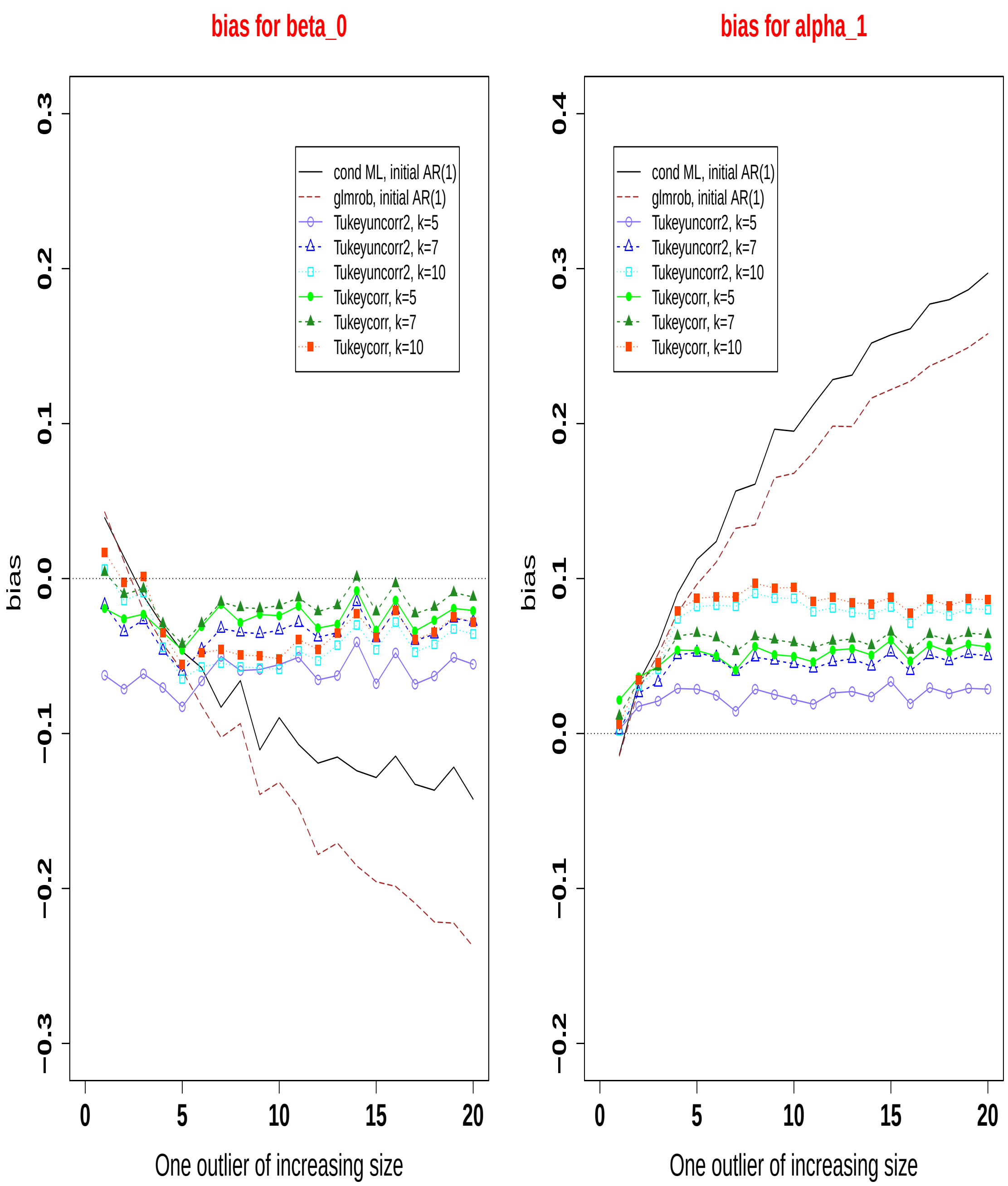


Figure 4.13: Simulated biases of the conditional maximum likelihood estimator, glmrob, corrected and uncorrected Tukey M-estimators with different tuning constants k for β_0 (left) and α_1 (right) in case of one transient outlier of increasing size with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

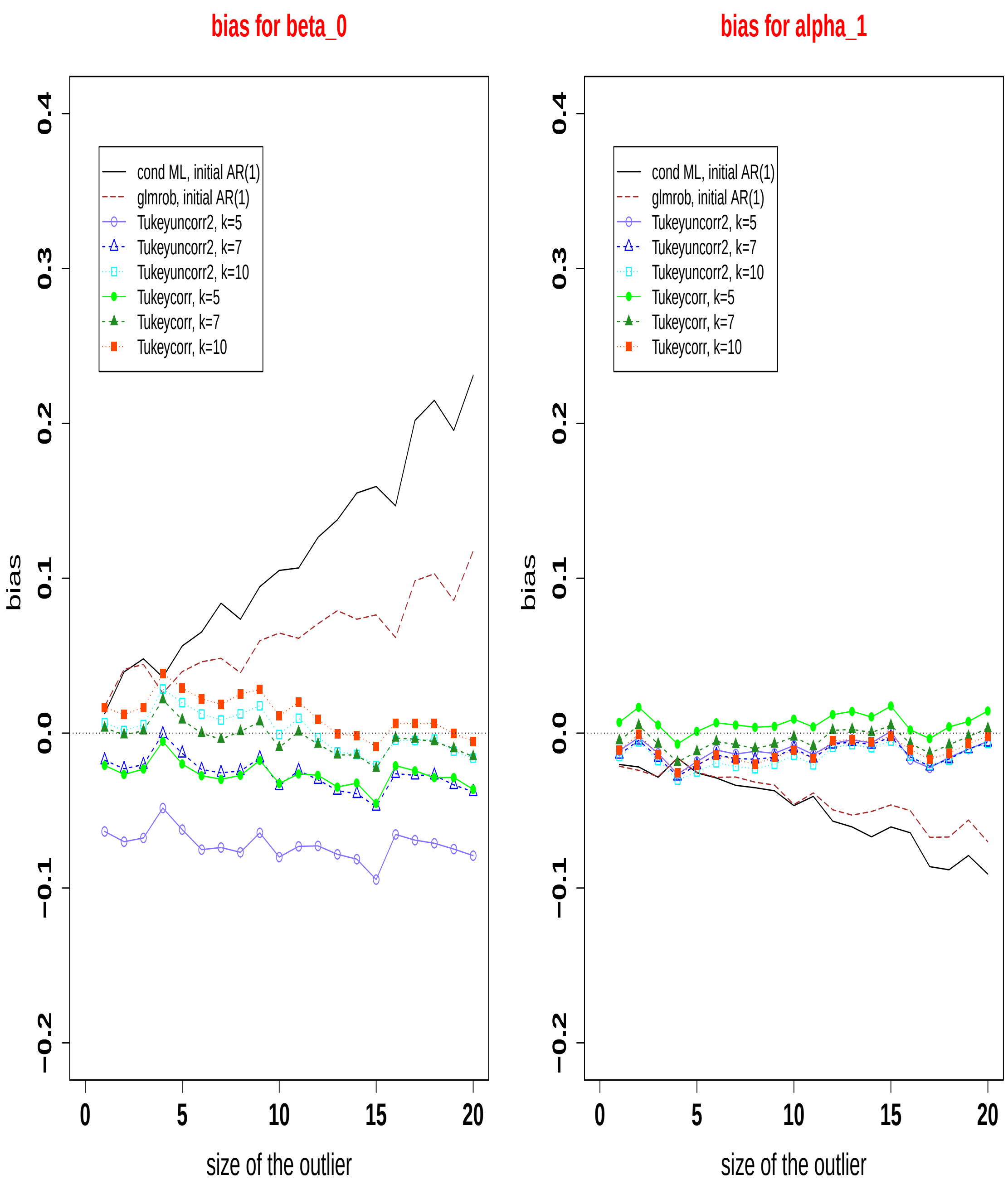


Figure 4.14: Simulated biases of the conditional maximum likelihood estimator, glmrob, corrected and uncorrected Tukey M-estimators with different tuning constants k for β_0 (left) and α_1 (right) in case of an additive outlier of increasing size with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

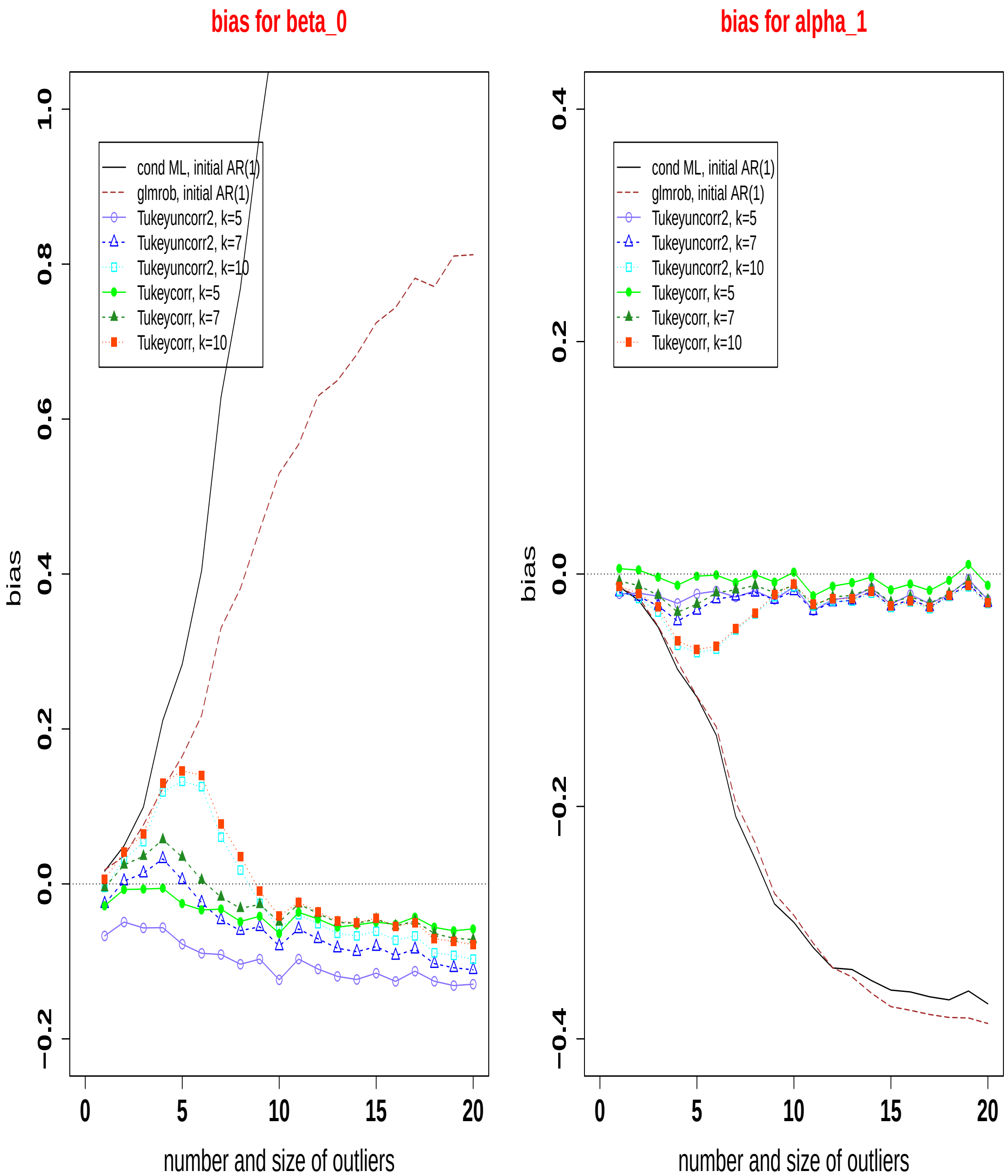


Figure 4.15: Simulated biases of the conditional maximum likelihood estimator, glmrob, corrected and uncorrected Tukey M-estimators with different tuning constants k for β_0 (left) and α_1 (right) in case of increasing numbers of additive outliers of increasing sizes with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

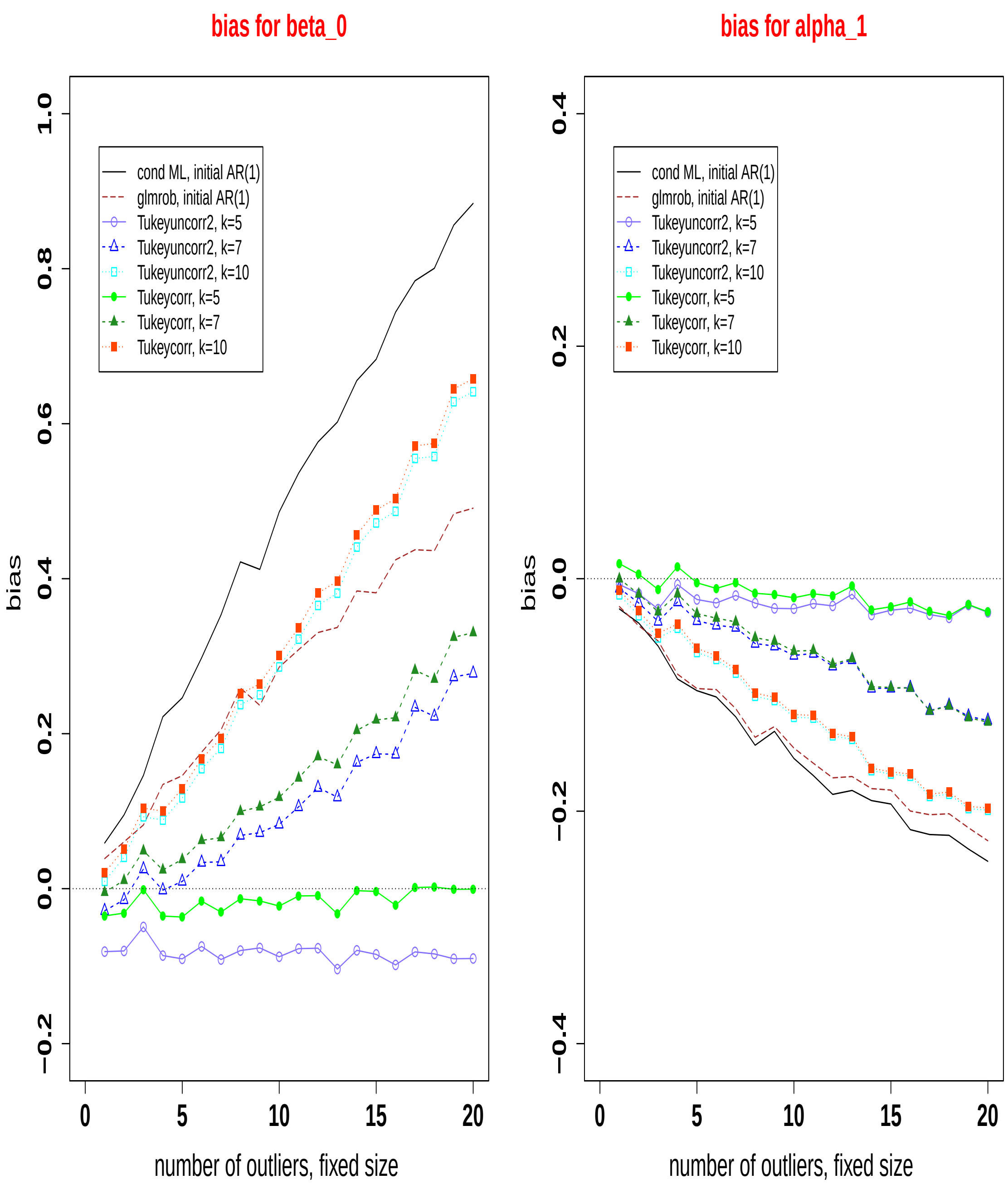


Figure 4.16: Simulated biases of the conditional maximum likelihood estimator, glmrob, corrected and uncorrected Tukey M-estimators with different tuning constants k for β_0 (left) and α_1 (right) in case of increasing numbers of additive outlier of fixed size $[4\sigma]$ with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

4.7 M-estimation for INGARCH model

In this section, we extend robust M-estimation with and without bias correction to general INGARCH(p,q) models, focusing on the case p=1, q=1.

1. **M-estimation for INGARCH(1,1) model without bias correction**
   Returning to (4.25) with $\frac{\partial \lambda_t}{\partial \theta_i}$ defined in (4.22), (4.23) and (4.24), we downweight the influence of unusual observations with large standardized residuals $(y_t - \lambda_t)/\sqrt{\lambda_t}$ using Huber's or Tukey's ψ function, and do the same with regressors y_{t-1} which are outlying w.r.t. the marginal distribution. This leads to the following set of estimating equations:

$$\sum_{t=2}^{n} \psi\left(\frac{y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \frac{1}{\sqrt{\lambda_t}} \begin{pmatrix} 1 + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \beta_0} \\ \sigma\psi\left(\frac{y_{t-1}-\lambda}{\sigma}\right) + \lambda + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \alpha_1} \\ \lambda_{t-1} + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \beta_1} \end{pmatrix} = 0, \tag{4.38}$$

   where $\lambda = \frac{\beta_0}{(1-\alpha_1-\beta_1)}$ is the marginal mean and $\sigma^2 = \frac{\lambda[1-(\alpha_1+\beta_1)^2+\alpha_1^2]}{1-(\alpha_1+\beta_1)^2}$ is the marginal variance of the stationary process, see Fokianos et al. (2009).

2. **M-estimation for INGARCH(1,1) model with bias correction**
   After adding a bias correction term, (4.38) becomes

$$\sum_{t=2}^{n} \left(\psi\left(\frac{y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \frac{1}{\sqrt{\lambda_t}} \begin{pmatrix} 1 + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \beta_0} \\ \sigma\psi\left(\frac{y_{t-1}-\lambda}{\sigma}\right) + \lambda + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \alpha_1} \\ \lambda_{t-1} + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \beta_1} \end{pmatrix} - \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \right) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{4.39}$$

   with bias correction $B = (a_0, a_1, a_2)'$, such that the expectation of the term on the left hand side of (4.39) equals $(0, 0, 0)'$. We approximate the bias correction, B, as follows:

$$\frac{1}{N-1} \sum_{t=2}^{N} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi\left(\frac{j - \lambda_t}{\sqrt{\lambda_t}}\right) \frac{1}{\sqrt{\lambda_t}} \begin{pmatrix} 1 + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \beta_0} \\ \sigma\psi\left(\frac{i-\lambda}{\sigma}\right) + \lambda + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \alpha_1} \\ \lambda_{t-1} + \beta_1 \frac{\partial \lambda_{t-1}}{\partial \beta_1} \end{pmatrix} P\left(Y_t = j, Y_{t-1} = i\right) \tag{4.40}$$

   using a large simulated time series of length N. The probability $P\left(Y_t = j, Y_{t-1} = i\right)$ can be calculated from $P\left(Y_t = j | Y_{t-1} = i\right)$, which is the conditional probability of $Y_t = j$ given $Y_{t-1} = i$ (derived from a Poisson with parameter $\lambda_t = \beta_0 + \alpha_1 i + \beta_1 \lambda_{t-1}$), and $P(Y_{t-1} = i)$, which is the marginal probability of $Y_{t-1} = i$ (estimated from the observed relative frequencies from a long trajectory of an INGARCH series with the corresponding parameters.

4.8 Computation

In this section we describe some new functions, which we implemented in R to compute

1. conditional maximum likelihood estimates for the parameters of the INGARCH(1,1) model using function **condlikest11**.
2. robust M-estimates for the parameters of the INGARCH(1,1) model without bias correction and with robust initialization under the assumption of independence using function **robPoisTStukey11**.
3. robust M-estimates for the parameters of the INGARCH(1,1) model without bias correction and with robust initialization from the ARMA(1,1) model using function **robPoisTStukey22**.
4. robust M-estimates for the parameters of the INGARCH(1,1) model with bias correction and with robust initialization from the ARMA(1,1) model using function **robPoisTStukey33**.

1. **condlikest11**
   We can describe function condlikest11 in the same way as we describe the function condlikest in case of the INARCH(1) model by the following steps:
   (a) We define the function "liklinear.poisson.un" as a function in theta and data, where theta is our parameter vector (β_0, α_1, β_1). This function gives the summation of log-likelihood function (4.25) with $\frac{\partial \lambda_t}{\partial \theta_i}$ defined in (4.22), (4.23) and (4.24).
   (b) We define the function "condlikest11" as a function in y and theta. This function gives classical estimates for the parameters of the INGARCH(1,1) model, β_0, α_1 and β_1. These estimates are obtained using the function constrOptim.
   (c) constrOptim is a function in theta, f, grad, ui, ci, mu = 1e-04, control = list (), method = if (is.null(grad)) "Nelder-Mead" else "BFGS", outer.iterations = 100, outer.eps = 1e-05, ...
   Here, theta is our parameter vector (β_0, α_1, β_1), where we initialize theta with β_0 = mean of our data (y), $\alpha_1 = 0.001$ and $\beta_1 = 0.001$,
   f is our score function "liklinear.poisson.un", defined above in (a),
   grad is the function "scorelinear.poisson.un", which we implemented to compute the derivatives of the score function with respect to its parameters,
   ui and ci are a matrix and a vector for our constraints $\beta_0 > 0$, $\alpha_1, \beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$, method is "BFGS", see the function optim in R for more details, outer.iterations is set to 100 and outer.eps to 1e-05.

2. **robPoisTStukey11**
   This function is based on Tukey's ψ function and we can describe it like the function robPoisTStukey1 in case of the INARCH(1) model as follows:

   (a) We define the function "robPoistukeyscores11" as a function in param, y, k, which calculates the left hand side of (4.38) using Tukey's ψ function.

   (b) We define the function "robPoisTStukey11" as a function in y, k, iter=200. This function gives robust estimates for the parameters of the INGARCH(1,1), β_0, α_1 and β_1. These estimates are obtained using the function constrOptim.

   (c) constrOptim is a function in theta, f, grad, ui, ci, mu = 1e-04, control = list (), method = if (is.null(grad)) "Nelder-Mead" else "BFGS", outer.iterations = 100, outer.eps = 1e-05, ...
   Here, theta is our parameter vector (β_0, α_1 and β_1). We initialize theta under the assumption of independence with a robust estimate of β_0 using the function tukeypois, α_1 =0.001 and β_1 = 0.001,
   f is our score function "robPoistukeyscores11", defined above in (a),
   grad is set to null,
   ui and ci are a matrix and a vector describing our constraints $\beta_0 > 0$, $\alpha_1, \beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$,
   method is "Nelder-Mead", see the function optim in R for more details.
   outer.iterations is set to 200 and outer.eps to 1e-06.

3. **robPoisTStukey22**
   This function is a modified version of the function robPoisTStukey11, where we use a robust initialization for β_0, α_1 and β_1 to start the iteration in constrOptim using robust estimates in the ARMA(1,1) model. According to the properties of INGARCH(1,1), we conclude from (4.12) and (4.13) that $\beta_1 = -\theta_1$ and $\alpha_1 = \phi_1 - \beta_1$. To estimate the parameters ϕ_1 and θ_1 of the ARMA(1,1) model robustly, we estimate the autocorrelations at lags 1 and 2 robustly using the method of Ma and Genton (2000) as in case of the INARCH(1) model. Then we derive Yule-Walker type estimates of ϕ_1 and θ_1 by solving (4.14) for these parameters and plugging in the robust autocorrelation estimates. Afterwards we estimate the marginal mean using the function tukeypois and use the formula of the marginal mean $\beta_0/(1 - \alpha_1 - \beta_1)$ to obtain an initial estimate of β_0.

4. **robPoisTStukey33**.
   This function is a modified version of the function robPoisTStukey22, where we use the same robust initialization for β_0, α_1 and β_1. However, here we correct our score function robPoistukeyscores33 by subtracting suitable bias correction terms. These correction terms have been computed based on (4.40). The estimates are obtained using the function constrOptim. constrOptim, defined before in the function robPoisTStukey11, iterates between estimating β_0, α_1 and β_1 for a given value of bias corrections for β_0, α_1 and β_1 and the specification of bias corrections based

on the current estimate of β_0, α_1 and β_1. First we choose a value of the correction terms. Second we estimate β_0, α_1 and β_1, then we derive the values of our correction terms which are suitable for these estimates. Then we re-estimate β_0, α_1 and β_1. We iterate this process until the absolute differences of our estimates for two consecutive iterations do not exceed our desired accuracy 0.0001, which we use to determine convergence.

4.9 Simulations

In this section, we perform some simulation experiments to compare the performance of conditional maximum likelihood estimators, computed using the function condlikest11, and the generalized M-estimators, computed using the different versions of the function robPoisTStukey (robPoisTStukey11, robPoisTStukey22, robPoisTStukey33). Namely, we consider the following M-estimators:

- without bias correction and with robust initialization under the assumption of independence, see the results of robPoisTStukey11.
- with and without bias correction and with robust initialization using robust estimates in an ARMA(1,1) model, see the results of robPoisTStukey22 and robPoisTStukey33.

These estimates are compared in terms of bias and root of mean square error (RMSE), in finite samples from an INGARCH(1,1) model, with several tuning constants and with and without outliers.

4.9.1 Results in case of clean data

We compare the results of the conditional maximum likelihood estimator and the generalized M-estimator for the parameters of the INGARCH(1,1) model in situations without outliers. Namely, we compare

- condlikest11
- robPoisTStukey11 with different tuning constants $k \in \{7, 10, 12\}$.
- robPoisTStukey22 with different tuning constants $k \in \{7, 10, 12\}$.
- robPoisTStukey33 with different tuning constants $k \in \{7, 10, 12\}$.

We generate data from an INGARCH(1,1) model with parameters $\beta_0 = 1$, $\alpha_1 = 0.3$, $\beta_1 = 0.4$. The results are based on 500 data sets of size 200 each. We have the following findings:

- For β_0 and β_1, using robust initialization from ARMA(1,1) model gives better results than using robust initialization under the assumption of independence.

- The corrected Tukey M-estimators (robPoisTStukey33) give smaller biases and RMSEs than the uncorrected ones (robPoisTStukey11, robPoisTStukey22) for β_0 and β_1, whereas the uncorrected ones with initialization under the assumption of independence give smaller biases and RMSEs for α_1.

- The conditional maximum likelihood estimator gives the smallest biases and RMSEs for α_1 and β_1, and it also gives the smallest RMSE for β_0. We note that the conditional maximum likelihood estimator gives similar results as for the uncorrected Tukey M-estimators with initialization under the assumption of independence (robPoisTStukey11) for α_1, and it gives similar bias as for the corrected Tukey M-estimator with tuning constants $k = 12$ for β_0.

- Increasing the tuning constant k reduces the biases and RMSEs for α_1 and β_1, but not necessarily for β_0.

Figure 4.17 depicts boxplots of the conditional maximum likelihood estimates and the Tukey M-estimates with tuning constant $k = 7$ for β_0, α_1 and β_1 estimated from INGARCH(1,1) with true values $\beta_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.4$, 500 data sets of size 200. The boxplots indicate that the distribution for β_0 of robPoisTStukey11 (Tukey11) is symmetric but with a large number of outliers, which can explain its strange bias results.

Table 4.1: Results of the Tukey M-estimators and the conditional maximum likelihood estimator in case of clean data with $\beta_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.4$.

Estimators	Values	β_0	α_1	β_1
condlikest11	bias	0.1610	-0.0023	-0.0455
	RMSE	0.5028	0.0700	0.1728
robPoisTStukey11, $k = 7$	bias	0.4107	-0.0059	-0.1261
	RMSE	0.7464	0.0723	0.2163
robPoisTStukey11, $k = 10$	bias	0.4276	-0.0053	-0.1272
	RMSE	0.7510	0.0718	0.2171
robPoisTStukey11, $k = 12$	bias	0.4295	-0.0052	-0.1263
	RMSE	0.7536	0.0717	0.2166
robPoisTStukey22, $k = 7$	bias	0.0998	0.1328	-0.1675
	RMSE	0.7534	0.3116	0.2600
robPoisTStukey22, $k = 10$	bias	0.2169	0.0606	-0.1286
	RMSE	0.7068	0.2137	0.2310
robPoisTStukey22, $k = 12$	bias	0.2440	0.0443	-0.1203
	RMSE	0.6901	0.1849	0.2230
robPoisTStukey33, $k = 7$	bias	-0.0174	0.1289	-0.1173
	RMSE	0.6916	0.3124	0.2490
robPoisTStukey33, $k = 10$	bias	0.1030	0.0827	-0.1112
	RMSE	0.6672	0.2538	0.2290
robPoisTStukey33, $k = 12$	bias	0.1552	0.0559	-0.1012
	RMSE	0.6568	0.2090	0.2210

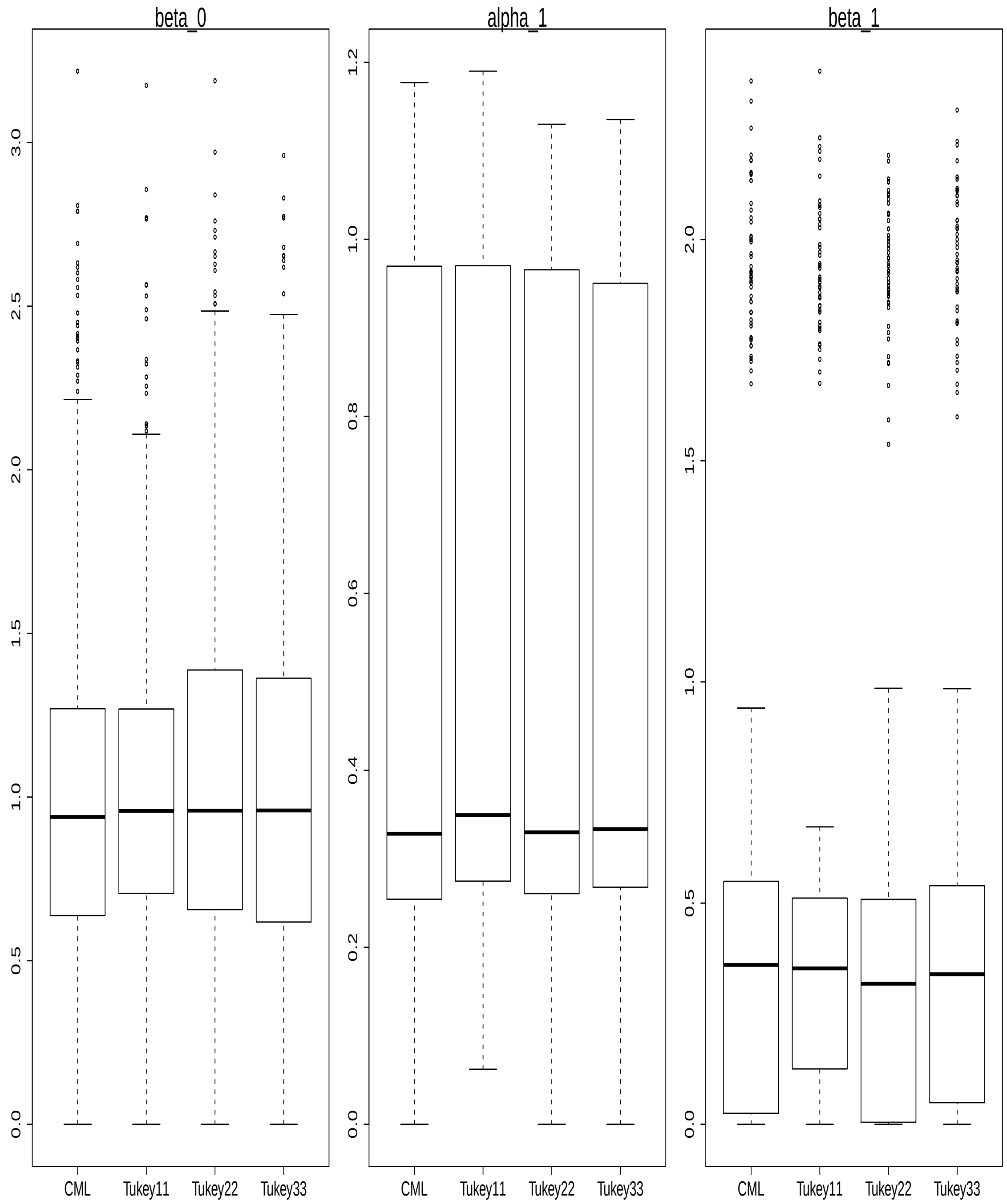


Figure 4.17: Boxplots of the conditional maximum likelihood estimator and Tukey M-estimators with tuning constant $k = 7$ for β_0, α_1 and β_1 (from left to right) estimated from INGARCH(1,1) with true values $\beta_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.4$, 500 data sets of size 200.

4.9.2 Results in case of contaminated data

We compare the results of the predefined estimators in the presence of outliers. We consider the following three outlier scenarios:

1. Three additive outliers of fixed size 20 at times 50, 100, 150.
2. Six additive outliers of fixed size 20 at times 40, 80, 100, 120, 160, 200.
3. Ten additive outliers of fixed size 20 at times 10, 20, ..., 100.

We have the following findings:

- The conditional maximum likelihood estimator shows strange positive bias for β_0 and substantial biases for for α_1 and β_1 in the presence of isolated additive outliers considered here.
- The Tukey M-estimators reduce these biases substantially in case of β_0 and α_1 and have thus smaller RMSEs for these parameters. However these are even more biased for β_1.
- Increasing the tuning constant k reduces biases and RMSEs for α_1 and β_1 slightly, as in case of clean data.
- In case of using less than 5% (3 or 6) additive outliers, the corrected Tukey M-estimators give the smallest biases and RMSEs for β_0. The uncorrected Tukey M-estimators with initialization under the assumption of independence give the smallest biases and RMSEs for α_1, whereas the conditional maximum likelihood estimator gives the smallest bias for β_1. We note that the sum of the dependence parameters α_1 and β_1 is estimated with less bias when using the bias correction, but it seems difficult to distinguish these two influences.
- In case of 5% (10) additive outliers, the corrected Tukey M-estimator with tuning constant $k = 12$ gives the smallest biases and RMSEs for α_1 and β_1, and with tuning constant $k = 7$ gives the smallest bias for β_0.

4.9.3 General conclusions

The Tukey M-estimators provide robustness for the estimation of β_0 and α_1, but not for β_1, in the presence of additive outliers. The bias-corrected versions with initialization from an ARMA(1,1) fit and a large tuning constant additionally perform almost as good as the conditional maximum likelihood in case of clean data.

Table 4.2: Results of the Tukey M-estimators and the conditional maximum likelihood estimator in case of 3 additive outliers with $\beta_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.4$.

Estimators	Values	β_0	α_1	β_1
condlikest11	bias	1.1138	-0.1735	-0.0941
	RMSE	1.5074	0.1947	0.2770
robPoisTStukey11, $k = 7$	bias	0.7017	0.0351	-0.2705
	RMSE	0.8059	0.0898	0.2949
robPoisTStukey11, $k = 10$	bias	0.7196	0.0347	-0.2702
	RMSE	0.8233	0.0885	0.2940
robPoisTStukey11, $k = 12$	bias	0.7304	0.0289	-0.2645
	RMSE	0.8399	0.0854	0.2902
robPoisTStukey22, $k = 7$	bias	0.4471	0.1228	-0.2812
	RMSE	0.8368	0.2571	0.3029
robPoisTStukey22, $k = 10$	bias	0.5847	0.0820	-0.2782
	RMSE	0.8369	0.1992	0.3013
robPoisTStukey22, $k = 12$	bias	0.6436	0.0542	-0.2653
	RMSE	0.8435	0.1558	0.2951
robPoisTStukey33, $k = 7$	bias	0.4090	0.1257	-0.2616
	RMSE	0.7976	0.2488	0.2911
robPoisTStukey33, $k = 10$	bias	0.5852	0.0727	-0.2618
	RMSE	0.7916	0.1652	0.2885
robPoisTStukey33, $k = 12$	bias	0.6474	0.0518	-0.2596
	RMSE	0.8044	0.1305	0.2876

Table 4.3: Results of the Tukey M-estimators and the conditional maximum likelihood estimator in case of 6 additive outliers with $\beta_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.4$.

Estimators	Values	β_0	α_1	β_1
condlikest11	bias	1.5163	-0.1949	-0.1441
	RMSE	1.8486	0.2091	0.2921
robPoisTStukey11, $k = 7$	bias	0.6388	0.0280	-0.2513
	RMSE	0.7481	0.0888	0.2784
robPoisTStukey11, $k = 10$	bias	0.6454	0.0271	-0.2471
	RMSE	0.7538	0.0867	0.2752
robPoisTStukey11, $k = 12$	bias	0.6763	0.0153	-0.2381
	RMSE	0.7943	0.0806	0.2682
robPoisTStukey22, $k = 7$	bias	0.4451	0.0886	-0.2520
	RMSE	0.7825	0.2096	0.2855
robPoisTStukey22, $k = 10$	bias	0.5671	0.0556	-0.2508
	RMSE	0.7813	0.1530	0.2877
robPoisTStukey22, $k = 12$	bias	0.6317	0.0295	-0.2378
	RMSE	0.8200	0.1172	0.2827
robPoisTStukey33, $k = 7$	bias	0.3911	0.1149	-0.2509
	RMSE	0.7965	0.2378	0.2920
robPoisTStukey33, $k = 10$	bias	0.5088	0.0825	-0.2537
	RMSE	0.7852	0.1943	0.2904
robPoisTStukey33, $k = 12$	bias	0.6056	0.0490	0.1515
	RMSE	0.8175	0.1515	0.2852

Table 4.4: Results of the Tukey M-estimators and the conditional maximum likelihood estimator in case of 10 additive outliers with $\beta_0 = 1$, $\alpha_1 = 0.3$ and $\beta_1 = 0.4$.

Estimators	Values	β_0	α_1	β_1
condlikest11	bias	2.7535	-0.2671	-0.2994
	RMSE	2.9580	0.2703	0.3801
robPoisTStukey11, $k = 7$	bias	0.7247	0.0464	-0.3127
	RMSE	0.8051	0.0943	0.3255
robPoisTStukey11, $k = 10$	bias	0.7405	0.0464	-0.3116
	RMSE	0.8206	0.0925	0.3242
robPoisTStukey11, $k = 12$	bias	0.8213	0.0252	-0.3041
	RMSE	0.9137	0.0806	0.3186
robPoisTStukey22, $k = 7$	bias	0.6029	0.0784	-0.3079
	RMSE	0.8154	0.1798	0.3212
robPoisTStukey22, $k = 10$	bias	0.6590	0.0646	-0.3059
	RMSE	0.8261	0.1464	0.3206
robPoisTStukey22, $k = 12$	bias	0.7924	0.0293	-0.3013
	RMSE	0.9061	0.1048	0.3167
robPoisTStukey33, $k = 7$	bias	0.5768	0.0940	-0.3041
	RMSE	0.8221	0.1934	0.3190
robPoisTStukey33, $k = 10$	bias	0.6581	0.0665	-0.2998
	RMSE	0.8120	0.1360	0.3154
robPoisTStukey33, $k = 12$	bias	0.7896	0.0291	-0.2959
	RMSE	0.8962	0.0890	0.3117

Chapter 5

Real data applications in the medical field

This chapter applies our proposed methods from Chapter 4 to two real data examples in the medical field. The first example is the poliomyelitis (briefly: polio) data. The second is the campylobacterosis (briefly: campy) data. We start by an analysis of the polio data. We give a description of the polio data, then we fit an INGARCH(1,0) model to these data using conditional maximum likelihood as a non robust method and Tukey M-estimation as a robust method to estimate the model parameters. After that, we fit an INGARCH(1,0) model using the same methods but after having cleaned the data from outliers using the approaches given in Fokianos and Fried (2010), which concern the detection of sudden shifts and outliers. Second, to verify the reliability of our proposed methods, we analyse an artificial data example generated to resemble the polio data. We fit an INGARCH(1,0) model to artificial data using the same methods as for the polio data, then we fit an INGARCH(1,0) model again after having cleaned the artificial data from outliers. For the campy data, we repeat again what we did for the polio data.

5.1 Analysis of the poliomyelitis data

5.1.1 Description of the poliomyelitis data

Figure 5.1 shows the polio data. These data have been published by the U.S. Centers for Disease and Control and list the monthly number of poliomyelitis cases in the United States for the period 1970 to 1983. This data set has become a standard example in the field of count time series data. The data consist of 168 observations with some strongly deviating data points and periods of higher level and variability. This data set has been studied by Zeger (1988) using the Poisson log-linear regression model. Davis et al. (2000) developed a practical approach to diagnosing the existence of a latent process in Poisson log-linear regression models and derived the asymptotic properties of the generalized linear model estimator when an autocorrelated latent process is present. This data set has been studied further by Fahrmeir and Tutz (2001, Chap. 6.1) in the framework of gener-

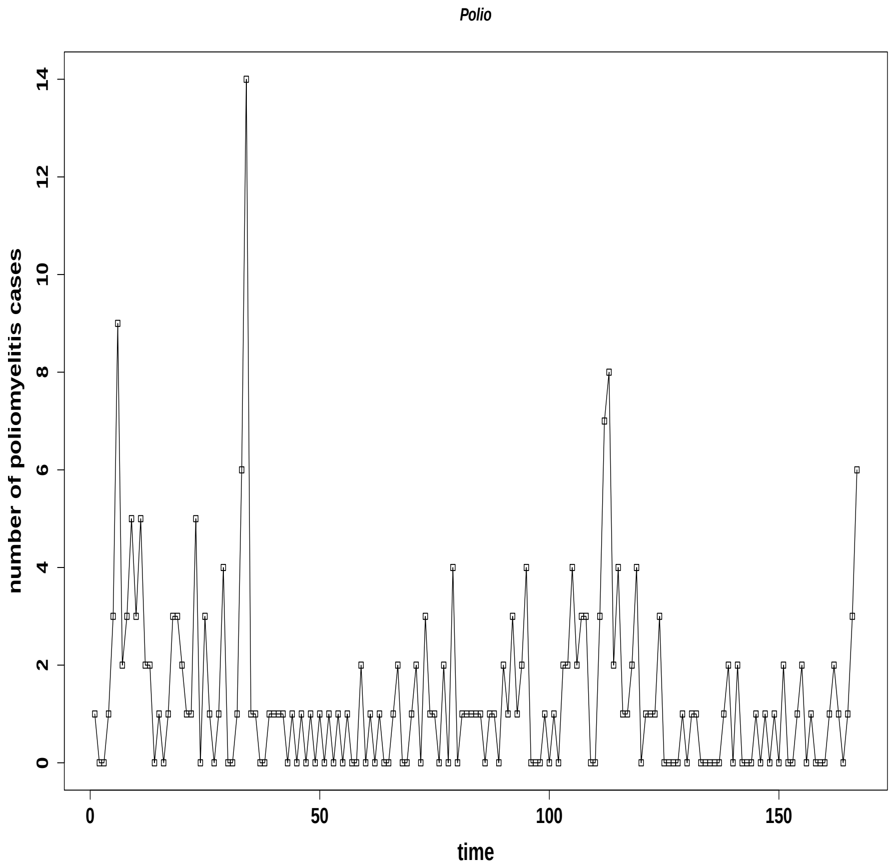


Figure 5.1: Monthly number of poliomyelitis cases in the United States for the period 1970 to 1983.

alized linear models for time series, where the authors use both conditional and marginal models to analyse these data. When using marginal models the long-term decrease in the rate of polio infections becomes more obvious. In contrast, this effect is attenuated with conditional models. Jowaheer and Sutradhar (2005) used AR(1) and MA(1) models to fit these data and suggested to search a suitable robust estimation technique to fit time series of counts in the possible presence of outliers. More recently Zheng et al. (2006) proposed the p-th order random coefficient integer valued process [RCINAR(p)] for these data. Mora et al. (2009) used integer valued AR (INAR) processes as an alternative to the Poisson regression model given by Zeger (1988) and Davis et al. (2000). Davis and Wu (2009) extended the asymptotic results of Davis et al. (2000), modelling the polio data by a negative binomial regression model. Kang and Lee (2009) proposed a cumulative sum test for identifying change points in a first order random coefficients integer autoregressive process [RCINAR(1)] to analyse the polio data. Fokianos and Fried (2011) investigated whether there are intervention effects corresponding to unusual events in this time series, which was modelled as a log-linear Poisson autoregression.
In this chapter, we consider that the polio data follow an INGARCH(1,0) model. Then we fit this model non robustly and robustly.

Table 5.1: Parameter estimates for the polio data (left) and for the cleaned polio data (right)

Estimators \ estimates	unclean		clean	
	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\alpha}_1$
condlikest	0.8522	0.3739	0.9036	0.0571
robPoisTStukey3	0.8433	0.2415	0.8771	0.0333

5.1.2 INGARCH(1,0) fit to the polio data

Here we use conditional maximum likelihood as a non robust estimation method, which is computed by the function condlikest, and we use Tukey M-estimators with bias correction and with robust initialization from an AR(1) model with $k = 7$ as a robust estimator, which is computed by the function robPoisTStukey3.
The left hand side of Table 5.1 gives the parameter estimates using the functions condlikest and robPoisTStukey3 for the polio data. For this data set, the difference between condlikest and robPoisTStukey3 is negligible for β_0, but we have a larger difference for α_1. So in this situation α_1 seems to absorb the outliers effects.

5.1.3 INGARCH(1,0) fit to the cleaned polio data

We clean the polio data from outliers using the approach given in the paper of Fokianos and Fried (2010). We use the third scenario in their paper, which concerns the detection of multiple interventions when the type and the time are both unknown. Following their lines, the INGARCH(1,0) with intervention effects can be defined by:

$$Z_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\kappa_t),$$
$$\kappa_t = \beta_0 + \alpha_1 Z_{t-1} + \nu X_t, \tag{5.1}$$

for $t \geq 1$, where $\mathcal{F}_t$ is the σ-field generated by $\{Z_0, ..., Z_t\}$ representing the whole information up to time t, κ_t is the new conditional mean process after adding the intervention effects, Z_t is the contaminated precess, ν is the size of the intervention effect, and $\{X_t\}$ is a sequence of deterministic covariates, which determines the intervention effect included in the observation equation. It has the following form

$$X_t = \xi(B)I_t(\tau), \tag{5.2}$$

where $I_t(\tau)$ is an indicator function with $I_t(\tau) = 1$ if t equals a specific time τ and $I_t(\tau) = 0$ if $t \neq \tau$, B is the shift operator such that $B^i X_t = X_{t-i}$, and $\xi(B)$ is a polynomial operator of the form

$$\xi(B) = (1 - \delta(B))^{-1}. \tag{5.3}$$

The choice of $\xi(B)$ determines the kind of intervention effect: $\delta = 0$ for a spiky outlier (SO), $\delta = 1$ for a level shift (LS) and $\delta \in \{0.7, 0.8, 0.9\}$ is a predefined constant for a

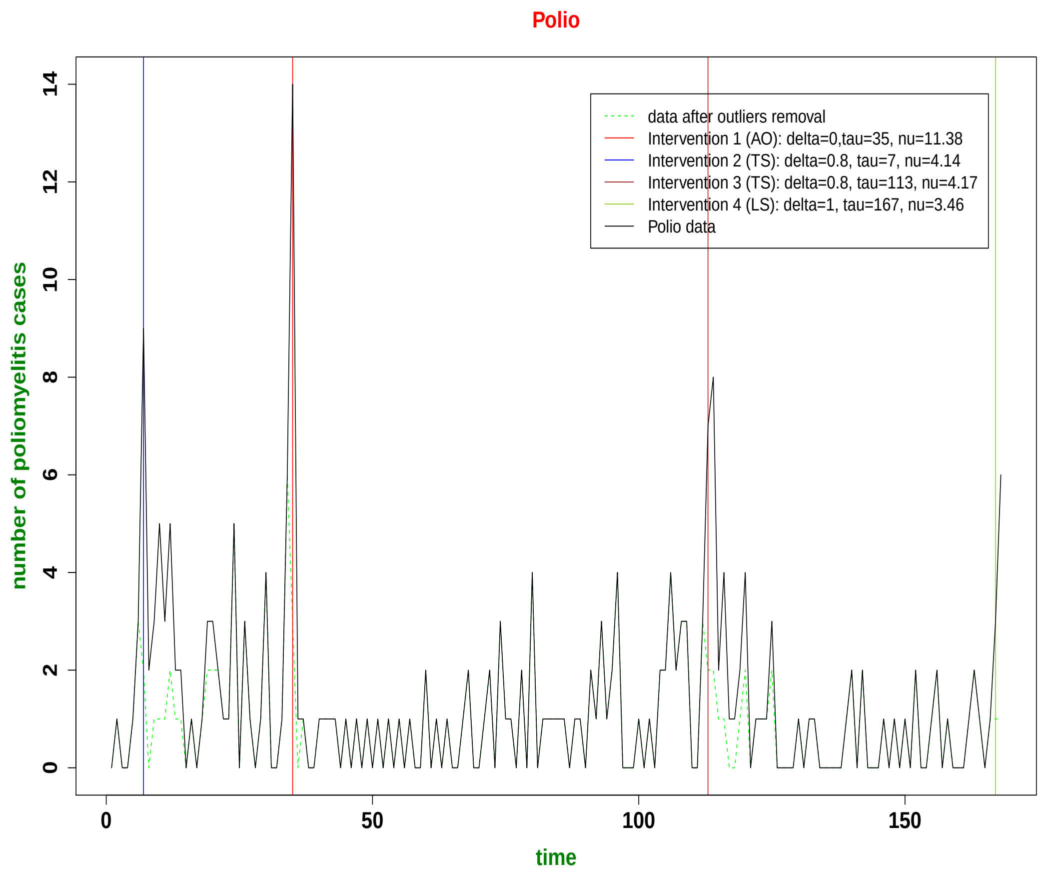


Figure 5.2: Polio data (solid black line) and data after outliers removal (dashed green line). Step 1: AO at time 35 (vertical red line), step 2: TS at time 7 (vertical blue line), step 3: TS at time 113 (vertical brown line), step 4: LS at time 167 (vertical yellowgreen line).

transient shift (TS). Note that for $\nu = 0$ the processes $\{\lambda_t\}$ and $\{\kappa_t\}$ are identical as well as $\{Y_t\}$ and $\{Z_t\}$. For more details, see Fokianos and Fried (2010, Section 6).

Figure 5.2 illustrates the corrections performed in the iterative procedure for the polio data. We detect (in this order) a spiky outlier (SO) at time $\tau = 35$ of size $\nu = 11.38$ (recall that δ is set to 0), transient shifts (TSs) at times 7 and 113 of sizes 4.14 and 4.17, respectively, (recall that δ is set to 0.8), and finally a level shift (LS) at time 167 of size 3.46 (recall that δ is set to 1).
The right hand side of Table 5.1 gives the parameter estimates using the functions condlikest and robPoisTStukey3 for the cleaned polio data (after removing intervention effects). We find that the differences between condlikest and robPoisTStukey3 are small for both β_0 and α_1 after cleaning this data set from outliers. Note that the robust estimate for α_1 in case of the original (contaminated) data is closer to the estimate for α_1 for the cleaned data than the non-robust estimate for α_1.

Table 5.2: Parameter estimates for the artificial polio data (left) and for the cleaned artificial polio data (right)

Estimators \ estimates	unclean		clean	
	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\alpha}_1$
condlikest	0.9906	0.2480	0.9976	0.0855
robPoisTStukey3	0.9594	0.0778	1.017	0.0824

5.2 Analysis of artificial poliomyelitis data

To know whether the INGARCH(1,0) model is a good model for the polio data, we generate a data set from the final INGARCH(1,0) model with outliers fitted to the polio data. Then we compare the parameter estimates using the functions condlikest and robPoisTStukey3. Second, we clean the artificial polio data from the outliers using the same approach as for the polio data, and compare our parameters estimates again.

5.2.1 INGARCH(1,0) fit to the artificial polio data

We generate a data set with sample size 168 from the final INGARCH(1,0) model with parameters $\beta_0 = 0.9$ and $\alpha_1 = 0.05$ and with a SO at time 35, TSs at times 7 and 113. We ignore the LS, which is located at the end of the time series. The left hand side of Table 5.2 gives the parameter estimates using the functions condlikest and robPoisTStukey3 for the artificial polio data generated from the final INGARCH(1,0) model with outliers. We find that the difference between condlikest and robPoisTStukey3 is small for β_0, but we have a larger difference for α_1 with the robust estimate recovering the true value well.

5.2.2 INGARCH(1,0) fit to the cleaned artificial polio data

We clean the artificial polio data from outliers using the same approach as for the real polio data. From Figure 5.3, we detect an additive outlier (AO) at time 35 and transient shifts (TSs) at times 7 and 113. This means that all outliers are identified correctly and at the correct time points. The right hand side of Table 5.2 illustrates the results for the cleaned artificial polio data. The differences between condlikest and robPoisTStukey3 are negligible for β_0 and α_1 and close to the true parameter values used for generating the data. Note that the robust estimates for the contaminated data are close to the estimates after cleaning.

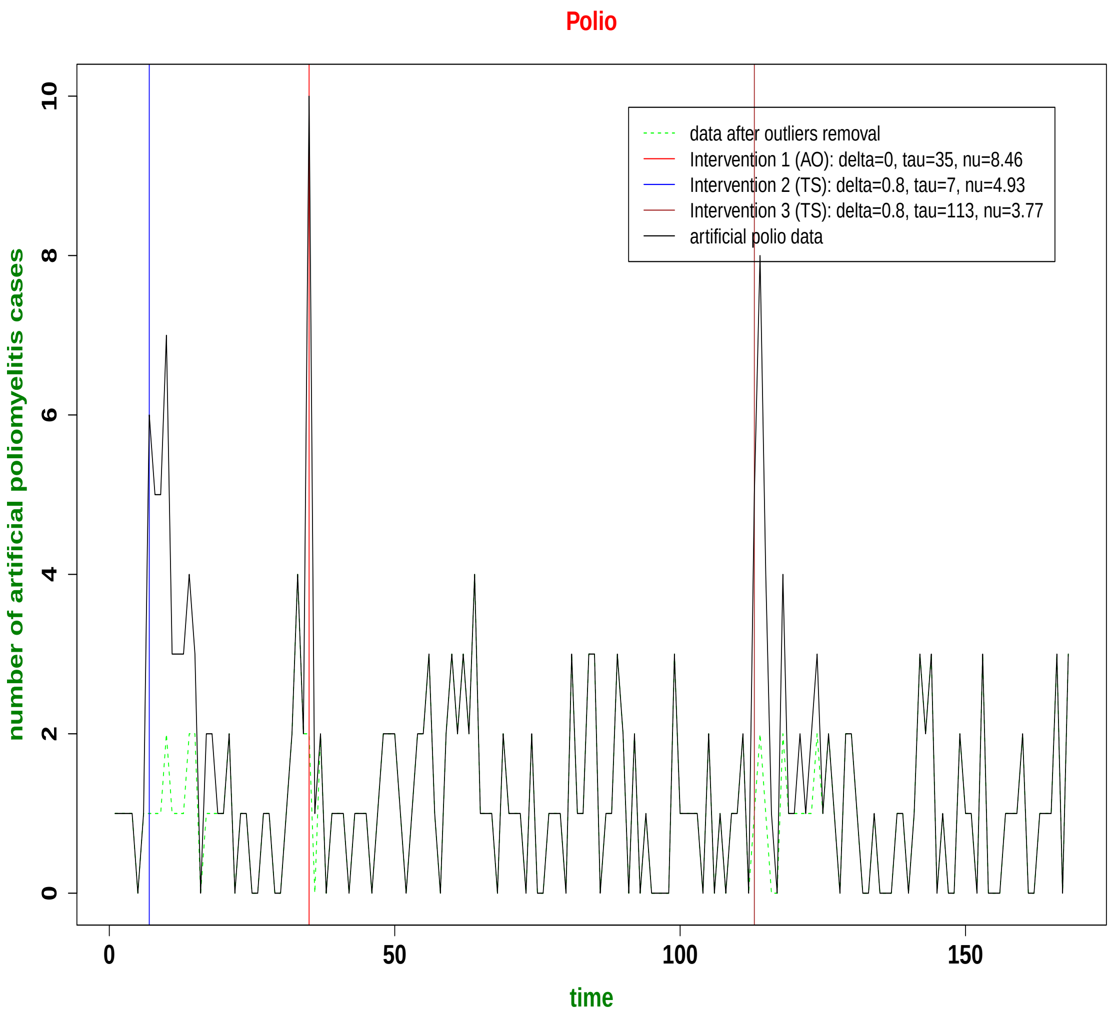


Figure 5.3: Artificial Polio data (solid black line) and data after removal of outliers (dashed green line). Step 1: AO at time 35 (vertical red line), step 2: TS at time 7 (vertical blue line), step 3: TS at time 113 (vertical brown line).

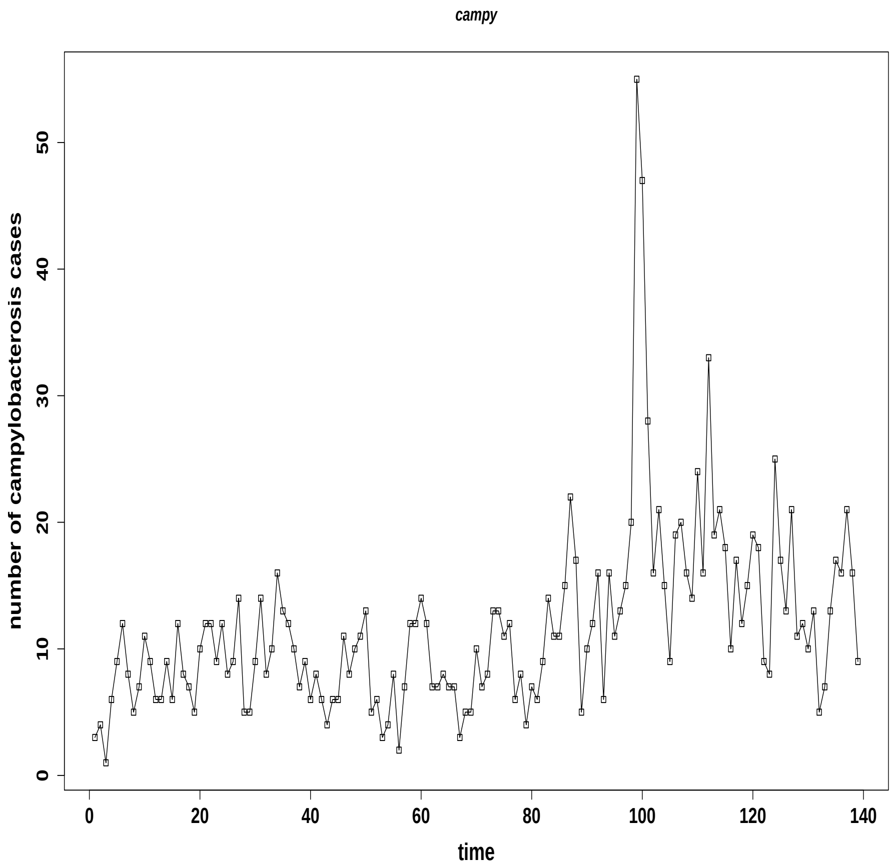


Figure 5.4: Monthly number of cases of campylobacterosis infections from January 1990 to the end of October 2000 in the north of the Province of Quebec, Canada.

5.3 Analysis of the campylobacterosis data

5.3.1 Description of the campylobacterosis data

Figure 5.4 shows the campy data. These data refer to the monthly number of cases of campylobacterosis infections from January 1990 to the end of October 2000 in the north of the Province of Quebec, Canada, see Ferland et al. (2006). The data were recorded every 28 days for a total number of 13 times per year. Ferland et al. (2006) proposed an INGARCH(1,1) model to fit these data and they used maximum likelihood to estimate the model parameters. Fokianos and Fried (2010) studied the problem of detection and estimation of sudden shifts and outliers in linear Poisson GARCH type models (INGARCH) by application of their approach to the campy data. They find a level shift around time point 84 and a spiky outlier at time 100. Fokianos and Fried (2011) reexamined these data but by employing a log-linear model instead of a linear model for count time series. Intervention effects in linear INGARCH are additive, but for the log-linear model they are multiplicative. They noted that in most cases the intervention effects detected in a time series when using the linear or the log-linear approach closely agreed.

Table 5.3: Parameter estimates for the campy data (left) and for the cleaned campy data (right)

Estimators \ estimates	unclean		clean	
	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\alpha}_1$
condlikest	4.144	0.6474	4.533	0.4139
robPoisTStukey3	4.495	0.5720	4.524	0.4167

In our chapter as for the polio data, we consider that the campy data follow INGARCH(1,0) model and fit this model non robustly and robustly to these data sets.

5.3.2 INGARCH(1,0) fit to the campy data

Here we use conditional maximum likelihood as a non robust estimation method, which is computed by the function condlikest, and we use Tukey M-estimators with bias correction and robust initialization from an AR(1) model with $k = 7$ for robust estimation, which is computed by the function robPoisTStukey3.
The left hand side of Table 5.3 gives the parameter estimates using the functions condlikest and robPoisTStukey3 for the campy data. For this data set, we have slightly larger differences between condlikest and robPoisTStukey3 for β_0 and α_1.

5.3.3 INGARCH(1,0) fit to the cleaned campy data

We clean the campy data using the same approach as for the polio data. For more details, see Fokianos and Fried (2010, Section 6).
Figure 5.5 illustrates the corrections performed in the iterative procedure for the campy data. We detect a level shift (LS) of size 4.51 at time 84 (recall that δ is set to 1), and a transient shift (TS) of size 8.32 at time 100 (recall that δ is set to 0.8).
The right hand side of Table 5.3 gives the parameter estimates using the functions condlikest and robPoisTStukey3 for the cleaned campy data (after removing intervention effects). There are negligible differences between condlikest and robPoisTStukey3 for both β_0 and α_1 after we clean this data set from outliers. Also note that the robust estimates for the contaminated data are closer to the estimates for the cleaned data than the non robust estimates.

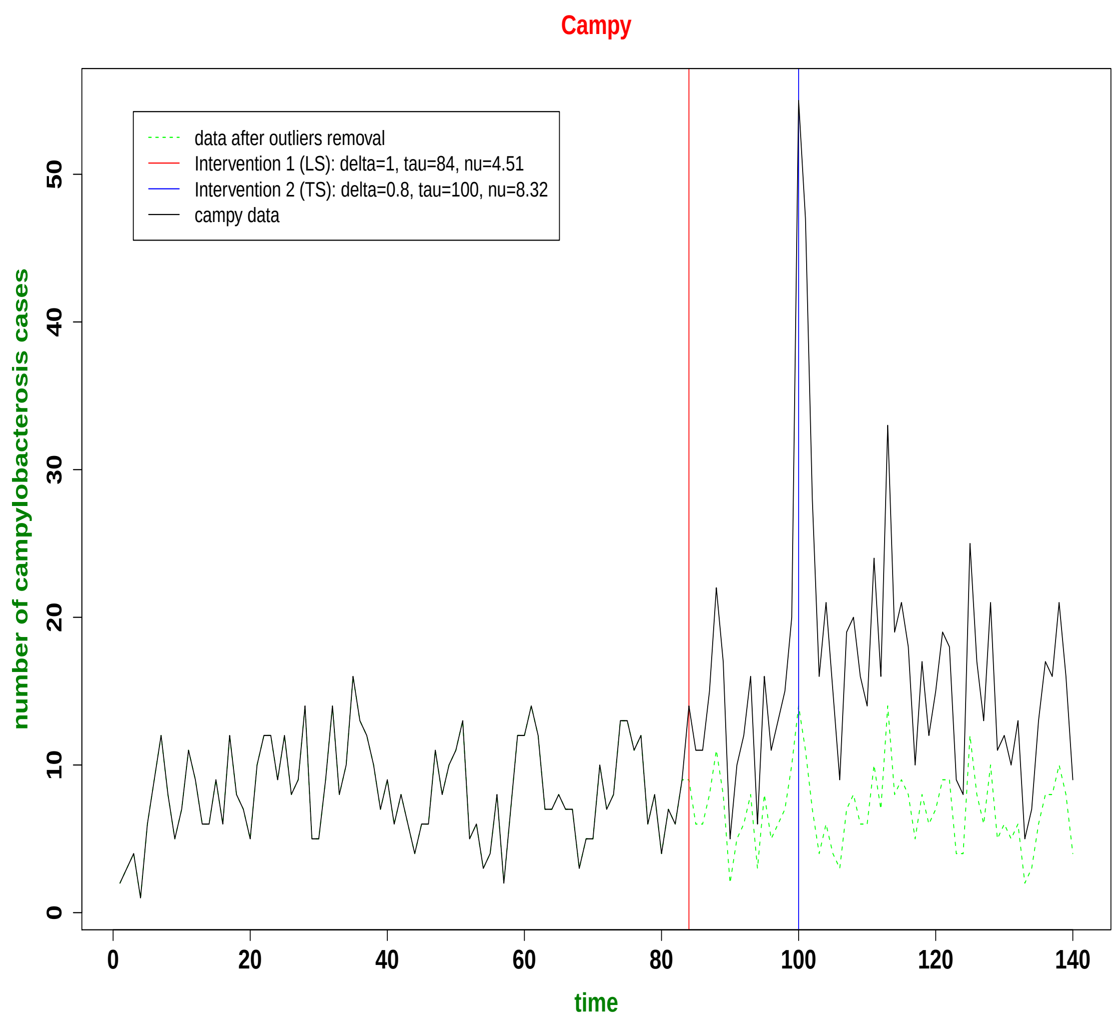


Figure 5.5: Campy data (solid black line) and data after removal of outliers (dashed green line). Step 1: LS at time 84 (vertical red line), step 2: TS at time 100 (vertical blue line).

Table 5.4: Parameter estimates for the artificial campy data (left) and for the cleaned artificial campy data (right)

Estimators \ estimates	unclean		clean	
	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\alpha}_1$
condlikest	2.0927	0.8264	4.4239	0.4218
robPoisTStukey3	2.3496	0.7891	4.3944	0.4252

5.4 Analysis of artificial campylobacterosis data

In order to know whether the INGARCH(1,0) model is a good model for the campy data, we generate a data set from the final INGARCH(1,0) model with outliers fitted to the campy data. We compare the parameter estimates using the functions condlikest and robPoisTStukey3. Then, we clean the artificial polio data from the outliers using the same approach as for the campy data, and compare again our parameter estimates.

5.4.1 INGARCH(1,0) fit to the artificial campy data

We generate a data set with sample size 140 from the final INGARCH(1,0) model with parameters $\beta_0 = 4.5$ and $\alpha_1 = 0.4$, and with an LS at time 84 and a TS at time 100.
The left hand side of Table 5.4 gives the parameter estimates using the functions condlikest and robPoisTStukey3 for the artificial campy data generated from the final INGARCH(1,0) model with outliers. The difference between condlikest and robPoisTStukey3 is small for α_1, but we have larger differences for β_0.

5.4.2 INGARCH(1,0) fit to the cleaned artificial campy data

We clean the artificial campy data from outliers using the same approach as for the campy data. From Figure 5.6, we detect a transient shift (TS) at time 100 and a level shift (LS) at time 84. We find exactly the intervention effects introduced into the data. The right hand side of Table 5.4 illustrates the results for the cleaned artificial campy data. The differences between condlikest and robPoisTStukey3 are negligible for β_0 and α_1. Here, the differences between the results for the contaminated and for the cleaned data are rather large, what can be explained by the large effects of level shifts on the estimates.

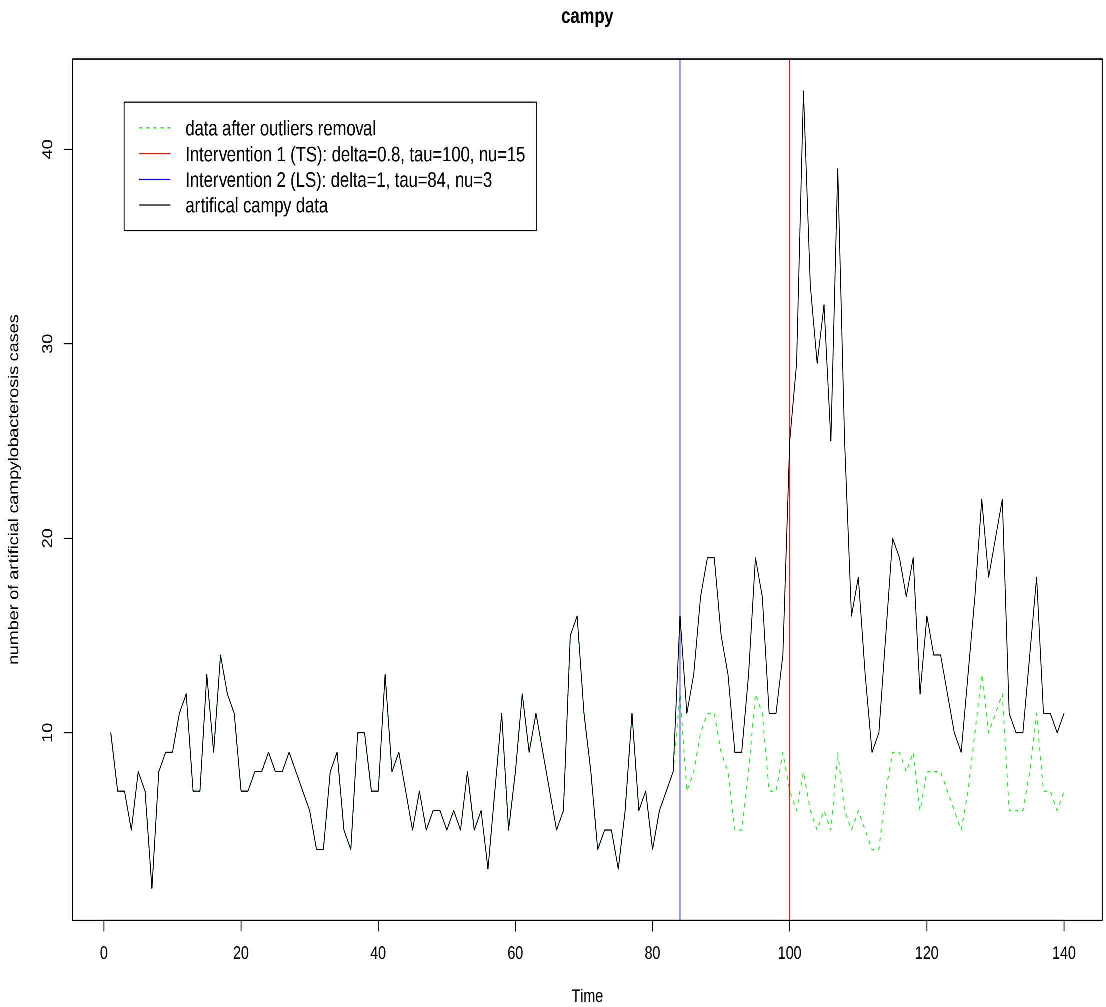


Figure 5.6: Artificial campy data (solid black line) and data after outliers removal (dashed green line). Step 1: TS at time 100 (vertical red line), step 2: LS at time 84 (vertical blue line).

Chapter 6

Summary, conclusions and outlook

The goal of our thesis was robust modelling of count data. Poisson models provide a standard framework for the analysis of this type of data. We constructed new M-estimators based on the Tukey function as modified versions of maximum likelihood estimators to estimate the parameters of such models in the case of independent data and INGRACH models for time series. To achieve this goal, we organized our thesis into six chapters, with Chapter 6 being this summary. Chapter 1 and Chapter 2 were introduction and basic concepts for location M-estimators. The main three chapters are Chapters 3 to 5. Chapter 3 and Chapter 4 presented M-estimation in the case of independent and dependent Poisson data, respectively. Chapter 5 gave two real data application examples in the medical field, and two artificial data examples.
This chapter is devoted to give a summary and some conclusions for our three main chapters. Also this chapter gives our ambition to what we want to do afterwards.

In Chapter 3, we introduced modified Tukey M-estimators with bias correction for estimating the mean of the Poisson distribution. We have developed a new algorithm for this estimation, which is based on the Tukey function. We modified the R-function glmrob by adding a bias correction term and external weights. We considered an alternative estimator based on adaptive trimming as a solution to the problem of combining high robustness against outliers and high efficiency relatively to the sample mean when the true mean is small.
Our simulation results indicated that the modified version of glmrob with external weights gives the best robustness properties among all estimation procedures based on the Huber function. The modified Tukey M-estimator provided improvements over the other procedures which depend on the Tukey function and also those which depend on the Huber function, particularly in case of moderately large outliers and very large outliers. An adaptive trimming estimator provided even better results at small Poisson means.

In Chapter 4, we treated robust M-estimation for the parameters of INGARCH(1,0) and INGARCH(1,1) models as special variants of INGARCH models, where we modified the classical estimation approach by giving robust estimators for the parameters of these models with bias correction and with robust initialization from the AR(1) model in case

of the INGARCH(1,0) model and with robust initialization from the ARMA(1,1) model in case of the INGARCH(1,1) model. We conjectured the asymptotic normality of the robust M-estimator in the INGARCH(1,0) model. We implemented some new functions in R, which are based on the Huber function and the Tukey function, respectively, to compute these estimates.
Our simulation results indicated that in case of the INGARCH(1,0) model, the bias-corrected Tukey M-estimators give better results than those which are based on the Huber function. The bias-corrected Tukey M-estimators with robust initialization from the AR(1) model show good performances relatively to other functions. In case of the INGARCH(1,1) model, our simulation results indicated that the Tukey M-estimators provide robustness for the estimation of β_0 and α_1, but not for β_1, in the presence of additive outliers. The bias-corrected versions with initialization from an ARMA(1,1) fit and a large tuning constant additionally performed almost as good as the conditional maximum likelihood in case of clean data.

Chapter 5 presented two real data examples in the medical field and two artificial data examples generated to resemble the two real data examples, so that we could compare the results to the ground truth. We concluded that the robust estimators give similar results as the non-robust estimators in case of clean data. The robust estimates gave better results than the non-robust ones in the presence of outliers. Using the artificial data examples indicated that the INGARCH(1,0) model with intervention effects is adequate for these data.

Future work should be to modify Tukey M-estimators or to develop other robust estimators for the INGARCH(1,1) model or higher order models. The asymptotic properties of these estimators should be derived. Applications to other real data examples should be given.

Appendix A

Asymptotic properties of M-estimators for INARCH(1) Parameters

Here we investigate the background of Conjecture 4.1, which is given in Chapter 4.
Let:
$\theta = (\beta_0, \alpha_1)'$ denote the vector of the parameters of the INARCH(1) model,
θ_0 be the value of θ for which the expectation of the left hand side of (4.31) equals 0,
$\hat{\theta} = (\hat{\beta_0}, \hat{\alpha_1})'$ denote the vector of the estimators,
$G_n(\hat{\theta}_n)$ denote the left hand side of (4.30) with bias correction $B = (a_0, a_1)'$

$$G_n(\theta) = \begin{pmatrix} G_{n1}(\theta) \\ G_{n2}(\theta) \end{pmatrix} = \sum_{t=2}^{n} X_t(\theta) \tag{A.1}$$

and let

$$G(\theta) = -E_{\theta_0}(X_t(\theta)) \tag{A.2}$$

where

$$X_t(\theta) = \psi\left(\tfrac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right)\tfrac{1}{\sqrt{\lambda_t}}\begin{pmatrix} 1 \\ \sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right) + \lambda \end{pmatrix} - \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$G_{n1}(\theta) = \sum_{t=2}^{n} \psi\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}} - (n-1)a_0, \tag{A.3}$$

$$G_{n2}(\theta) = \sum_{t=2}^{n} \psi\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\sigma\psi\left(\frac{Y_{t-1} - \lambda}{\sigma}\right) + \lambda\right) - (n-1)a_1, \tag{A.4}$$

with ψ being a bounded, continuous and monotone function as for the Huber ψ function. To proof this, we note first that when $\alpha_1 < 1$, (Y_t) is stationary and ergodic, see Doukhan et al. (2012).

By using Theorem 3.5.8 in Stout (1974), we can prove that (X_t) is stationary and ergodic, namely:
Theorem 3.5.8: Let $\{Y_i, i \geqslant 1\}$ be stationary ergodic and ϕ be a measurable function $\phi : \mathbb{R}_\infty \rightarrow \mathbb{R}_1$. Let $X_i = \phi(Y_i, Y_{i+1}, ...)$ define $\{X_i, i \geqslant 1\}$. Then $\{X_i, i \geqslant 1\}$ is stationary ergodic.
Under this theorem, X_t as a function of (Y_t, Y_{t-1}) is also stationary and ergodic.
Second to prove (a) and (b) in Conjecture 4.1, we need to establish all the conditions given in Theorem 4.1.2 and Theorem 4.1.3 in Amemiya (1985).

- Check the assumptions as in Theorem 4.1.2
  $(A1)$: For model (4.2), the parametric space Θ is compact with

$$\Theta = \{\delta \leq \beta_0 \leq \kappa,\, 0 \leq \alpha_1 \leq \kappa^* < 1\}$$

  where δ, κ are finite positive constants and θ_0 is an interior point in Θ.

  $(A2)$: $G_n(\theta)$ is a measurable and continuous function (this follows under $(A1)$ for every continuous ψ function, since the composition of continuous functions is also continuous).

  $(A3)$: $-\frac{1}{n-1}G_n(\theta)$ converges to $G(\theta)$ in probability uniformly in Θ for all $\theta \in \Theta$, and $G(\theta)$ has a unique root.

  First: we can establish the validity of the uniform convergence in two steps

  1. Proving $E(\sup_{\theta\in\Theta} |X_t(\theta)|) < \infty$

$$E\Big(\sup_{\theta\in\Theta} |X_t(\theta)|\Big) = E\Big(\sup_{(\delta\leq\beta_0\leq\kappa,0\leq\alpha_1<\kappa^*)} |X_t(\theta)|\Big)$$
$$= E\Big(\sup_{(\delta\leq\beta_0\leq\kappa,0\leq\alpha_1<\kappa^*)} \Big|\psi\left(\tfrac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\tfrac{1}{\sqrt{\lambda_t}}\begin{pmatrix} 1 \\ \sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda \end{pmatrix} - \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}\Big|\Big)$$

  We have the following inequalities, which are to be read componentwise:

$$\leq E\Big(\sup_{(\delta\leq\beta_0\leq\kappa,0\leq\alpha_1<\kappa^*)} \Big[\psi_{\max}\tfrac{1}{\sqrt{\beta_0}}\begin{pmatrix} 1 \\ \sigma_{\max}\psi_{\max}+\lambda_{\max} \end{pmatrix} - \begin{pmatrix} a_{0\max} \\ a_{1\max} \end{pmatrix}\Big]\Big)$$
$$\leq E\Big(\Big[\psi_{\max}\tfrac{1}{\sqrt{\delta}}\begin{pmatrix} 1 \\ \sigma_{\max}\psi_{\max}+\lambda_{\max} \end{pmatrix} - \begin{pmatrix} a_{0\max} \\ a_{1\max} \end{pmatrix}\Big]\Big) < \infty$$

  where

$$\lambda_{\max} = \max_{\theta\in\Theta}\lambda = \max_{\theta\in\Theta}[\beta_0/(1-\alpha_1)] < \infty$$

$$\sigma_{\max} = \max_{\theta\in\Theta}\sigma = \max_{\theta\in\Theta}\sqrt{\lambda(1+\alpha_1^2/(1-\alpha_1^2))} < \infty$$

$$a_{0\max} = \max\{|a_0(\theta)| : \theta \in \Theta\}$$

$$a_{1\max} = \max\{|a_1(\theta)| : \theta \in \Theta\}$$

and

$$\psi_{\max} = \max_{r\in\mathbb{R}} |\psi(r)| < \infty$$

(this follows for bounded ψ functions)

2. From 1, we have $E(\sup_{\theta\in\Theta} |X_t(\theta)|) < \infty$. Using the ergodic theorem in Jensen and Rahbek (2007) and arguments given in Section 2.2 in Straumann (2005), we have $-\frac{1}{n-1}G_n(\theta)$ converges to $G(\theta)$ in probability uniformly in Θ for all $\theta \in \Theta$, see Lemma 3 in Zhu and Wang (2009).

Second: To prove that $G(\theta)$ has a unique root, it is sufficient to prove that $0 = W_t(\theta) = -X_t(\theta)$ has a unique root, and for simplicity we exclude the bias correction.
Proof:
To prove this, we investigate the monotonicity of $W_t(\theta) = -X_t(\theta)$
Let

$$W_t(\theta) = \begin{pmatrix} W_{t1}(\theta) \\ W_{t2}(\theta) \end{pmatrix} = \begin{pmatrix} -X_{t1}(\theta) \\ -X_{t2}(\theta) \end{pmatrix},$$

$$= \begin{pmatrix} \psi\left(\frac{y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}} \\ \\ \psi\left(\frac{y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\sigma\psi\left(\frac{y_{t-1}-\lambda}{\sigma}\right)+\lambda\right) \end{pmatrix}. \tag{A.5}$$

Let $\theta_1 = (\beta_{01}, \alpha_{11})$ and $\theta_2 = (\beta_{02}, \alpha_{12})$ be any two parameter vectors with $\theta_2 \geq \theta_1 \quad (\beta_{02} \geq \beta_{01}, \alpha_{12} \geq \alpha_{11})$.
We have

$$W_{t1}(\theta_2) - W_{t1}(\theta_1) = X_{t1}(\theta_1) - X_{t1}(\theta_2) \tag{A.6}$$

$$= \frac{1}{\sqrt{(\beta_{01}+\alpha_{11}y_1)}}\psi\left(\frac{y_2-(\beta_{01}+\alpha_{11}y_1)}{\sqrt{(\beta_{01}+\alpha_{11}y_1)}}\right) - \frac{1}{\sqrt{(\beta_{02}+\alpha_{12}y_1)}}\psi\left(\frac{y_2-(\beta_{02}+\alpha_{12}y_1)}{\sqrt{(\beta_{02}+\alpha_{12}y_1)}}\right) \tag{A.7}$$

Let $a = \frac{y_2-(\beta_{01}+\alpha_{11}y_1)}{\sqrt{(\beta_{01}+\alpha_{11}y_1)}}$ and $b = \frac{y_2-(\beta_{02}+\alpha_{12}y_1)}{\sqrt{(\beta_{02}+\alpha_{12}y_1)}}$. If $b \geq 0$ then $a \geq b \geq 0$.
let ψ be Huber's ψ function. Then we need to distinguish several cases:

1. $a, \quad b \in [-k, k]$.
2. $a, \quad b \in \ ]-\infty, -k[$.
3. If $b \geq k$, then necessarily $y_2 - (\beta_{02}+\alpha_{12}y_1) > 0$ and $a > b \geq k$, so $a, b \in [k, \infty[$.
4. $b \in [-k, k]$ and $a \in [k, \infty[$.
5. $b \in]-\infty, -k[$ and $a \in [k, \infty[$.
6. $b \in [-k, k]$ and $a \in]-\infty, -k[$.

7. $b \in]-\infty, -k[$ and $a \in [-k, k]$.

1. $a, \quad b \in [-k, k]$
   For Huber's ψ function, we have $\psi(a) = a$ and $\psi(b) = b$, and

$$W_{t1}(\theta_2) - W_{t1}(\theta_1) = \frac{a}{\sqrt{(\beta_{01} + \alpha_{11} y_1)}} - \frac{b}{\sqrt{(\beta_{02} + \alpha_{12} y_1)}}, \tag{A.8}$$

$$= \left(\frac{y_2 - (\beta_{01} + \alpha_{11} y_1)}{(\beta_{01} + \alpha_{11} y_1)} \right) - \left(\frac{y_2 - (\beta_{02} + \alpha_{12} y_1)}{(\beta_{02} + \alpha_{12} y_1)} \right)$$

$$= \frac{[(\beta_{02} + \alpha_{12} y_1) - (\beta_{01} + \alpha_{11} y_1)] y_2}{(\beta_{01} + \alpha_{11} y_1)(\beta_{02} + \alpha_{12} y_1)}$$

$$= \frac{[(\beta_{02} - \beta_{01}) + (\alpha_{12} - \alpha_{11}) y_1] y_2}{(\beta_{01} + \alpha_{11} y_1)(\beta_{02} + \alpha_{12} y_1)}$$

.

   $W_{t1}(\theta_2) - W_{t1}(\theta_1) > 0$ if $\beta_{02} \geq \beta_{01}$, $\alpha_{12} \geq \alpha_{11}$, $(\beta_{01}, \alpha_{11}) \neq (\beta_{02}, \alpha_{12})$, except if $y_1 = 0$ or $y_2 = 0$, but then at least $W_{t1}(\theta_2) - W_{t1}(\theta_1) \geq 0$.

2. $a, \quad b \in \;]-\infty, -k[$
   Here we have $\psi(a) = -k$ and $\psi(b) = -k$, and

$$W_{t1}(\theta_2) - W_{t1}(\theta_1) = \frac{-k}{\sqrt{(\beta_{01} + \alpha_{11} y_1)}} - \frac{-k}{\sqrt{(\beta_{02} + \alpha_{12} y_1)}} \tag{A.9}$$

$$= \frac{k\left(\sqrt{(\beta_{02} + \alpha_{12} y_1)} - \sqrt{(\beta_{01} + \alpha_{11} y_1)}\right)}{\sqrt{(\beta_{01} + \alpha_{11} y_1)}\sqrt{(\beta_{02} + \alpha_{12} y_1)}} < 0$$

From 2, we have $W_{t1}(\theta_2) < W_{t1}(\theta_1)$, then $W_t(\theta)$ is not monotone and we do not need to check the other cases. Also to be sure that $W_t(\theta)$ is not monotone, we plot equation (A.5) considering the following cases:

1. $y_1 = 2$, $y_2 = 3$, fixed $\beta_0 = 1$ and $\alpha_1 \in [0.1, 0.2, ..., 0.9]$, see Figure A.1 (top left)
2. $y_1 = 2$, $y_2 = 3$, fixed $\alpha_1 = 0.4$ and $\beta_0 \in [1, 2, ..., 9]$, see Figure A.1 (top right)
3. $y_1 = 3$, $y_2 = 2$, fixed $\beta_0 = 1$ and $\alpha_1 \in [0.1, 0.2, ..., 0.9]$, see Figure A.1 (bottom left)
4. $y_1 = 3$, $y_2 = 2$, fixed $\alpha_1 = 0.4$ and $\beta_0 \in [1, 2, ..., 9]$, see Figure A.1 (bottom right)

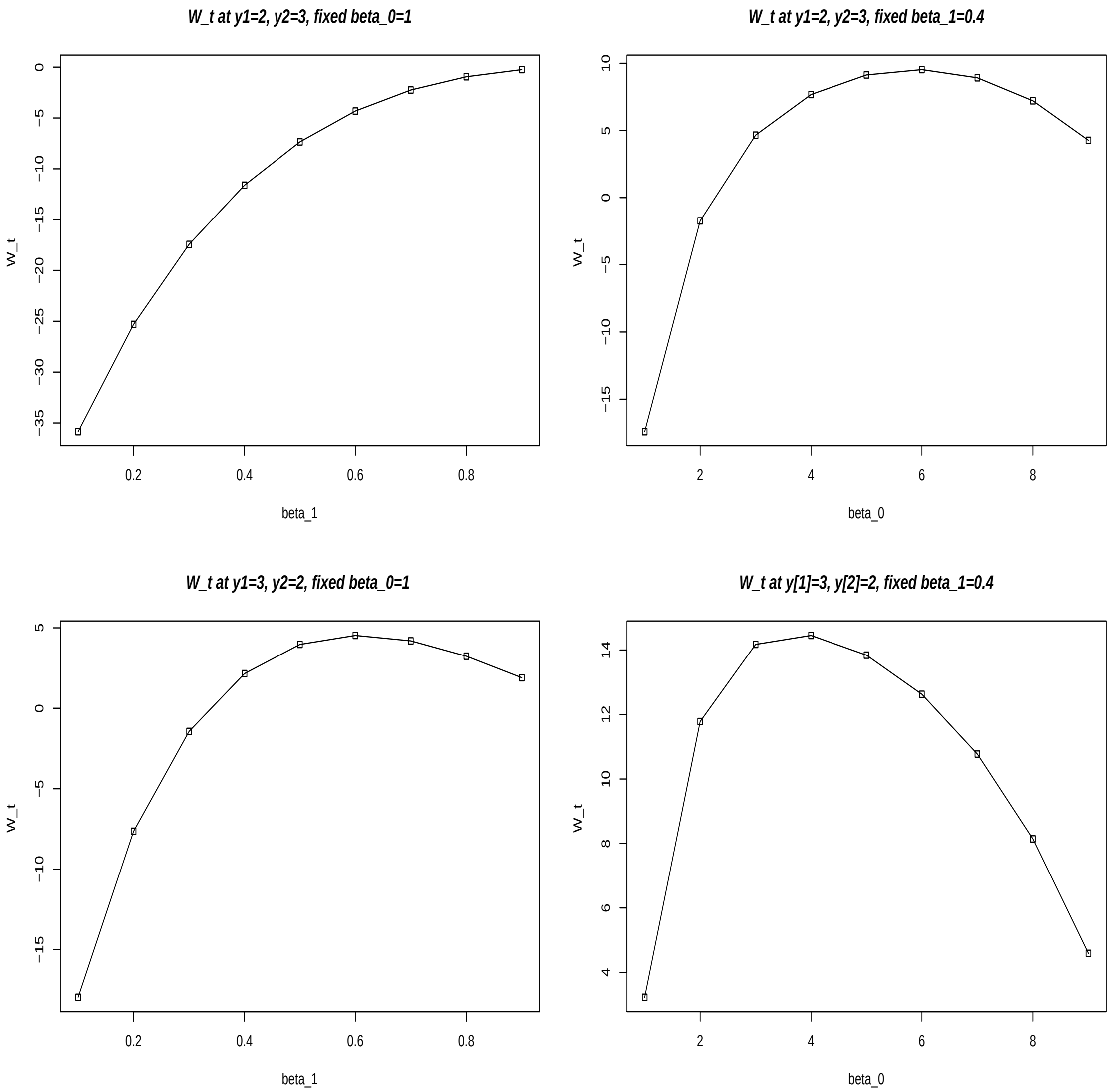


Figure A.1: Equation (A.5)

From Figure A.1, we conclude that $W_t(\theta)$ is not monotone, so that (4.27) can have more than one solution. Therefore a good choice of the initial estimate in the iterative numerical algorithm used to solve these estimating equations is essential.

Thus, we need to assume existence of a unique root of $G(\theta)$ for establishing the conditions in Theorem 4.1.2 in Amemiya (1985), so that part (a) of Conjecture 4.1 holds ($\hat{\theta}_n$ converges to θ_0 in probability).

- Check the assumptions as in Theorem 4.1.3

$(A4)$ $U_n(\hat{\theta}_n) = \frac{\partial G_n(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_n}$ exists and is continuous. The continuity follows under $(A2)$ and we can prove the existence as follows:

$$U_n(\hat{\theta}_n) = \begin{pmatrix} \dfrac{\partial G_{n1}(\theta)}{\partial \beta_0}\Big|_{\theta=\hat{\theta}_n} & \dfrac{\partial G_{n1}(\theta)}{\partial \alpha_1}\Big|_{\theta=\hat{\theta}_n} \\ \\ \dfrac{\partial G_{n2}(\theta)}{\partial \beta_0}\Big|_{\theta=\hat{\theta}_n} & \dfrac{\partial G_{n2}(\theta)}{\partial \alpha_1}\Big|_{\theta=\hat{\theta}_n} \end{pmatrix} \tag{A.10}$$

(A.10) is finite, where by direct differentiation without including the bias correction we have

$$\frac{\partial G_{n1}(\theta)}{\partial \beta_0}\Big|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n} \frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}} \tag{A.11}$$

Let

$$Q_t = \frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}$$

then

$$Q_t = -\frac{1}{2}\lambda_t^{\frac{-3}{2}}\psi\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right) + \frac{1}{\sqrt{\lambda_t}}\frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right)$$

when we use Huber's ψ function with $|(Y_t - \lambda_t)/\sqrt{\lambda_t}| \leq k$ then

$$\frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right) = \frac{\partial}{\partial \beta_0}\left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right) I_{\{-k \leq \left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \leq k\}}$$

$$= \left[\frac{-\lambda_t^{\frac{1}{2}} - \left[\frac{1}{2}(Y_t - \lambda_t)\lambda_t^{\frac{-1}{2}}\right]}{\lambda_t}\right] I_{\{-k \leq \left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \leq k\}} = \left[-\frac{1}{2}\lambda_t^{\frac{-1}{2}} - \frac{Y_t}{2}\lambda_t^{\frac{-3}{2}}\right] I_{\{-k \leq \left(\frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}\right) \leq k\}}$$

so

$$Q_t = -\frac{1}{2}\lambda_t^{\frac{-3}{2}}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right) - \left[\frac{1}{2\lambda_t}+\frac{Y_t}{2\lambda_t^2}\right] I_{\{-k\le\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\le k\}} \tag{A.12}$$

and hence

$$\left.\frac{\partial G_{n1}(\theta)}{\partial\beta_0}\right|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n} Q_t \tag{A.13}$$

Similarly, we calculate

$$\left.\frac{\partial G_{n1}(\theta)}{\partial\alpha_1}\right|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n} \frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}} \tag{A.14}$$

Let

$$R_t = \frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}$$

then

$$R_t = -\frac{1}{2}\lambda_t^{\frac{-3}{2}}Y_{t-1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right) + \frac{1}{\sqrt{\lambda_t}}\frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)$$

when we use Huber's ψ function, then

$$\frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right) = \frac{\partial}{\partial\alpha_1}\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right) I_{\{-k\le\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\le k\}}$$

$$= \left[-\frac{1}{2}\lambda_t^{\frac{-1}{2}}Y_{t-1} - \frac{Y_tY_{t-1}}{2}\lambda_t^{\frac{-3}{2}}\right] I_{\{-k\le\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\le k\}}$$

then

$$R_t = -\frac{1}{2}\lambda_t^{\frac{-3}{2}}Y_{t-1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right) - \left[\frac{Y_{t-1}}{2\lambda_t}+\frac{Y_tY_{t-1}}{2\lambda_t^2}\right] I_{\{-k\le\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\le k\}} = Y_{t-1}Q_t \tag{A.15}$$

and hence

$$\left.\frac{\partial G_{n1}(\theta)}{\partial\alpha_1}\right|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n} R_t = \sum_{t=2}^{n} Y_{t-1}Q_t \tag{A.16}$$

In the same way

$$\left.\frac{\partial G_{n2}(\theta)}{\partial\beta_0}\right|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n} \frac{\partial}{\partial\beta_0}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right) \tag{A.17}$$

Let

$$M_t = \frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)$$

then

$$M_t = \left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)Q_t + \psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\frac{\partial}{\partial \beta_0}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right) \tag{A.18}$$

where
$Q_t = \frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}$ is defined in (A.12).
Let

$$S_t = \frac{\partial}{\partial \beta_0}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)$$

then

$$S_t = \left[\sigma\frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\frac{\partial \sigma}{\partial \beta_0}\right]+\frac{\partial \lambda}{\partial \beta_0}$$

where
$\frac{\partial \sigma}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\sqrt{\frac{\beta_0}{(1-\alpha_1)(1-\alpha_1^2)}} = \frac{\beta_0^{\frac{-1}{2}}}{2\left((1-\alpha_1)(1-\alpha_1^2)\right)^{\frac{1}{2}}} = \frac{1}{2\left(\beta_0(1-\alpha_1)(1-\alpha_1^2)\right)^{\frac{1}{2}}}$

and

$$\frac{\partial \lambda}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\frac{\beta_0}{(1-\alpha_1)} = \frac{1}{(1-\alpha_1)}$$

Let

$$w_t = \frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)$$

when we use Huber's ψ function, then

$$w_t = \frac{\partial}{\partial \beta_0}\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right) = \frac{\partial}{\partial \beta_0}\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)I_{\{-k\le\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\le k\}}$$

$$= \frac{1}{\sigma^2}\left[\left(\frac{-\sigma}{1-\alpha_1}\right)-\left(\frac{Y_{t-1}-\lambda}{2\left(\beta_0(1-\alpha_1)(1-\alpha_1^2)\right)^{\frac{1}{2}}}\right)\right]I_{\{-k\le\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\le k\}}$$

then
$S_t = \left[\sigma w_t + \psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\left(\frac{1}{2\left(\beta_0(1-\alpha_1)(1-\alpha_1^2)\right)^{\frac{1}{2}}}\right)\right] + \frac{1}{(1-\alpha_1)}$

using S_t and Q_t, then

$$M_t = \left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)Q_t + \psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}S_t \tag{A.19}$$

and hence

$$\frac{\partial G_{n2}(\theta)}{\partial \beta_0}\bigg|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n} M_t. \tag{A.20}$$

Similarly, we calculate

$$\frac{\partial G_{n2}(\theta)}{\partial \alpha_1}\bigg|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n}\frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right) \tag{A.21}$$

Let

$$N_t = \frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)$$

then
$N_t = \left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)R_t + \psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}\frac{\partial}{\partial\alpha_1}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)$

where
$R_t = \frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}$ is defined in (A.15).

Let

$$Z_t = \frac{\partial}{\partial\alpha_1}\left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)$$

then

$$Z_t = \left[\sigma\frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\frac{\partial\sigma}{\partial\alpha_1}\right]+\frac{\partial\lambda}{\partial\alpha_1}$$

where

$$\frac{\partial\lambda}{\partial\alpha_1} = \frac{\partial}{\partial\alpha_1}\Big(\frac{\beta_0}{1-\alpha_1}\Big) = \frac{\beta_0}{(1-\alpha_1)^2}$$

and

$$\frac{\partial\sigma}{\partial\alpha_1} = \frac{\partial}{\partial\alpha_1}\Big(\frac{\beta_0}{(1-\alpha_1)(1-\alpha_1^2)}\Big)^{\frac{1}{2}}$$

$$= \frac{1}{2}\Big(\frac{\beta_0}{(1-\alpha_1)(1-\alpha_1^2)}\Big)^{\frac{-1}{2}}\Big[\frac{-\beta_0(3\alpha_1^2-2\alpha_1-1)}{\Big((1-\alpha_1)(1-\alpha_1^2)\Big)^2}\Big]$$

then

$$\frac{\partial\sigma}{\partial\alpha_1} = \frac{-\beta_0^{\frac{1}{2}}(3\alpha_1^2-2\alpha_1-1)}{2\Big((1-\alpha_1)(1-\alpha_1^2)\Big)^{\frac{3}{2}}}$$

Let

$$H_t = \frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)$$

when we use Huber's ψ function, then

$$H_t = \frac{\partial}{\partial\alpha_1}\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right) = \frac{\partial}{\partial\alpha_1}\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)I_{\{-k\leq\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\leq k\}}$$

$$= \frac{1}{\sigma^2}\Big[\Big(\frac{-\sigma\beta_0}{(1-\alpha_1)^2}\Big)+\Big(\frac{\beta_0^{\frac{1}{2}}(Y_{t-1}-\lambda)(3\alpha_1^2-2\alpha_1-1)}{2\Big((1-\alpha_1)(1-\alpha_1^2)\Big)^{\frac{3}{2}}}\Big)\Big]I_{\{-k\leq\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\leq k\}}$$

then

$Z_t = \sigma H_t + \psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)\left[\frac{-\beta_0^{\frac{1}{2}}(3\alpha_1^2-2\alpha_1-1)}{2\left((1-\alpha_1)(1-\alpha_1^2)\right)^{\frac{3}{2}}}\right]+\left(\frac{\beta_0}{(1-\alpha_1)^2}\right)$

using Z_t and R_t, then

$$N_t = \left(\sigma\psi\left(\frac{Y_{t-1}-\lambda}{\sigma}\right)+\lambda\right)R_t + \psi\left(\frac{Y_t-\lambda_t}{\sqrt{\lambda_t}}\right)\frac{1}{\sqrt{\lambda_t}}Z_t$$

and hence

$$\frac{\partial G_{n2}(\theta)}{\partial\alpha_1}\bigg|_{\theta=\hat{\theta}_n} = \sum_{t=2}^{n}N_t. \tag{A.22}$$

Using (A.13), (A.16), (A.20) and (A.22), then

$$U_n(\hat{\theta}_n) = \begin{pmatrix} \sum_{t=2}^n Q_t & \sum_{t=2}^n R_t \\ \sum_{t=2}^n M_t & \sum_{t=2}^n N_t \end{pmatrix}. \tag{A.23}$$

(A5): $n^{-1}U_n(\theta_n^*)$ converges to a finite nonsingular matrix $A(\theta_0) = \lim En^{-1}U_n(\theta_0)$ in probability for any sequence θ_n^* such that $p\lim\theta_n^* = \theta_0$. This assumption follows

by using the ergodic theorem in Jensen and Rahbek (2007).
We can estimate $A(\theta_0)$ by $n^{-1}U_n(\hat{\theta}_n)$ using (A.23) as follows:

$$\widehat{A(\theta_0)} = \begin{pmatrix} n^{-1}\sum_{t=2}^{n} Q_t & n^{-1}\sum_{t=2}^{n} R_t \\ \\ n^{-1}\sum_{t=2}^{n} M_t & n^{-1}\sum_{t=2}^{n} N_t \end{pmatrix} \tag{A.24}$$

$(A6)$: $n^{-\frac{1}{2}}G_n(\theta_0) \rightarrow N(0, B(\theta_0))$, where $B(\theta_0) = \lim_{n\longrightarrow\infty} En^{-1}G_n(\theta_0)G_n(\theta_0)'$. This assumption follows using the results of Martingale Central Limit Theorem (MCLT), see Theorem 3 in Sethuraman (2002).

Remark 1: To estimate $B(\theta_0)$, we use Theorem A in Serfling (1980, Page 14), namely:
If we suppose that $Z_n \xrightarrow{d} N(0, B(\theta_0))$, then

$$Z_n Z_n' \xrightarrow{d} B(\theta_0)\chi_1^2$$

If the sequence $\{Z_n Z_n', n \in \mathbb{N}\}$ is uniformly integrable, then

$$E(Z_n Z_n') \rightarrow E(B(\theta_0)\chi_1^2) = B(\theta_0)$$

Using $(A6)$ and Remark 1, we can estimate $B(\theta_0)$ as follows:

$$\widehat{B(\theta_0)} = n^{-1}G_n(\hat{\theta})G_n(\hat{\theta})' \tag{A.25}$$

Now assuming part (a), we can prove part (b) of Conjecture 4.1 as follows:
Using a first-order Taylor expansion, we obtain

$$0 = G_n(\hat{\theta}_n) = G_n(\theta_0) + U_n(\theta_n^*)(\hat{\theta}_n - \theta_0) \tag{A.26}$$

where θ^* lies in between $\hat{\theta}_n$ and θ_0.

Rewrite (A.26) as follows:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \left(n^{-1}\frac{\partial U_n(\theta^*)}{\partial \theta'}\right)^{+}(n^{-\frac{1}{2}}G_n(\theta_0)) \tag{A.27}$$

where $+$ denotes the Moore-Penrose generalized inverse, see Amemiya (1985, Page 112). Under assumptions $(A5)$ and $(A6)$, apply Slutsky's theorem and the results of the MCLT, then we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\Big(0, A(\theta_0)^{-1}B(\theta_0)(A(\theta_0)^{-1})^t\Big) \tag{A.28}$$

The asymptotic covariance matrix $A(\theta_0)^{-1}B(\theta_0)(A(\theta_0)^{-1})^t$ can be estimated by $\widehat{A(\theta_0)}^{-1}\widehat{B(\theta_n)}(\widehat{A(\theta_0)^{-1}})^t$, where this estimate is guaranteed to be at least positive-semidefinite by construction.

Appendix B

Results of the functions robPoisTShuber2 and robPoisTShuber3

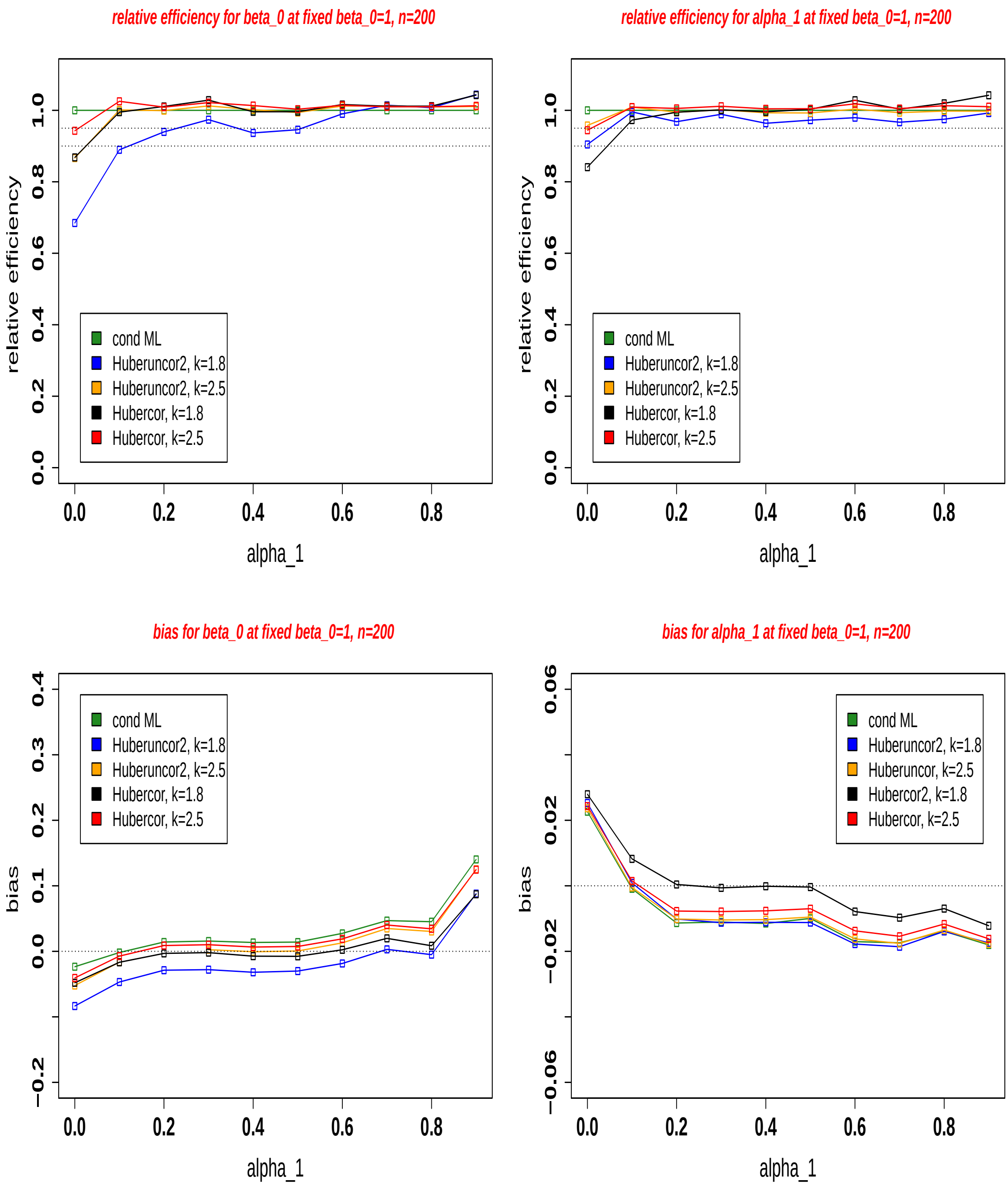


Figure B.1: Simulated biases (bottom) and relative efficiencies (top) of corrected and uncorrected Huber M estimator with different tuning constants k relatively to the conditional maximum likelihood estimator for β_0 (left) and α_1(right) as a function of the true α_1, for $\beta_0 = 1$ and n=200.

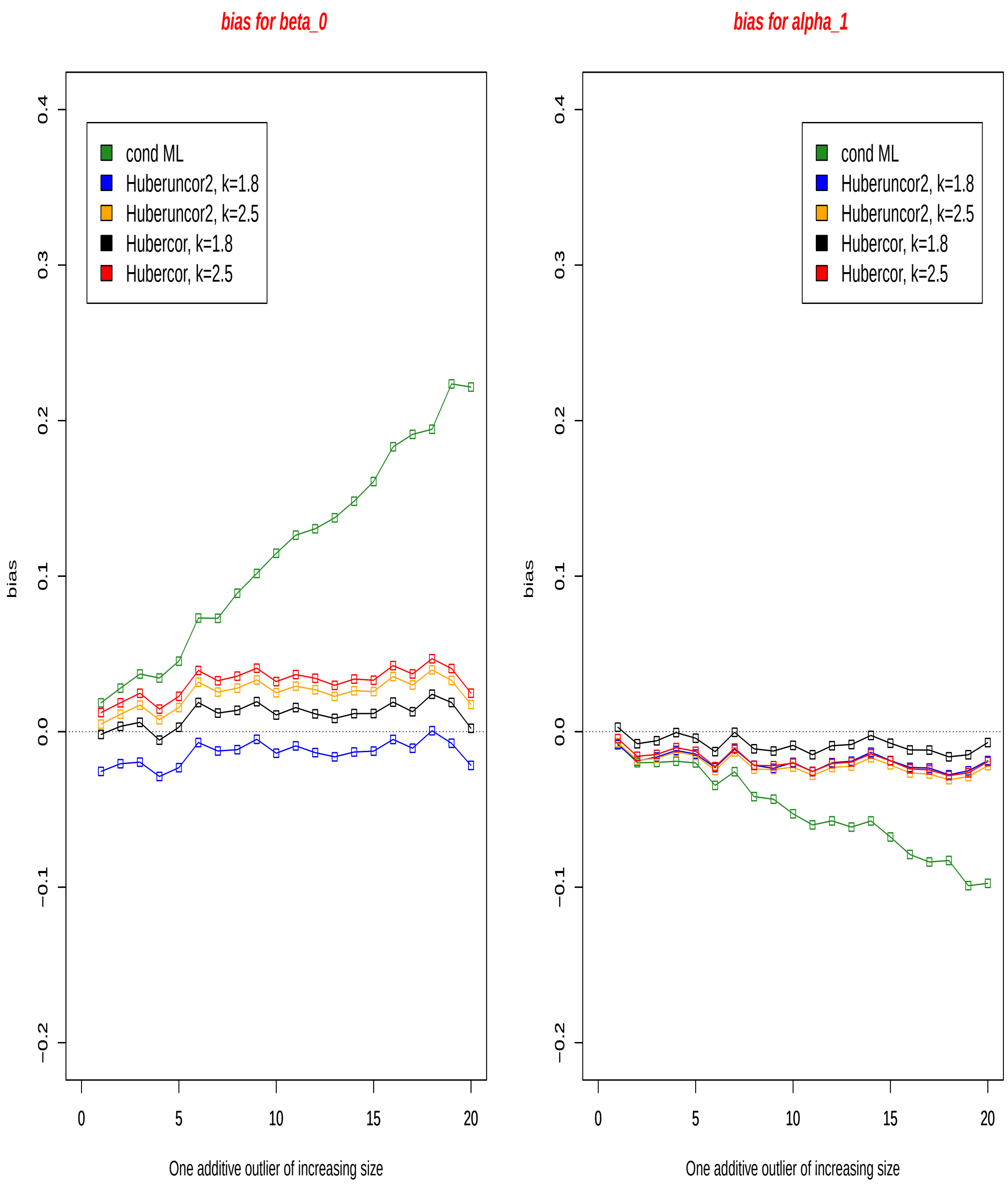


Figure B.2: Simulated biases of the conditional maximum likelihood estimator and of corrected and uncorrected Huber M-estimators with different tuning constants k for β_0 (left) and α_1(right) in case of an additive outlier of increasing size with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

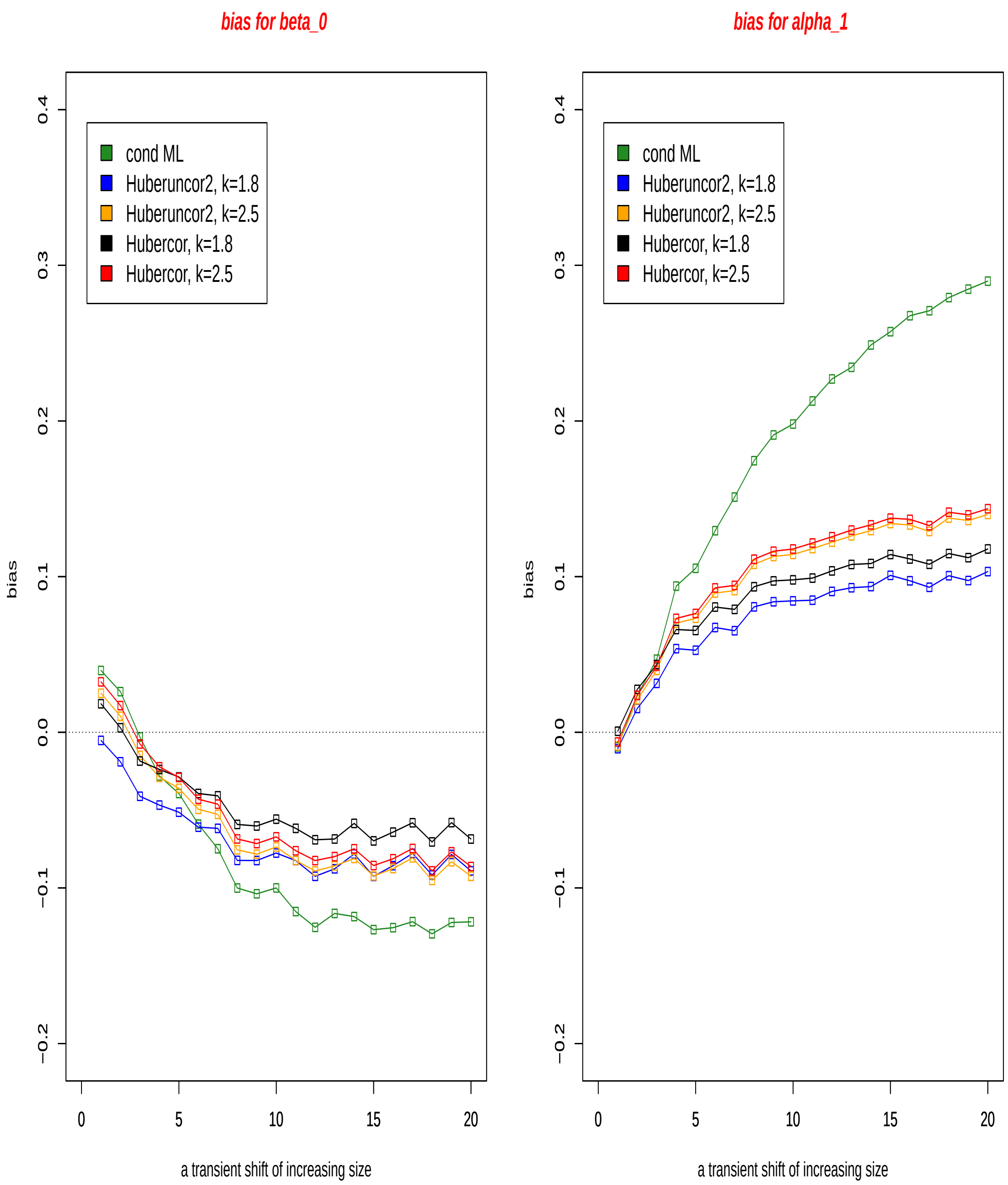


Figure B.3: Simulated biases of the conditional maximum likelihood estimator and of corrected and uncorrected Huber M-estimators with different tuning constants k for β_0 (left) and α_1(right) in case of a transient outlier of increasing size with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

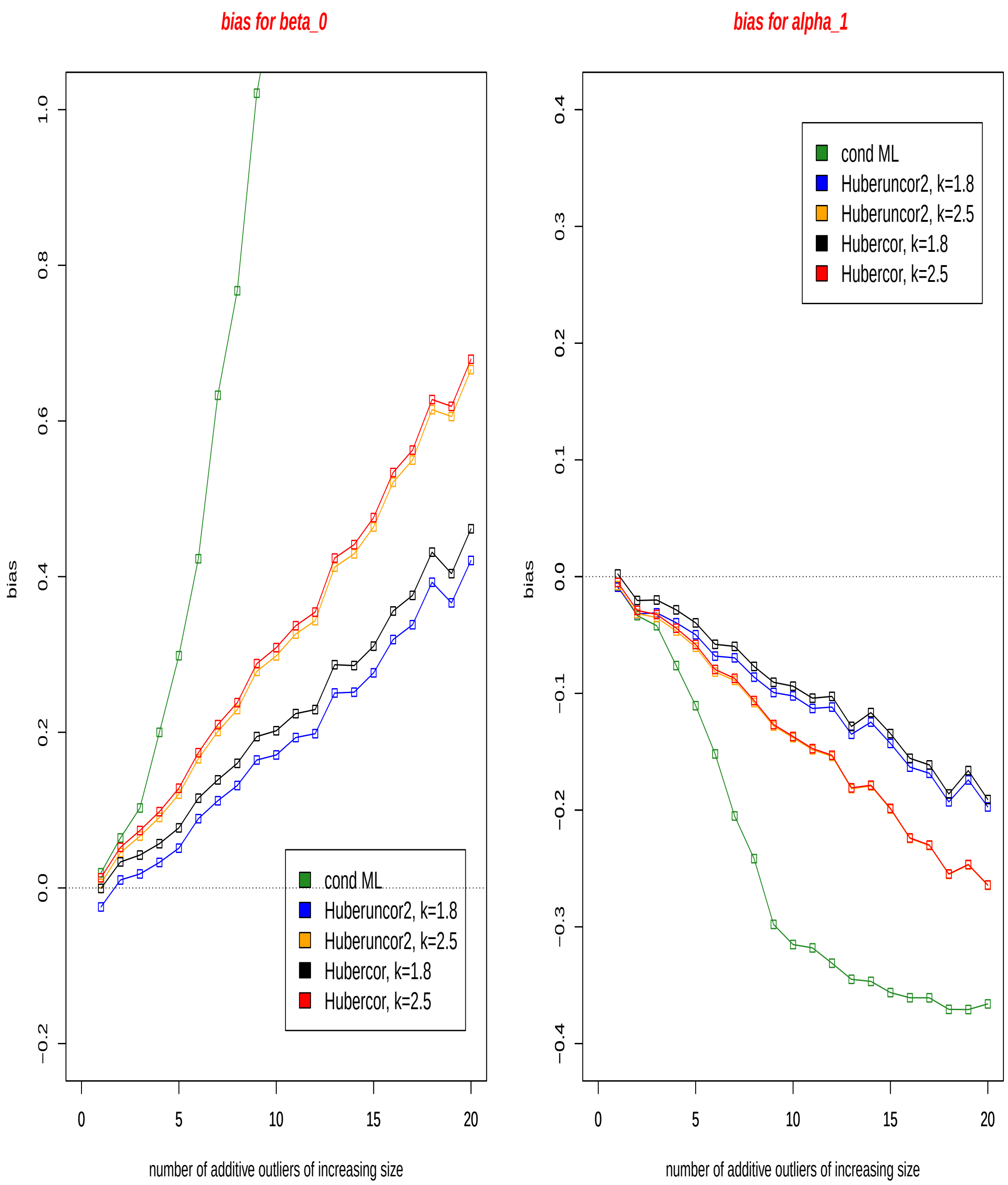


Figure B.4: Simulated biases of the conditional maximum likelihood estimator and of corrected and uncorrected Huber M-estimators with different tuning constants k for β_0 (left) and α_1(right) in case of increasing numbers of additive outliers of increasing sizes with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

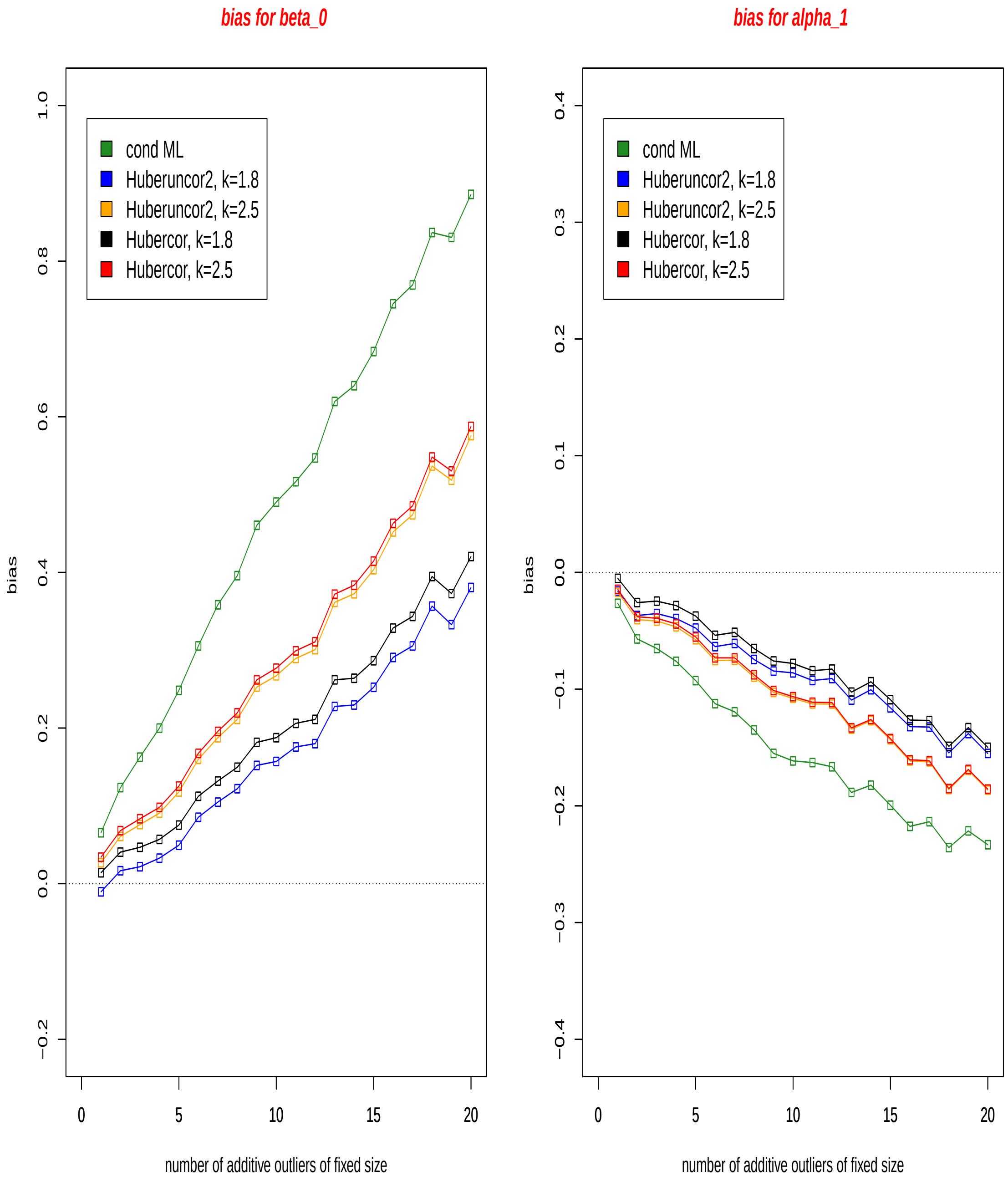


Figure B.5: Simulated biases of the conditional maximum likelihood estimator and of corrected and uncorrected Huber M-estimators with different tuning constants k for β_0 (left) and α_1(right) in case of increasing numbers of additive outlier of fixed size with true values $\beta_0 = 1$ and $\alpha_1 = 0.4$, sample size n=200.

Bibliography

Amemiya, T. (1985). *Advanced econometrics.* Harvard University Press, Cambridge.

Basawa, I., R. Huggins, and R. Staudte (1985). Robust tests for time series with an application to first-order autoregressive processes. *Biometrika 72*, 559–571.

Bickel, P. and E. Lehmann (1975). Descriptive statistics for nonparametric models, I and II. *Annals of Statistics 3*, 1038–1069.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Econometrics 31*, 307–327.

Cadigan, N. and J. Chen (2001). Properties of robust M-estimators for poisson and negative binomial data. *Journal of Statistical Computation and Simulation 70*, 273–288.

Cantoni, E. and E. Ronchetti (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association 96*, 1022–1030.

Chen, C. and L. Liu (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association 88*, 284–297.

Croux, C. (1998). Limit behavior of the empirical influence function of the median. *Statistics and Probability Letters 37*, 331–340.

Davis, R. and R. Wu (2009). A negative binomial model for time series of counts. *Biometrika 96*, 735–749.

Davis, R. A., W. T. M. Dunsmuir, and Y. Wang (2000). On autocorrelation in a poisson regression model. *Biometrika 87*, 491–505.

Denby, L. and R. Martin (1979). Robust estimation of the first order autoregressive parameter. *Journal of the American Statistical Association 74*, 140–146.

Doukhan, P., K. Fokianos, and D. Tjøstheim (2012). On weak dependence conditions for poisson autoregressions. *Statistics and Probability Letters 82*, 942–948.

Elsaied, H. and R. Fried (2010). M-estimation in INARCH models with a special focus on small means. *Proceedings in Computational Statistics COMPSTAT 2010, Lechevallier, Y., Saporta, G.(eds.)*, 967–974.

Elsaied, H., R. Fried, and K. Fokianos (2011). Outliers and interventions in INGARCH time series. *Proceedings in 26th International Workshop on Statistical Modelling IWSM Universitat de València*, 234–239.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica 50*, 987–1008.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer, New York.

Ferland, R., A. Latour, and D. Oraichi (2006). Integer-valued GARCH processes. *Journal of Time Series Analysis 27*, 923–942.

Fokianos, K. and R. Fried (2010). Interventions in INGARCH processes. *Journal of Time Series Analysis 31*, 210–225.

Fokianos, K. and R. Fried (2011). Interventions in log-linear poisson autoregression. tentatively accepted in Statistical Modelling.

Fokianos, K. and B. Kedem (2002). *Regression Models for Time Series Analysis*. John Wiley and Sons, New York.

Fokianos, K. and B. Kedem (2004). Partial likelihood inference for time series analysis following generalized linear models. *Journal of Time Series Analysis 25*, 173–197.

Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *Journal of the American Statistical Association 104*, 1430–1439.

Gastwirth, J. and H. Rubin (1975). The behaviour of robust estimators on dependent data. *Annals of Statistics 3*, 1070–1100.

Hampel, F. (1968). *Contributions to the theory of robust estimation*. Ph. D. thesis, University of California, Berkeley.

Hampel, F. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association 69*, 383–393.

Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel (1986). *Robust Statistics: The Approach Based on Influence Function*. John Wiley and Sons, New York.

Huber, P. (1964). Robust estimation of location parameter. *Annals of Mathematical Statistics 35*, 73–101.

Huber, P. (1972). Robust statistic: A review. *Annals of Mathematical Statistics 43*, 1041–1067.

Huber, P. (1973). Robust regression: asymptotics, conjectures and monte carlo. *Annals of Statistics 1*, 799–821.

Huber, P. (1981). *Robust Statistic of location parameter*. John Wiley and Sons, New York.

Jensen, S. and A. Rahbek (2007). A note on the law of large numbers for functions of geometrically ergodic time series. *Econometric Theory 23*, 761–766.

Jowaheer, V. and B. C. Sutradhar (2005). On AR(1) versus MA(1) models for non-stationary time series of poisson counts: Part II (application). *Proceedings of the 8th WSEAS International Conference on APPLIED MATHEMATICS, Tenerife, Spain*, 363–366.

Kang, J. and S. Lee (2009). Parameter change test for random coefficient integervalued autoregressive processes with application to polio data analysis. *Journal of Time Series Analysis 30*, 239–258.

Kohl, M. (2005). *Numerical Contributions to the Asymptotic Theory of Robustness.* Ph. D. thesis, University of Beyreuth, Germany.

Kohl, M. and P. Ruckdeschel (2010). R package distrMod: Object-oriented implementation of probability models. *Journal of Statistical Software 35*, 1–27.

Kordzakhia, N., G. Mishra, and L. Reiersølmoen (2001). Robust estimation in the logistic regression model. *Journal of Statistical Planning and Inference 98*, 211–223.

Ma, Y. and M. Genton (2000). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis 21*, 663–684.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and other Models for Discrete Valued Time Series.* Chapman and Hall.

Maronna, R., R. Martin, and V. Yohai (2006). *Robust Statistics: Theory and Methods.* John Wiley and Sons, New York.

Martin, R. (1978). Robust estimation of autoregressive models. *In Directions in Time Series, D.R. Billinger and G.C. Tiao (eds.). Institute of Mathematical Statistics, Haywood, Calif. May 1-3*, 228–254.

Martin, R. and V. Yohai (1985). *Robustness in time series and estimating ARMA models.* In Handbook of Statistics, 4, Ed. D.R. Brillinger and P.R. Krishnaiah. Elsevier, New York.

Mora, V. E., P. Neal, and T. S. Rao (2009). Integer valued AR processes with explanatory variables. *The Indian Journal of Statistics 71*, 248–263.

Mukherjee, K. (2008). M-estimation in GARCH models. *Journal of Econometric Theory 24*, 1530–1553.

Muler, N. and V. Yohai (2002). Robust estimates for ARCH processes. *Journal of Time Series Analysis 23*, 341–375.

Muler, N. and V. Yohai (2008). Robust estimates for GARCH models. *Journal of Statistical and Inference 138*, 2918–2940.

Peracchi, F. (1990). Robust M-estimators. *Econometric Reviews 9*, 1–30.

Rousseeuw, P. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association 88*, 1273–1283.

Serfling, R. (1980). *Approximation Theorem of Mathematical Sratistics.* Wiley, New York.

Sethuraman, S. (2002). A martingale central limit theorem. Iowa State University.

Simpson, D., R. Carroll, and D. Ruppert (1987). M-estimation for discrete data: asymptotic distribution theory and implications. *Annals of Statistics 15*, 657–669.

Staudte, R. and S. Sheather (1990). *Robust estimation and Testing.* John Wiley and Sons, New York.

Stout, W. F. (1974). *Almost Sure Convergence. Probability and Mathematical Statistics.* New York: Academic Press.

Straumann, D. (2005). *Estimation in conditionally heteroscedastic time series models.* Springer, New York.

Weiss, C. (2009). Modelling time series of counts with over dispersion. *Statistical Methods and Applications 18*, 507–519.

Weiss, C. (2010). The INARCH(1) model for overdispersed time series of counts. *Communications in Statistics Simulation and Computation 39*, 1269–1291.

Yohai, V. (1987). High breakdown piont and high efficiency robust estimates for regression. *Annals of Statistics 15*, 642–658.

Zeger, L. S. (1988). A regression model for time series of counts. *Biometrika 75*, 621–629.

Zheng, H., I. V. Basawa, and S. Datta (2006). Inference for the pth-order random coefficient integervalued process. *Journal of Time Series Analysis Analysis 27*, 411–440.

Zhu, F. and D. Wang (2009). Estimation and testing for a poisson autoregressive model. *Metrika 73*, 211–230.