Robust estimation of marginal regression parameters in clustered data

Somnath Datta[1] and **James D Beck**[2]

[1]University of Louisville, Louisville, KY 40202, USA

[2]University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Abstract

We develop robust methods for analyzing clustered data where estimation of marginal regression parameters is of interest. Inverse cluster size reweighting in the objective function to be minimized is incorporated to handle the issue of informative cluster size. Performance of the resulting estimators is studied by simulation. Large sample inference and variance estimation is carried out. The methodology is illustrated using a periodontal disease dataset.

Keywords

Informative cluster size; random cluster size; R estimator; dental data

1 Introduction

The method of least squares is applied most often to estimate the parameters of a linear regression model. It is equivalent to maximum likelihood estimation if one assumes that the model errors are normally distributed. However, these estimates are sensitive to the presence of outliers in the data. More robust estimators of the regression parameters can be obtained by partly replacing the residual by its rank in the objective function of the least squares criterion leading to the so called R-estimator (Jurecková, 1971; Jaeckel, 1972; McKean and Hettmansperger, 1978). The classical regression setup assumes that the data (response) are independent, though not identically distributed.

We consider R-estimation of regression parameters in a linear model when the data are clustered so that the observations (responses) within a cluster are correlated, but data belonging to different clusters are independent. Let M denote the number of clusters and Y_{ij} denote the jth value of the response variable in the ith cluster $1 \leq j \leq n_i$, $1 \leq i \leq M$, where for each cluster i, n_i denotes its size. In addition, suppose we have an observed covariate vector X_{ij} which affects the distribution of Y_{ij} through a marginal linear model

$$Y_{ij} = \alpha + \beta^T X_{ij} + \epsilon_{ij}, \quad (1.1)$$

Address for correspondence: Somnath Datta, University of Louisville, Louisville, KY 40202, USA. somnath.datta@louisville.edu.

where $\in_{ij}$ are the model errors within cluster i with a common distribution F_i. A common mathematically convenient assumption in modelling clustered data is that the errors $\in_{ij}$ are exchangeable. Although for many applications such an assumption may hold, we avoid making a within cluster exchangeability assumption on $\in$ for the broadest possible applicability of the proposed methodology (including dental data). Most of the existing approaches treat n_i to be non-random and assumptions are made on each F_i; see, e.g., Jung and Ying (2003); Wang and Zhu (2006); Wang and Zhao (2008); Hettmansperger and McKean (2011).

We consider the possibility that the cluster size is random and informative (Hoffman *et al.*, 2001; Williamson *et al.*, 2003; Gansky and Neuhaus, 2009; Nevalainen *et al.*, 2013) in which case Y_{ij} may be statistically correlated with the cluster size n_i. In practice, this could arise when the cluster size depends on a cluster level latent factor (such as a random effect), a cluster level covariate, or both. For a more formal definition of informative cluster size in a regression setting, see Nevalainen *et al.* (2013). We give an example of this in Section 5 where the use of standard R estimator leads to substantially biased estimators of certain parameters.

The rest of the article is organized as follows: The R-estimating functions are introduced in the next section. Theoretical (large sample) results are developed in Section 3. Simulation results are presented in Section 4 where we compare the R-estimators with a weighted least squares estimator and a mixed effects model based estimator. We illustrate the use of our estimator with the dental data in Section 5 which motivated the statistical methodology developed in this article. We study the marginal effects of covariates such as smoking on attachment loss in a sample of periodontal disease patients. The clusters in this application are all remaining teeth belonging to the same individual. Since the number of remaining teeth may be indicative of a patient's overall oral health, the cluster size is potentially informative. Also, the quantitative outcomes in this dataset were non-normal, making the use of robust methods more appealing. The article concludes with a discussion section (Section 6).

2 Methods

We propose to estimate the vector of marginal regression parameters β by minimizing the following inverse cluster size weighted objective function:

$$D_M(\beta) = \sum_{i=1}^{M} n_i^{-1} \sum_{j=1}^{n_i} \varphi(\overline{F}(e_{ij}(\beta);\beta)) e_{ij}(\beta), \quad (2.1)$$

i.e.,

$$\hat{\beta} = argmin \, D_M(\beta),$$

where $e_{ij}(\beta)=Y_{ij}-\beta^T X_{ij}, \overline{F}(\cdot;\beta)=(M+1)^{-1}\sum_{i=1}^{M}F_i(\cdot;\beta)$, and $F_i(\cdot;\beta)$ is the empirical

distribution of $\{e_{ij}(\beta), 1 \quad j \quad n_i\}$; that is $F_i(u;\beta)=n_i^{-1}\sum_{j=1}^{ni}I(e_{ij}(\beta)\leq u)$ for $u\in\Re$ Here ϕ defined on $(0, 1)$ such that $\int\phi=0$ and $\int\phi^2<\infty$. Note that due to the presence of the factor of n_i^{-1}, the resulting estimator differs from the traditional R-estimator for clustered data (Hettmansperger and McKean, 2011, Ch. 5) which is obtained by minimizing the objective function:

$$D_M^*(\beta)=\sum_{i=1}^{M}\sum_{j=1}^{n_i}\varphi\left(\frac{R(e_{ij}(\beta))}{M+1}\right)e_{ij}(\beta), \quad (2.2)$$

where $1 \quad R(e_{ij}(\beta) \quad n$ is the rank $e_{ij}(\beta)$ in the pooled sample of model residuals $\{e_{11}(\beta),\ldots,$

$e_{MnM}(\beta)\}$ and $n=\sum_{i=1}^{M}ni$ is the total sample size. The readers may consult Hettmansperger and McKean (2011) for obtaining the necessary insights for the workings of the R-estimators in the case of independent (e.g., non-clustered) data. As we shall see from the simulation results, the estimators obtained from (2.2) could be seriously biased in an informative cluster size setup. In order to differentiate between the two sets of estimators, we call our estimators derived from (2.1) 'reweighted R-estimators'. The reason for using the inverse cluster size weighting is that each cluster (e.g., each patient in a dental study) should contribute the same amount to the marginal estimating function irrespective of its size. While these resulting estimators will be consistent (and asymptotically unbiased) irrespective of whether the cluster size is non-informative or not, methods that do not balance the weight of each cluster may lead to inconsistency and may exhibit substantial bias when the cluster size is informative (Williamson *et al.*, 2003; Wang *et al.*, 2011).

Once the regression parameter β is estimated, the marginal location parameter α can be estimated by

$$\hat{\alpha}=\inf\{t:\overline{F}_{\hat{\beta}}(t)\geq 1/2\}. \quad (2.3)$$

Once again, this is different from the traditional R-estimator of α which is taken to be the sample median of the residuals $\{e_{11}(\hat{\beta_R}),\ldots, e_{MnM}(\hat{\beta_R})\}$, where $\hat{\beta_R}$ is the R-estimator of β.

We undertake an extensive simulation study in Section 4 comparing the performances of these two sets of R-estimators.

3 Large sample inference

A careful formulation of the estimation problem and technical arguments for its asymptotic analysis will be necessary since a zero median (or mean) property for the ϵ_{ij} conditioning on the cluster size n_i may not hold when the cluster size is informative. This necessitates us to formulate our assumptions on the overall marginal distribution of the errors given by:

$$F(x) = E\{M^{-1}\sum_{i=1}^{M} n_i^{-1}\sum_{j=1}^{n_i} I(\in_{ij} \leq x)\}. \quad (3.1)$$

Note that F can be regarded as the distribution of the model error associated with a typical measurement (i.e., chosen at random from all units in that cluster) of a typical cluster (i.e., chosen at random from all available clusters). Mathematically speaking, consider two random indices I and J such that $I \sim$ uniform $\{1, \ldots, M\}$ and, given $I = i$, $J \sim$ uniform $\{1, \ldots, n_i\}$. Then $F(t) = \Pr\{\in_{IJ} \quad t\}$, for $t \in \Re$.

We assume $T(F) = 0$, where T is the median; other location functionals can be used as well which will lead to the corresponding estimators of the intercept parameter α.

Let, without loss of generality, the true β be 0. One can show that D_M is almost everywhere differentiable and $\hat{\beta}$ satisfies the estimating equation:

$$S_M(\beta) := \sum_{i=1}^{M} n_i^{-1}\sum_{j=1}^{n_i} \varphi(\overline{F}(e_{ij}(\beta);\beta))X_{ij} = 0. \quad (3.2)$$

By similar argument as in Datta *et al.* (2012), $M^{-1/2}S_M(0) \xrightarrow{d} N(0, \sum)$, where

$$\sum = \lim_{M \to \infty} M^{-1}\sum_{i=1}^{M}\sum_{i}, \quad (3.3)$$

with $\sum_i = Var(n_i^{-1}\sum_{i=1}^{ni} \varphi(\overline{F}(e_{ij}(\beta);\beta))X_{ij})$ Next, mimicking the expansions for R-estimators from Hettmansperger and McKean (2011, Ch. 3), we can obtain the following expansion under our setup:

$$M^{-1/2}S_M(\beta) = M^{-1/2}S_M(0) - \tau^{-1}\left(M^{-1}\sum_{i=1}^{M} n_i^{-1}\sum_{j=1}^{n_i} X_{ij}X_{ij}^T\right)\sqrt{M}\beta + _{OP}(1.1),$$

in a local neighborhood of 0 (the true β) and hence

$$\sqrt{M}\hat{\beta} \xrightarrow{d} N(0, \tau^2\Gamma^{-1}\Sigma\Gamma^{-1}), \quad (3.4)$$

with

$\Gamma = \text{plim}_{M \to \infty} M^{-1}\sum_{i=1}^{M} n_i^{-1}\sum_{j=1}^{ni} X_{ij}X_{ij}^T, \tau = [\int \varphi(u)\varphi_f(u)du]^{-1}, \varphi_f = -(f'\circ F^{-1})/(f'\circ F^{-1})$, where f and f' are the first and the second derivatives, respectively, of F given by (3.1) The details of the technical arguments (cf. Datta *et al.*, 2012), which we omit here, show that besides moment conditions for certain cluster averages needed for an application of the central limit theorem we also need an L_1 equicontinuity condition to hold (which is satisfied by the identity score $\phi(x) = x$). This may be regarded as somewhat more stringent than what is required in the classical i.i.d. case and may not hold for certain unbounded scores. A

practical way around in such cases would be to use a truncated (bounded) version of the score if the use of such scores is desired.

For making statistical inferences based on this asymptotic normality result, we need to estimate the asymptotic variance covariance matrix. To that end, we will first discuss the estimation of τ. Our estimator, when specialized to the case of independent (i.e., non-clustered) data, is different from (and arguably simpler than) the estimator proposed by Koul *et al*. (1987). Note that:

$$\tau^{-1} = -E(\varphi(F(\in_{IJ}))S(\in_{IJ})),$$

where $S = (\log f)'$ and $\in_{ij}$ was defined as before. Therefore, if F and $\in_{ij}$ were known, a

consistent estimator of τ^{-1} would be given by $-M^{-1}\sum_{i=1}^{M} n_i^{-1}\sum_{j=1}^{ni}\varphi(F(\epsilon_{ij}))S(\epsilon_{ij})$. We, however, can replace them by their observed data based counterparts to obtain

$$\hat{\tau} = \left[-\frac{1}{M}\sum_{i=1}^{M}\frac{1}{n_i}\sum_{j=1}^{n_i}\varphi(\overline{F}(\hat{\in}_{ij};\hat{\beta}))\hat{S}(\hat{\in}_{ij}) \right]^{-1}, \quad (3.5)$$

where $\hat{\in}_{ij} = e_{ij}(\hat{\beta}) - \hat{\alpha}$,

$$\hat{S}(e) = (2h)^{-1}\log\left(\frac{\hat{f}(e+h)}{\hat{f}(e+h)} \right),$$

and

$$\hat{f}(e) = \frac{1}{Mh}\sum_{i=1}^{M}\frac{1}{n_i}\sum_{j=1}^{n_i}K\left(\frac{\hat{e}_{ij} - e}{h} \right).$$

Here h ($\downarrow 0$) is a bandwidth sequence and K is a density kernel.

Finally, the asymptotic variance–covariance matrix of $\hat{\beta}$ can be estimated from data by:

$$\hat{A}Var(\hat{\beta}) = \hat{\tau}^2\hat{\Gamma}^{-1}\hat{\Sigma}\hat{\Gamma}^{-1}/M, \quad (3.6)$$

with

$$\hat{\Gamma} = \frac{1}{M}\sum_{i=1}^{M}n_i^{-1}\sum_{j=1}^{n_i}X_{ij}X_{ij}^{T},$$

and

$$\hat{\Sigma} = \frac{1}{M} \sum_{i=1}^{M} \left\{ n_i^{-1} \sum_{j=1}^{n_i} \varphi(\overline{F}(e_{ij}(\hat{\beta}); \hat{\beta})) X_{ij} \right\} \left\{ n_i^{-1} \sum_{j=1}^{n_i} \varphi(\overline{F}(e_{ij}(\hat{\beta}); \hat{\beta})) X_{ij} \right\}^T.$$

Next, combining the arguments of Datta *et al.* (2012) with Hettmansperger and McKean (2011, equation 3.5.22), we get:

$$\sqrt{M}(\hat{\alpha} - \alpha_0) = \tau_0 M^{-12} \sum_{i=1}^{M} n_i^{-1} \sum_{j=1}^{n_i} \mathrm{sgn}(\in_{ij}) + _{O_p}(1.1),$$

where α_0 is the true value of α, and

$$\tau_0 = (2F'(F^{-1}(0.5)))^{-1}.$$

Therefore,

$$\sqrt{M}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \sigma_\alpha^2), \quad (3.7)$$

where

$$\sigma_\alpha^2 = \tau_0^2 \sigma^2,$$

with $\sigma^2 = \lim_{M \to \infty} M^{-1} \sum_{i=1}^{M} var(n_i^{-1} \sum_{j=1}^{ni} \mathrm{sgn}(\epsilon_{ij}))$, assuming this limit exists.

An estimator of the asymptotic variance of $\hat{a}$ is given by:

$$\hat{A}Var(\hat{\alpha}) = \hat{\tau}_0^2 \hat{\sigma}^2 / M, \quad (3.8)$$

where

$$\hat{\tau}_0^2 = \left\{ \frac{2}{M\tilde{h}} \sum_{i=1}^{M} \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{K}(\frac{\hat{\alpha} - \hat{\in}_{ij}}{\tilde{h}}) \right\}^{-2}$$

and

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^{M} \left\{ n_i^{-1} \sum_{j=1}^{n_i} \mathrm{sgn}(\hat{\in}_{ij}) \right\}^2.$$

Here, $\hat{\epsilon}_{ij} = Y_{ij} - \hat{\alpha} - X'_{ij}\hat{\beta}$, $\tilde{K}$ is a density kernel and $\tilde{h}$ is another bandwidth sequence.

Theoretical investigation of the issue of optimal selection of h and $\tilde{h}$ is beyond the scope of the present article. In addition, one may have to make additional assumptions beyond the marginal model for this purpose. It may be possible to obtain a data-based selector minimizing a criterion function computed via resampling. In this article we have used $h = M^{-1/7}$ and $\tilde{h} = M^{-1/5}$, respectively, in order to reduce the computational burden for the Monte Carlo simulations conducted next. Assuming an extreme form of cluster dependence, these would correspond to asymptotically optimal rates of L_2 estimation of a density and its derivative respectively (Singh, 1979).

4 Simulation

We consider a data generation scheme with M clusters. Two choices of M (50 and 100) were considered. First we generate a cluster specific random effects term μ_i from a mean zero normal distribution with standard deviations ranging from 1 to 5. More specifically, let $\mu_i \sim N(0, \sigma_i^2)$, where $\sigma_i = 5$ if i is divisible by 5, and $= i$ mod 5, otherwise, $1 \leq i \leq M$. We also generate a cluster level binary covariate Z_i taking values ± 1 with equal probabilities. Informative cluster size is generated by relating it with both the latent variable μ_i and the cluster level covariate as follows:

$$n_i = \begin{cases} 2, & \text{if } \mu_i < 0.5 \text{ and } Z_i = 1, \\ 8, & \text{if } \mu_i < 0.5 \text{ and } Z_i = 1, \\ 15, & \text{otherwise.} \end{cases}$$

Another individual level covariate W_{ij} is generated independently for each individual following the standard normal distribution. Model errors η_{ij} are generated following a distribution $G(x) = G_0(10x)$, where we make three choices for G_0. The measurements with a given cluster i are then generated using the following linear model $Y = \mu_i + 3Z_i + W_{ij} + \eta_{ij}$. Note that data generated this way satisfy the marginal model (1.1), with $\alpha = 0$, $\beta_1 = 3$, $\beta_2 = 1$, $X_{ij} = (Z_i, W_{ij})^T$, $\epsilon_{ij} = \mu_i \eta_{ij}$. For this investigation we use the normalized identity score

$$\varphi(x) = \sqrt{12}(x - \frac{1}{2}).$$

In Table 1, we report the bias and the standard deviation of four sets of competing estimators including two R-estimators, one incorporating the inverse cluster size weighting (weighted R) and one without (naive R). Each of these entries is computed using a Monte Carlo sample size of 500. As can be seen from this table, the naive R-estimators for the intercept term α and β_1, the coefficient of Z, are biased whereas that of β_2 is not. This is due to the fact that in these simulations, the cluster size is related both to the random effects and the covariate Z but not to W. This phenomenon was observed in literature with estimators obtained from generalized estimating equations (Neuhaus and McCulloch, 2011; Wang et al., 2011). The weighted R-estimators, on the other hand, are nearly unbiased.

We have also investigated the behaviour of the least squares estimator (LSE) and the one obtained from a classical mixed effects modelling of the data. To be fair, we use the same inverse cluster size reweighting in computing the least squares estimator since without these weights, the LSE for α and β_1 will be biased just like the naive R-estimator (details not

shown). It turns out that both these estimators perform fairly well for the normal and double exponential errors under the cluster size distribution considered here. However, they exhibit substantial bias and/or variance in case of a heavy tailed error (Pareto) distribution. Note that for Pareto errors, the response does not have a finite first or second moments and hence these estimators fail to be consistent and asymptotically normal. Thus, overall, the weighted R-estimator is the only one to have low bias and variance in all cases.

We also investigate the performance of the variance estimates for our weighted R-estimator. In Column 7, we report the average estimated standard errors over the Monte Carlo samples. These values seem to be in good agreement with the empirical standard errors reported in parentheses of Column 6. Finally, the last column reports the true (empirical) coverage of a nominal *95%* large sample confidence interval obtained using the weighted R-estimator and the corresponding estimated standard errors. This is the empirical proportion of the 500 confidence intervals containing the true parameter. The coverage appears to be reasonable and generally improves with the number of clusters.

5 Real data example

We illustrate our methodology on a periodontal dataset extracted from the Piedmont *65* + Dental Study (Beck *et al*., 1990). The Piedmont Health Study of the Elderly (Blazer and George, 2004), which is the parent study for this Piedmont *65* + Dental Study, is a longitudinal study of the health status of people aged *65* and over in five contiguous North Carolina counties. The Piedmont *65* + Dental Study takes advantage of the data available from the parent study while collecting additional information by means of an interview, oral examination, and microbiological and salivary assays.

Our response here is the total attachment loss which is measured at the tooth level. We apply our robust regression technique to study the effect of two potentially important covariates, tobacco use and socioeconomic status (SEIRSP), on the periodontal condition of a patient as measured by the total attachment loss of a typical remaining tooth. Since the sampling proportions for the blacks and whites were different for the parent study, the data for the two races should be analyzed separately in order to avoid any potential bias. For this reporting, we only use the data for the white patients at sixty months from study enrolment.

Here, the teeth belonging to each patient form a cluster and the cluster size is the number of remaining teeth at sixty months. It ranged from 1 to 32; 8 was the mode and 14 was the median. Since the tooth level data are clustered within each patient and the patients with fewer remaining teeth during the study tend to have greater attachment loss (possibly linked by health style choice such as tobacco use and possible latent factors such as oral hygiene), the cluster size for this data is potentially informative. This is clearly visible from the boxplots of patients grouped according to the extent of tooth loss (low = greater than 18 remaining teeth, between 8 and 18 remaining teeth, high = less than 8 remaining teeth) where the median attachment loss tends to be greater for individuals with fewer remaining teeth. (Figure 1)

The use of tobacco was measured by a binary variable TOBUSE (=1, if user and = 2, if non user) and the socioeconomic status of the participant (in combination with that of his/her

spouse) was measured by a continuous variable SEIRSP (with higher SEIRSP value indicating better socioeconomic condition). We fit a marginal regression model of the form (1.1) with these two covariables. The parameter estimates are reported in Table 2 along with 95% confidence intervals. In fact, we recompute the standard error and the confidence intervals with different bandwidths to ensure that our results are not greatly affected by the bandwidth selector. Based on this analysis, we can conclude that tobacco use is a statistically significant predictor of attachment loss. The effect of socioeconomic status on attachment loss was (borderline) significant as well.

We also report the results of naive (e.g., unweighted) R-estimation for comparison. There seems to be substantial differences between the estimated effects of the covariates between the two methods which is perhaps a reflection of the informative cluster size. In particular, note that the point estimate of the intercept term based on naive R lies outside the 95% confidence intervals constructed using the weighted R estimation methodology, suggesting that the naive R estimators may be severely biased. We have also reported the weighted least squares estimators which also differ substantially from the weighted R-estimators for this non-normal data (Figure 2). In particular, the intercept term is estimated to be much higher (as compared to the robust weighted R estimator) due to the long right tail of the error distribution; this in turn might have affected the estimator of the effect of tobacco use.

Finally we inspect the model residuals computed using the weighted R fit of the marginal regression model. In Figure 2, we display the inverse cluster size weighted histogram of the residuals. The shape of the histogram suggests a substantially non-normal error distribution. Thus, rank based methods may be more appropriate for this data than normal distribution based methods.

6 Discussion

Clustered data methods are becoming increasingly popular and useful in applied research. Most clustered data approaches incorporate correlations in order to improve efficiency of the resulting inference. However, the issue of informative cluster size is less understood and often ignored.

As shown here, contrary to popular belief, the classical estimators of a marginal linear model may be biased, not just for the intercept term, but also for certain covariate effects when the cluster size is informative. A simple inverse cluster size reweighting at the correct place is capable of rectifying the problem. Theoretical development, albeit more difficult, is possible including asymptotic variance estimation that exploits independence of the clusters. However, appropriate methods, when the number of clusters is small, may involve more complex joint modelling of the cluster size, covariate and response.

Robust methods are a useful and appropriate choice for many practical applications. Examples include nonparametric rank type tests for clustered data problems with informative cluster size developed in Datta and Satten (2005, 2008) and Datta *et al*. (2012). Overall, the methodology developed in this article may avoid many pitfalls faced by standard analyses (e.g., least squares, mixed effects model, etc.) when the data are non-

normal, clustered, and the cluster size is correlated with the cluster response, either through a latent factor or through one or more measured covariates. In addition, the methodology does not require specification of a correlation structure which may be complex and non-verifiable for certain applications. For the dental data application, the proposed method yielded reasonable and statistically significant estimates of various effects.

The estimator proposed here is fairly easy to code in R (R Core Team, 2012). However, since it involves numerical optimization, the computation could be somewhat time consuming in the case of large sample sizes and several covariates.

Acknowledgements

References

Beck JD, Koch GC, Rozier RG, Tudor GE. Prevalence and risk indicators for periodontal attachment loss in a population of older community-dwelling blacks and whites. Journal of Periodontology. 1990; 61:521–528. [PubMed: 2391631]

Blazer, DG.; George, LK. ICPSR02744-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]; 2004. Established populations for epidemiologic studies of the elderly, 1996–1997: Piedmont health survey of the elderly, fourth in-person survey Durham, Warren, Vance, Granville, and Franklin Counties, North Carolina [Computer file].

Datta S, Satten GA. Rank-sum tests for clustered data. Journal of American Statistical Association. 2005; 100:908–915.

Datta S, Satten GA. A signed-rank test for clustered data. Biometrics. 2008; 64:501–507. [PubMed: 17970820]

Datta S, Nevalainen J, Oja H. A general class of signed rank tests for clustered data when the cluster size is potentially informative. Journal of Nonparametric Statistics. 2012; 24:797–808. [PubMed: 23074359]

Gansky, SA.; Neuhaus, JM. Missing data and informative cluster sizes. In: Lesaffre, E.; Feine, J.; Leroux, B.; Declerck, D., editors. Statistical and methodological aspects of oral health research. Chichester: John Wiley & Sons; 2009. p. 241-258.

Hettmansperger, TP.; McKean, JW. Robust nonparametric statistical methods. 2nd ed.. New York: Chapman & Hall; 2011.

Hoffman EB, Sen PK, Weinberg CR. Within-cluster resampling. Biometrika. 2001; 88:1121–1134.

Jaeckel LA. Estimating regression coefficients by minimizing the dispersion of the residuals. The Annals of Mathematical Statistics. 1972; 43:1449–1458.

Jung SH, Ying Z. Rank-based regression with repeated measurements data. Biometrika. 2003; 90:732–740.

Jureckova J. Nonparametric estimate of regression coefficients. The Annals of Mathematical Statistics. 1971; 42:1328–1338.

Koul HL, Sievers G, McKean JW. An estimator of the scale parameter for the rank analysis of linear models under general score functions. Scandinavian Journal of Statistics. 1987; 14:131–141.

McKean J, Hettmansperger T. A robust analysis of the general linear model based on one step R-estimates. Biometrika. 1978; 65:571–579.

Neuhaus JM, McCulloch CE. Estimation of covariate effects in generalized linear mixed models with informative cluster sizes. Biometrika. 2011; 98:147–162. [PubMed: 23049125]

Nevalainen J, Datta S, Oja H. Inference on the marginal distribution of clustered data with informative cluster size. Statistical Papers. 2013

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna; Austria: 2012. Available at http://www.Rproject.org/

Singh RS. Mean squared errors of a density and its derivatives. Biometrika. 1979; 66:177–180.

Wang M, Kong M, Datta S. Inference for marginal linear models with clustered longitudinal data with potentially informative cluster sizes. Statistical Methods in Medical Research. 2011; 20:347–367. [PubMed: 20223781]

Wang Y-G, Zhao Y. Weighted rank regression for clustered data analysis. Biometrics. 2008; 64:34–45.

Wang Y-G, Zhu M. Rank-based regression for analysis of repeated measures. Biometrika. 2006; 93:459–464.

Williamson JM, Datta S, Satten GA. Marginal analyses of clustered data when cluster size is informative. Biometrics. 2003; 59:36–42. [PubMed: 12762439]
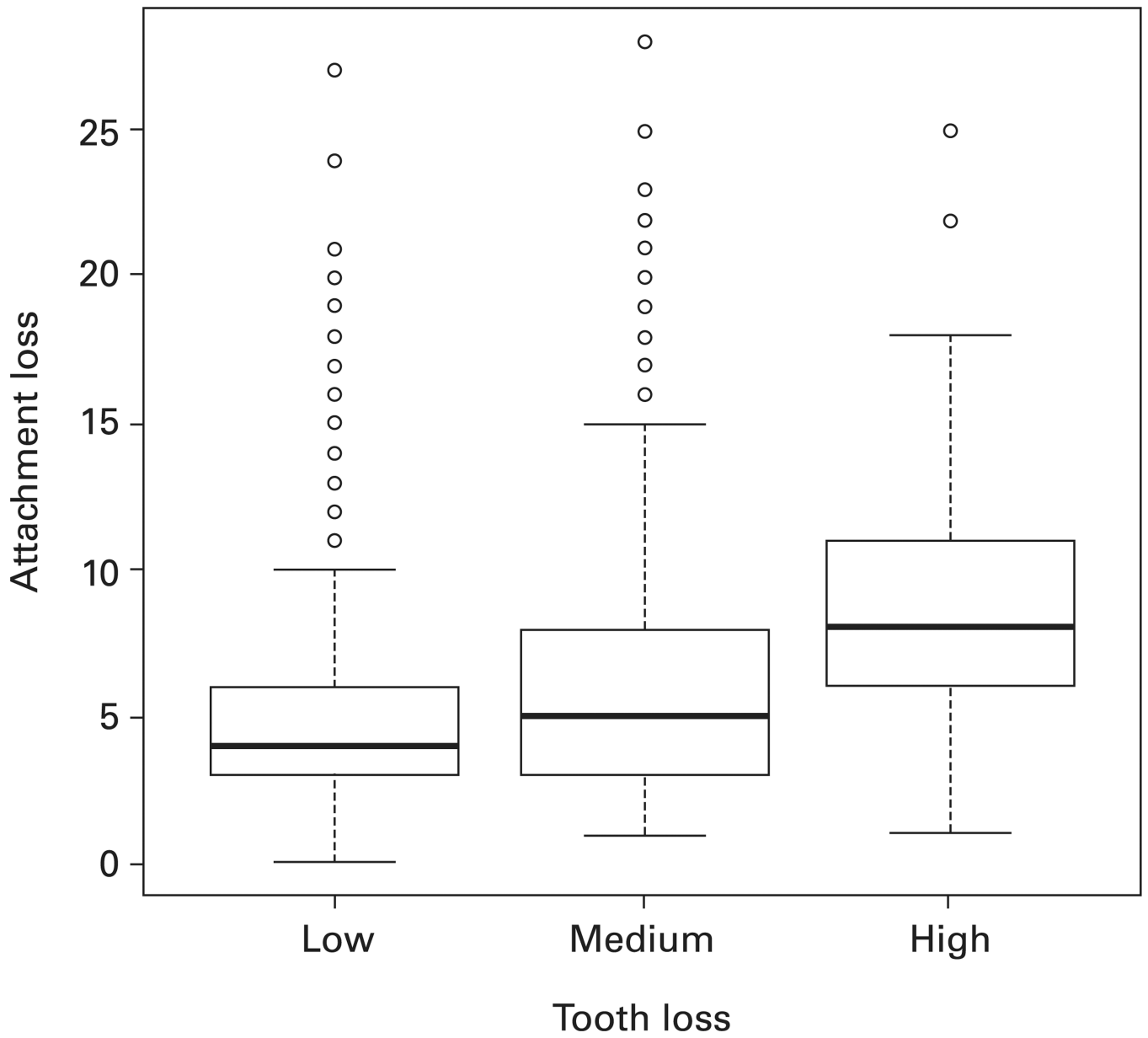
Figure 1.
Boxplot of attachment loss values grouped by individuals according to their tooth loss
Source: Authors' own.
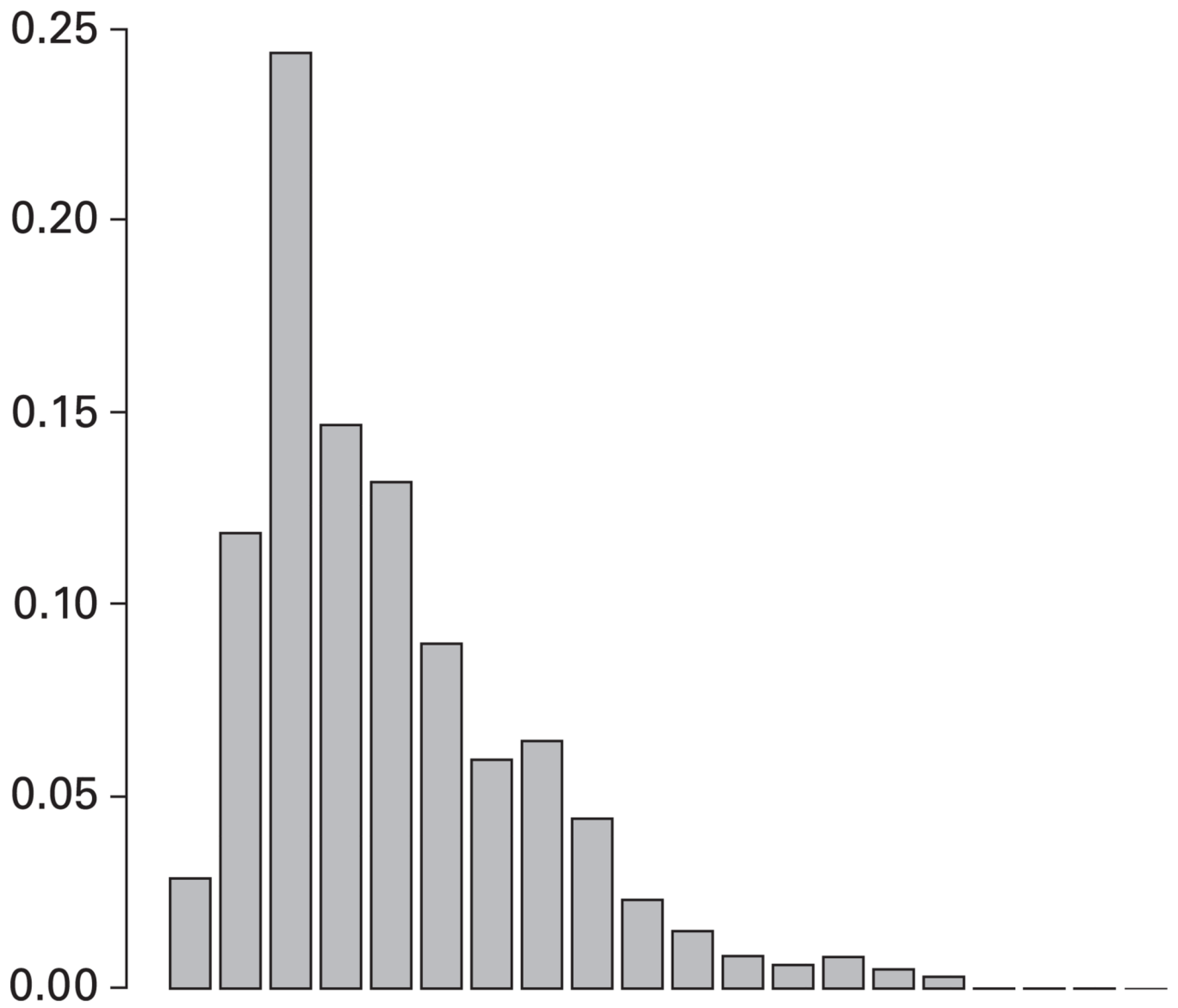
Figure 2.
Inverse cluster size weighted histogram of the model residuals.
Source: Authors' own.

Table 1

Monte Carlo estimates of bias and standard deviation (within parentheses) of various estimators under three different error distributions; each entry is based on 500 iterations. The high bias values are indicated in bold

Error distribution G_0	Parameter	Bias (sd) of weighted least squares estimators	Bias (sd) of linear mixed effects model based estimators	Bias (sd) of naive R estimators	Bias (sd) of weighted R estimators (with n_i^{-1} weight)	Estimated standard error of weighted R (averaged)	Coverage(%) of 95% CIs using weighted R
				Number of clusters = 50			
	α	0.049 (0.469)	0.045 (0.469)	**1.343** (0.433)	0.062 (0.403)	0.403	93.8
Standard	β₁	0.031 (0.482)	0.029 (0.482)	**0.656**(0.441)	0.037 (0.426)	0.455	97.0
Normal	β₂	0.008(0.187)	0.000(0.005)	0.000(0.092)	0.003(0.146)	0.196	98.0
	α	0.013(0.448)	0.007 (0.448)	1.295 (0.412)	0.037 (0.380)	0.397	93.4
Double	β₁	0.051 (0.466)	0.046 (0.467)	0.651 (0.416)	0.045 (0.404)	0.449	97.0
Exponential	β₂	−0.015(0.202)	−0.000(0.007)	−0.002(0.100)	−0.007(0.160)	0.195	98.4
	α	−0.008(2.119)	0.702 (2.743)	**1.359** (0.421)	0.069 (0.408)	0.413	92.2
Pareto	β₁	−0.049(2.100)	0.343 (2.692)	**0.660** (0.447)	0.038 (0.437)	0.446	95.8
(two sided)	β₂	−0.023(1.715)	0.069(1.502)	0.008(0.123)	0.007(0.184)	0.211	97.4
				Number of clusters = 100			
	α	0.026 (0.337)	0.002 (0.337)	**1.299**(0.313)	0.026(0.266)	0.280	95.4
Standard	β₁	−0.015(0.330)	−0.002 (0.330)	**0.594**(0.282)	−0.007 (0.284)	0.308	96.4
Normal	β₂	0.005(0.151)	−0.000(0.003)	−0.003 (0.076)	0.003(0.121)	0.135	96.6
	α	−0.009(0.341)	−0.001 (0.341)	**1.301** (0.315)	0.004 (0.284)	0.280	93.4
Double	β₁	−0.006(0.321)	−0.001 (0.321)	**0.602** (0.299)	−0.002 (0.279)	0.308	97.2
Exponential	β₂	−0.001 (0.143)	−0.000(0.005)	0.001(0.073)	−0.000(0.114)	0.134	97.0
	α	**0.400** (8.594)	**1.212**(8.695)	**1.313**(0.297)	0.036 (0.274)	0.287	92.0
Pareto	β₁	**−0.309** (8.828)	**0.129** (8.555)	**0.602**(0.291)	−0.003 (0.298)	0.308	95.0
(two sided)	β₂	0.028(9.000)	−0.466 (9.830)	−0.002 (0.094)	0.002(0.144)	0.148	95.6

Source: Authors' own.

Table 2

Parameter estimates for the periodontal disease data

Parameter	Weighted LSE	Naive R–estimates	Weighted R–estimates	95% CI based on weighted R		
				$h = M^{-1/7}h = m^{-1/5}$	bandwidths halved	bandwidths doubled
Intercept	8.467	5.534	6.648	(5.679, 7.616)	(5.905, 7.390)	(5.849, 7.446)
tobacco use	−1.673	−0.842	−1.161	(−1.875, −0.446)	(−1.973, −0.348)	(−1.974, −0.347)
socioeconomic status ($\times 10^{-2}$)	−0.293	−0.156	−0.208	(−0.362, −0.056)	(−0.383, −0.034)	(−0.383, −0.034)

Source: Authors' own.