Charles University in Prague, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

M-ESTIMATIONS IN NONLINEAR REGRESSION FOR LONGITUDINAL DATA

Doctoral Thesis

Mgr. Martina Orsáková

Advisor:
Doc. RNDr. Petr Lachout, CSc.

Prague, 2006

Branch m5 - Econometrics and Operations Research

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky

M-ODHADY V NELINEÁRNÍ REGRESI PRO LONGITUDINÁLNÍ DATA

Disertační práce

Mgr. Martina Orsáková

Školitel:
Doc. RNDr. Petr Lachout, CSc.

Praha, 2006

Obor m5 - Ekonometrie a operační výzkum

Hereby I declare that I elaborated this thesis on my own under the leadership of my advisor and that the references include all the materials I have used.

This thesis is freely available for study purposes.

Prague, 26 May 2006 .....

Mgr. Martina Orsáková

Acknowledgement

First of all, I would like to express my sincere thanks to my advisor, Mr. Petr Lachout, for his leadership throughout my doctoral study, his valuable advice and comments during my research and very careful correction of this manuscript.

Further, I would like to thank to all members of Department of Probability and Mathematical Statistics and members of Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic for their assistance and partnership.

A very special support was given by my family and friends, with love and understanding.

I dedicate this work to my parents.

Prague, 26 May 2006 .....

Mgr. Martina Orsáková

Contents

Introduction

Before attending the doctoral study, I faced together with my advisor question what topic shall I choose for my research. Since I was not specialized in a specific area we decided to select sufficiently general theme. Soon I realized that the topic "Nonlinear Methods in Econometrics" was too broad. So I started to explore a specific area of nonlinear regression models and later I focused on regression models with repeated measurements.

As the terminology used in the literature relating to models with repeated measurements can be sometimes confusing we would like to start by clarifying the most common terms. What does it mean repeated measurements?

One area of statistical methodology dealing with repeated measurements are longitudinal studies. The distinguishing feature of a longitudinal data is that the outcome of interest is measured repeatedly over time in the same subjects, with the general objective of characterizing change in the outcome over time, and studying factors which contribute to the mean level and to the change.

We deal with longitudinal studies where the number of subjects is generally large relative to the number of time points and the objective of the analysis is to characterize the mean response on the i-th subject at the j-th occasion, say $E(Z_i^j)$ as a function of time, as well as covariates thought to influence mean response, the interaction of time and these covariates. We shall refer to models for the mean response as longitudinal regression models, where the regression function is known function with unknown parameters. With repeated values, we can borrow strength across time for the individual as well as across people. Given longitudinal data, we can acknowledge the mutually occurring differences among subjects when estimating a person's current value or predicting his future one. There are several different approaches how to cope with repeated measurements. A simple and often effective strategy is to reduce the repeated values into one or two summaries and analyze each summary variable as a function of covariates. There are three distinct techniques in the literature: marginal models, random effects models and transitions models.

The marginal models are similar to a cross-sectional study, where the marginal mean is modeled. Since repeated values are not likely to be independent, the marginal analysis must also include assumptions about the form of the correlation. For example, in the linear model we can assume $E(Z_i^j) = (\mathbb{X}_i^j)'\beta$, and $Var(Z_i) = V_i(\alpha)$ where β, α must be estimated.

A second approach, the random effects model, assumes that correlation arises among repeated responses because the regression coefficients vary across individuals. Here, it is assumed $E(Z_i^j|\beta_i) = (\mathbb{X}_i^j)'\beta_i$. Because there is too little data on a single subject to estimate β_i from $(Z_i, \mathbb{X}_i)$ alone, it is further assumed that the β_i's are independent realizations from some distribution.

The final approach, the transition model focuses on the conditional expectation of Z_i^j given past outcomes $Z_i^{j-1}, \dots Z_i^1$. Transition models combine the assumptions about the dependence of Z on $\mathbb{X}$ and the correlation among repeated Z's into a single equation. For comprehensive survey see Lindsey [17] or Diggle and others [5].

Until now we were speaking mostly within a paradigm which either treats time of measurements as fixed by the study design or, in an observational setting, assuming that

the measurement times are unrelated to the interest, and from a statistical modeling perspective can therefore be treated as if they were fixed in advance. However there are some cases in which the outcome of interest is the time when a certain event occurs characterized by such data, known under different terms, like point process data, event history data or recurrent event data. We consider an event that can occur at irregularly spaced intervals in time, known as waiting times, such that no two events occur simultaneously. The sequence of events is known as a point process. The number of events in given time intervals, the waiting time between successive events, or, more generally, the intensity of occurrence of events may be of interest. Other variables may be measured each time an event occurs, yielding a marked point process.

In this study we deal with regression model with repeated measurements in time of event occurrence. We consider a transition model that models the conditional mean of the current response given past outcomes. This dissertation does not contain mathematical discoveries. It is an applied work based on theory of counting processes. We have followed closely the approach of Scheike [23] and extended it for M-estimator. Scheike derived a consistent and asymptotic normal estimator for the unknown parameter in nonlinear regression model for longitudinal data with counting process measurement times. He used weighted least square method. The standard least square method is unstable if there are outliers present in the data. Outlying data give an effect so strong in the minimization that the parameters thus estimated are distorted. The M-estimators try to reduce the effect of outliers by replacing the square function by another.

Content of my dissertation is as follows. In order to facilitate further reading, selected definitions and known results in theory of counting processes are described in chapter 1. Formal definitions and theorems are in the appendix. In chapter 2 we define precisely model of nonlinear regression with longitudinal data. Then the assumptions which guarantee the consistency and asymptotic normality of M-estimator are described. Further the properties of M-estimator are shown. Some theoretical results are numerically demonstrated by short simulation study in chapter 3. Also studentized versions of estimators are computed. All simulations were programmed in Splus. The code of simulations is enclosed in the appendix.

Chapter 1

Overview of Known Results

There is a number of papers dealing with the estimators of unknown parameter θ_0 in nonlinear regression model $Z_i(\omega) = m(\theta_0, \mathbb{X}_i(\omega)) + \varepsilon_i(\omega), \;\; i = 1, 2, \dots n$. Rubio and Víšek [22] studied an asymptotic properties of M-estimators in a classical nonlinear regression model with independent and identically distributed observation. They established the conditions according which the $\sqrt{n}$-consistent estimator exists. Their contribution has been derived by generalization the results for linear models which have been proved by Jurečková and Sen in [12].

The paper Scheike [23] deals with more general model $Z_i^k(\omega) = m(\theta_0, \mathbb{X}_i(T_i^k)(\omega)) + \varepsilon_i^k(\omega), \quad k = 1, \dots K_i, \quad i = 1, 2 \dots n$. The model is based on observing n individuals over a certain time period. For the i-th individual there are K_i observations available. The covariates $\mathbb{X}_i(T_i^k)$ may include qualitative as well quantitative variables and also can depend on the events prior to T_i^k, i.e. on the (Z_i^j, T_i^j) for $T_i^j < T_i^k$. A quite general type of dependence among the observation is allowed. Scheike [23] demonstrates the consistency and asymptotic normality of estimator for the parameter θ_0. The underlying true parameter in the regression model is estimated by weighted conditional least squares. This method meshes well with the modelling in terms of the conditional distribution of the current observation given the past. The analysis of the estimator is based on the theory of marked point processes as in e.g. Boel et al. [4], Andersen et al. [2], Last and Brandt [14]. Sheike [23] showed that he optimal choice of the weight function is the inverse of the conditional variance of error term.

Sheike and Zhang [24] proposed a Nadaraya-Watson type estimator of the regression function which is uniformly consistent. In the paper they found the asymptotic distribution of the cumulative regression function and presented a nonparametric test to compare the regression functions for two groups of longitudinal data. They also presented uniformly consistent estimator for the conditional variance of error term that has been used later in [23]. For properties of this estimator see Theorem A.21.

1.1 Time Event Data

This brief introductory chapter has goal to provide background and description of a data set used in this study.

To give a specific example that can be modeled in this framework, let T_i^k be the time at

which the i-th patient arrives for the k-th measurement at a clinic, and let the Z_i^k be the measurements obtained. It is assumed that the measurement times T_i^k can be modeled as a counting process with random intensity. This restricts the model to the continuous time parameter case. The counting process formulation allows censoring schemes for the measurement times T_i^k. It is furthermore, permissible for the time until the next measurement T_i^{k+1}, to depend on the current measurement Z_i^k; e.g. when a measurement indicates that there is a disorder, and thus makes it necessary for a follow-up examination. Another case fitting into the framework presented is a model where the current measurement Z_i^k is a function of the previous measurement and the time since the measurement.

The analysis of data recording the time to an event has origins in the construction of population life tables and, more recently, in the study of chronic diseases and in reliability and life-testing in engineering. As a result, censored data are often referred to generically as survival data or failure time data. Obviously, the event of interest need not be death or failure; the subject could be more accurately called the analysis of censored event time data.

The simplest kind of event history analysis is a clasicall survival analysis, where a collection of individuals are observed from some entry time until a particular event A happens. Often it is impossible to wait for the event to happen for all individuals; for some, it is only known that the event had not yet happened at some specified time and in this case the observation of the time to the occurrence of the event is right-censored. For example this model could be applied in study of probability of default for a client from given segment. The time variable is time since giving loan. Among the possible risk factors screened for significance could be the salary, age of client, type of living, etc. Note that the time event is known only for those clients who defaulted before the end of study. The rest of clients were censored at the duration of the study. See paper Orsáková [20] for further models.

1.2 An Informal Introduction to the Basic Concepts

We should mention the excellent textbook by Fleming and Harrington [6]. These authors provided a comprehensive introduction to the basic mathematical theory needed for the counting process approach. Other very useful textbook is Andersen and others [2] where readers could find a survey of theory and topics developed since the first journal publication of the survival theory in 1978. Some basic preliminaries for marked point processes could be find in Gasparra [7]. Carefully developed context of the martingales on jump processes is given in Boel and others [4]. Last and Brandt [14] gives an introduction the dynamic martingale approach to marked point processes.

The methodology to be discussed later is a relevant to a fairly broad class of event history analysis. Very informal spoken the counting process is a stochastic point process registering random events and counting their number. In most cases, it is studied a collection of individuals, each moving between a finite (usually small) number of states. But also more than a finite number of states can occur.

Formal introduction to the theory of counting processes is given in the appendix.

1.2.1 Counting Processes

A one-dimensional counting process $\{N_t,\ t \geq 0\}$ may be thought of as a stochastic

process recording at any given time t the number of certain events having occurred before time t, see Definition A.11.

The most natural filtrations are histories of stochastic processes $\mathcal{F}_t = \sigma(N_s, 0 \leq s \leq t)$. In this case $\mathcal{F}_t$ "contains the information" generated by the process N on $\langle 0, t\rangle$. Write $\mathcal{F}_{t-}$ correspondingly for the available data just before time at t, see Definition A.1. We also have then that $\mathcal{F}_{t-}$ is generated by $N_s, s < t$. Many results from general theory of stochastic processes rely on completeness and right-continuity of filtrations, see Definition A.2). Since completing a filtration changes no important features of P, we can assume that all filtrations encountered are complete. More care must be taken with right-continuity. Heuristically, for the history of a stochastic process X to be a right-continuous filtration, the new information at time s should be in $\mathcal{F}_s$ and there should be no new information infinitesimally after time s, for every $s \geq 0$. This structure should hold when N is right-continuous and has sample paths which are step functions. We rely on theorem that implies that the history $\{\mathcal{F}_t : t \geq 0\}$ of a counting process N is a right-continuous filtration (see Theorem A.4).

If T is thought of as the time an event occurs, then T will be a stopping time (see Definition A.7) if the information in $\mathcal{F}_T$ specifies whether or not the event has happened by the time T, see Theorem A.5 and Theorem A.16. One can define σ-algebras $\mathcal{F}_T$ and $\mathcal{F}_{T-}$ having the interpretation as being all events which have occurred up to and including the stopping time T or strictly before stopping time. The former, $\mathcal{F}_T$, can be characterized as the σ-algebra generated by T together with all random variables $X(T)$, for any adapted right-continuous with left-limits process X. The process X is called "cadlag" (continuo à droite, limité à gauche) if its sample paths $X(t, \omega), t \geq 0$, for almost all ω, are right-continuous with left-hand limits.

Suppose variable T has a continuous probability distribution with density $f(t)$, where t is a realization of T and the distribution function $F(t)$. The survival function is the probability of an individual surviving beyond time t: $S(t) = 1 - F(t) = P(T \geq t)$. The hazard function is defined as

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t | T \geq t),$$

i.e. in survival analysis $\alpha(t)dt$ is the probability that an individual dies in the small time interval from t to $t + \Delta t$, given that the individual is alive at time t.

A **multivariate counting process** $\{(N_t^1, N_t^2, \ldots, N_t^n),\ t \geq 0\}$ is a stochastic process which can be thought of as registering the occurrences in time of a number of types of disjoint, discrete events, see Definition A.13.

The development in time of a process N is governed by its (random) intensity process $\lambda = [\{\lambda_1(t), ...\lambda_k(t)\},\ \ t \geq 0]$, which is given as follows. Let I_{dt} be a small time interval of length dt around time t, then $\lambda_h(t)dt$ is the conditional probability that N_h jumps in I_{dt} given all that has happened till just before time t. If we let $dN_h(t)$ denote the increment of N_h over I_{dt}, and let $\mathcal{F}_{t-}$ denote everything that has happened up to, but not including t, then we can write

$$\lambda_h(t)dt = P\{dN_h(t) = 1 | \mathcal{F}_{t-}\}. \tag{1.1}$$

Example 1.1. *Survival analysis applied to a fixed term credit portfolio*
Let us suppose that we have a group of n applicants (indexed by i) accepted for a credit with

fixed repayment time. An account is deemed to have failed if three consecutive payments have been missed. All other accounts are considered to be censored since failure has not yet occurred. Thus for each applicant i we observe a failure time $\tilde{T}_i$ which is either his true failure time T_i, i.e. the length of time from granting of credit to the failure, or a censoring time, i.e. the length of repayment. Let $D_i = 1$ if $\tilde{T}_i$ is a true failure time, $D_i = 0$ otherwise. Moreover, we assume that the pairs (T_i, D_i); $i = 1, ...n$ are independent.

We can define a multivariate counting process N by

$$N_i(t) = I(\tilde{T}_i \leq t, D_i = 1), \quad i = 1, ..n, \tag{1.2}$$

where $I(.)$ is the indicator function. Thus $N_i = 0$ before $\tilde{T}_i$ and jumps to 1 at $\tilde{T}_i$ if $\tilde{T}_i$ is a true failure time; otherwise N_i does not jump at all. At any time t we know that either the ith applicant has been observed to failed, or has been censored, or he is still paying off in installments and is uncensored. For the first two cases the conditional probability of observing N_i to jump in the interval I_{dt} is 0. For the latter case this conditional probability is $\alpha_i(t)dt$, where $\alpha_i(t)$ is the hazard function for the true failure time T_i for this applicant.

$$\boxed{\textit{good case}_i} \xrightarrow{\alpha_i} \boxed{\textit{bad case}_i}$$

Thus if we define $Y_i(t) = I(\tilde{T}_i \geq t)$, then we have that

$$P\{dN_i(t) = 1 | \mathcal{F}_{t-}\} = \alpha_i(t) Y_i(t) dt. \tag{1.3}$$

By (1.1) and (1.3) we see that the multivariate counting process $N = (N_1, ...N_n)$, given (1.2), has an intensity process λ with components λ_i given by

$$\lambda_i(t) = \alpha_i(t) Y_i(t), \quad i = 1, ...n. \tag{1.4}$$

In some situations it is reasonable to assume that the intensities α_i are the same for all individuals, so that we have a homogeneous population. Denote $\alpha_i = \alpha$. Then we get an univariate counting process N by aggregating the "individual" counting process (1.2), i.e.

$$N(t) = \sum_{i=1}^{n} N_i(t). \tag{1.5}$$

This process counts the total number of failures in $[0, t]$. By (1.1), (1.4), and the fact that no two individuals fail at the same time, it follows, that (1.5) has intensity process λ given by

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i(t) = \alpha(t) Y(t), \tag{1.6}$$

where $Y(t) = \sum_{i=1}^{n} Y_i(t)$.

□

Sometimes more than a finite number of different types of events are possible though the process counting all of them is still a CP. The different types may even vary continuously. It is now no longer convenient to count each type separately. Rather, one counts over aggregates of types; for instance, measurements in certain intervals or, more generally, in any given Borel set. We speak about the **marked point process**, see Definition A.15.

1.2.2 Stochastic Integration

For Y an increasing process and $X \geq 0$ a real-valued measurable process, the stochastic integral of X with respect to Y is defined as $\int_0^t X(s)dY(s)$. Here, Y is assumed to be a right-continuous process with paths of a bounded variation; i.e. $\int_0^t |dY(s)|$ is finite for all $t \in \langle 0, T\rangle$, for almost all ω. If N is a counting process, f is some possibly random function of time, and $0 \leq s < t \leq \infty$, then $\int_s^t f(u)dN(u)$, is the Stieltjes integral representation of the sum of the values of f at the jump times of N in the interval $(s, t]$. The process combining the counting process with the process of random increments is called the compound counting process, or sometimes the cumulative process. This model is suitable for the description of many technological, environmental, biological and also financial process (especially in the insurance). Description of the compound process with the aid of characteristics of both its components can be found e.g. in Volf [26]. In particular, $\int_s^t XdY$ will always be a random variable, and hence $\{\int_s^t XdY : \ s \leq t < \infty\}$ will be a stochastic process. A proof may be found in Jacobsen [9], page 100.

However, this integral has special and valuable properties when the integrand X is a predictable, bounded process and we integrate with respect to a process Y which is a martingale (see martingale transform theorem A.7).

1.2.3 Martingale Framework

Two kinds of stochastic processes play important and complementary roles in the general theory of processes and, hence, in the theory of stochastic integration and of counting processes: martingales and predictable processes, especially finite variation predictable processes (compensators). Predictable processes arise in two settings.

The first is in the unique Doob-Meyer Decomposition of a right-continuous nonnegative submartingale X into the sum of a right-continuous martingale M and an increasing right-continuous predictable process A (see the theorem A.1). Since any adapted nonnegative increasing process with finite expectation is a submartingale, there is a unique process A, so that for any counting process N with finite expectation, $N-A$ is a martingale. This is summarized in Corollary A.1. The process A in the Doob-Meyer Decomposition is called compensator for the submartingale N. If M is a square integrable martingale, Jensen's inequality implies that M^2 is a submartingale. The Doob-Meyer Decomposition gives that there exists an increasing right-continuous predictable process $\langle M, M\rangle$ such that $M^2 - \langle M, M\rangle$ is a right-continuous martingale, see Corollary A.2. The process $\langle M, M\rangle$ is called the predictable variation process of M. A predictable covariation process $\langle M_1, M_2\rangle$ can be defined similarly, see Theorem A.6, and Corollary A.3. The predictable quadratic variation $\langle M, M\rangle$ for counting process N is also the integrated hazard function if A is continuous and has a finite expectation (i.e. the compensator for N and $(N - A)^2$ are identical when A is continuous), see Theorem A.15.

The second is in the martingale transform $L \equiv \int HdM$ where H is left-continuous and

adapted or, more generally, predictable and $M = N - A$. If H is a bounded predictable process and N is a counting process with finite expectation then the process $\int HdM$ is a martingale, see Theorem A.7.

The martingale decomposition, stochastic integrals and predictable quadratic variation processes were studied under some regularity conditions strong enough to allow straightforward proofs of an important properties. Localization of stochastic processes allows these boundedness conditions to be relaxed. Basic general settings for local martingales you could find in the appendix. Extended version of the Doob-Meyer decomposition for nonnegative local submartingales is formalized in Theorem A.3. Extended Doob-Meyer Decomposition can be used to represent an arbitrary counting process as the sum of a local martingale and a predictable increasing process. When the increasing process is locally bounded the local martingale will be locally square integrable (Theorem A.8). Theorem A.9 says that each counting process has a locally bounded compensator. The predictable variation and covariation processes for martingales with finite second moments can be extended to local square integrable martingales. The existence of the predictable variation process $\langle M, M\rangle$ for a local square integrable martingale is established in Theorem A.10 while the existence of the predictable covariation process for two local square integrable martingales is given in the theorem A.11. Weaker condition that H is locally bounded and $\mathcal{F}_t$-predictable is used and the assumption $EN(t) < \infty$ is removed in the local martingale transform, see Theorem A.12.

There is briefly outlined also generalization of these results for marked point processes in the appendix. MPP version of Theorem A.4 and Theorem A.5 is summarized in Theorem A.16. This theorem says how filtration of MPP looks like and that the history of MPP is a right continuous filtration. Doob-Meyer Decomposition for MPP is stated in Theorem A.17. A predictable covariation process $\langle M(t, B_1), M(t, B_2)\rangle$ is defined in Theorem A.18. The subsection is concluded by martingale transform theorem for MPP A.19.

We continue by section with martingale central limit theorem which will be used in development of large sample properties of our M-estimator. Theorem A.20 says under which assumptions a vector of k local square integrable martingales converges to a continuous Gaussian martingale.

Chapter 2

Own Research and Methods

2.1 Notation, Model and Assumptions

2.1.1 Model Specification

Let $(\Omega, \mathcal{A}, P)$ be a probability space. (In what follows that all $o_p(.)$ and $O_p(.)$ are understood with respect to this P.) We shall consider the nonlinear regression model

$$Z_i^k = m(\theta_0, \mathbb{X}_i(T_i^k)) + \varepsilon_i^k, \quad i = 1, \dots n, \quad k = 1, \dots N_t^i, \tag{2.1}$$

where $m(.)$, the regression function, is a known function of an unknown parameter θ_0. For fixed i we define $N_t^i = \sum_k I(T_i^k \leq t)$ (I is the indicator function), N_t^i counts number of observations for i-th subject in the time interval $[0, t]$. Further $Z_i^k : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ represents responses of the model, $\varepsilon_i^k : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a noise, $T_i^k : (\Omega, \mathcal{A}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ is a random time of observation and $\mathbb{X}_i(s) : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is a process of covariates. Next define $R_t^i = \sum_{k=1}^{N_t^i} Z_i^k$ and a filtration

$$\mathcal{F}_u = \sigma\big(\sigma(R_s^i, N_s^i : s \leq u,\ i = 1, ..n) \cup \mathcal{C}\big),$$

the history of the process R_u^i and N_u^i for $i = 1, ..n$. A σ-algebra $\mathcal{C}$ is assumed to be independent of $\sigma(R_s^i, N_s^i : s \leq u,\ i = 1, ..n)$, and represents knowledge prior to time 0. Note that observing $\{(Z_i^k, T_i^k),\ i = 1, \dots n,\ k = 1, \dots N_u^i\}$ is equivalent to observing the process $\{(R_s^1, N_s^1, \dots R_s^n, N_s^n),\ s \in [0, u]\}$. Denote $\mathcal{F}_{S-} = \sigma\big(\mathcal{F}_u \cup \{T_i^{N_u^i+1} = S\} : u < S\big)$, where S is a time when the last $N_u^i + 1$-th event has occurred and $\mathcal{F}_{S-}$ is a σ-algebra generated by all events strictly before time S.

The M-estimator is defined as

$$\underset{\theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{N_t^i} \rho\left(Z_i^j - m(\theta, \mathbb{X}_i(T_i^j))\right) \right\}, \tag{2.2}$$

where $\rho(.)$ is a real valued continuous function. For the general definition of the M-estimator see Definition A.16. In the case of the differentiability of functions, the mini-

mization problem (2.2) leads to the equations

$$\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi\left(Z_i^j - m(\theta, \mathbb{X}_i(T_i^j))\right)\cdot\left(\frac{\partial}{\partial\theta_k}m(\theta, \mathbb{X}_i(T_i^j))\right) = 0, \quad k = 1, \dots d, \tag{2.3}$$

where $\psi(.) = \rho'(.)$. However, this system may have more roots, while only one of them leads to a global minimum of (2.2). We will prove that there exists at least one root of (2.3) which is a consistent estimator of θ_0.

2.1.2 Assumptions

To facilitate reading we divided the assumptions into several groups.

(A) The noise term have conditional mean and variance given by

$$E(\varepsilon_i^{N_u^i+1}|\mathcal{F}_{S-}) = 0, \tag{2.4}$$

$$E((\varepsilon_i^{N_u^i+1})^2|\mathcal{F}_{S-}) = \sigma^2(\mathbb{X}_i(s)), \tag{2.5}$$

where $\sigma^2(.)$ is deterministic, continuous and bounded function. Let $\mathbb{X}_i(s)$ is predictable process with respect to the filtration $\mathcal{F}_s$ and is "cadlag". The conditional distribution of $\frac{\varepsilon_i^k}{\sigma(\mathbb{X}_i(T_i^k))}$ is denoted

$$F_s(z) = P\left(\frac{\varepsilon_i^{N_u^i+1}}{\sigma(\mathbb{X}_i(T_i^{N_u^i+1}))} \le z \middle| \mathcal{F}_{S-}\right), \tag{2.6}$$

therefore, the conditional distribution of Z_i^k is

$$P(Z_i^{N_u^i+1} \le z|\mathcal{F}_{S-}) = F_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right). \tag{2.7}$$

We shall assume that $E\left(\psi\left(\frac{\varepsilon_i^{N_u^i+1}}{\sigma(\mathbb{X}_i(T_i^{N_u^i+1}))}\right)\middle|\mathcal{H}_s\right) < \infty$.
Without loss of generality we may assume

$$\begin{aligned} E\left(\psi\left(\frac{\varepsilon_i^{N_u^i+1}}{\sigma(\mathbb{X}_i(T_i^{N_u^i+1}))}\right)\middle|\mathcal{H}_s\right) &= \\ \int\psi\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)dF_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right) &= 0. \end{aligned} \tag{2.8}$$

The process $\{(N_s^1, N_s^2, \dots, N_s^n), s \ge 0\}$ is assumed to be n-variate counting process. The process N_s^i has a random intensity $\lambda_s^i \ge 0$, which is "cadlag". One particular form for the intensity is Aalen's multiplicative model: $\lambda_s^i = \alpha(s)Y_s^i$, where $\alpha(.)$ is deterministic function and Y_s^i is $\mathcal{F}_s$-predictable process. In practice for example, censoring indicators play the role of Y_s^i, i.e. the variable is 1 if the i-th subject is at risk and 0 otherwise. Notice that $\mathcal{F}_u$ is according to Theorem A.16 right continuous σ-algebra and time S defined as $T_i^{N_u^i+1} = S$ is a stopping time.

(B) Let $m(\theta, x)$ is three times differentiable in a neighborhood $B(\theta_0, K)$ around parameter θ_0, i.e. $\theta \in B(\theta_0, K) \Leftrightarrow ||\theta - \theta_0|| < K$). Assume that $\psi(.)$ is two times differentiable and $\psi''(.)$ is absolutely continuous function. Define

$$H_i(\theta, s, z) = \rho\left(\frac{z - m(\theta, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right).$$

Assume further

$$E\left(\int_0^t \int [\frac{\partial}{\partial\theta_k} H_i(\theta_0, s, z)]^2 \lambda_s^i dF_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right) ds\right) < \infty, \quad (2.9)$$

$$E\left(\int_0^t \int [\frac{\partial^2}{\partial\theta_k \partial\theta_l} H_i(\theta_0, s, z)]^2 \lambda_s^i dF_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right) ds\right) < \infty \quad (2.10)$$

for all $k, l = 1, \dots d$.

(C) Let us set

$$\begin{aligned}
\gamma_1(\theta_0, \mathbb{X}_i(s)) &:= \int \psi'\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right) dF_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right), \\
\gamma_2(\theta_0, \mathbb{X}_i(s)) &:= \int \left[\psi'\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\right]^2 dF_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right), \\
\gamma_3(\theta_0, \mathbb{X}_i(s)) &:= \int \left[\psi\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\right]^2 dF_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)
\end{aligned}$$

and assume that

$$\gamma_2(\theta_0, \mathbb{X}_i(s)) < \infty, \quad (2.11)$$

$$\gamma_3(\theta_0, \mathbb{X}_i(s)) < \infty. \quad (2.12)$$

Further we assume that there exist non-negative definite symmetric matrices Σ_I, Σ_U such that as $n \to \infty$

$$\left(\frac{1}{n}\sum_{i=1}^n \int_0^t m_k'(\theta_0, \mathbb{X}_i(s)) \cdot m_l'(\theta_0, \mathbb{X}_i(s)) \cdot \frac{\lambda_s^i}{\sigma(\mathbb{X}_i(s))} \cdot \gamma_1(\theta_0, \mathbb{X}_i(s)) ds\right)_{k,l=1}^d \xrightarrow[n\to\infty]{p} \Sigma_I, \quad (2.13)$$

$$\left(\frac{1}{n}\sum_{i=1}^n \int_0^t m_k'(\theta_0, \mathbb{X}_i(s)) \cdot m_l'(\theta_0, \mathbb{X}_i(s)) \cdot \frac{\lambda_s^i}{\sigma^2(\mathbb{X}_i(s))} \cdot \gamma_3(\theta_0, \mathbb{X}_i(s)) ds\right)_{k,l=1}^d \xrightarrow[n\to\infty]{p} \Sigma_U \quad (2.14)$$

and

$$\frac{1}{n^2}\sum_{i=1}^n \int_0^t [m_k'(\theta_0, \mathbb{X}_i(s)) \cdot m_l'(\theta_0, \mathbb{X}_i(s))]^2 \cdot \lambda_s^i \cdot \frac{1}{\sigma^4(\mathbb{X}_i(s))} \cdot \gamma_2(\theta_0, \mathbb{X}_i(s)) ds +$$

$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t [m''_{k,l}(\theta_0,\mathbb{X}_i(s))]^2\cdot\lambda_s^i\cdot\frac{1}{\sigma^2(\mathbb{X}_i(s))}\cdot\gamma_3(\theta_0,\mathbb{X}_i(s))ds\xrightarrow[n\to\infty]{p}0 \qquad (2.15)$$

for all $k,l=1,\dots d$.

(D) Let G is a such function that for $\theta\in B(\theta_0,r)$

$$\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}H_i(\theta,t,z)\right|\le G(t,z),$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}\int_0^t\int G(s,z)\cdot\lambda_s^i\cdot dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds\right)=O_p(1),$$

$$\left(\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\int G^2(s,z)\cdot\lambda_s^i\cdot dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds\right)\xrightarrow[n\to\infty]{p}0.$$

2.2 Properties of M-estimation

2.2.1 Properties of M-estimation for Known $\sigma^2(.)$

An M-estimator is not scale invariant in case of linear regression. For the discussion on the role of rescaling residuals in the nonlinear regression see [22]. Due to the importance of the rescaled residuals, we examine a studentized version of the M-estimator in the model (2.1). Define

$$L_n(\theta,t):=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\rho\left(\frac{Z_i^j-m(\theta,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right). \qquad (2.16)$$

The estimator $\widehat{\theta}_n$ of the parameter θ is given as a solution of the equation:

$$U_n(\theta,t)=0, \qquad (2.17)$$

where

$$\begin{aligned}U_n(\theta,t) &= \left(\frac{\partial}{\partial\theta_k}L_n(\theta,t)\right)_{k=1}^{d}\\ &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi\left(\frac{Z_i^j-m(\theta,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right)\cdot\left(m'_k(\theta,\mathbb{X}_i(T_i^j))\right)_{k=1}^{d}\cdot\frac{1}{\sigma(\mathbb{X}_i(T_i^j))}.\end{aligned}$$

Theorem 2.1. *Assume that the assumptions (A),... (D) are satisfied. Then there exists a consistent solution of (2.17) (a sequence of the solutions of the equation (2.17) ,$\{\widehat{\theta}_n\}$, such that $||\widehat{\theta}_n-\theta_0||=O_p(1)$ as $n\to\infty$) and provides a local minimum of (2.16) with probability tending to one.*

Proof. (The proof follows as in Theorem 1 of Scheike [23].) Start by making a Taylor expansion of the first derivative of $L_n(\theta,t)$ for $\theta \in B(\theta_0,r)$ around true value θ_0

$$\begin{aligned} U_n(\theta,t) &= U_n(\theta_0,t) + d_{\theta-\theta_0}U_n(\theta_0,t) + \frac{1}{2}\cdot d^2_{\theta-\theta_0}U_n(\theta^*,t) \\ &= U_n(\theta_0,t) + (\theta-\theta_0)^T\cdot I_n(\theta_0,t) + \frac{1}{2}\cdot d^2_{\theta-\theta_0}U_n(\theta^*,t) \end{aligned}$$

where $\theta^* \in$ line (θ,θ_0) (line (θ,θ_0) denotes the line segment between θ and θ_0), $I_n(\theta,t) := \left(\frac{\partial^2}{\partial\theta_k\partial\theta_l}L_n(\theta_0,t)\right)^d_{k,l=1}$ and $d^n_h f(x) = \sum\limits^{k_1+\ldots k_d=n}_{k_1\geq 0,\ldots,k_d\geq 0} \frac{n!}{k_1!\ldots k_d!}\frac{\partial^n f(x)}{\partial x_1^{k_1}\ldots\partial x_d^{k_d}}h_1^{k_1}\ldots h_d^{k_d}$.

To prove proposition, the plan is to show that

$$U_n(\theta_0,t) \xrightarrow[n\to\infty]{p} 0, \tag{2.18}$$

$$I_n(\theta_0,t) \xrightarrow[n\to\infty]{p} \Sigma_I, \tag{2.19}$$

$$d^2_{\theta-\theta_0}U_n(\theta^*,t) \xrightarrow[n\to\infty]{p} 0 \tag{2.20}$$

for all $\theta \in B(\theta_0,r)$. Then it follows that $\widehat{\theta}_n$ provides a local minimum with probability tending to one, and is a consistent solution of (2.17).

To show (2.18) we need to establish that the process $U_n(\theta_0,t)$ is a martingale. Note that due to the assumption (2.8) you can write

$$\begin{aligned} & U_n(\theta_0,t) = \\ = & -\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^{N_t^i}\psi\left(\frac{Z_i^j - m(\theta_0,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right)\cdot\frac{1}{\sigma(\mathbb{X}_i(T_i^j))}\cdot m_k'(\theta_0,\mathbb{X}_i(T_i^j))\right)^d_{k=1} = \\ = & -\frac{1}{n}\left(\sum_{i=1}^n\sum_{j=1}^{N_t^i}\psi\left(\frac{Z_i^j - m(\theta_0,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right)\cdot\frac{1}{\sigma(\mathbb{X}_i(T_i^j))}\cdot m_k'(\theta_0,\mathbb{X}_i(T_i^j)) - \right. \\ & \left. -\sum_{i=1}^n\int_0^t\int\psi\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\cdot\frac{m_k'(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\cdot\lambda_s^i dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds\right)^d_{k=1}. \end{aligned}$$

According to Theorem A.19 $U_n(\theta_0,t)$ is a martingale with $E(U_n(\theta_0,t)) = 0$ (define further 0-martingale to be a martingale with mean 0) with predictable quadratic variation

process

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\int\left(\frac{\partial}{\partial\theta_k}H_i(\theta_0,s,z)\right)\cdot\left(\frac{\partial}{\partial\theta_l}H_i(\theta_0,s,z)\right)\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\int\psi^2\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\cdot\frac{m_k'(\theta_0,\mathbb{X}_i(s))\cdot m_l'(\theta_0,\mathbb{X}_i(s))}{\sigma^2(\mathbb{X}_i(s))}\cdot$$

$$\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\gamma_3(\theta_0,\mathbb{X}_i(s))\cdot\frac{m_k'(\theta_0,\mathbb{X}_i(s))\cdot m_l'(\theta_0,\mathbb{X}_i(s))}{\sigma^2(\mathbb{X}_i(s))}\cdot\lambda_s^i ds\xrightarrow[n\to\infty]{p}0,$$

by assumption (C). Then Lenglart's inequality A.2 gives that

$$\sup_{s\le t}|U_n(\theta_0,s)|\xrightarrow[n\to\infty]{p}0 \tag{2.21}$$

and (2.18) follows.
To establish (2.19), write

$$I_n(\theta_0,t) = \frac{1}{n}\sum_{i=1}^{n}\Bigg(M^i\left(\frac{\partial^2}{\partial\theta_k\partial\theta_l}L_n(\theta_0,s)\right)_t + \int_0^t\int\frac{\partial^2}{\partial\theta_k\partial\theta_l}L_n(\theta_0,s)\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds\Bigg)_{k,l=1}^{d} \tag{2.22}$$

where $M^i\left(\frac{\partial^2}{\partial\theta_k\partial\theta_l}L_n(\theta_0,s)\right)_t = \frac{\partial^2}{\partial\theta_k\partial\theta_l}L_n(\theta_0,t) - \int_0^t\int\frac{\partial^2}{\partial\theta_k\partial\theta_l}L_n(\theta_0,s)\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds$.
Note that the first term in (2.22) is a 0-martingale with a compensator that is equal to

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t\int\psi'\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\cdot\frac{m_k'(\theta_0,\mathbb{X}_i(s))\cdot m_l'(\theta_0,\mathbb{X}_i(s))}{\sigma^2(\mathbb{X}_i(s))}\cdot$$

$$\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds -$$

$$-\frac{1}{n}\sum_{i=1}^{n}\int_0^t\int\psi\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\cdot\frac{m_{kl}''(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^t\int\psi'\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\cdot\frac{m_k'(\theta_0,\mathbb{X}_i(s))\cdot m_l'(\theta_0,\mathbb{X}_i(s))}{\sigma^2(\mathbb{X}_i(s))}$$

$$\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds$$

and quadratic predictable variation process that is given as

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\int\left[\psi'\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\right]^2\cdot\frac{[m'_k(\theta_0,\mathbb{X}_i(s))\cdot m'_l(\theta_0,\mathbb{X}_i(s))]^2}{\sigma^4(\mathbb{X}_i(s))}\cdot$$
$$\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds+$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\int\left[\psi\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)\right]^2\cdot\frac{[m''_{kl}(\theta_0,\mathbb{X}_i(s))]^2}{\sigma^2(\mathbb{X}_i(s))}\cdot$$
$$\cdot\lambda_s^i\ dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds=$$
$$\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\gamma_2(\theta_0,\mathbb{X}_i(s))\cdot\frac{1}{\sigma^4(\mathbb{X}_i(s))}\cdot[m'_k(\theta_0,\mathbb{X}_i(s))\cdot m'_l(\theta_0,\mathbb{X}_i(s))]^2\cdot\lambda_s^i\ ds+$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\gamma_3(\theta_0,\mathbb{X}_i(s))\cdot\frac{1}{\sigma^2(\mathbb{X}_i(s))}\cdot[m''_{kl}(\theta_0,\mathbb{X}_i(s))]^2\cdot\lambda_s^i\ ds\xrightarrow[n\to\infty]{p}0$$

by assumption (C). So by Lenglarts's inequality A.2 the first term of (2.22) converges in probability to zero. The second term is equal to

$$\left(\frac{1}{n}\sum_{i=1}^{n}\int_0^t\gamma_1(\theta_0,\mathbb{X}_i(s))\cdot\frac{1}{\sigma^2(\mathbb{X}_i(s))}\cdot m'_k(\theta_0,\mathbb{X}_i(s))\cdot m'_l(\theta_0,\mathbb{X}_i(s))\cdot\lambda_s^i\ ds\right)_{k,l=1}^{d}\xrightarrow[n\to\infty]{p}\Sigma_I.$$

Finally consider the last proposition (2.20)

$$d^2_{\theta-\theta_0}U_n(\theta^*,t)\xrightarrow[n\to\infty]{p}0.$$

$$\frac{1}{2}d^2_{\theta-\theta_0}U_n(\theta^*,t)=\sum_{k_1\geq0,\ \ldots k_d\geq0}^{k_1+\ \ldots k_d=2}\frac{1}{k_1!\ldots k_d!}\frac{\partial^2U_n(\theta^*,t)}{\partial\theta_1^{k_1}\ldots\partial\theta_d^{k_d}}(\theta_1-\theta_{01})^{k_1}\ldots(\theta_d-\theta_{0d})^{k_d}.$$

The assumption (D) gives that

$$\left|\frac{\partial^2U_{nl}(\theta^*,t)}{\partial\theta_j\theta_k}\right|:=\left|\frac{\partial^3L_n(\theta^*,t)}{\partial\theta_j\theta_k\theta_l}\right|\leq\frac{1}{n}\sum_{i=1}^{n}\int_0^t\int G(s,z)N^i(ds\times dz).$$

Now, again as above, one gets:

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t\int G(s,z)N^i(ds\times dz)-\frac{1}{n}\sum_{i=1}^{n}\int_0^t\int G(s,z)\lambda_s^i dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds$$

is a 0-martingale with predictable quadratic variation process that is equal to

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\int G^2(s,z)\lambda_s^i dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds$$

and according to assumption (D) converges in probability to zero. Finally Lenglart's inequality gives that

$$\sup_{u\le t}\left(\frac{1}{n}\sum_{i=1}^{n}\int_0^u\int G(s,z)N^i(ds\times dz)\right)\xrightarrow[n\to\infty]{p}0.$$

□

The next theorem gives asymptotic normality for a consistent solution of the equations (2.17).

Theorem 2.2. *Under the assumptions (A),... (D) and with $\widehat{\theta}_n$ consistent solution of (2.17), one has*

$$\sqrt{n}\,(\widehat{\theta}_n-\theta_0)\xrightarrow[n\to\infty]{\mathcal{D}}N(0,\Sigma),$$

where $\Sigma=\Sigma_I^{-1}\Sigma_U\Sigma_I^{-1}$.

Further $\hat{\Sigma}_U:=\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi^2\left(\frac{Z_i^j-m(\widehat{\theta}_n,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right)\cdot\frac{m_k'(\widehat{\theta}_n,\mathbb{X}_i(T_i^j))\cdot m_l'(\widehat{\theta}_n,\mathbb{X}_i(T_i^j))}{\sigma^2(\mathbb{X}_i(T_i^j))}\right)_{k,l=1}^{d}$ and $I(\widehat{\theta}_n,t)$ provide consistent estimates of Σ_U and Σ_I respectively.

Proof. Making a Taylor expansion around the true parameter θ_0 one gets

$$U_n(\widehat{\theta}_n,t)=U_n(\theta_0,t)+(\widehat{\theta}_n-\theta_0)^T\cdot I_n(\theta^*,t)$$

where $\theta^*\in$ line $(\widehat{\theta}_n,\theta_0)$, so since $\widehat{\theta}_n$ is a solution, this equation states that

$$-I_n(\theta^*,t)^{-1}\cdot U_n(\theta_0,t)=\widehat{\theta}_n-\theta_0.$$

Theorem follows if one can show

$$\sqrt{n}\,U_n(\theta_0,t)\xrightarrow[n\to\infty]{\mathcal{D}}N(0,\Sigma_U),\tag{2.23}$$

$$I_n(\theta^*,t)\xrightarrow[n\to\infty]{p}\Sigma_I\ \text{ for }\widehat{\theta}_n\xrightarrow[n\to\infty]{p}\theta_0.\tag{2.24}$$

The first condition (2.23) follows by the martingale convergence theorem, see Theorem A.20. The quadratic predictable variation process of $\sqrt{n}U_n(\theta_0,t)$ is equal to

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t\gamma_3(\theta_0,\mathbb{X}_i(s))\cdot\frac{m_k'(\theta_0,\mathbb{X}_i(s))\cdot m_l'(\theta_0,\mathbb{X}_i(s))}{\sigma^2(\mathbb{X}_i(s))}\cdot\lambda_s^i ds$$

which converges to Σ_U according to the assumption (2.14). Then together with the result (2.21) one gets that

$$\langle\sqrt{n}U_n(\theta_0,t)I\{|U_n(\theta_0,t)|\ge\varepsilon\}\rangle\xrightarrow[n\to\infty]{p}0.$$

Finally, condition (2.24) follows from the Taylor expansion (similarly as in the last part of previous proof). □

Further follows result which can be used for testing a simple hypothesis H: $\theta = \theta_0$ against the composite hypothesis G: $\theta \in \Theta$. Let $W_p(\Sigma)$ denote the Wishart distribution corresponding to a p-dimensional normal distribution $N_p(0, \Sigma)$. Define $R_n = n(L_n(\widehat{\theta}_n, t) - L_n(\theta, t))$ one has following theorem.

Theorem 2.3. *Under the hypothesis H and under the assumptions of Theorem 2.2 one has*

$$R_n \xrightarrow[n\to\infty]{\mathcal{D}} W_p(1/2\Sigma_I^{-1/2}\Sigma_U\Sigma_I^{-1/2}).$$

Proof. The proof can be again carry out by repeating the steps of the proof of Theorem 3 of [23]. □

2.2.2 Properties of M-estimation for Unknown $\sigma^2(.)$

The conditional variance $\sigma^2(.)$ is usually unknown. The natural choice is replacing $\sigma^2(.)$ by its estimator $\widehat{\sigma_n^2}(.)$. To be able to apply asymptotic properties of the M-statistics on the estimator $\widehat{\theta}_n$, we need to know something about the behaviour of the sum

$$\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi\left(\frac{Z_i^j - m(\theta, \mathbb{X}_i(T_i^j))}{\widehat{\sigma_n}(\mathbb{X}_i(T_i^j))}\right)\cdot\left(m_k'(\theta, \mathbb{X}_i(T_i^j))\right)_{k=1}^{d}\cdot\frac{1}{\widehat{\sigma_n}(\mathbb{X}_i(T_i^j))}. \tag{2.25}$$

It seems to be obviously if the derivative of the function ρ exists and is sufficiently smooth that for consistent estimator of $\sigma^2(.)$ we can rely on properties proved in the previous section. Indeed under some additional assumptions the theorems will be still valid. For the non-parametric estimation of conditional variance which is uniformly consistent, see Theorem A.21 in the appendix.

Before stating the theorem, let us redefining of the notation to make the dependency on σ more explicit. Redefine $L_n(\theta, t)$ to $L_n(\sigma, \theta, t)$, $U_n(\theta, t)$ to $U_n(\sigma, \theta, t)$, $I_n(\theta, t)$ to $I_n(\sigma, \theta, t)$ and $R(\theta, t)$ to $R(\sigma, \theta, t)$.

Theorem 2.4. *Under the assumptions*

(i) *(A),*

(ii) $\mathbb{X}_i(s)$ *belongs to some compact set* C *almost surely for all* s*,*

(iii) $\widehat{\sigma_n^2}(.)$ *is a uniformly consistent estimator of* $\sigma^2(.)$*, i.e.* $\sup_{x\in C}|\widehat{\sigma_n^2}(x) - \sigma^2(x)| \xrightarrow[n\to\infty]{p} 0$*, and further there exists* $\varepsilon > 0$ *such that* $\widehat{\sigma_n}(x) > \varepsilon$ *for all* $x \in C$*,*

(iv) $m(\theta, x)$ *is three times differentiable in a neighborhood* $B(\theta_0, K)$ *around the parameter* θ_0 *and* $\psi(.)$ *is two times differentiable and* $\psi''(.)$ *is absolutely continuous function,*

(v) $\gamma_4(\theta_0, \mathbb{X}_i(s)) := \int\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)^4 dF_s\left(\frac{z - m(\theta_0, \mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right) < \infty$,

(vi) $E\frac{1}{n}\sum\int_0^t \lambda_s^i ds = O_p(1)$

there exists a consistent solution of $U(\widehat{\sigma_n}, \theta, t) = 0$*, that provides a local minimum of* $L(\widehat{\sigma_n}, \theta, t)$*.*

Proof. The proof proceeds similarly as in Theorem 2.1. To prove the proposition, we need to show that

$$(U_n(\widehat{\sigma_n},\theta_0,t) - U_n(\sigma,\theta_0,t)) \xrightarrow[n\to\infty]{p} 0, \tag{2.26}$$

$$(I_n(\widehat{\sigma_n},\theta,t) - I_n(\sigma,\theta_0,t)) \xrightarrow[n\to\infty]{p} 0, \tag{2.27}$$

$$\left(d^2_{\theta-\theta_0}U_n(\widehat{\sigma_n},\theta^*,t) - d^2_{\theta-\theta_0}U_n(\sigma,\theta^*,t)\right) \xrightarrow[n\to\infty]{p} 0, \tag{2.28}$$

where $\theta^* \in \text{line}\ (\theta,\theta_0)$.

To show (2.26), start by making a Taylor expansion of the $\psi(\widehat{\sigma_n},\theta_0,s)$ for all $\widehat{\sigma_n}(\mathbb{X}_i(s))$ satisfying the assumption (iii) around true value $\sigma(\mathbb{X}_i(s))$:

$$\psi(\widehat{\sigma_n},\theta_0,s) = \psi(\sigma,\theta_0,s) - \psi'(\sigma_*,\theta_0,s)\cdot\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma^2_*(\mathbb{X}_i(s))}\cdot(\widehat{\sigma_n}(\mathbb{X}_i(s)) - \sigma(\mathbb{X}_i(s))) \tag{2.29}$$

where $\sigma_*(x) \in \text{line}\ (\widehat{\sigma_n}(x),\sigma(x))$ for all $x \in C$. Let us apply (2.29) in the formulae

$$U_n(\widehat{\sigma_n},\theta_0,t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi\left(\frac{Z_i^j - m(\theta_0,\mathbb{X}_i(T_i^j))}{\widehat{\sigma_n}(\mathbb{X}_i(T_i^j))}\right)\cdot\left(m'_k(\theta_0,\mathbb{X}_i(T_i^j))\right)_{k=1}^{d}\cdot\frac{1}{\widehat{\sigma_n}(\mathbb{X}_i(T_i^j))}$$

and study both summands in the difference $\Delta U_n(\theta_0,t) := U_n(\widehat{\sigma_n},\theta_0,t) - U_n(\sigma,\theta_0,t)$ separately. Define the k-th component in the $\Delta U_n(\theta_0,t)$ as $\Delta U_n(k,\theta_0,t)$ and define the first summand in the $\Delta U_n(k,\theta_0,t)$ as $\Delta U1_n(k,\theta_0,t)$. Then

$$\Delta U1_n(k,\theta_0,t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi(\sigma,\theta_0,T_i^j)\cdot m'_k(\theta,\mathbb{X}_i(T_i^j))\cdot\frac{1}{\sigma(\mathbb{X}_i(T_i^j))}\cdot\left(\frac{\sigma(\mathbb{X}_i(T_i^j))}{\widehat{\sigma_n}(\mathbb{X}_i(T_i^j))} - 1\right).$$

Since the $\widehat{\sigma^2_n}(x)$ is a consistent estimator of $\sigma^2(x)$, and $\widehat{\sigma_n}(x) > \varepsilon > 0$ it holds that $\sup_{x\in C}\left|\frac{\sigma(x)}{\widehat{\sigma_n}(x)}\right| \xrightarrow[n\to\infty]{p} 1$ and there exists $K > 0$ such that $\sup_{x\in C}\left|\frac{\sigma(x)}{\widehat{\sigma_n}(x)} - 1\right| < K$. It follows that $|\Delta U1_n(k,\theta_0,t)| \le K|U_n(k,\theta_0,t)|$ and similarly as in (2.21) we get $\Delta U1_n(k,\theta_0,t) \xrightarrow[n\to\infty]{p} 0$.

Let us consider the second summand of $\Delta U_n(k,\theta_0,t)$:

$$\Delta U2_n(k,\theta_0,t) =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi'(\sigma_*,\theta_0,T_i^j)\cdot\frac{Z_i^j - m(\theta_0,\mathbb{X}_i(T_i^j))}{\sigma^2_*(\mathbb{X}_i(T_i^j))}\cdot m'_k(\theta_0,\mathbb{X}_i(T_i^j))\cdot\frac{\widehat{\sigma_n}(\mathbb{X}_i(T_i^j)) - \sigma(\mathbb{X}_i(T_i^j))}{\widehat{\sigma_n}(\mathbb{X}_i(T_i^j))}.$$

Since $\sigma(x)$, $\sigma_*(x)$, $\widehat{\sigma_n}(x)$ and $m'_k(\theta_0,x)$ are bounded functions and $\psi'(.)$ is absolutely continuous function we can write

$$|\Delta U2_n(k,\theta_0,t)| \le \frac{K}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\left|\frac{Z_i^j - m(\theta_0,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right|\cdot\left|\widehat{\sigma_n}(\mathbb{X}_i(T_i^j)) - \sigma(\mathbb{X}_i(T_i^j))\right|.$$

We need to show that $|\Delta U2_n(k,\theta_0,t)| \xrightarrow[n\to\infty]{p} 0$. Since the $\widehat{\sigma_n^2}(x)$ is a consistent estimator of $\sigma^2(x)$ it is sufficient to prove that $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\left|\frac{Z_i^j - m(\theta_0,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right| < \infty$. Write

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\left|\frac{Z_i^j - m(\theta_0,\mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right| = \frac{1}{n}\sum_{i=1}^{n}\left(M^i\left(\left|\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right|\right)_t + \right. \tag{2.30}$$
$$\left. + \int_0^t\int\left|\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right|\lambda_s^i dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds\right)$$

and note that due to the assumptions (v) and (vi) the first term is a 0-martingale with a compensator that is equal to the second term. The quadratic predictable variation process of the martingale is given as

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\int\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)^2\lambda_s^i dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds = \frac{1}{n^2}\sum_{i=1}^{n}\int_0^t\lambda_s^i ds \xrightarrow[n\to\infty]{p} 0.$$

So by Lenglarts's inequality the first term of (2.30) converges in probability to zero. The second term is equal to

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t\int\left|\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right|\lambda_s^i dF_s\left(\frac{z-m(\theta_0,\mathbb{X}_i(s))}{\sigma(\mathbb{X}_i(s))}\right)ds \le \frac{1}{n}\sum_{i=1}^{n}\int_0^t\lambda_s^i ds = O_p(1).$$

Now it is easy to show (2.27) and (2.28). Since we would just repeat the corresponding technique as above we omit the proof. □

Theorem 2.5. *Under the assumption of Theorem 2.4, theorems 2.2 and 2.3 remain true if $\sigma^2(.)$ is replaced by its uniformly consistent estimator $\widehat{\sigma_n^2}(.)$.*

Proof. The proof follows similarly as in previous theorems. □

2.3 The Discussion of the Assumptions

Let us look at the assumptions more closely. The most strict limitation was given on the form of error penalty function $\rho(.)$ in the M-estimation in (2.2). Since we used Taylor's expansion for proving characteristics of estimator we required $\rho(.)$, $\psi(.)$, $\psi'(.)$,$\psi''(.)$ to be absolutely continuous functions. For example ordinary least square estimator meets these requirements but we are seeking for function which is less increasing than square. Since the influence function of an M-estimate is proportional to $\psi(x)$, the function $\psi(x)$ (roughly speaking) measures the influence of a datum on the value of the parameter estimation. For the least square estimator with $\rho(x) = \frac{x^2}{2}$, the influence function is $\psi(x) = x$, that is, the influence of a datum on the estimation increases linearly with the size of its error, which confirms the non-robustness of the least square estimator. Although the set of sufficiently smooth functions of $\psi(.)$ is limited, we still can use various functions such

as Geman and McClure function, Hebert and Leahy function, $L_1 - L_2$ function, Cauchy function or Welsch function. For the definition of the most commonly used M-estimators see Table A.1 in the appendix.

Area for further research can be to enhance a set of minimization functions ψ, that fulfill the equations

$$\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\psi\left(\frac{Z_i^j - m(\theta, \mathbb{X}_i(T_i^j))}{\sigma(\mathbb{X}_i(T_i^j))}\right)\cdot\left(m_k'(\theta, \mathbb{X}_i(T_i^j))\right)_{k=1}^{d}\cdot\frac{1}{\sigma(\mathbb{X}_i(T_i^j))} = 0.$$

The results should be extended also for non-smooth penalty functions and further research can be inspired with the results and methods of Rubio and Víšek [22].

Chapter 3

Numerical Simulations and Examples

3.1 Numerical Simulations

3.1.1 Simulation Study 1

Let the covariate $\mathbb{X}_i(T_i^j) = T_i^j$ be the random observation time. The regression function has the form $m(t) = 3 + t^5$. In the simulation study ε_i^j is Gaussian white noise with mean 0 and a standard deviation 0.1. The observation times for each individual were generated from the Poisson process over the period $\langle 0, 1\rangle$ with parameter λ, where $\lambda = 15$ if the previous response value $Z_i^{j-1} \leq 3$ and $\lambda = 3$ otherwise. It means, that a short follow-up time was set up for small values of response variable. To observe an influence of outlying data we changed the generation of every fifth observation. We raised mean of Gaussian white noise to the value of 1 and standard deviation remained unchanged.

We generated data for 20 individuals. On the average there were about 5 measurements for each individual. Together we generated 99 measurements (out of that 11 measurements were generated as outlying). We estimated unknown parameters by methods of ordinary least squares (L_2), Geman-Mc Clure, Hebert and Leahy, $L_2 - L_1$ and L_1. Although L_1 does not meet assumptions (B), we applied L_1 as well to compare also "non-smooth" estimator. For the results see Table 3.1. As you can see, there is significant difference among estimators. All M-estimators gave better results than ordinary L_2 estimator. The estimations of standard errors of estimated parameters are given in Table 3.2. We did not compute standard errors for L_1, since L_1 does not meet assumptions (B), we can not apply Theorem 2.2.

Further we computed also studentized versions of M-estimators. We used Scheike's estimator $\widehat{V}(.)$ of conditional variance $\sigma^2(.)$, see Theorem A.21. We chose a uniform kernel function with bandwidth $b = 0.08$. For the results see Table 3.3, the estimations of standard errors and correlation of estimated parameters are given in Table 3.4. Figure 3.1 shows the plots of raw data. Symbol $\times$ stands for outliers. The red line is true regression function, green line is least square estimate of the regression function, orange line is Geman-Mc Clure estimation, grey line is Hebert and Leahy estimation, blue line is $L_2 - L_1$ estimation and finally pink line is L_1 estimation. Non-parametric estimation of the regression function is depicted by black line. In case of studentized version of estimators, all M-estimators

gave better results than studentized L_2 estimator in this limited simulation study. Notice also that studentized version of L_2 estimation did not reduce the influence of outlying measurements.

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$	L_1
$a = 3$	3.044980	3.035793	3.038786	3.042684	3.030482
$b = 5$	2.914834	4.051153	3.442478	3.047699	4.568847

Table 3.1: M-estimations of parameters in $Z_i^j = a + t^b + \varepsilon_i^j$

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$
$std(\hat{a})$	0.02617415	0.0158056	0.01998707	0.02415176
$std(\hat{b})$	0.59239153	0.6058654	0.64576934	0.61151856
$corr(\hat{a}, \hat{b})$	0.57447129	0.5277010	0.55122972	0.56319328

Table 3.2: Standard errors of estimated parameters in $Z_i^j = a + t^b + \varepsilon_i^j$

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$	L_1
$a = 3$	3.033489	3.026034	3.027146	3.030077	3.011145
$b = 5$	2.116830	4.292186	2.846265	2.262145	5.023877

Table 3.3: Studentized M-estimations of parameters in $Z_i^j = a + t^b + \varepsilon_i^j$

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$
$std(\hat{a})$	0.01548364	0.02635398	0.01833361	0.0158429
$std(\hat{b})$	0.73807954	0.58523977	0.58495028	0.7558261
$corr(\hat{a}, \hat{b})$	0.42277296	0.81915634	0.63817315	0.4733446

Table 3.4: Standard errors of estimated parameters in $Z_i^j = a + t^b + \varepsilon_i^j$

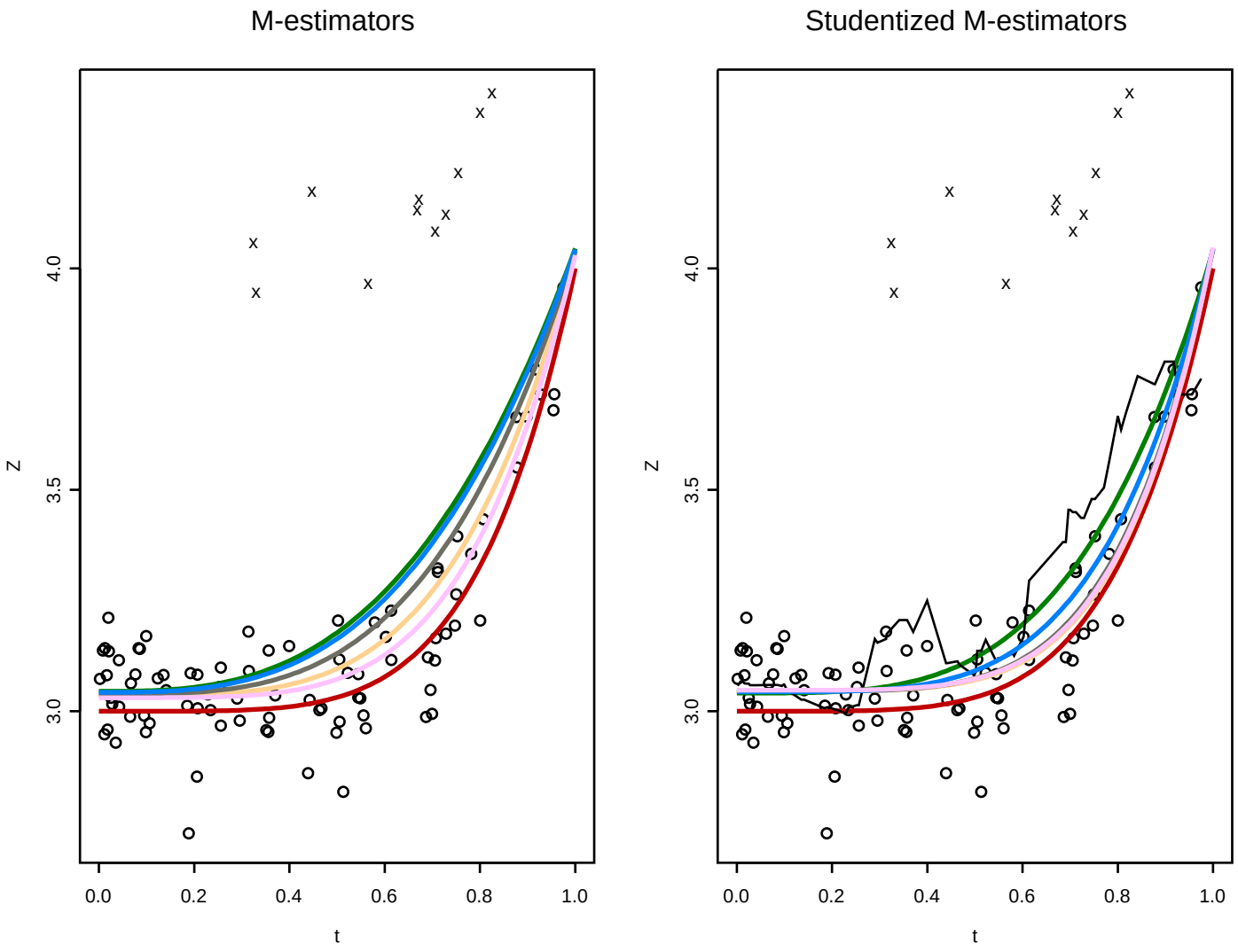

Figure 3.1: $Z_i^j = a + t^b + \varepsilon_i^j$, $a = 3$, $b = 5$

3.1.2 Simulation Study 2

In this study we examined model with two covariates $\mathbb{X}_i(T_i^j) = (T_i^j, X_i^j)$. We observed the regression function in the form $m(t, x) = 1 + 2 \cdot e^{-3 \cdot x} + 0.5 \cdot \ln t$. In the simulation study ε_i^j is Gaussian white noise with mean 0 and a standard deviation 0.6. The observation times for each individual were generated from the Poisson process over the period $\langle 0, 1 \rangle$ with parameter λ, where $\lambda = 25$ if the previous response value $Z_i^{j-1} \leq 0$ and $\lambda = 10$ otherwise. The covariate process X_i^j was generated from univariate distribution on $\langle 0, 1 \rangle$. To observe an influence of outlying data we changed the generation of every first observation. We raised mean of Gaussian white noise to the value of 10 and standard deviation remained unchanged.

We generated data for 20 individuals. On the average there were about 11 measurements for each individual. Together we generated 216 measurements (out of that 20 measurements were generated as outlying). Figure 3.2 shows the raw data. Symbol $\times$ stands for outliers. The surface indicates the true regression function. To improve the depth cues in a 3D plot, we added drop-line to each of the generated points.

As before we estimated unknown parameters by methods of ordinary least squares (L_2), Geman-Mc Clure, Hebert and Leahy, $L_2 - L_1$ and L_1. For the results see Table 3.5. As you can see, ordinary L_2 estimator totally failed while other M-estimators were quite successful. The estimations of standard errors of estimated parameters are given in Table 3.6. Figure 3.3 shows results of L_2 estimation and Figure 3.4 Geman-Mc Clure estimation.

Further we computed also studentized versions of M-estimators. This time we chose a uniform kernel function with bandwidth $b = 0.15$. For the results see Table 3.7, the estimations of standard errors and correlation of estimated parameters are given in Table 3.8. Figure 3.5 shows generated data and kernel estimation of regression function. Studentized L_2 estimation of regression function is depicted in Figure 3.6 while studentized Geman-Mc Clure estimation is in Figure 3.7. The difference between studentized L_2 estimator and other studentized M-estimators is not as evident as in Simulation study 1. Other M-estimators still produce better results then L_2 estimator.

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$	L_1
$a = 1$	-0.1456128	1.0036151	0.9315337	0.04063408	0.4664059
$b = 2$	2.2529190	1.8626411	1.8657871	2.37611516	2.2270274
$c = 3$	3.2322148	3.0246388	3.1449683	3.09096237	3.0439006
$d = 0.5$	-1.3089531	0.4469881	0.3268518	-0.91402020	-0.3335623

Table 3.5: M-estimations of parameters in $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$
$std(\hat{a})$	0.356560194	0.26513101	0.19442881	0.40799412
$std(\hat{b})$	0.546932800	0.27232437	0.24461602	0.43791624
$std(\hat{c})$	1.657271186	1.54747311	1.23880172	1.21545393
$std(\hat{d})$	0.167258963	0.08051531	0.06969169	0.40314785
$corr(\hat{a}, \hat{b})$	-0.443320592	-0.05598768	-0.08555453	-0.55121394
$corr(\hat{a}, \hat{c})$	0.861153950	0.92433949	0.88093121	0.68277720
$corr(\hat{a}, \hat{d})$	0.198247991	0.38596980	0.26760740	0.75805482
$corr(\hat{b}, \hat{c})$	-0.068027311	0.20531533	0.25323151	0.01176079
$corr(\hat{b}, \hat{d})$	0.003991285	1.20829642	-0.12261235	0.42621921

Table 3.6: Standard errors of estimated parameters in $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$	L_1
$a = 1$	0.81637031	0.9146266	0.8804381	0.846989233	0.8974326
$b = 2$	1.74800463	1.9290291	1.8765349	1.805459099	1.8849015
$c = 3$	3.26698844	2.6427983	3.0591629	3.276621757	2.6383915
$d = 0.5$	-0.07355174	0.3190015	0.1354888	-0.003984783	0.2343083

Table 3.7: Studentized M-estimations of parameters in $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

Estimator	L_2	Geman-Mc Clure	Hebert and Leahy	$L_2 - L_1$
$std(\hat{a})$	0.13063111	0.3940867	0.21940854	0.1500206
$std(\hat{b})$	0.17829043	0.3264160	0.20270702	0.1763022
$std(\hat{c})$	0.82663774	1.0363719	0.97084349	0.8812946
$std(\hat{d})$	0.13146290	0.1825487	0.15837683	0.1440156
$corr(\hat{a}, \hat{b})$	-0.34619481	-0.9275856	-0.76718507	-0.4818247
$corr(\hat{a}, \hat{c})$	0.73715940	0.9355505	0.83081588	0.7542663
$corr(\hat{a}, \hat{d})$	0.27453797	0.7085942	0.40201587	0.3057476
$corr(\hat{b}, \hat{c})$	0.04575273	-0.8398843	-0.55673391	-0.1252128
$corr(\hat{b}, \hat{d})$	-1.00258153	1.7794665	-0.06351819	-0.9830419

Table 3.8: Standard errors of estimated parameters in $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

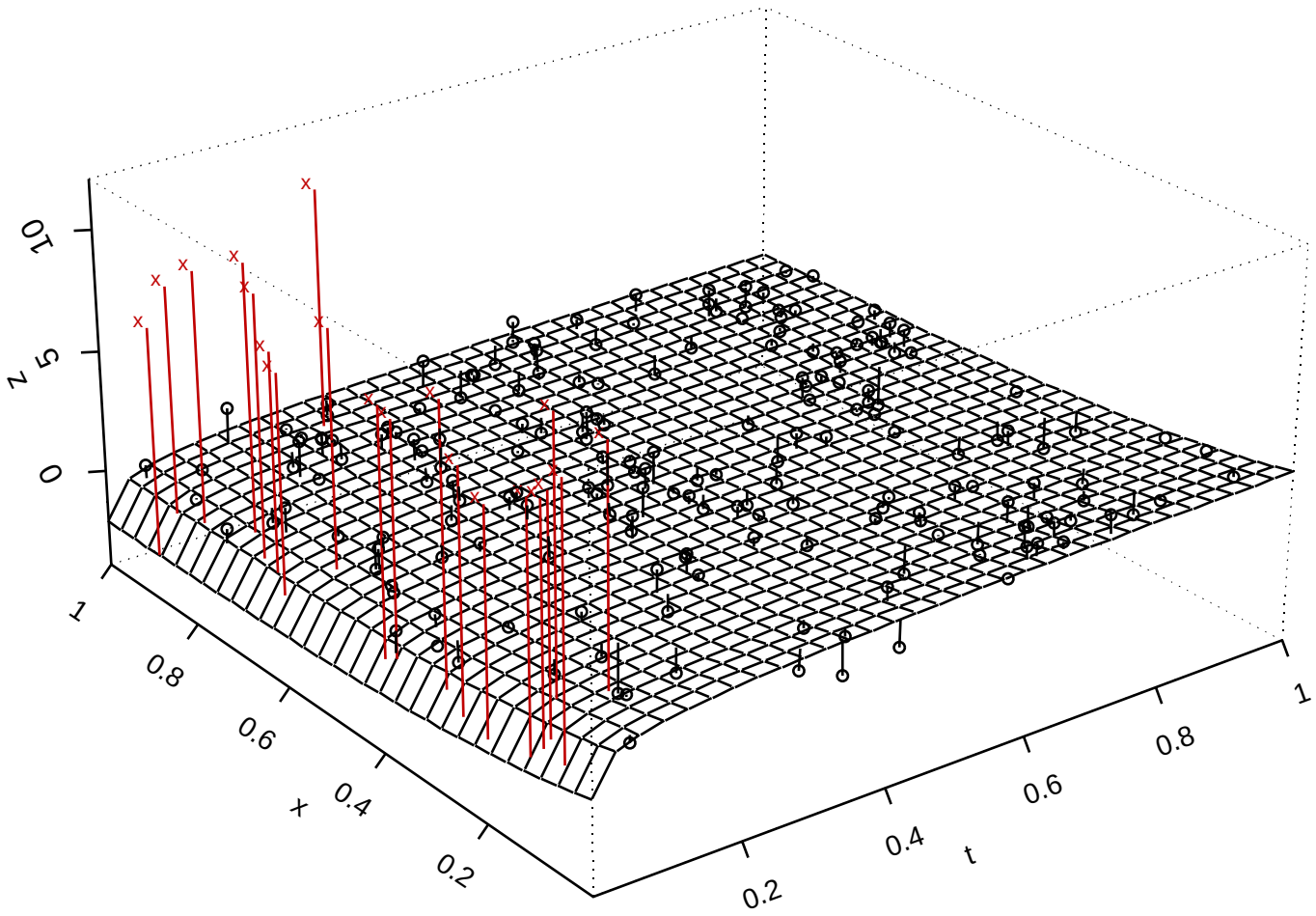

Figure 3.2: Generated data with true regression function $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$, $a = 1$, $b = 1$, $c = 3$, $d = 0.5$

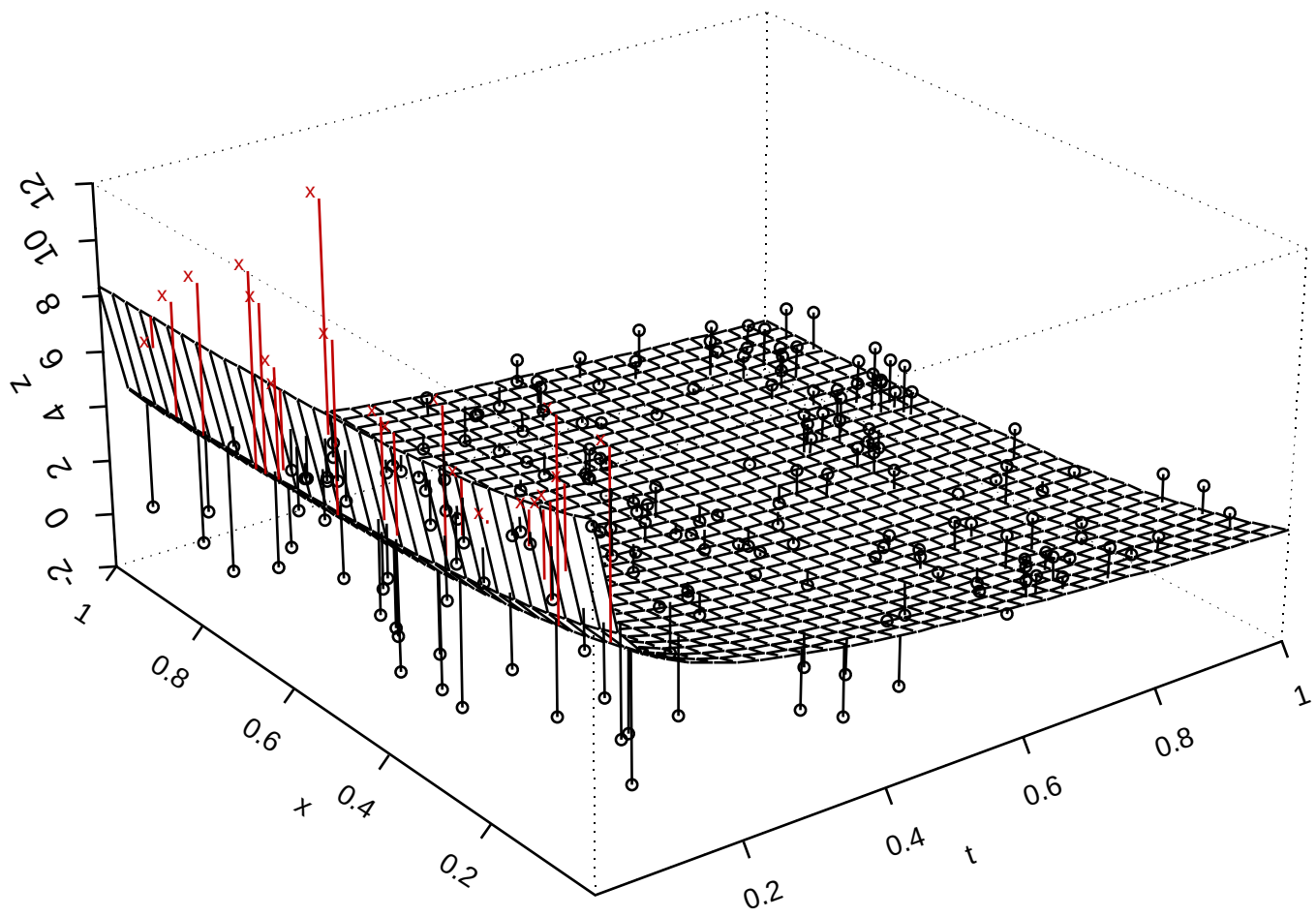

Figure 3.3: Ordinary L_2 estimation of the regression function $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

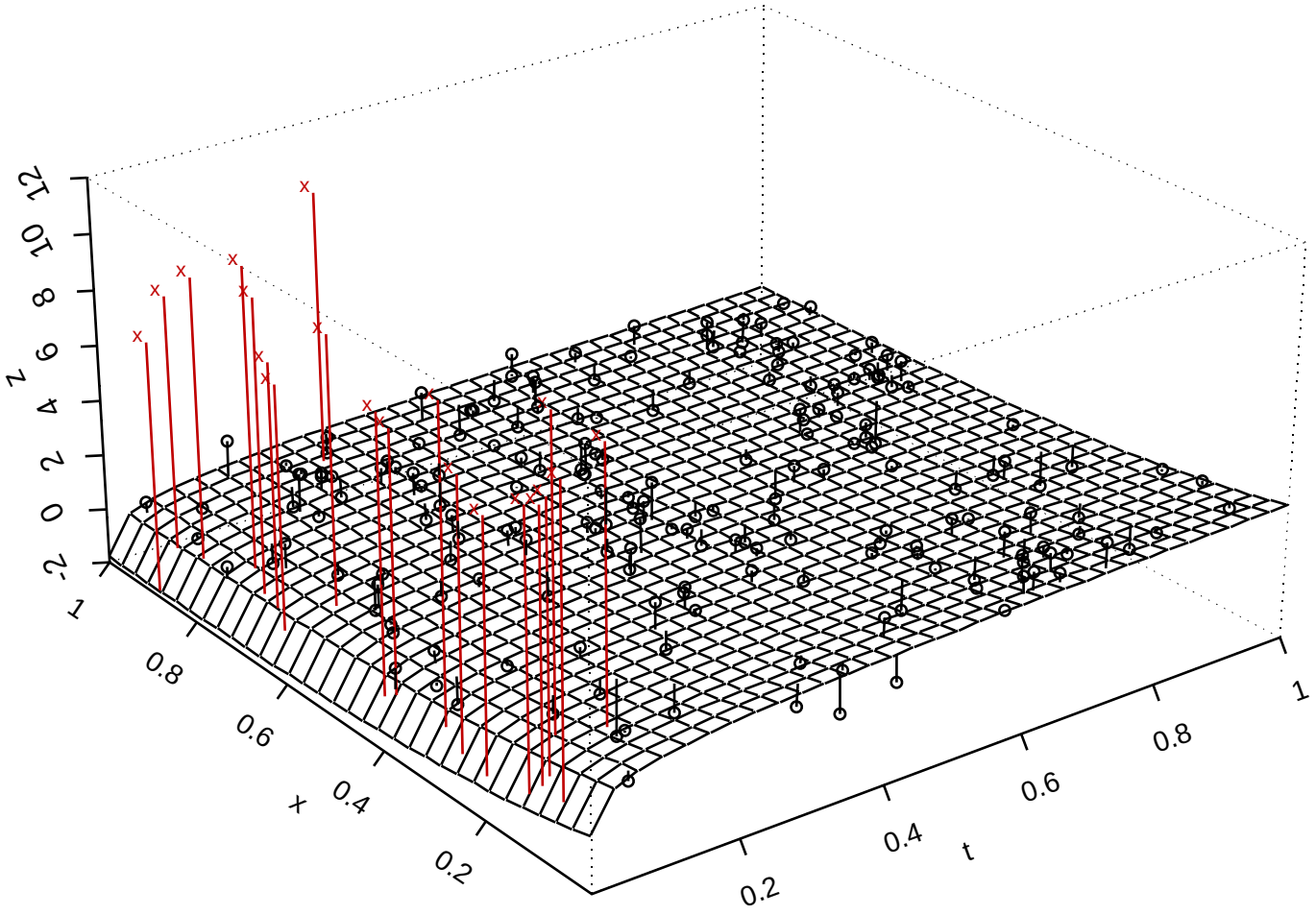

Figure 3.4: Geman - Mc Clure estimation of the regression function $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

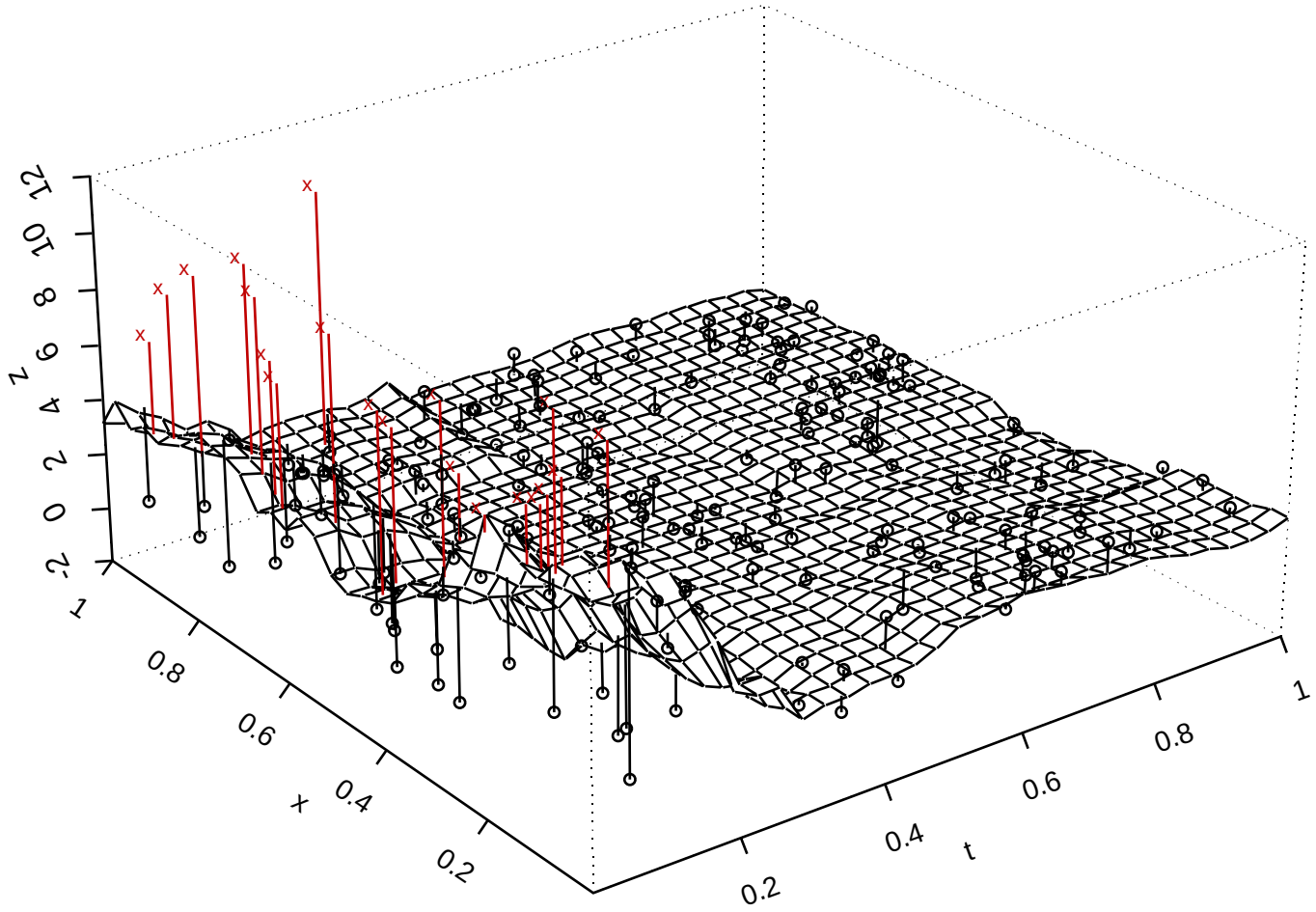

Figure 3.5: Kernel estimation of regression function $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

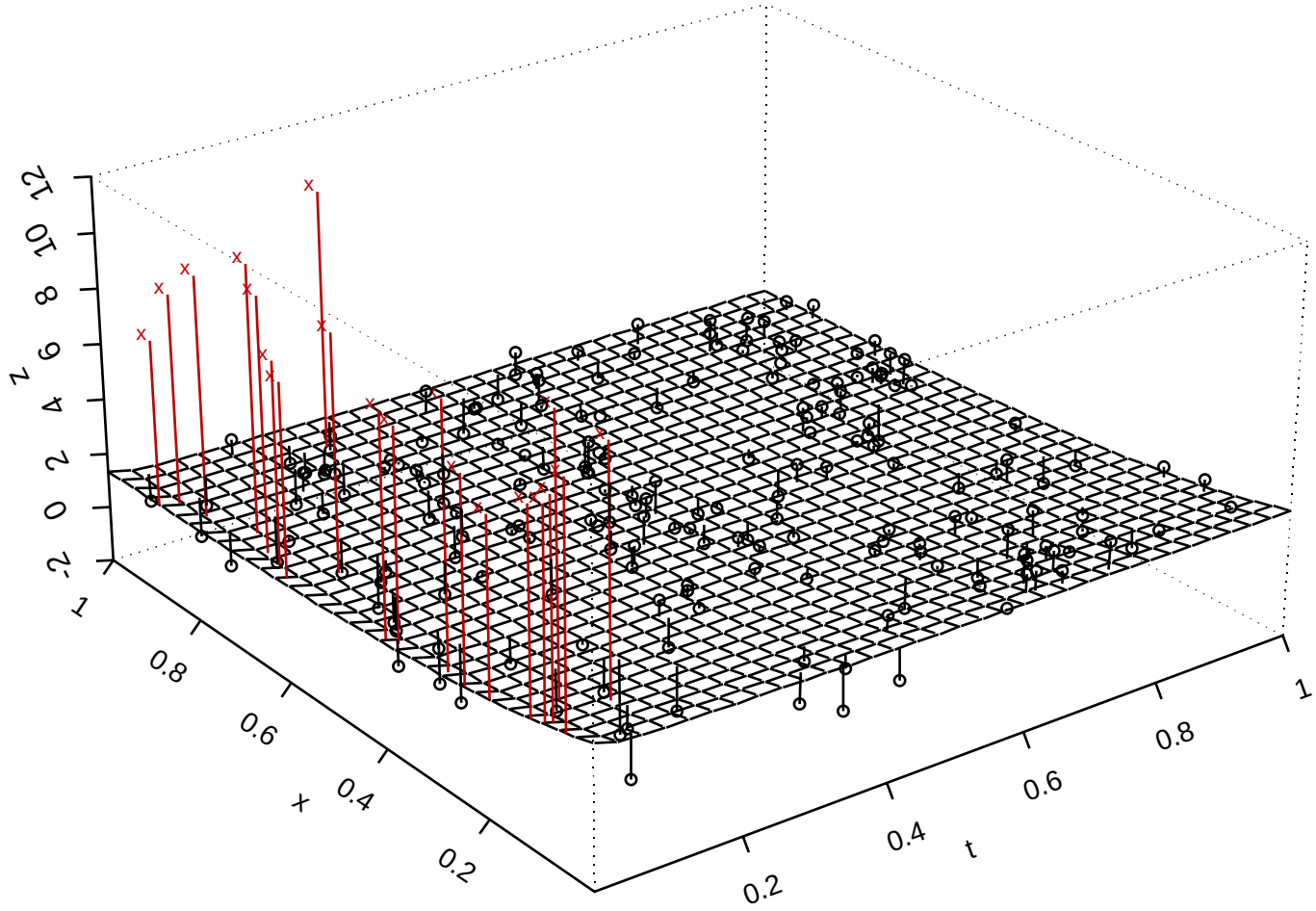

Figure 3.6: Sudentized L_2 estimation of regression function $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

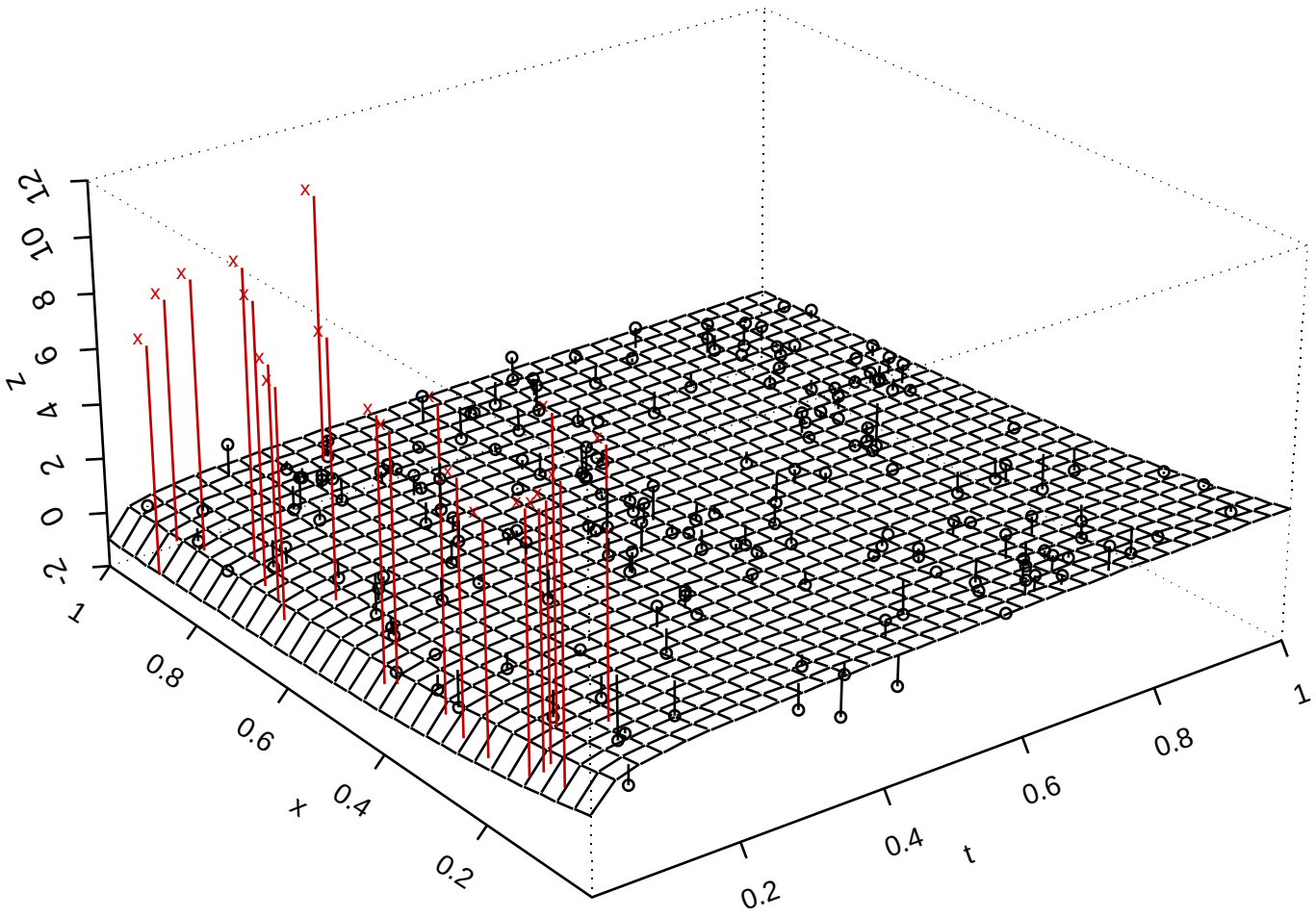

Figure 3.7: Studentized Geman - Mc Clure's estimation of regression function $Z_i^j = a + b \cdot e^{-c \cdot X_i^j} + d \cdot \ln(T_i^j) + \varepsilon_i^j$

Conclusion

I have presented in this thesis the assumptions under which a consistent and asymptotic normal M-estimation of unknown parameter in model with longitudinal data exists. I believe that proposed M-estimators can provide better results than ordinary least square estimation.

The presented methodology of M-estimations (or studentized M-estimations) is useful especially observing outliers (data which are far away from the bulk of the data). The most frequent reasons for occurrence of outliers are errors caused by source of deviations that act only occasionally but are quite powerful. A single outlier which is sufficiently far away can ruin a least squares analysis completely as was demonstrated in simulations studies, e.g. in simulation study 1, 11 % error caused underestimation of parameter b in a case when ordinary least square estimation was used. If you look more closely at Figure 3.1, x-coordinates of outliers are placed mainly in the center of an interval $\langle 0, 1\rangle$. Their high value caused underestimation of parameter b. Ordinary L_2 estimation of parameter b is the worst while ordinary M-estimations decrease an undesirable influence of outliers. If the outliers were generated as the first measurements, they would have effected more parameter a than parameter b. This situation occurred in simulation study 2, were L_2 estimator totally distorted a location parameter a. When I used weighted L_2 estimator according to Scheike [23], the influence of outliers was a little bit reduced, but outliers still effect the estimations of parameters (especially in simulation study 2). Final applying of weighted M-estimations gives the best results in both simulations studies.

Whenever a distant outlier occurs, any robust method is definitely needed. On the other hand, for high quality data, there is usually only a small increase in efficiency by the use of robust methods. Moreover, even for high-quality data, the improvements possible by the use of good robust methods may still be important. The key point is that even in highest quality of data settings, there may be a tiny fraction of outliers sometimes hard to be found. This fraction may upset parts of the statistical results if left untreated.

In order to expand the range of application I suggest to research properties of M-estimations also for non-smooth minimization functions $\rho(.)$. As I used Taylor's expansion to prove demand properties of estimations, behaviour for non-smooth functions remains to be unknown. It can be envisioned that the estimator will be consistent under some regularity conditions. For example it was shown in the simulations studies, that L_1 estimation gives reasonable results however rigorous proof is missing.

Bibliography

[1] Andersen, P.K., Borgan, Ø. (1985): Counting Process for Life History Data: A Review. Scand. J. Statistist. 12, p. 97 - 158

[2] Andersen, P.K., Borgan, Ø, Gill, R.D., Keiding, N. (1991): Statistical Models Based on Counting Processes. Springer - Verlag.

[3] Billingsley, P. (1999): Covergence of Probability Measures. Wiley.

[4] Boel, R., Varaiya, P., Wong, E. (1975): Martingales on Jump Processes. Siam J. Control 13, 999-1061.

[5] Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L. (2002): Analysis of Longitudinal Data. Oxford Univerity Press.

[6] Flemming T.R., Harrington, D.P. (1991): Counting Processes and Survival Analysis. Wiley - Sons.

[7] Gasbarra, D. (1997): Bayesian inference for models based on point processes, by using Markov chain methods. Acta Universitatis Ouluensis.

[8] Huber, P.J. (1981): Robust Statistics. John Wiley & Sons.

[9] Jacobsen M. (1982): Statistical Analysis of Counting Processes. Springer - Verlag.

[10] Jacod J., Shiryaev, A. N. (2003): Limit Theorems for Stochastic Processes. Springer.

[11] Jarník, V. (1984): Diferenciální počet II. Academia.

[12] Jurečková, J., Sen, P.K. (1989): Uniform second order asymptotic linearity of *M*-statistics in linear models. Statist. Decisions 7, 263-276.

[13] Klein, J.P., Moeschberger, M.L. (1997): Survival Analysis. Springer.

[14] Last, G., Brandt, A. (1995): Marked Point Processes on the Real Line, The Dynamic Approach. Springer.

[15] Lehmann, E. (1983): Theory of Point Estimation. Wiley.

[16] Liang, K.Y., Zeger, S.L. (1986): Longitudinal data analysis using generalized linear models. Biometrika, 73, 13-22.

[17] Lindsey, J.K. (1999): Models for Repeated Measurements. Oxford University Press.

[18] Martinussen, T., Scheike, T.H. (1998): A Non-Parametric Dynamic Additive Regression Model for Longitudinal Data. Research Report, Copenhagen.

[19] Martinussen, T., Scheike, T.H. (1998): A Semi-Parametric Additive Regression Model for Longitudinal Data. Research Report. Copenhagen.

[20] Orsáková, M. (1999): Models for censored data. Proceedings of the 8^{th} Annual Conference of Doctoral Students WDS'99, Faculty of Mathematics and Physics, Charles University, Prague.

[21] Orsáková, M. (2001): Regression models for longitudinal data. Robust'2000, JČMF Prague, 210-216.

[22] Rubio, A.M., Víšek, J.Á. (1996): A Note On Asymptotic Linearity of M-Statistics in Nonlinear Models. Kybernetika 32, 353-374.

[23] Scheike, T.H. (1994): Parametric Regression for Longitudinal Data with Counting Process Measurement Times. Scandinavian Journal of Statistics 21, 245-263.

[24] Scheike, T.H., Zhang, M. (1998): Cumulative Regression Tests for Longitudinal Data. Annals of Statistics.

[25] Štěpán, J. (1987): Teorie pravděpodobnosti. Matematické základy. Academia.

[26] Volf, R. (2001): Statistical models and analysis of cumulated damage processes. Robust'2000, JČMF Prague, 357-367.

Appendix A

Related Definitions and Theorems

It will be useful to recall some definitions. This will be followed by some remarks and a reproduction of some known results which were used in the discussion above. The present chapter surveys and summarizes the theory of counting processes and their intensity processes, the theory of stochastic integrals with respect to counting processes or martingale, and martingale central limit theory. Definitions of minimization functions used in M-estimators follows. Nadaraya-Watson non-parametric estimation is described in the very end of chapter.

A.1 Martingale Theory

Definition A.1. *When $\{\mathcal{F}_t : t \geq 0\}$ is a filtration, the σ-algebra $\cap_{h>0}\mathcal{F}_{t+h}$ is denoted by $\mathcal{F}_{t+}$. The corresponding limit from the left, $\mathcal{F}_{t-}$, is the smallest σ-algebra containing all the sets in $\cup_{h>0}\mathcal{F}_{t-h}$.*

Definition A.2. *A filtration $\{\mathcal{F}_t : t \geq 0\}$ is a right-continuous, if for any t, $\mathcal{F}_{t+} = \mathcal{F}_t$.*

Definition A.3. *Let $M = \{M(t) : t \geq 0\}$ be a right-continuous stochastic process with left-hand limits (cadlag process) and $\{\mathcal{F}_t : t \geq 0\}$ a filtration, defined on a common probability space $(\Omega, \mathcal{F}, P)$. M is a called martingale with respect to $\{\mathcal{F}_t : t \geq 0\}$ if*

i) M is adapted to $\{\mathcal{F}_t : t \geq 0\}$,

ii) M is integrable (i.e. $E(|M(t)|) < \infty$) for all $t < \infty$,

iii) $E(M(t+s)|\mathcal{F}_t) = M(t)$ a.s. for all $s \leq 0$, $t \leq 0$.

M is called a submartingale if iii) is replaced by $E(M(t+s)|\mathcal{F}_t) \geq M(t)$ a.s.

Definition A.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $\{\mathcal{F}_t : t \geq 0\}$. The σ-algebra on $[0, \infty) \times \omega$ generated by all sets of the form*

$$[0] \times A, \quad A \in \mathcal{F}_0,$$

and

$$(a, b] \times A, \quad 0 \leq a < b < \infty, \quad A \in \mathcal{F}_a,$$

is called the predictable σ-algebra for the filtration $\{\mathcal{F}_t : t \geq 0\}$.

Definition A.5. *A process X is called predictable with respect to a filtration if, as a mapping from $[0,\infty)\times\Omega$ to $\mathbb{R}$, it is measurable with respect to the predictable σ-algebra generated by that filtration. We call X an $\mathcal{F}_t$-predictable process.*

Definition A.6. *A collection of random variables $\{X(t) : t \in \tau\}$, where τ is an arbitrary index set, is uniformly integrable if*

$$\lim_{n\to\infty} \sup_{t\in\tau} E(|X(t)| \cdot I_{|X(t)>n|}) = 0.$$

Theorem A.1. *Doob-Meyer Decomposition, 1966:*
Let X be a right-continuous nonnegative submartingale with respect to a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$. Then there exists a right-continuous martingale M and an increasing right-continuous predictable process A (compensator) such that $EA(t) < \infty$ and

$$X(t) = M(t) + A(t) \quad \text{a.s.}$$

for any $t \geq 0$. If $A(0) = 0$ a.s., and if $X = M' + A'$ is another such decomposition with $A'(0) = 0$, then for any $t \geq 0$,

$$P\{M'(t) \neq M(t)\} = 0 = P\{A'(t) \neq A(t)\}.$$

If in addition X is bounded, then M is uniformly integrable and A is integrable.

Definition A.7. *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $\{\mathcal{F}_t : t \geq 0\}$. A nonnegative random variable T is a stopping time with respect to $\mathcal{F}_t$ if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.*

Theorem A.2. *Lenglart's Inequality:*
Let X be a right-continuous adapted process, and Y a nondecreasing predictable process with $Y(0) = 0$. Suppose, for all bounded stopping times T, $E|X(T)| \leq EY(T)$. Then for any stopping time T, and any $\varepsilon, \eta > 0$,

$$P\{\sup_{t\leq T} |X(t)| \geq \varepsilon\} \leq \frac{\eta}{\varepsilon} + P\{Y(T) \geq \eta\}.$$

Proof. See Jacod and Shiryaev [10], p.35. □

Definition A.8. *An increasing sequence of random times τ_n, $n = 1, 2, \cdots$ is called a localizing sequence with respect to a filtration if the following hold is true:*

i) Each τ_n is a stopping time relative to the filtration, and

ii) $\lim_{n\to\infty} \tau_n = \infty$ a.s.

Definition A.9. *A stochastic process $M = \{M(t) : t \geq 0\}$ is called a local martingale (submartingale) with respect to filtration $\{\mathcal{F}_t : t \geq 0\}$ if there exists a localizing sequence τ_n such that, for each n, $M_n = \{M(t \wedge \tau_n) : 0 \leq t < \infty\}$ is an $\mathcal{F}_t$-martingale (submartingale).*

Definition A.10. *A stochastic process $X = \{X(t) : t \geq 0\}$ is called a locally bounded if, for a suitable localizing sequence τ_n, $X_n = \{X(t \wedge \tau_n) : 0 \leq t < \infty\}$ is a bounded process for each n.*

Theorem A.3. *Extended Doob-Meyer Decomposition:*
Let X be a right-continuous nonnegative local $\mathcal{F}_t$-submartingale with localizing sequence τ_n, where $\{\mathcal{F}_t, t \geq 0\}$ is a right-continuous filtration. Then there exists a unique increasing right-continuous predictable process A such that $A(0) = 0$ a.s., $P\{A(t) < \infty\} = 1$ for all $t > 0$, and $X - A$ is a right-continuous predictable local martingale. At each t, $A(t)$ may be taken as the a.s. $\lim_{n\to\infty} A_n(t)$, where A_n is the compensator of the stopped submartingale $X(\cdot \wedge \tau_n)$.

Proof. See Fleming and Harington [6], p.58. □

A.2 Counting Processes

Definition A.11. *A counting process (CP) is a stochastic process $\{N_t,\ t \geq 0\}$ adapted to a complete and right-continuous filtration $\{\mathcal{F}_t,\ t \geq 0\}$ with $N_0 = 0$ and $N_t < \infty$ a.s., and whose paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with jump of size +1.*

Definition A.12. *The full counting process path space W is the space of all paths $\omega : [0, \infty) \to \{0, 1, 2, \cdots \infty\}$ with $\omega(0) = 0$ which are everywhere right-continuous and non-decreasing, increasing only in jumps of size 1.*

Theorem A.4. *For $t \geq 0$ define $x_t : W \to \{0, 1, 2, \cdots, \infty\}$ by $x_t(\omega) = \omega(t)$ and let $\mathcal{F} = \sigma((x_t = n), n \in \mathbb{N}_0\ t \geq 0)$, further $\mathcal{F}_t = \sigma(x_s : s \leq t)$. Then we have $\mathcal{F}_{t+} = \mathcal{F}_t$.*

Proof. See Jacobsen [9], p.7-8. □

Theorem A.5. *Define inductively the function T_n:*

$$\begin{aligned} T_0 &\equiv 0 \\ T_n(\omega) &= \inf\{t : t \geq 0,\ x_t = n\}, \\ T_\infty &= \lim_{n\to\infty} T_n. \end{aligned}$$

where the infimum over an empty set is taken to be $+\infty$. Then

i) T_n is a stopping time for all $n = 1, \cdots \infty$,

ii) $\mathcal{F}_{T_n} = \sigma(T_i; 0 \leq i \leq n)$,

iii) $\mathcal{F}_{T_\infty} = \mathcal{F}$.

Proof. See Jacobsen [9], p.12-14. □

Corollary A.1. *Doob-Meyer Decomposition for the Counting Process:*
Let $N(t) : t \geq 0$ be a counting process adapted to a right-continuous filtration $\{\mathcal{F}_t : t \geq 0\}$ with $EN(t) < \infty$ for any t. Then there exists a unique increasing right-continuous $\mathcal{F}_t$-predictable process A such that $A(0) = 0$ a.s., $EA(t) < \infty$ for any t, and $\{M(t) = N(t) - A(t) : t \geq 0\}$ is a right-continuous $\mathcal{F}_t$-martingale.

Corollary A.2. *Let M be a right-continuous martingale with respect to a right-continuous filtration $\{\mathcal{F}_t : t \geq 0\}$ and assume $EM^2(t) < \infty$ for any $t \geq 0$. Then there exists a unique increasing right-continuous predictable process $\langle M, M\rangle$, called the predictable quadratic variation of M, such that $\langle M, M\rangle(0) = 0$ a.s., $E\langle M, M\rangle(0) < \infty$ for each t, and $\{M^2(t) - \langle M, M\rangle(t) : t \geq 0\}$ is a right-continuous martingale.*

Theorem A.6. *Let M_1 and M_2 be two right-continuous martingales with respect to a right-continuous filtration $\mathcal{F}_t : t \geq 0\}$ and assume $EM_i^2(t) < \infty$ for any $t \geq 0$ and $i = 1, 2$. Then there exists a unique increasing right-continuous predictable process $\langle M_1, M_2\rangle$, called the predictable covariation process, with $\langle M_1, M_2\rangle(0) = 0$ a.s., $E\langle M_1, M_2\rangle(0) < \infty$ for each t, such that*

i) $\langle M_1, M_2\rangle$ is the difference of two increasing right-continuous predictable processes, and

ii) $\{M_1M_2 - \langle M_1, M_2\rangle(t) : t \geq 0\}$ is a right-continuous martingale.

Proof. See Fleming and Harington [6], p.40. □

Corollary A.3. *If M_1 and M_2 are two right-continuous $\mathcal{F}_t$-martingales with $EM_i^2(t) < \infty$ for any $t \geq 0$. Then the right-continuous process M_1M_2 is a martingale if and only if $\langle M_1, M_2\rangle(0) \equiv 0$. In this case, M_1 and M_2 are said to be orthogonal.*

Theorem A.7. *The martingale transform:*
Let N be a counting process with $EN(t) \leq \infty$ for any t. Let $\{\mathcal{F}_t : t \geq 0\}$ be a right-continuous filtration such that
1. $M = N - A$ is an $\mathcal{F}_t$-martingale, where $A = \{A(t), t \geq 0\}$ is an increasing $\mathcal{F}_t$-predictable process with $A(0) = 0$;
2. H is a bounded , $\mathcal{F}_t$-predictable process.
Then the process $L(t) = \int_0^t H(u)dM(u)$ is an $\mathcal{F}_t$-martingale.

Proof. See Fleming and Harington [6], p.47. □

Theorem A.8. *Let N be an arbitrary counting process.*

i) Then there exists a unique right-continuous predictable increasing process A such that $A(0) = 0$ a.s., $A(t) < \infty$ a.s. for any t, and the process $M = N - A$ is a local martingale.

ii) If A in (i) is locally bounded, M is a local square integrable martingale.

Proof. See Fleming and Harington [6], p.61. □

Theorem A.9. *Let N be a counting process and let A be its unique compensator in the Extended Doob-Meyer Decomposition Theorem. Then A is a locally bounded process.*

Theorem A.10. *Suppose M is a right-continuous local square integrable martingale on $[0, \infty)$. Then there exists a unique predictable right-continuous increasing process $\langle M, M\rangle$ with $\langle M, M\rangle(0) = 0$ a.s. and $\langle M, M\rangle(t) < \infty$ a.s. for any t, such that $M^2 - \langle M, M\rangle$ is a right-continuous local martingale. If $\{\tau_n\}$ is a localizing sequence for M, then*

$$\langle M, M\rangle(t) = \lim_{n\to\infty} \langle M(\cdot \wedge \tau_n), M(\cdot \wedge \tau_n)\rangle(t).$$

Proof. See Fleming and Harington [6], p.63. □

Theorem A.11. *Suppose M_1, M_2 are right-continuous local square integrable martingales on $[0,\infty)$. Then there exists a predictable right-continuous increasing process $\langle M_1, M_2\rangle$ with $\langle M_1, M_2\rangle(0) = 0$ a.s. and $\langle M_1, M_2\rangle(t) < \infty$ a.s. for any t, such that $M_1M_2 - \langle M_1, M_2\rangle$ is a right-continuous local martingale. In fact*

$$\langle M_1, M_2\rangle = \frac{1}{2}\{\langle M_1 + M_2, M_1 + M_2\rangle - \langle M_1, M_1\rangle - \langle M_2, M_2\rangle\}.$$

When the underlying filtration is complete and right-continuous, $\langle M_1, M_2\rangle$ is unique.

Proof. See Fleming and Harington [6], p.64.
Proof of uniqueness may be found in Jacobsen [9], p.96. □

Theorem A.12. *The local martingale transform:*
Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, P)$ be a stochastic basis with right-continuous filtration $\{\mathcal{F}_t; t \geq 0\}$, H a locally bounded $\mathcal{F}_t$-predictable process, and N a counting process. Let $M = N - A$ be the local square integrable $\mathcal{F}_t$-martingale whose existence is established in Theorems A.8 and A.9. Then $\int HdM$ is a local square integrable martingale.

Proof. See Fleming and Harington [6], p.66. □

Theorem A.13. *Consider two processes $U_l \equiv \sum_{i=1}^{n} \int H_{i,l} dM_i$, $l = 1, 2$. Let $\{\mathcal{F}_t; t \geq 0\}$ be a right-continuous filtration such that for each i and l $H_{i,l}$ is a locally bounded $\mathcal{F}_t$-predictable process, and $M_i = N_i - A_i$ is the local square integrable martingale corresponding to the arbitrary counting process N_i. Then*

i) U_l is a local square integrable martingale.

If $E\int_0^t H_{i,l}^2 d\langle M_i, M_i\rangle < \infty$ for any i, l then

ii) $EU_l(t) = 0$,

iii) $E\{U_1(t)U_2(t)\} = E\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^t H_{i,1}H_{j,2}d\langle M_i, M_j\rangle$,

iv) U_l is a martingale over $[0, t]$.

Proof. See Fleming and Harington [6], p.66, p.72-73. □

Definition A.13. *A n-variate process $\{(N_t^1, N_t^2, \ldots, N_t^n),\ t \geq 0\}$ is called a multivariate counting process if each $\{N_t^i,\ t \geq 0\}$, $i = 1, \ldots, n$ is a counting process, and no two component processes jump at the same time.*

Theorem A.14. *Let $\{N_1, N_2, \cdots N_k\}$ be a multivariate counting process and, $j = 1, \cdots k$, let A_j be the compensator of N_j. Assume that each A_j is a continuous process. Then*

i) $\langle M_j, M_j \rangle = A_j$, *that is,* A_j *is the unique predictable, right-continuous increasing process with* $A_j(0) = 0$ *a.s. and* $A_j(t) < \infty$ *a.s. for any* t, *such that* $M_j^2 - A_j$ *is a local martingale,* $j = 1, \cdots k$.

ii) If $i \neq j$, $\langle M_i, M_j \rangle(t) = 0$ *a.s., that is,* $M_i M_j$ *is a local martingale.*

Proof. See Fleming and Harington [6], p.75. □

Theorem A.15. *Let* N *be a counting process and* A *compensator. If* A *is continuous then* $EM^2(t) \leq EA(t)$ *for* $t \geq 0$. *If in addition,* $EA(t) < \infty$ *(or equivalently if* $EN(t) < \infty$*) for any* t, *then* $EM^2(t) = EA(t)$ *for any* t *and* $M^2 - A$ *is a martingale.*

Proof. See Fleming and Harington [6], p.77. □

A.3 Marked Point Processes

Definition A.14. *Let* T *be a stopping time.* T *is said to be totally inaccessible if for every increasing sequence of stopping times* $S_1 \leq S_2 \leq \cdots$, *and* $P\{S_k(\omega) < T(\omega),\ \lim_{k\to\infty} S_k(\omega) = T(\omega) < \infty\} = 0$.

Definition A.15. *Let* $(E, \mathcal{E})$ *be some measurable space of marks or types. We say that* $\mathbf{N}$ *is a marked point process (MPP) if* $\mathbf{N}$ *is a counting measure on the product space* $(\mathbb{R}^+ \times E, \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{E})$ *and* $N(t, B) = \mathbf{N}([0, t] \times B)$ *is a counting process for every* $B \in \mathcal{E}$.

Observation A.1. *The number* $N(t, B)$ *is a number of jumps which occur prior to* t *and which end in the set* B. *It follows from this that, for disjoint sets* $B_1, \ldots B_k \in \mathcal{E}$, *the process* $N(t, B_1), \ldots N(t, B_k)$ *is a multivariate counting process in the ordinary sense.*

The process $N(B_1), \cdots N(B_k)$, *therefore, has a compensator* $\Lambda(B_1), \cdots \Lambda(B_k)$. *It turns out that one can extract all these compensator from a single so-called predictable measure* $\mathbf{\Lambda}$ *on* $(\mathbb{R}^+ \times E, \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{E})$ *in the same way as is done for* $\mathbf{N} : \Lambda(t, B) = \mathbf{\Lambda}([0, t] \times B)$.

Letting (T_n, J_n) *be the points of the MPP, it turns out that* T_n *is a stopping time and* J_n *is* $\mathcal{F}_{T_n}$*-measurable [with values in* $(E, \mathcal{E})$*] for each* n.

Theorem A.16. *For* $t \geq 0$ *define* $x_t : W \to E$ *by* $x_t(\omega) = \omega(t)$. *Let* $\mathcal{F}_t$ *be the* σ*-algebra generated by the sets of the form* $\{x_s \in B\}$, $B \in E$, $s \leq t$. *Let* $\mathcal{F} = \cup_{t \in [0,\infty)} \mathcal{F}_t$. *Define inductively the function* T_n*:*

$$\begin{aligned} T_0 &\equiv 0 \\ T_{n+1}(\omega) &= \inf\{t : t \geq T_n(\omega),\ N(t, \omega) \neq N(T_n(\omega), \omega)\}, \end{aligned}$$

where the infimum over an empty set is taken to be $+\infty$. *Then*

i) T_n *is a stopping time for all* n,

ii) $\mathcal{F}_{T_n} = \sigma(x_{T_i}, T_i; 0 \leq i \leq n)$,

iii) $\mathcal{F}_t = \sigma(x_{T_i \wedge t}, T_i \wedge t; 0 \leq i < n)$,

iv) $\mathcal{F}_{T_n -} = \sigma(x_{T_i}, T_{i+1}; 0 \leq i \leq n-1)$.

v) Further for all $t \in [0,\infty)$ is $\mathcal{F}_{t+} = \mathcal{F}_t$.

Proof. See Boel [4], p.1001-1002. □

Theorem A.17. *Let $N(t,B)$ is a MPP and T_n are totally inaccessible for $n \geq 1$. There is a unique increasing continuous process $A(t,B)$ that is a locally integrable and the process $M(t,B) = N(t,B) - A(t,B)$ is locally square integrable $\mathcal{F}_t$-martingale.*

Proof. See Boel [4], p.1007. □

Theorem A.18. *Let $B_i \in E,\ i = 1,2$ and T_n are totally inaccessible for $n \geq 1$. Then $M(t,B_1)M(t,B_2) - A(t,B_1 \cap B_2)$ is locally integrable martingale, i.e. $\langle M(B_1), M(B_2)\rangle(t) = A(t,B_1 \cap B_2)$. In particular, $M(t,B_1)$ and $M(t,B_2)$ are orthogonal if $B_1 \cap B_2 = \emptyset$.*

Proof. See Boel [4], p.1007. □

Theorem A.19. *The martingale transform for MPP:*
Consider a process $U \equiv \int\int H(s,z)dM(s,z)$. Let T_n are totally inaccessible for $n \geq 1$ and $H(s,z)$ be $\mathcal{F}_s$-predictable process such that

$$E\left(\int\limits_E \int\limits_{\mathbb{R}^+} H^2(s,z)A(ds,dz)\right) < \infty$$

and $M = N - A$ is the local square integrable martingale corresponding to the arbitrary marked point process N. Then there is a unique process $\int\int H dM$ that is a square integrable martingale over $[0,t]$, called the "stochastic integral of H with respect to M". If G is another predictable process with property $E\left(\int\limits_E \int\limits_{\mathbb{R}^+} G^2(s,z)A(ds,dz)\right) < \infty$ then

$$\left\langle \int\int H dM, \int\int G dM \right\rangle (t) = \int\limits_E \int\limits_{\mathbb{R}^+} H(s,z)G(s,z)I_{(0,t]}(s)A(ds,dz),$$

and in particular

$$E\left(\int\int H dM\right)^2 = E\left(\int\limits_E \int\limits_{\mathbb{R}^+} H^2(s,z)A(ds,dz)\right).$$

Proof. See Boel [4], p.1010. □

A.4 Central Limit Theorems for Martingales

Theorem A.20. *For each $n = 1,2,\ldots$, let $\mathbf{M}^{(n)} = (M_1^{(n)}, \ldots M_k^{(n)})$ be a vector of k local square integrable martingales. Define $M_{\varepsilon,h}^{(n)}(t) = M_h^{(n)}(t) \cdot I\left\{|M_h^{(n)}(t)| \geq \varepsilon\right\}$ for each $\varepsilon > 0$. We write $\langle \mathbf{M}^{(n)} \rangle$ for the $k \times k$ matrix of processes $\langle M_h^{(n)}, M_{h'}^{(n)} \rangle$. Let $\mathcal{T}$ is a dense subset of $\mathbb{R}^+$. Next, let $\mathbf{M}^{\infty}$ be a continuous Gaussian vector martingale with $\langle \mathbf{M}^{(\infty)} \rangle(t) = V(t)$, a continuous deterministic $k \times k$ positive semidefinite matrix-valued function on $\mathcal{T}$ with*

positive semidefinite increments, zero at time zero. So $\mathbf{M}^{(\infty)}(t) - \mathbf{M}^{(\infty)}(s) \sim N(\mathbf{0}, \mathbf{V}(t) - \mathbf{V}(s))$ *(a multivariate normal distribution) and is independent of* $(\mathbf{M}^{(\infty)}(u); u \leq s)$ *for all* $0 \leq s \leq t$. *Consider the conditions*

(i) $\langle \mathbf{M}^{(n)} \rangle(t) \xrightarrow[n\to\infty]{p} \mathbf{V}(t)$ *for all* $t \in \mathcal{T}$,

(ii) $\langle M_{\varepsilon h}^{(n)} \rangle(t) \xrightarrow[n\to\infty]{p} 0$ *for all* $t \in \mathcal{T}$, h *and* $\varepsilon > 0$.

Then

$$\mathbf{M}^{(n)} \xrightarrow[n\to\infty]{\mathcal{L}} \mathbf{M}^{(\infty)}$$

on $(D(\mathcal{T}))^k$, *where* $(D(\mathcal{T}))^k$ *denotes the space of* $\mathbb{R}^k$ *valued functions on* $\mathcal{T}$ *which are cadlag, endowed with the Skorokhod topology;* $\xrightarrow[n\to\infty]{\mathcal{L}}$ *means weak convergence of the laws, relative to the Skorokhod topology.*

Proof. See Jacod and Shiryaev [10], p.311. □

A.5 The influence function of the M-estimate

Definition A.16. *Any estimate* T_n, *defined by a minimum problem of the form* $\min_{\mathrm{T_n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \rho(\mathrm{x_i}, \mathrm{T_n})$ *or by an implicit equation* $\sum_{i=1}^{n} \psi(x_i, T_n) = 0$, *where rho is an arbitrary function,* $\psi(x, \theta) = \frac{\delta}{\delta\theta}\rho(x, \theta)$, *is called an M-estimate or maximum likelihood type estimate.*

Let $X_1, ...X_n$ be i.i.d.r.v.'s with a d.f. F. We have a parametric model $\{F_\theta : \theta \in \Theta\}$ formed by a dominated system of distributions and wish to estimate θ for which F_θ is as close to F as possible. Let $T_n = T_n(X_1, ...X_n)$ be an estimator of θ which we express as a functional $T(F_n)$ of the empirical d.f. F_n of $X_1, ...X_n$. In this context, it is always assumed that the functionals under the study are Fisher consistent; that is $T(F_\theta) = \theta$, $\theta \in \Theta$.

Definition A.17. *The influence function of* T *at* F *is given by*

$$IF(x, T, F) = \lim_{t\downarrow 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t},$$

where Δ_x *is the masspoint 1 at* x.

Observation A.2. *The influence function allows us to assess the relative influence of individual observations toward the value of an estimate or test statistic. If it is unbounded, an outlier might cause trouble.*

A.6 The Non-Parametric Estimation in Regression Models with Counting Data

Let us consider the model in (2.1). Assume further that the assumptions (2.4), (2.5) and (2.6) are fulfilled. Let $K(.)$ be a kernel function with support on $[-1; 1]$, $\int K(u)du = 1$, and

Type	$\rho(x)$	$\psi(x)$
L_2	$\frac{x^2}{2}$	x
L_1	$\lvert x\rvert$	$\operatorname{sgn}(x)$
$L_2 - L_1$	$2\left(\sqrt{1+\frac{x^2}{2}}-1\right)$	$\frac{x}{\sqrt{1+\frac{x^2}{2}}}$
L_p	$\frac{\lvert x\rvert^\nu}{\nu}$	$\operatorname{sgn}(x)\lvert x\rvert^{\nu-1}$
Fair	$c^2\left(\frac{\lvert x\rvert}{c}-\log\left(1+\frac{\lvert x\rvert}{c}\right)\right)$	$\frac{x}{1+\frac{\lvert x\rvert}{c}}$
Huber	$\begin{cases}\frac{x^2}{2}, & \lvert x\rvert\le k\\ k(\lvert x\rvert-\frac{k}{2}), & \lvert x\rvert>k\end{cases}$	$\begin{cases}x\\ k\ \operatorname{sgn}(x)\end{cases}$
Cauchy	$\frac{c^2}{2}\log\left(1+\left(\frac{x}{c}\right)^2\right)$	$\frac{x}{1+(\frac{x}{c})^2}$
Geman-Mc Clure	$\frac{x^2/2}{1+x^2}$	$\frac{x}{(1+x^2)^2}$
Welsch	$\frac{c^2}{2}\left(1-e^{-(\frac{x}{c})^2}\right)$	$xe^{-(\frac{x}{c})^2}$
Tukey	$\begin{cases}\frac{c^2}{6}\left(1-\left(1-(\frac{x}{c})^2\right)^3\right), & \lvert x\rvert\le c\\ \frac{c^2}{6}, & \lvert x\rvert>c\end{cases}$	$\begin{cases}x\left(1-(\frac{x}{c})^2\right)^2\\ 0\end{cases}$
Blake and Zisserman	$\min\{\mathrm{x}^2,1\}$	$\begin{cases}2x, & \lvert x\rvert\le 1\\ 0, & \lvert x\rvert>1\end{cases}$
Hebert and Leahy	$\frac{1}{2}\log(1+x^2)$	$\frac{x}{1+x^2}$
Geman and Reynolds	$\frac{\lvert x\rvert}{1+\lvert x\rvert}$	$\frac{\operatorname{sgn}(x)(1+2\lvert x\rvert)}{(1+\lvert x\rvert)^2}$

Table A.1: M-estimators

let $b=(b_1,\dots b_d)$ be a d-dimensional bandwidth, $|b|=b_1\dots b_d$, $b\in(0,\infty)^d$. Define further $c_K=\int K^2(u)du$, $d_k=\int u^2K(u)du$ and $e_K=\int uK(u)du$. We assume $c_K=O_P(1)$, $d_K=O_P(1)$ and $e_K=0$. Finally, we assume that the kernel function satisfies a Lipschitz condition, that is, $|K(x)-K(y)|\le C|x-y|$. The Nadaraya-Watson type estimator, $\hat{m}(z)$ of $m(z)$ is defined by

$$\hat{m}(z)=\frac{\hat{r}(z)}{\hat{\alpha}(z)},$$

where

$$\hat{r}(z)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}Z_i^j\frac{1}{|b|}K(z-\mathbb{X}_i(T_i^j),b)$$

and

$$\hat{\alpha}(z)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}\frac{1}{|b|}K(z-\mathbb{X}_i(T_i^j),b).$$

Similarly, the estimator of the variance function, $\sigma^2(.)$, can be estimated by the squared-residual kernel estimator

$$\hat{V}(z)=\frac{V(z)}{\hat{\alpha}(z)}-(\hat{m}(z))^2,$$

where

$$V(z) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_t^i}(Z_i^j)^2\frac{1}{|b|}K(z - \mathbb{X}_i(T_i^j), b).$$

Define

$$\alpha^*(z) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{1}{|b|}K(z - \mathbb{X}_i(s), b)\lambda_s^i ds.$$

Theorem A.21. *Assume the following statements:*

(i) $m(.)$ *is twice continuously differentiable,* $\sigma^2(.)$ *is bounded and the kernel satisfies conditions described above.*

(ii) Conditional on the intensity, λ_s^i*, the covariates,* $\mathbb{X}_i(s)$*, have distribution* $f_s^i(v)dl_d$*, where the densities are bounded.*

(iii) $E(\int_0^t \lambda_s^i) = O_P(1)$.

(iv) $b \to 0$, $n \to \infty$, *such that* $n|b| \to \infty$.

(v) There exists $\alpha(z) > \delta > 0$ *and a compact set in d-dimensional space A, such that* $\sup_{z\in A}|\alpha^*(z) - \alpha(z)| \xrightarrow[n\to\infty]{p} 0$.

It then follows that

$$\sup_{z\in A}|\hat{\alpha}(z) - \alpha(z)| \xrightarrow[n\to\infty]{p} 0,$$

$$\sup_{z\in A}|\hat{m}(z) - m(z)| \xrightarrow[n\to\infty]{p} 0.$$

If the conditional fourth moment Z_i^j *is bounded, then*

$$\sup_{z\in A}|\hat{V}(z) - \sigma^2(z)| \xrightarrow[n\to\infty]{p} 0.$$

Proof. See Scheike and Zhang [24]. □

Appendix B

Splus Code of Simulations Studies

B.1 Simulation Study 1

```
#############################################################################
# REGRESSION FUNCTION m(parameters,t)

"model1" <- function (parameters,t)
{
    parameters[1] + t^parameters[2]
}

#############################################################################
# GENERATING DATA FOR PARAMETRIC REGRESSION MODELS

# Z_i^j  = m(parameters,T_i^j) + Eps_i^j
# Eps    ~ N(0,std)
# T      ~ Poisson's process with lambda.small  if Z_i^j-1 < w
#                                 lambda.big    otherwise
# T      in <0,1>

# Values:
#   t    ... vector of times of measurements for all n individuals
#   z    ... vector of values of model for all n individuals
#   N    ... vector of number of measurements for all n individuals
#   coor.error   ... vector of coordinates of outlying measurements

"generate" <- function (n,model,parameters,error)
{
    # Initialization of parameters of distributions
    std_0.1
    lambda.small_15
    lambda.big_3
    w_3
```

```
# order of outlying measurement for each individual
k_5

# Initialization of variables for 1st individual and his 1st measurement
t_rexp(1,lambda.small)
if (k==1)
{
    z_model(parameters,t)+rnorm(1,error,std)
}
else
    z_model(parameters,t)+rnorm(1,0,std)
r_1
N_rep(0,n)
coor.error_rep(0,n)

# Initialization of variables for 1st individual from his 2nd measurement
while (t[length(t)]<=1)
{
    r_r+1
    # T      ~ Poisson's process with lambda.small  if Z_iˆj-1 < w
    #                                 lambda.big          otherwise
    if (z[length(z)]<w)
    {
        t_c(t,t[length(t)]+rexp(1,lambda.small))
    }
    else
        t_c(t,t[length(t)]+rexp(1,lambda.big))
    # every k-th measurement is generated as outlying
    if (r==k)
    {
        z_c(z,model(parameters,t[length(t)])+rnorm(1,error,std))
    }
    else
        z_c(z,model(parameters,t[length(t)])+rnorm(1,0,std))
}
N[1]_length(t)-1
t_t[1:N[1]]
z_z[1:N[1]]
counter_N[1]
if (N[1]<k)
{
    coor.error[1]_0
}
else
    coor.error[1]_k
```

```
    if (n>=2)
    {
    for (i in 2:n)
    {
        t[sum(N)+1]_rexp(1,lambda.small)
        z[sum(N)+1]_model(parameters,t[length(t)])+rnorm(1,0,std)
        r_1
        while (t[length(t)]<=1)
        {
            if (z[length(z)]<w)
            {
                t_c(t,t[length(t)]+rexp(1,lambda.small))
            }
            else
                t_c(t,t[length(t)]+rexp(1,lambda.big))
            r_r+1
            if (r==k)
            {
                z_c(z,model(parameters,t[length(t)])+rnorm(1,error,std))
            }
            else
                z_c(z,model(parameters,t[length(t)])+rnorm(1,0,std))
        }
        N[i]_length(t)-1-sum(N)
        t_t[1:sum(N)]
        z_z[1:sum(N)]
    }
    }

    cumsum.N_cumsum(N)
    for (i in 2:n)
    {
        if (N[i]>=k)
        {
            coor.error[i]_cumsum.N[i-1]+k
        }
        else
            coor.error[i]_0
    }
    return(t,z,N,coor.error)
}

#########################################################################
# UNIFORM KERNEL FUNCTION
# K(y) = K(y-yi,b) = K((y-yi/b)) = 0.5 if |K|<1          yi ...data
#                                = 0    otherwise
```

```
"uniform.kernel" <- function(yi,y)
{
    bandwidth_0.08
    if (abs((y-yi)/bandwidth)<1)
    {
        return(0.5)
    }
    else return(0)
}

# KERNEL ESTIMATION OF THE REGRESSION FUNCTION
# m.hat(y) = r(y)/alpha(y)
# r.hat(y) = 1/n * sum_i sum_j (Z_i^j * K(y-T_i^j,bandwidth))
# alpha(y)  = 1/n * sum_i sum_j (K(y-T_i^j,bandwidth))

"kernel.m" <- function(y,data)
{
    k_apply(as.matrix(data$t), 1., uniform.kernel,y)
    r.hat_1/length(data$N)*t(data$z)%*%k
    alpha.hat_1/length(data$N)*sum(k)
    r.hat/alpha.hat
}

# KERNEL ESTIMATION OF THE CONDITIONAL VARIANCE
# m.hat(y)      = r(y)/alpha(y)
# r.hat(y)      = 1/n * sum_i sum_j (Z_i^j * K(y-T_i^j,bandwidth))
# v.hat(y)      = 1/n * sum_i sum_j ((Z_i^j)^2 * K(y-T_i^j,bandwidth))
# alpha(y)      = 1/n * sum_i sum_j (K(y-T_i^j,bandwidth))
# sigma^2(y)    = v.hat(y)/alpha(y) - m.hat(y)^2

"kernel.sigma" <- function(y,data)
{
    k_apply(as.matrix(data$t), 1., uniform.kernel,y)
    r.hat_1/length(data$N)*t(data$z)%*%k
    v.hat_1/length(data$N)*t(data$z^2)%*%k
    alpha.hat_1/length(data$N)*sum(k)
    m.hat_r.hat/alpha.hat
    v.hat/alpha.hat-m.hat^2
}

#########################################################################
# M ESTIMATIONs
# definitions of minimization function rho(.)
# rho[Z_i^j-m(T_i^j,parameters)) /sqrt(sigma^2(T_i^j))]
```

```
# L_2 estimator
"distanc.LS" <- function(parameters,data,sigma)
{
    sum(((data$z-model(parameters,data$t))^2)/sigma)/2
}

# Geman - Mc Clure's estimator
"distanc.Geman" <- function(parameters,data,sigma)
{
    sum((((data$z-model(parameters,data$t))^2)/sigma)/
    (1+((data$z-model(parameters,data$t))^2)/sigma))/2
}

# Hebert and Leahy's estimator
"distanc.HL" <- function(parameters,data,sigma)
{
    sum(log(1+((data$z-model(parameters,data$t))^2)/sigma))/2
}

# L_2-L_1 estimator
"distanc.L2L1" <- function(parameters,data,sigma)
{
    sum(sqrt(1+((data$z-model(parameters,data$t))^2)/(2*sigma))-1)*2
}

# L_1 estimator
"distanc.L1" <- function(parameters,data,sigma)
{
    sum(abs((data$z-model(parameters,data$t))/sqrt(sigma)))
}

#########################################################################
# ANALYSE

"analyze" <- function(data, param)
{
    total.N_sum(data$N)

    # Kernel estimation of the regression function
    order.data.t_order(data$t)
    sorted.data_list(t=sort(data$t), z=data$z[order.data.t], N=data$N)
    m_apply(as.matrix(sorted.data$t), 1., kernel.m, sorted.data)

    # Kernel estimation of the conditional variance ... sigma
    sigma_apply(as.matrix(sorted.data$t), 1., kernel.sigma, sorted.data)
```

```
# Estimations of parameters
par.LS_nlminb(rep(2,length(param)),distanc.LS, data, sigma=1)
par.LS.W_nlminb(par.LS$parameters,distanc.LS, sorted.data, sigma)

par.Geman_nlminb(par.LS$parameters,distanc.Geman, data, sigma=1)
par.Geman.W_nlminb(par.LS$parameters,distanc.Geman, sorted.data, sigma)

par.HL_nlminb(par.LS$parameters,distanc.HL, data, sigma=1)
par.HL.W_nlminb(par.LS$parameters,distanc.HL, sorted.data, sigma)

par.L2L1_nlminb(par.LS$parameters,distanc.L2L1, data, sigma=1)
par.L2L1.W_nlminb(par.LS$parameters,distanc.L2L1, sorted.data, sigma)

par.L1_nlminb(par.LS$parameters,distanc.L1, data, sigma=1)
par.L1.W_nlminb(par.LS$parameters,distanc.L1, sorted.data, sigma)

# Plots
x_seq(0,1,0.01)

# Values of regression function in true or estimated parameters
regr.true_model(parameters=param,x)

regr.LS_model(parameters=par.LS$parameters,x)
regr.LS.W_model(parameters=par.LS.W$parameters,x)

regr.Geman_model(parameters=par.Geman$parameters,x)
regr.Geman.W_model(parameters=par.Geman.W$parameters,x)

regr.HL_model(parameters=par.HL$parameters,x)
regr.HL.W_model(parameters=par.HL.W$parameters,x)

regr.L2L1_model(parameters=par.L2L1$parameters,x)
regr.L2L1.W_model(parameters=par.L2L1.W$parameters,x)

regr.L1_model(parameters=par.L1$parameters,x)
regr.L1.W_model(parameters=par.L1.W$parameters,x)

par(mfrow=c(1,2))
yrange_range(data$z,regr.true,
 regr.LS,regr.Geman,regr.HL, regr.L2L1,regr.L1)
yrange.W_range(data$z,m,
 regr.true,regr.LS.W,regr.Geman.W,regr.HL.W, regr.L2L1.W,regr.L1.W)
xrange_range(data$t,x)

# Plot of generated data (x - for outliers)
plot(data$t[-data$coor.error],data$z[-data$coor.error],
```

```
ylim=yrange,xlim=xrange, main="M-estimators",xlab="t",ylab="Z")
points(data$t[data$coor.error],data$z[data$coor.error], pch='x')

# True regression function (red)
lines(x,regr.true, col=8,lwd=3)

# LSE of the regression function (green)
lines(x,regr.LS,col=4,lwd=3)

# Geman -Mc Clure estimation of the regression function (orange)
lines(x,regr.Geman,col=12,lwd=3)

# Hebert Leahy estimation of the regression function (grey)
lines(x,regr.HL,col=16,lwd=3)

# L2 - L1 estimation of the regression function (blue)
lines(x,regr.L2L1,col=6,lwd=3)

# L1 estimation of the regression function (pink)
lines(x,regr.L1,col=10,lwd=3)

# Studentized estimations of the regression function
# Plot of generated data (x - for outliers)
plot(data$t[-data$coor.error],data$z[-data$coor.error], ylim=yrange.W,
 xlim=xrange, main="Studentized M-estimators",xlab="t",ylab="Z")
points(data$t[data$coor.error],data$z[data$coor.error], pch='x')

# True regression function (red)
lines(x,regr.true, col=8,lwd=3)

# LSE of the regression function (green)
lines(x,regr.LS.W,col=4,lwd=3)

# Kernel estimation of the regression function (black)
lines(sorted.data$t,m)

# Geman -Mc Clure estimation of the regression function (orange)
lines(x,regr.Geman.W,col=12,lwd=3)

# Hebert Leahy estimation of the regression function (shadow)
lines(x,regr.HL.W,col=16,lwd=3)

# L2 - L1 estimation of the regression function (blue)
lines(x,regr.L2L1.W,col=6,lwd=3)

# L1 estimation of the regression function (pink)
```

```
    lines(x,regr.L1.W,col=10,lwd=3)

    return(par.LS, par.LS.W, par.Geman, par.Geman.W, par.HL, par.HL.W,
     par.L2L1, par.L2L1.W, par.L1, par.L1.W,sigma)
}

# "est.parameter" gives estimated parameters from function "analyze"
# into the matrix form
"est.parameter" <- function(results)
{
    a_c(results$par.LS$parameters[1],results$par.Geman$parameters[1],
     results$par.HL$parameters[1],results$par.L2L1$parameters[1],
     results$par.L1$parameters[1])
    b_c(results$par.LS$parameters[2],results$par.Geman$parameters[2],
     results$par.HL$parameters[2],results$par.L2L1$parameters[2],
     results$par.L1$parameters[2])
    L_c(results$par.LS$objective,results$par.Geman$objective,
     results$par.HL$objective, results$par.L2L1$objective,
     results$par.L1$objective)
    result_matrix(c(a,b,L), ncol=5, byrow=T)
    dimnames(result)[[2]]_
     c("LSE", "GEMAN-McCLURE", "HEBERT and LEAHY", "L2-L1", "L1" )
    dimnames(result)[[1]]_c("a","b","Objective function")

    result
}

"est.parameter.W" <- function(results)
{
    a.W_c(results$par.LS.W$parameters[1],results$par.Geman.W$parameters[1],
     results$par.HL.W$parameters[1],results$par.L2L1.W$parameters[1],
     results$par.L1.W$parameters[1])
    b.W_c(results$par.LS.W$parameters[2],results$par.Geman.W$parameters[2],
     results$par.HL.W$parameters[2],results$par.L2L1.W$parameters[2],
     results$par.L1.W$parameters[2])
    L.W_c(results$par.LS.W$objective,results$par.Geman.W$objective,
     results$par.HL.W$objective,results$par.L2L1.W$objective,
     results$par.L1.W$objective)
    result.W_matrix(c(a.W,b.W,L.W), ncol=5, byrow=T)
    dimnames(result.W)[[2]]_
     c("LSE", "GEMAN-McCLURE", "HEBERT and LEAHY", "L2-L1", "L1" )
    dimnames(result.W)[[1]]_c("a","b","Objective function")

    result.W
}
```

```
##############################################################################
# VARIANCE OF ESTIMATED PARAMETERs for MODEL 1

##############################################################################
# 1. DERIVATIVES OF REGRESSION FUNCTION / sqrt(sigma^2)

"der.m" <- function(parameters, data, sigma)
{
    # dm/da
    m1_rep(1, length(data$t))
    # dm/db
    m2_((data$t)^(parameters[2])) * log(data$t)
    # d^2 m / da^2
    m11_rep(0, length(data$t))
    # d^2 m / db^2
    m22_m2 * log(data$t)
    # d^2 m / da db
    m12_m11

    sigma_sqrt(sigma)
    matrix(c(m1,m2,m11,m22,m12)/sigma, nrow=5, byrow=T)
}

##############################################################################
# 2. DERIVATIVES OF RHO

"psi.LS" <- function(parameters,data,sigma)
{
    # psi
    psi_(data$z-model(parameters,data$t))/sqrt(sigma)
    # psi'
    psi1_rep(1,length(psi))
    return(matrix(c(psi,psi1),nrow=2, byrow=T))
}

"psi.Geman" <- function(parameters,data,sigma)
{
    # psi
    psi_((data$z-model(parameters,data$t))/sqrt(sigma))/
     (1+(data$z-model(parameters,data$t))^2/sigma)^2
    # psi'
    psi1_(1-3*(data$z-model(parameters,data$t))^2/sigma)/
     (1+(data$z-model(parameters,data$t))^2/sigma)^3
    return(matrix(c(psi,psi1),nrow=2, byrow=T))
}
```

```
"psi.HL" <- function(parameters,data,sigma)
{
    # psi
    psi_((data$z-model(parameters,data$t))/sqrt(sigma))/
     (1+(data$z-model(parameters,data$t))^2/sigma)
    # psi'
    psi1_(1-(data$z-model(parameters,data$t))^2/sigma)/
     (1+(data$z-model(parameters,data$t))^2/sigma)^2
    return(matrix(c(psi,psi1),nrow=2, byrow=T))
}


"psi.L2L1" <- function(parameters,data,sigma)
{
    # psi
    psi_((data$z-model(parameters,data$t))/sqrt(sigma))/
     (1+(data$z-model(parameters,data$t))^2/(2*sigma))^(0.5)
    # psi'
    psi1_1/(1+(data$z-model(parameters,data$t))^2/(2*sigma))^(1.5)
    return(matrix(c(psi,psi1),nrow=2, byrow=T))
}


##############################################################################
# 3. VARIANCE MATRIX

"sigma.U" <- function(psi,m,n)
{
    sigma.U11_sum(psi[1,]^2*m[1,]*m[1,])
    sigma.U22_sum(psi[1,]^2*m[2,]*m[2,])
    sigma.U12_sum(psi[1,]^2*m[1,]*m[2,])
    matrix(c(sigma.U11,sigma.U12,sigma.U12,sigma.U22), ncol=2, byrow=T)/n
}

"sigma.I" <- function(psi,m,n)
{
    sigma.I11_sum(psi[2,]*m[1,]*m[1,]) - sum(psi[1,]*m[3,])
    sigma.I22_sum(psi[2,]*m[2,]*m[2,]) - sum(psi[1,]*m[4,])
    sigma.I12_sum(psi[2,]*m[1,]*m[2,]) - sum(psi[1,]*m[5,])
    matrix(c(sigma.I11,sigma.I12,sigma.I12,sigma.I22), ncol=2, byrow=T)/n
}

##############################################################################
# 4. VARIANCE OF M-ESTIMATORs
```

```
"var.parameters.LS" <- function(parameters,data,sigma,m)
{
    psi_psi.LS(parameters,data,sigma)
    sigmaI_sigma.I(psi,m,n=length(data$N))
    sigmaU_sigma.U(psi,m,n=length(data$N))

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$N)
}

"var.parameters.Geman" <- function(parameters,data,sigma,m)
{
    psi_psi.Geman(parameters,data,sigma)
    sigmaI_sigma.I(psi,m,n=length(data$N))
    sigmaU_sigma.U(psi,m,n=length(data$N))

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$N)
}

"var.parameters.HL" <- function(parameters,data,sigma,m)
{
    psi_psi.HL(parameters,data,sigma)
    sigmaI_sigma.I(psi,m,n=length(data$N))
    sigmaU_sigma.U(psi,m,n=length(data$N))

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$N)
}

"var.parameters.L2L1" <- function(parameters,data,sigma,m)
{
    psi_psi.L2L1(parameters,data,sigma)
    sigmaI_sigma.I(psi,m,n=length(data$N))
    sigmaU_sigma.U(psi,m,n=length(data$N))

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$N)
}

# Standard error estimation of estimated parameters

# data              ... data generated by function "generate"
# mtx.est.parameter ... matrix with estimated parameters computed by
#                       function "est.parameter"

"std.parameters" <- function(mtx.est.parameter,data,sigma)
{
    # Matrix Sigma /n
    order.data.t_order(data$t)
```

```
    sorted.data_list(t=sort(data$t), z=data$z[order.data.t], N=data$N)

    m_der.m(parameters=c(mtx.est.parameter[1,1],mtx.est.parameter[2,1]),
     sorted.data,sigma)
    var.LS_var.parameters.LS(parameters=c(mtx.est.parameter[1,1],
     mtx.est.parameter[2,1]),sorted.data,sigma,m)

    m_der.m(parameters=c(mtx.est.parameter[1,2],mtx.est.parameter[2,2]),
     sorted.data,sigma)
    var.Geman_var.parameters.Geman(parameters=c(mtx.est.parameter[1,2],
     mtx.est.parameter[2,2]),sorted.data,sigma,m)

    m_der.m(parameters=c(mtx.est.parameter[1,3],mtx.est.parameter[2,3]),
     sorted.data,sigma)
    var.HL_var.parameters.HL(parameters=c(mtx.est.parameter[1,3],
     mtx.est.parameter[2,3]),data=sorted.data,sigma,m)

    m_der.m(parameters=c(mtx.est.parameter[1,4],mtx.est.parameter[2,4]),
     sorted.data,sigma)
    var.L2L1_var.parameters.L2L1(parameters=c(mtx.est.parameter[1,4],
     mtx.est.parameter[2,4]),sorted.data,sigma,m)

    # Standard errors, correlation of estimated parameters
    stda_sqrt(c(var.LS[1,1],var.Geman[1,1],var.HL[1,1],var.L2L1[1,1]))
    stdb_sqrt(c(var.LS[2,2],var.Geman[2,2],var.HL[2,2],var.L2L1[2,2]))
    corab_(c(var.LS[1,2],var.Geman[1,2],var.HL[1,2],var.L2L1[1,2])) /
     (stda * stdb )

    std.result_matrix(c(stda,stdb,corab), ncol=4, byrow=T)
    dimnames(std.result)[[2]]_
     c("LSE", "GEMAN-McCLURE", "HEBERT and LEAHY", "L2-L1" )
    dimnames(std.result)[[1]]_c("stda","stdb","corab")

    return(std.result)
}

##############################################################################
# MODEL1 - ANALYSE

model_model1

sim.data.1.error_generate(n=20,model,parameters=c(3,5), error=1)
result1.error_analyze(data=sim.data.1.error, param=c(3,5))

# Table with estimation of parameters
param1_est.parameter(result1.error)
```

```
param1.W_est.parameter.W(result1.error)

# Total number of measurements
TotalN.1_sum(sim.data.1.error$N)

# Total number of outliers
Total.errors.1_
 length(sim.data.1.error$coor.error[sim.data.1.error$coor.error>0])

# Standard errors of estimated parameters
std.result1_std.parameters(mtx.est.parameter=param1,
 sim.data.1.error, sigma=1)
std.result1.W_std.parameters(mtx.est.parameter=param1.W,
 sim.data.1.error, sigma=result1.error$sigma)
```

B.2 Simulation Study 2

```
##############################################################################
# REGRESSION FUNCTIONS m(parameters,t,x)

"model2" <- function (parameters,t,x)
{
    parameters[1]+parameters[2]*exp(-parameters[3]*x)+parameters[4]*log(t)
}

##############################################################################
# GENERATING DATA FOR PARAMETRIC REGRESSION MODELS

# Z_i^j  = m(parameters,X(T_i^j)) + Eps_i^j
# Eps    ~ N(0,std)
# T      ~ Poisson's process with lambda.small  if Z_i^j-1 < w
#                                 lambda.big    otherwise
#          in <0,1>

# Values:
#   t   ... vector of times of measurements for all n individuals
#   z   ... vector of values of model for all n individuals
#   N   ... vector of number of measurements for all n individuals
#   coor.error   ... vector of coordinates of outlying measurements

"generate.2" <- function (n,model,parameters,error)
{
    std_0.6
    lambda.small_25
```

```
lambda.big_10
w_0
# order of outlying measurement for each individual
k_1

# Initialization of variables for 1st individual and his 1st measurement
t_rexp(1,lambda.small)
x_runif(1,0,1)
if (k==1)
{
    z_model(parameters,t,x)+rnorm(1,error,std)
}
else
    z_model(parameters,t,x)+rnorm(1,0,std)
r_1
N_rep(0,n)
coor.error_rep(0,n)

# Initialization of variables for 1st individual, from his 2nd measurement
while (t[length(t)]<=1)
{
    r_r+1
    if (z[length(z)]<w)
    {
        t_c(t,t[length(t)]+rexp(1,lambda.small))
    }
    else
        t_c(t,t[length(t)]+rexp(1,lambda.big))
    x_c(x,runif(1,0,1))
    if (r==k)
    {
        z_c(z,model(parameters,t[length(t)],x[length(t)])+
         rnorm(1,error,std))
    }
    else
        z_c(z,model(parameters,t[length(t)],x[length(t)])+
         rnorm(1,0,std))
    x_c(x,runif(1,0,1))
}
N[1]_length(t)-1
t_t[1:N[1]]
z_z[1:N[1]]
x_x[1:N[1]]
counter_N[1]
if (N[1]<k)
{
```

```
    coor.error[1]_0
}
else
    coor.error[1]_k

if (n>=2)
{
for (i in 2:n)
{
    t[sum(N)+1]_rexp(1,lambda.small)
    x[sum(N)+1]_runif(1,0,1)
    if (k==1)
    {
        z[sum(N)+1]_model(parameters,t[length(t)],x[length(t)])+
         rnorm(1,error,std)
    }
    else
        z[sum(N)+1]_model(parameters,t[length(t)],x[length(t)])+
         rnorm(1,0,std)
    r_1
    while (t[length(t)]<=1)
    {
        if (z[length(z)]<w)
        {
            t_c(t,t[length(t)]+rexp(1,lambda.small))
        }
        else
            t_c(t,t[length(t)]+rexp(1,lambda.big))
        r_r+1
        x_c(x,runif(1,0,1))
        if (r==k)
        {
            z_c(z,model(parameters,t[length(t)],x[length(t)])+
             rnorm(1,error,std))
        }
        else
            z_c(z,model(parameters,t[length(t)],x[length(t)])+
             rnorm(1,0,std))
        x_c(x,runif(1,0,1))
    }
    N[i]_length(t)-1-sum(N)
    t_t[1:sum(N)]
    z_z[1:sum(N)]
    x_x[1:sum(N)]
}
}
```

```
    cumsum.N_cumsum(N)
    for (i in 2:n)
    {
        if (N[i]>=k)
        {
            coor.error[i]_cumsum.N[i-1]+k
        }
        else
            coor.error[i]_0
    }
    return(t,x,z,N,coor.error)
}

#############################################################################
# KERNEL ESTIMATION

# GRID FUNCTION
# Produce the grid used by  persp,  .. as  N x 2 matrix
# Arguments: x,y : any vectors of same mode

"xy.grid" <- function(x,y)
{
  nx <- length(x)
  ny <- length(y)
  cbind(rep(x, rep.int(ny,nx)),rep(y, nx))
}

# UNIFORM KERNEL FUNCTION
# K(y) = K(y-yi,b) = K((y-yi/b))    = 0.5 if |K|<1          yi ...data
#                                   = 0   otherwise

uniform.kernel <- function(yi, y)
{
    bandwidth <- 0.15
    if(abs((y - yi)/bandwidth) < 1)
    {
        return(0.5)
    }
    else return(0)
}

# KERNEL ESTIMATION OF THE REGRESSION FUNCTION
# Produce kernel estimation of regression function with 2 variables
# m.hat(y)  = r(y)/alpha(y)
# r.hat(y)  = 1/n * sum_i sum_j (Z_i^j * K(y-V_i^j,bandwidth))
```

```
# alpha(y)  = 1/n * sum_i sum_j (K(y-V_i^j,bandwidth))

# Arguments:    t,x : any vectors of same mode
#               data : a list of data containing 2 vector of variables X(s),
#                      and 1 vector with measured value Z

"kernel.m.2" <- function(t,x,data)
{
    k1_apply(t(data$t), 2., uniform.kernel,t)
    k2_apply(t(data$x), 2., uniform.kernel,x)
    k_k1+k2
    k[k==0.5]_0
    k_k/2
    r.hat_1/length(data$N)*t(data$z)%*%k
    alpha.hat_1/length(data$N)*sum(k)
    r.hat/alpha.hat
}


# KERNEL ESTIMATION OF THE CONDITIONAL VARIANCE
# Produce kernel estimation of variance function sigma^2()
#  in regression models with 2 variables
# Arguments:    t,x : any vectors of same mode
#               data : a list of data containing 2 vector of variables X(s)
#                      and 1 vector with measured value Z

"kernel.sigma.2" <- function(t,x,data)
{
    k1_apply(t(data$t), 2., uniform.kernel,t)
    k2_apply(t(data$x), 2., uniform.kernel,x)
    k_k1+k2
    k[k==0.5]_0
    k_k/2
    r.hat_1/length(data$N)*t(data$z)%*%k
    v.hat_1/length(data$N)*t(data$z^2)%*%k
    alpha.hat_1/length(data$N)*sum(k)
    m.hat_r.hat/alpha.hat
    v.hat/alpha.hat-m.hat^2
}

##########################################################################
# M ESTIMATIONs (for regression with 2 variables)

"distanc.LS.2" <- function(parameters,data,sigma)
{
    sum(((data$z-model(parameters,data$t,data$x))^2)/sigma)/2
```

```
}

"distanc.Geman.2" <- function(parameters,data,sigma)
{
    sum((((data$z-model(parameters,data$t,data$x))^2)/sigma)/
     (1+((data$z-model(parameters,data$t,data$x))^2)/sigma))/2
}

"distanc.HL.2" <- function(parameters,data,sigma)
{
    sum(log(1+((data$z-model(parameters,data$t,data$x))^2)/sigma))/2
}

"distanc.L2L1.2" <- function(parameters,data,sigma)
{
    sum(sqrt(1+((data$z-model(parameters,data$t,data$x))^2)/(2*sigma))-1)*2
}

"distanc.L1.2" <- function(parameters,data,sigma)
{
    sum(abs((data$z-model(parameters,data$t,data$x))/sqrt(sigma)))
}

# Plots
# Adding the drop-lines from each point to the surface of model
"plot.surface" <- function(parameters, x,y,data)
{
    xy.kernel_xy.grid(x,y)
    m_model(parameters,t=xy.kernel[,1], x=xy.kernel[,2])
    M.plot_persp(x,y,matrix(m,nrow=length(x),byrow=T),xlab="t",ylab="x",
     zlab="z",zlim=range(c(data$z,m)))
    M.line.start <- perspp(data$t,data$x,data$z,M.plot)
    points(M.line.start$x[data$coor.error], M.line.start$y[data$coor.error],
     pch='x',cex=0.75,col=8)
    points(M.line.start$x[-data$coor.error],
     M.line.start$y[-data$coor.error],cex=0.75 )
    M.line.end_perspp(data$t,data$x,model(parameters,t=data$t, x=data$x),M.plot)
    segments(x1=M.line.start$x[data$coor.error],
     y1=M.line.start$y[data$coor.error], x2=M.line.end$x[data$coor.error],
     y2=M.line.end$y[data$coor.error], col=8)
    segments(x1=M.line.start$x[-data$coor.error],
     y1=M.line.start$y[-data$coor.error], x2=M.line.end$x[-data$coor.error],
     y2=M.line.end$y[-data$coor.error])
}

##############################################################################
```

```
# ANALYSE

"analyze.2" <- function(data, param)
{
    total.N_sum(data$N)

    # Kernel estimation of the regression function
    order.data.t_order(data$t)
    rank.data.t_rank(data$t)
    sorted.data_list(t=sort(data$t), x=data$x[order.data.t],
     z=data$z[order.data.t], N=data$N)

    m_rep(0,total.N)
    for (i in 1:total.N)
        # Kernel smooth of regression function in generated data
        m[i]_kernel.m.2(t=sorted.data$t[i], x=sorted.data$x[i],sorted.data)

    # Kernel smooth of regression function computed on grid of possible data
    x_seq(min(data$t)-0.0000001,1,length=33)
    y_seq(min(data$x),1,length=33)
    xy.kernel_xy.grid(x,y)
    N.grid.kernel_nrow(xy.kernel)
    m.grid.kernel_rep(0,N.grid.kernel)
    for (i in 1:N.grid.kernel)
        m.grid.kernel[i]_kernel.m.2(t=xy.kernel[i,1], x=xy.kernel[i,2],
         sorted.data)

    # Kernel estimation of the conditional variance ... sigma^2
    sigma_rep(0,total.N)
    for (i in 1:total.N)
        # Kernel smooth of regression function in generated data
        sigma[i]_kernel.sigma.2(t=sorted.data$t[i], x=sorted.data$x[i],
         sorted.data)

    par.LS_nlminb(rep(1,length(param)),distanc.LS.2, data, sigma=1)
    par.LS.W_nlminb(par.LS$parameters,distanc.LS.2, sorted.data, sigma)

    par.Geman_nlminb(par.LS$parameters,distanc.Geman.2, data, sigma=1)
    par.Geman.W_nlminb(par.LS$parameters,distanc.Geman.2, sorted.data, sigma)

    par.HL_nlminb(par.LS$parameters,distanc.HL.2, data=data, sigma=1)
    par.HL.W_nlminb(par.LS$parameters,distanc.HL.2, sorted.data, sigma)

    par.L2L1_nlminb(par.LS$parameters,distanc.L2L1.2, data, sigma=1)
    par.L2L1.W_nlminb(par.LS$parameters,distanc.L2L1.2, sorted.data, sigma)
```

```
    par.L1_nlminb(par.LS$parameters,distanc.L1.2, data=data, sigma=1)
    par.L1.W_nlminb(par.LS$parameters,distanc.L1.2, sorted.data, sigma)

    # Plots

    # TRUE regression function
    plot.surface(parameters=param, x,y,data)

    # KERNEL estimation of regression function
    # if bandwidth is too small, kernel estimation is NA
    if (length(m.grid.kernel[is.na(m.grid.kernel)==T])==0)
    {
        M.plot <- persp(x,y,matrix(m.grid.kernel,nrow=length(x),byrow=T),
         xlab="t",ylab="x",zlab="z", zlim=range(c(data$z,m.grid.kernel)))
        M.line.start <- perspp(data$t,data$x,data$z,M.plot)
        points(M.line.start$x[data$coor.error], M.line.start$y[data$coor.error],
         pch='x',cex=0.75,col=8)
        points(M.line.start$x[-data$coor.error],
         M.line.start$y[-data$coor.error],cex=0.75 )
        M.line.end <- perspp(data$t,data$x,m[rank.data.t],M.plot)
        segments(x1=M.line.start$x[data$coor.error],
         y1=M.line.start$y[data$coor.error], x2=M.line.end$x[data$coor.error],
         y2=M.line.end$y[data$coor.error], col=8)
        segments(x1=M.line.start$x[-data$coor.error],
         y1=M.line.start$y[-data$coor.error], x2=M.line.end$x[-data$coor.error],
         y2=M.line.end$y[-data$coor.error])
    }

    # LS estimation
    plot.surface(parameters=par.LS$parameters, x,y,data)

    # Geman estimation
    plot.surface(parameters=par.Geman$parameters, x,y,data)

    # LS.W estimation
    plot.surface(parameters=par.LS.W$parameters, x,y,data)

    # Geman.W estimation
    plot.surface(parameters=par.Geman.W$parameters, x,y,data)

    return(par.LS, par.LS.W, par.Geman, par.Geman.W, par.HL, par.HL.W, par.L2L1,
     par.L2L1.W, par.L1, par.L1.W,sigma)
}

# "est.parameter" gives estimated parameters from "analyze" into the matrix form
"est.parameter.2" <- function(results)
```

```
{
    a_c(results$par.LS$parameters[1],results$par.Geman$parameters[1],
     results$par.HL$parameters[1],results$par.L2L1$parameters[1],
     results$par.L1$parameters[1])
    b_c(results$par.LS$parameters[2],results$par.Geman$parameters[2],
     results$par.HL$parameters[2],results$par.L2L1$parameters[2],
     results$par.L1$parameters[2])
    c_c(results$par.LS$parameters[3],results$par.Geman$parameters[3],
     results$par.HL$parameters[3],results$par.L2L1$parameters[3],
     results$par.L1$parameters[3])
    d_c(results$par.LS$parameters[4],results$par.Geman$parameters[4],
     results$par.HL$parameters[4],results$par.L2L1$parameters[4],
     results$par.L1$parameters[4])
    L_c(results$par.LS$objective,results$par.Geman$objective,
     results$par.HL$objective,results$par.L2L1$objective,
     results$par.L1$objective)
    result_matrix(c(a,b,c,d,L), ncol=5, byrow=T)
    dimnames(result)[[2]]_
     c("LSE", "GEMAN-McCLURE", "HEBERT and LEAHY", "L2-L1", "L1" )
    dimnames(result)[[1]]_c("a","b","c","d","Objective function")

    result
}

"est.parameter.W.2" <- function(results)
{
    a.W_c(results$par.LS.W$parameters[1],results$par.Geman.W$parameters[1],
     results$par.HL.W$parameters[1],results$par.L2L1.W$parameters[1],
     results$par.L1.W$parameters[1])
    b.W_c(results$par.LS.W$parameters[2],results$par.Geman.W$parameters[2],
     results$par.HL.W$parameters[2],results$par.L2L1.W$parameters[2],
     results$par.L1.W$parameters[2])
    c.W_c(results$par.LS.W$parameters[3],results$par.Geman.W$parameters[3],
     results$par.HL.W$parameters[3],results$par.L2L1.W$parameters[3],
     results$par.L1.W$parameters[3])
    d.W_c(results$par.LS.W$parameters[4],results$par.Geman.W$parameters[4],
     results$par.HL.W$parameters[4],results$par.L2L1.W$parameters[4],
     results$par.L1.W$parameters[4])
    L.W_c(results$par.LS.W$objective,results$par.Geman.W$objective,
     results$par.HL.W$objective,results$par.L2L1.W$objective,
     results$par.L1.W$objective)
    result.W_matrix(c(a.W,b.W,c.W,d.W,L.W), ncol=5, byrow=T)
    dimnames(result.W)[[2]]_
     c("LSE", "GEMAN-McCLURE", "HEBERT and LEAHY", "L2-L1", "L1" )
    dimnames(result.W)[[1]]_c("a","b","c","d","Objective function")
```

```
    result.W
}

##############################################################################
# VARIANCE OF ESTIMATED PARAMETERs for MODEL 2

##############################################################################
# 1. DERIVATIVES OF REGRESSION FUNCTION / sqrt(sigma^2)

"der.m.2" <- function(parameters,data,sigma)
{
    # dm/da
    m1_rep(1, length(data$t))
    # dm/db
    m2_exp(-1*(parameters[3]) * data$x)
    # dm/dc
    m3_-data$x * parameters[2] * m2
    # dm/dd
    m4_log(data$t)
    # d^2 m / da^2
    m11_rep(0,length(data$t))
    # d^2 m / da db
    m12_m11
    # d^2 m / da dc
    m13_m11
    # d^2 m / da dd
    m14_0
    # d^2 m / db^2
    m22_m11
    # d^2 m / db dc
    m23_-data$x * exp(-parameters[3] * data$x)
    # d^2 m / db dd
    m24_m11
    # d^2 m / dc^2
    m33_data$x^2 * parameters[2] * exp(-parameters[3] * data$x)
    # d^2 m / dc dd
    m34_m11
    # d^2m / dd^2
    m44_m11

    sigma_sqrt(sigma)
    return(rbind(m1,m2,m3,m4,m11,m12,m13,m14,m22,m23,m24,m33,m34,m44)/sigma)
}

##############################################################################
# 2. DERIVATIVES OF RHO
```

```
"psi.LS.2" <- function(parameters,data,sigma)
{
    psi_(data$z-model(parameters,data$t,data$x))/sqrt(sigma)
    psi1_rep(1,length(psi))
    return(rbind(psi,psi1))
}

"psi.Geman.2" <- function(parameters,data,sigma)
{
    psi_((data$z-model(parameters,data$t,data$x))/sqrt(sigma))/
     (1+(data$z-model(parameters,data$t,data$x))^2/sigma)^2
    psi1_(1-3*(data$z-model(parameters,data$t,data$x))^2/sigma)/
     (1+(data$z-model(parameters,data$t,data$x))^2/sigma)^3
    return(rbind(psi,psi1))
}


"psi.HL.2" <- function(parameters,data,sigma)
{
    psi_((data$z-model(parameters,data$t,data$x))/sqrt(sigma))/
     (1+(data$z-model(parameters,data$t,data$x))^2/sigma)
    psi1_(1-(data$z-model(parameters,data$t,data$x))^2/sigma)/
     (1+(data$z-model(parameters,data$t,data$x))^2/sigma)^2
    return(rbind(psi,psi1))
}


"psi.L2L1.2" <- function(parameters,data,sigma)
{
    psi_((data$z-model(parameters,data$t,data$x))/sqrt(sigma))/
     (1+(data$z-model(parameters,data$t,data$x))^2/(2*sigma))^(0.5)
    psi1_1/(1+(data$z-model(parameters,data$t,data$x))^2/(2*sigma))^(1.5)
    return(rbind(psi,psi1))
}


##############################################################################
# 3. VARIANCE MATRIX

"sigma.U.2" <- function(psi,m)
{
    n_length(psi[1,])
    sigma.U11_sum(psi[1,]^2*m[1,]*m[1,])
    sigma.U22_sum(psi[1,]^2*m[2,]*m[2,])
    sigma.U33_sum(psi[1,]^2*m[3,]*m[3,])
```

```
    sigma.U44_sum(psi[1,]^2*m[4,]*m[4,])
    sigma.U12_sum(psi[1,]^2*m[1,]*m[2,])
    sigma.U13_sum(psi[1,]^2*m[1,]*m[3,])
    sigma.U14_sum(psi[1,]^2*m[1,]*m[4,])
    sigma.U23_sum(psi[1,]^2*m[2,]*m[3,])
    sigma.U24_sum(psi[1,]^2*m[2,]*m[4,])
    sigma.U34_sum(psi[1,]^2*m[3,]*m[4,])

    matrix(c(sigma.U11,sigma.U12,sigma.U13,sigma.U14, sigma.U12,sigma.U22,
     sigma.U23,sigma.U24, sigma.U13,sigma.U23,sigma.U33,sigma.U34, sigma.U14,
     sigma.U24,sigma.U34,sigma.U44), ncol=4, byrow=T)/n
}

"sigma.I.2" <- function(psi,m)
{
    n_length(psi[1,])
    sigma.I11_sum(psi[2,]*m[1,]*m[1,] - psi[1,]*m[5,])
    sigma.I22_sum(psi[2,]*m[2,]*m[2,] - psi[1,]*m[9,])
    sigma.I33_sum(psi[2,]*m[3,]*m[3,] - psi[1,]*m[12,])
    sigma.I44_sum(psi[2,]*m[4,]*m[4,] - psi[1,]*m[14,])
    sigma.I12_sum(psi[2,]*m[1,]*m[2,] - psi[1,]*m[6,])
    sigma.I13_sum(psi[2,]*m[1,]*m[3,] - psi[1,]*m[7,])
    sigma.I14_sum(psi[2,]*m[1,]*m[4,] - psi[1,]*m[8,])
    sigma.I23_sum(psi[2,]*m[2,]*m[3,] - psi[1,]*m[10,])
    sigma.I24_sum(psi[2,]*m[2,]*m[4,] - psi[1,]*m[11,])
    sigma.I34_sum(psi[2,]*m[3,]*m[4,] - psi[1,]*m[13,])
    matrix(c(sigma.I11,sigma.I12,sigma.I13,sigma.I14, sigma.I12,sigma.I22,
     sigma.I23,sigma.I24, sigma.I13,sigma.I23,sigma.I33,sigma.I34,
     sigma.I14,sigma.I24,sigma.I34,sigma.I44), ncol=4, byrow=T)/n
}

##############################################################################
# 4. VARIANCE OF M-ESTIMATORs

"var.parameters.LS.2" <- function(parameters,data,sigma,m)
{
    psi_psi.LS.2(parameters,data,sigma)
    sigmaI_sigma.I.2(psi,m)
    sigmaU_sigma.U.2(psi,m)

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$t)
}

"var.parameters.Geman.2" <- function(parameters,data,sigma,m)
{
    psi_psi.Geman.2(parameters,data,sigma)
```

```
    sigmaI_sigma.I.2(psi,m)
    sigmaU_sigma.U.2(psi,m)

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$t)
}

"var.parameters.HL.2" <- function(parameters,data,sigma,m)
{
    psi_psi.HL.2(parameters,data,sigma)
    sigmaI_sigma.I.2(psi,m)
    sigmaU_sigma.U.2(psi,m)

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$t)
}

"var.parameters.L2L1.2" <- function(parameters,data,sigma,m)
{
    psi_psi.L2L1.2(parameters,data,sigma)
    sigmaI_sigma.I.2(psi,m)
    sigmaU_sigma.U.2(psi,m)

    (solve(sigmaI) %*% sigmaU %*% solve(sigmaI))/length(data$t)
}

# Standard error estimation of estimated parameters

# data              ... data generated by function "generate.2"
# mtx.est.parameter ... matrix with estimated parameters computed by
#                       function "est.parameter.2" or "est.parameter.W.2"

"std.parameters.2" <- function(mtx.est.parameter,data,sigma)
{
    # matrix Sigma /n
    order.data.t_order(data$t)
    sorted.data_list(t=sort(data$t), x=data$x[order.data.t],
     z=data$z[order.data.t], N=data$N)

    m_der.m.2(data=sorted.data,parameters=c(mtx.est.parameter[1,1],
     mtx.est.parameter[2,1],mtx.est.parameter[3,1],mtx.est.parameter[4,1]),sigma)
    var.LS_var.parameters.LS.2(parameters=c(mtx.est.parameter[1,1],
     mtx.est.parameter[2,1],mtx.est.parameter[3,1],mtx.est.parameter[4,1]),
     sorted.data,sigma,m)

    m_der.m.2(data=sorted.data,parameters=c(mtx.est.parameter[1,2],
     mtx.est.parameter[2,2],mtx.est.parameter[3,2],mtx.est.parameter[4,2]),sigma)
    var.Geman_var.parameters.Geman.2(parameters=c(mtx.est.parameter[1,2],
```

```
     mtx.est.parameter[2,2],mtx.est.parameter[3,2],mtx.est.parameter[4,2]),
     sorted.data,sigma,m)

    m_der.m.2(data=sorted.data,parameters=c(mtx.est.parameter[1,3],
     mtx.est.parameter[2,3],mtx.est.parameter[3,3],mtx.est.parameter[4,3]),sigma)
    var.HL_var.parameters.HL.2(parameters=c(mtx.est.parameter[1,3],
     mtx.est.parameter[2,3],mtx.est.parameter[3,3],mtx.est.parameter[4,3]),
     sorted.data,sigma,m)

    m_der.m.2(data=sorted.data,parameters=c(mtx.est.parameter[1,4],
     mtx.est.parameter[2,4],mtx.est.parameter[3,4],mtx.est.parameter[4,4]),sigma)
    var.L2L1_var.parameters.L2L1.2(parameters=c(mtx.est.parameter[1,4],
     mtx.est.parameter[2,4],mtx.est.parameter[3,4],mtx.est.parameter[4,4]),
     sorted.data,sigma,m)

    # Standard errors, correlation of estimated parameters
    stda_sqrt(c(var.LS[1,1],var.Geman[1,1],var.HL[1,1],var.L2L1[1,1]))
    stdb_sqrt(c(var.LS[2,2],var.Geman[2,2],var.HL[2,2],var.L2L1[2,2]))
    stdc_sqrt(c(var.LS[3,3],var.Geman[3,3],var.HL[3,3],var.L2L1[3,3]))
    stdd_sqrt(c(var.LS[4,4],var.Geman[4,4],var.HL[4,4],var.L2L1[4,4]))

    corab_(c(var.LS[1,2],var.Geman[1,2],var.HL[1,2],var.L2L1[1,2])) /
     (stda * stdb )
    corac_(c(var.LS[1,3],var.Geman[1,3],var.HL[1,3],var.L2L1[1,3])) /
     (stda * stdc )
    corad_(c(var.LS[1,4],var.Geman[1,4],var.HL[1,4],var.L2L1[1,4])) /
     (stda * stdd )
    corbc_(c(var.LS[2,3],var.Geman[2,3],var.HL[2,3],var.L2L1[2,3])) /
     (stdb * stdc )
    corbd_(c(var.LS[3,4],var.Geman[3,4],var.HL[3,4],var.L2L1[3,4])) /
     (stdb * stdd )


    std.result_matrix(
     c(stda,stdb,stdc,stdd,corab,corac,corad,corbc,corbd), ncol=4, byrow=T)
    dimnames(std.result)[[2]]_
     c("LSE", "GEMAN-McCLURE", "HEBERT and LEAHY", "L2-L1" )
    dimnames(std.result)[[1]]_
     c("stda","stdb","stdc","stdd","corab","corac","corad","corbc","corbd")

    return(std.result)
}

#########################################################################
# MODEL2 - ANALYSE
model_model2
```

```
sim.data.2.error_generate.2(n=20,model,parameters=c(1,2,3,0.5), error=10)
result2.error_analyze.2(data=sim.data.2.error, param=c(1,2,3,0.5))

param2_est.parameter.2(result2.error)
param2.W_est.parameter.W.2(results=result2.error)

TotalN.2_sum(sim.data.2.error$N)
Total.errors.2_
 length(sim.data.2.error$coor.error[sim.data.2.error$coor.error>0])

std.result.2_std.parameters.2(mtx.est.parameter=param2,sim.data.2.error,sigma=1)
std.result.2.W_std.parameters.2(param2.W,sim.data.2.error,result2.error$sigma)
```