Robust negative binomial regression

Michel Amiguet, Alfio Marazzi, Marina S. Valdora, Víctor J. Yohai

september 2016

Negative Binomial Regression

Poisson regression is the standard method used to model count response data i.e. data for which the response variable takes on values in $\{0, 1, \ldots, n, \ldots\}$.

Negative Binomial Regression

Poisson regression is the standard method used to model count response data i.e. data for which the response variable takes on values in $\{0, 1, \dots, n, \dots\}$.

However, Poisson distribution assumes that its variance equals its mean, property which is not always satisfied in real data sets.

Negative Binomial Regression

Poisson regression is the standard method used to model count response data i.e. data for which the response variable takes on values in $\{0, 1, \ldots, n, \ldots\}$.

However, Poisson distribution assumes that its variance equals its mean, property which is not always satisfied in real data sets.

Negative binomial regression can be used to model overdispersed count data, that is to say, count data whose variance is larger than their mean.

Negative Binomial Regression

Poisson regression is the standard method used to model count response data i.e. data for which the response variable takes on values in $\{0, 1, \ldots, n, \ldots\}$.

However, Poisson distribution assumes that its variance equals its mean, property which is not always satisfied in real data sets.

Negative binomial regression can be used to model overdispersed count data, that is to say, count data whose variance is larger than their mean.

An important application of negative binomial regression is the analysis of hospital length of stay.

The maximum likelihood method(ML) provides fully efficient estimators when all the observations follow the model.

The maximum likelihood method(ML) provides fully efficient estimators when all the observations follow the model.

Unfortunately, the ML estimator is extremely sensitive to the presence of outliers in the sample.

The maximum likelihood method(ML) provides fully efficient estimators when all the observations follow the model.

Unfortunately, the ML estimator is extremely sensitive to the presence of outliers in the sample.

In certain applications, such as the analysis of hospital length of stay, the proportion of outliers may be as high as 10%.

The maximum likelihood method(ML) provides fully efficient estimators when all the observations follow the model.

Unfortunately, the ML estimator is extremely sensitive to the presence of outliers in the sample.

In certain applications, such as the analysis of hospital length of stay, the proportion of outliers may be as high as 10%.

Aeberhard, Cantoni, and Heritier (2014) proposed a generalized M (GM) estimator of NB regression.

The maximum likelihood method(ML) provides fully efficient estimators when all the observations follow the model.

Unfortunately, the ML estimator is extremely sensitive to the presence of outliers in the sample.

In certain applications, such as the analysis of hospital length of stay, the proportion of outliers may be as high as 10%.

Aeberhard, Cantoni, and Heritier (2014) proposed a generalized M (GM) estimator of NB regression.

GM estimators have several drawbacks (Maronna, Martin, Yohai, 2006). In particular, they cannot be simultaneously very efficient and very robust.

A convenient parametrization of the negative binomial density function is

$$f_{\mu,\alpha}(y) = \frac{\Gamma(1/\alpha + y)}{\Gamma(y+1)\Gamma(1/\alpha)} \left(\frac{1/\alpha}{1/\alpha + \mu}\right)^{1/\alpha} \left(\frac{\mu}{1/\alpha + \mu}\right)^{y}$$

A convenient parametrization of the negative binomial density function is

$$f_{\mu,\alpha}(y) = \frac{\Gamma(1/\alpha + y)}{\Gamma(y+1)\Gamma(1/\alpha)} \left(\frac{1/\alpha}{1/\alpha + \mu}\right)^{1/\alpha} \left(\frac{\mu}{1/\alpha + \mu}\right)^{y}$$

If $Y \sim NB(\mu, \alpha)$ then $E(Y) = \mu$ and $V(Y) = \mu + \alpha\mu$.

A convenient parametrization of the negative binomial density function is

$$f_{\mu,\alpha}(y) = \frac{\Gamma(1/\alpha + y)}{\Gamma(y+1)\Gamma(1/\alpha)} \left(\frac{1/\alpha}{1/\alpha + \mu}\right)^{1/\alpha} \left(\frac{\mu}{1/\alpha + \mu}\right)^{y}$$

If $Y \sim NB(\mu, \alpha)$ then $E(Y) = \mu$ and $V(Y) = \mu + \alpha\mu$.

When $\alpha \to 0$, $f_{\mu,\alpha}(y)$ converges to the Poisson probability density function.

Let $F_{\mu,\alpha}$ denote the NB cdf.

Let $F_{\mu,\alpha}$ denote the NB cdf.
We observe a response Y and a vector $X = (X_1, ...X_p)^\mathsf{T}$ of covariates so that the distribution of Y given $X = x$ is

$$Y \mid x \sim F_{h(x^\mathsf{T}\beta_0),\alpha_0}, \qquad (1)$$

Let $F_{\mu,\alpha}$ denote the NB cdf.
We observe a response Y and a vector $X = (X_1, ...X_p)^\mathsf{T}$ of covariates so that the distribution of Y given $X = x$ is

$$Y \mid x \sim F_{h(x^\mathsf{T}\beta_0),\alpha_0}, \tag{1}$$

where $h(t) = \exp(t)$, and $\beta_0 = (\beta_{01}, ...\beta_{0p})^\mathsf{T}$ is a vector of coefficients.

Let $F_{\mu,\alpha}$ denote the NB cdf.
We observe a response Y and a vector $X = (X_1, ...X_p)^\mathsf{T}$ of covariates so that the distribution of Y given $X = x$ is

$$Y \mid x \sim F_{h(x^\mathsf{T}\beta_0),\alpha_0}, \qquad (1)$$

where $h(t) = \exp(t)$, and $\beta_0 = (\beta_{01}, ...\beta_{0p})^\mathsf{T}$ is a vector of coefficients.
We assume that X_1 is constantly equal to one, that is, β_{01} is an intercept.

We use the notations

1. $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$

We use the notations

1. $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$
2. $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$

We use the notations

1. $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$
2. $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$
3. $\gamma_0 = \beta_0^* / \|\beta_0^*\|$

We use the notations

1. $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$
2. $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$
3. $\gamma_0 = \beta_0^* / \|\beta_0^*\|$
4. $\eta_0 = \|\beta_0^*\|$

We use the notations

1. $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$
2. $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$
3. $\gamma_0 = \beta_0^* / \|\beta_0^*\|$
4. $\eta_0 = \|\beta_0^*\|$
5. $\theta = (\beta, \alpha)$

Robust Estimation in Negative Regression Models.

Following the ideas in Gervini and Yohai(2002) and Marazzi and Yohai(2004) we consider a three step procedure.

Step 1 Compute a highly robust but possibly inefficient initial estimator.

Robust Estimation in Negative Regression Models.

Following the ideas in Gervini and Yohai(2002) and Marazzi and Yohai(2004) we consider a three step procedure.

Step 1 Compute a highly robust but possibly inefficient initial estimator.

Step 2 Identify outliers using the initial estimator.

Robust Estimation in Negative Regression Models.

Following the ideas in Gervini and Yohai(2002) and Marazzi and Yohai(2004) we consider a three step procedure.

Step 1 Compute a highly robust but possibly inefficient initial estimator.

Step 2 Identify outliers using the initial estimator.

Step 3 Compute a conditional maximum likelihood estimator, where observations are constrained to belong to a subsample without outliers.

The final estimator is

- Fully efficient

The final estimator is

- Fully efficient
- As robust as the initial estimator

Initial Estimator

First we compute an initial consistent and robust estimator of $\widetilde{\gamma}$ of $\gamma_0 = \beta_0^* / \|\beta_0^*\|$

Initial Estimator

First we compute an initial consistent and robust estimator of $\widetilde{\gamma}$ of $\gamma_0 = \beta_0^* / \|\beta_0^*\|$

Second, we compute robust estimators $\widetilde{\eta}$ and $\widetilde{\alpha}$ of $\eta_0 = \|\beta_0^*\|$ and α respectively.

Initial Estimator

First we compute an initial consistent and robust estimator of $\widetilde{\gamma}$ of $\gamma_0 = \beta_0^* / \|\beta_0^*\|$

Second, we compute robust estimators $\widetilde{\eta}$ and $\widetilde{\alpha}$ of $\eta_0 = \|\beta_0^*\|$ and α respectively.

Third, define an initial estimator of β_0^* as $\widetilde{\beta} = \widetilde{\gamma}\widetilde{\eta}$.

Initial Estimator

First we compute an initial consistent and robust estimator of $\widetilde{\gamma}$ of $\gamma_0 = \beta_0^* / \|\beta_0^*\|$

Second, we compute robust estimators $\widetilde{\eta}$ and $\widetilde{\alpha}$ of $\eta_0 = \|\beta_0^*\|$ and α respectively.

Third, define an initial estimator of β_0^* as $\widetilde{\beta} = \widetilde{\gamma}\widetilde{\eta}$.

Initial estimator of γ. The Maximum Rank Correlation Estimator

Given a bivariate sample $(w_1, y_1), \ldots, (w_n, y_n)$, we say that the observations (w_i, y_i) and (w_j, y_j) are concordant if

$$(y_i - y_j)(w_i - w_j) > 0.$$

Initial estimator of γ. The Maximum Rank Correlation Estimator

Given a bivariate sample $(w_1, y_1), \ldots, (w_n, y_n)$, we say that the observations (w_i, y_i) and (w_j, y_j) are concordant if

$$(y_i - y_j)(w_i - w_j) > 0.$$

The Kendall rank correlation between $y = (y_1, ..., y_n)$ and $w = (w_1, ..., w_n)$ is defined as the mean number of concordant pairs.

Initial estimator of γ. The Maximum Rank Correlation Estimator

Given a bivariate sample $(w_1, y_1), \ldots, (w_n, y_n)$, we say that the observations (w_i, y_i) and (w_j, y_j) are concordant if

$$(y_i - y_j)(w_i - w_j) > 0.$$

The Kendall rank correlation between $y = (y_1, ..., y_n)$ and $w = (w_1, ..., w_n)$ is defined as the mean number of concordant pairs.
That is to say

$$\tau(y, w) = \frac{1}{(n-1)n} \sum_{i \neq j} I((w_j - w_i)(y_j - y_i) > 0).$$

For a given coefficient vector $\gamma = (\gamma_2, ..., \gamma_p)^\mathsf{T}$ Kendall's correlation between the responses and the linear predictor $\gamma^\mathsf{T} x^*$ is

$$\tau(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} I\left[(\gamma^\mathsf{T} x_j^* - \gamma^\mathsf{T} x_i^*)(y_j - y_i) \geq 0\right].$$

For a given coefficient vector $\gamma = (\gamma_2, ..., \gamma_p)^\mathsf{T}$ Kendall's correlation between the responses and the linear predictor $\gamma^\mathsf{T} x^*$ is

$$\tau(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} I\left[(\gamma^\mathsf{T} x_j^* - \gamma^\mathsf{T} x_i^*)(y_j - y_i) \geq 0\right].$$

The MRC estimator of γ_0 is defined by

$$\tilde{\gamma} = \arg\max_{\|\gamma\|=1} \tau(\gamma). \qquad (2)$$

For a given coefficient vector $\gamma = (\gamma_2, ..., \gamma_p)^\mathsf{T}$ Kendall's correlation between the responses and the linear predictor $\gamma^\mathsf{T} x^*$ is

$$\tau(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} I \left[(\gamma^\mathsf{T} x_j^* - \gamma^\mathsf{T} x_i^*)(y_j - y_i) \geq 0 \right].$$

The MRC estimator of γ_0 is defined by

$$\tilde{\gamma} = \arg \max_{\|\gamma\|=1} \tau(\gamma). \tag{2}$$

Note that $\tau(\gamma) = \tau(\lambda\gamma)$ for every $\lambda > 0$.
That is why we normalize the parameters $\gamma = \beta^* / \|\beta^*\|$ and search for the MRC estimator on the unit circle.

For a given coefficient vector $\gamma = (\gamma_2, ..., \gamma_p)^\mathsf{T}$ Kendall's correlation between the responses and the linear predictor $\gamma^\mathsf{T} x^*$ is

$$\tau(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} I\left[(\gamma^\mathsf{T} x_j^* - \gamma^\mathsf{T} x_i^*)(y_j - y_i) \geq 0 \right].$$

The MRC estimator of γ_0 is defined by

$$\tilde{\gamma} = \arg \max_{\|\gamma\|=1} \tau(\gamma). \tag{2}$$

Note that $\tau(\gamma) = \tau(\lambda\gamma)$ for every $\lambda > 0$.
That is why we normalize the parameters $\gamma = \beta^* / \|\beta^*\|$ and search for the MRC estimator on the unit circle.
Under very general conditions on the covariates, this estimate is consistent if all the observations follow the model.

Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$, Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$, $\eta_0 = \|\beta_0^*\|$,

Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$, $\eta_0 = \|\beta_0^*\|$, $\gamma_0 = \beta_0^* / \|\beta_0^*\|$

Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$, $\eta_0 = \|\beta_0^*\|$, $\gamma_0 = \beta_0^* / \|\beta_0^*\|$
then
$\beta_0^T x = \beta_{01} + \eta_0 \gamma_0 x^*$

Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$, $\eta_0 = \|\beta_0^*\|$, $\gamma_0 = \beta_0^* / \|\beta_0^*\|$
then
$\beta_0^T x = \beta_{01} + \eta_0 \gamma_0 x^*$ and

$$y|x \sim NB(\beta_{01} + \eta_0 \gamma_0 x^*, \alpha_0)$$

Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$, $\eta_0 = \|\beta_0^*\|$, $\gamma_0 = \beta_0^* / \|\beta_0^*\|$ then
$\beta_0^T x = \beta_{01} + \eta_0 \gamma_0 x^*$ and

$$y|x \sim NB(\beta_{01} + \eta_0 \gamma_0 x^*, \alpha_0)$$

Therefore, if we know γ_0, we can think of the observations following a simple negative binomial regression model.

Since $x^{\mathsf{T}} = (1, x^{*\mathsf{T}})$, $\beta_0^{\mathsf{T}} = (\beta_{01}, \beta_0^{*\mathsf{T}})$, $\eta_0 = \|\beta_0^*\|$, $\gamma_0 = \beta_0^* / \|\beta_0^*\|$ then
$\beta_0^T x = \beta_{01} + \eta_0 \gamma_0 x^*$ and

$$y|x \sim NB(\beta_{01} + \eta_0 \gamma_0 x^*, \alpha_0)$$

Therefore, if we know γ_0, we can think of the observations following a simple negative binomial regression model.
Since we already have an estimator for γ_0 we can write $z = \widetilde{\gamma} x^*$ and fit the following simple negative regression model:

$$y|x \sim NB(\beta_{01} + \eta_0 z, \alpha_0).$$

Robust initial estimates for simple negative binomial regression models.

A robust and consistent initial estimator for (β_{01}, η_0) can be computed using an weighted M estimator based on transformations (WMT estimator) defined in Valdora and Yohai (2014).

Robust initial estimates for simple negative binomial regression models.

A robust and consistent initial estimator for (β_{01}, η_0) can be computed using an weighted M estimator based on transformations (WMT estimator) defined in Valdora and Yohai (2014).

A robust and consistent estimator of the univariate parameter α can be defined using a Hampel type estimator following the ideas introduced in Marazzi and Yohai (2010).

Robust initial estimates for simple negative binomial regression models.

A robust and consistent initial estimator for (β_{01}, η_0) can be computed using an weighted M estimator based on transformations (WMT estimator) defined in Valdora and Yohai (2014).

A robust and consistent estimator of the univariate parameter α can be defined using a Hampel type estimator following the ideas introduced in Marazzi and Yohai (2010).

Combining these two estimators, we construct $(\widetilde{\beta}_1, \widetilde{\eta}, \widetilde{\alpha})$ a robust and consistent estimator for the simple negative binomial regression model.

Robust initial estimates for simple negative binomial regression models.

A robust and consistent initial estimator for (β_{01}, η_0) can be computed using an weighted M estimator based on transformations (WMT estimator) defined in Valdora and Yohai (2014).

A robust and consistent estimator of the univariate parameter α can be defined using a Hampel type estimator following the ideas introduced in Marazzi and Yohai (2010).

Combining these two estimators, we construct $(\widetilde{\beta}_1, \widetilde{\eta}, \widetilde{\alpha})$ a robust and consistent estimator for the simple negative binomial regression model.

The initial estimator of $\beta_0 = (\beta_{01}, \beta_0^*)$ is then $(\widetilde{\beta}_1, \widetilde{\eta}\widetilde{\gamma})$ and the initial estimator for α is $\widetilde{\alpha}$.

Step 2: Outlier Detection

Since the residual distribution depends on the covariates, residuals cannot be used in the usual way for the purpose of highlighting outliers.

Step 2: Outlier Detection

Since the residual distribution depends on the covariates, residuals cannot be used in the usual way for the purpose of highlighting outliers.

We overcome this problem by transforming the variables y_i so that the conditional distribution of trasnformed variables are identical.

Step 2: Outlier Detection

Since the residual distribution depends on the covariates, residuals cannot be used in the usual way for the purpose of highlighting outliers.

We overcome this problem by transforming the variables y_i so that the conditional distribution of trasnformed variables are identical.

Suppose that y has a continuos distribution F and let $z = F(y)$, then it is known that $z \sim U[0,1]$

Suppose that y has a continuos distribution F and let $z = F(y)$, then it is known that $z \sim U[0, 1]$

Consider now a discrete random variable y which has a distribution funcion F and a density f and let $z = F(y) - uf(y)$.

Suppose that y has a continuos distribution F and let $z = F(y)$, then it is known that $z \sim U[0, 1]$

Consider now a discrete random variable y which has a distribution funcion F and a density f and let $z = F(y) - uf(y)$.

Take $u \sim U[0, 1]$ independent of y . Then $z \sim U[0, 1]$

Let $u_1, ..., u_n$ be a sample from a uniform distribution $U[0, 1]$ and define $z_1, ..., z_n$ as

$$z_i = F_{\tilde{\mu}_{x_i},\tilde{\alpha}}(y_i) - u_i f_{\tilde{\mu}_{x_i},\alpha}(y_i). \tag{3}$$

Let $u_1, ..., u_n$ be a sample from a uniform distribution $U[0, 1]$ and define $z_1, ..., z_n$ as

$$z_i = F_{\tilde{\mu}_{x_i}, \tilde{\alpha}}(y_i) - u_i f_{\tilde{\mu}_{x_i}, \alpha}(y_i). \tag{3}$$

$z_1, ..., z_n$ is approximately a random sample from a $U[0, 1]$ distribution.

Let $u_1, ..., u_n$ be a sample from a uniform distribution $U[0, 1]$ and define $z_1, ..., z_n$ as

$$z_i = F_{\tilde{\mu}_{x_i},\tilde{\alpha}}(y_i) - u_i f_{\tilde{\mu}_{x_i},\alpha}(y_i). \tag{3}$$

$z_1, ..., z_n$ is approximately a random sample from a $U[0, 1]$ distribution.

Let F_n be the empirical distribution of z_i.

Let $u_1, ..., u_n$ be a sample from a uniform distribution $U[0,1]$ and define $z_1, ..., z_n$ as

$$z_i = F_{\tilde{\mu}_{x_i},\tilde{\alpha}}(y_i) - u_i f_{\tilde{\mu}_{x_i},\alpha}(y_i). \tag{3}$$

$z_1, ..., z_n$ is approximately a random sample from a $U[0,1]$ distribution.

Let F_n be the empirical distribution of z_i.

Then we can detect outliers comparing the tails of F_n with the tails of F_0, the $U[0,1]$ distribution.

Let ϵ_1 be a small positive number.

Let ϵ_1 be a small positive number.

Let

$$F^{-}_{n,k}(z) = \frac{1}{n-k} \sum_{i=1}^{n} I\left(z_{(i)} \leq z\right).$$

Let ϵ_1 be a small positive number.

Let

$$F_{n,k}^{-}(z) = \frac{1}{n-k} \sum_{i=1}^{n} I\left(z_{(i)} \leq z\right).$$

We want to eliminate outliers so that the left tail and right tails of the remaining observations are not heavier than those of F_0.

Let ϵ_1 be a small positive number.

Let

$$F^{-}_{n,k}(z) = \frac{1}{n-k} \sum_{i=1}^{n} I\left(z_{(i)} \le z\right).$$

We want to eliminate outliers so that the left tail and right tails of the remaining observations are not heavier than those of F_0.

Then the minimum number of outliers to the left that we should eliminate is

$$k_1 = \min\left\{k : \max_{z \le \epsilon_1}\left(F^{-}_{n,k}(z) - F_0(z)\right) \le 0\right\}.$$

Let ϵ_1 be a small positive number.

Let

$$F_{n,k}^{-}(z) = \frac{1}{n-k} \sum_{i=1}^{n} I\left(z_{(i)} \leq z\right).$$

We want to eliminate outliers so that the left tail and right tails of the remaining observations are not heavier than those of F_0.

Then the minimum number of outliers to the left that we should eliminate is

$$k_1 = \min\left\{k : \max_{z \leq \epsilon_1}\left(F_{n,k}^{-}(z) - F_0(z)\right) \leq 0\right\}.$$

In an analogous way we define k_2, the number of outliers to the right.

Let ϵ_1 be a small positive number.

Let

$$F^-_{n,k}(z) = \frac{1}{n-k} \sum_{i=1}^{n} I\left(z_{(i)} \leq z\right).$$

We want to eliminate outliers so that the left tail and right tails of the remaining observations are not heavier than those of F_0.

Then the minimum number of outliers to the left that we should eliminate is

$$k_1 = \min\left\{k : \max_{z \leq \epsilon_1}\left(F^-_{n,k}(z) - F_0(z)\right) \leq 0\right\}.$$

In an analogous way we define k_2, the number of outliers to the right.

Let $a = z_{(k_1+1)}$ and $b = z_{(n-k_1)}$

Lemma

$$sup_{z \in [0,1]} |F_n(z) - z| \to 0 \text{ in probability}$$

Theorem

$$a \to 0 \text{ and } b \to 1 \text{ in probability}$$

.

Lemma

$$sup_{z\in[0,1]}|F_n(z) - z| \to 0 \text{ in probability}$$

Theorem

$$a \to 0 \text{ and } b \to 1 \text{ in probability}$$

.

Observations such that $z \notin [a, b]$ will be identified as “outliers” and will be removed from the sample used by the final estimator.

Step 3. Conditional Maximum Likelihood Estimator

Conditional maximum estimators to define robust estimators were proposed by Gervini and Yohai (2002), Marazzi and Yohai (2004) and Cuesta-Albertos et al. (2008).

Step 3. Conditional Maximum Likelihood Estimator

Conditional maximum estimators to define robust estimators were proposed by Gervini and Yohai (2002), Marazzi and Yohai (2004) and Cuesta-Albertos et al. (2008).

We define the *CML estimator* $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$ given $z \in [a, b]$ by

Step 3. Conditional Maximum Likelihood Estimator

Conditional maximum estimators to define robust estimators were proposed by Gervini and Yohai (2002), Marazzi and Yohai (2004) and Cuesta-Albertos et al. (2008).

We define the *CML estimator* $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$ given $z \in [a, b]$ by

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta; \tilde{S}),$$

Step 3. Conditional Maximum Likelihood Estimator

Conditional maximum estimators to define robust estimators were proposed by Gervini and Yohai (2002), Marazzi and Yohai (2004) and Cuesta-Albertos et al. (2008).

We define the *CML estimator* $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$ given $z \in [a, b]$ by

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta; \tilde{S}),$$

where $\tilde{S}$ denotes the subsample $\tilde{S} = \{(x, y)/z \in [a, b]\}$

Step 3. Conditional Maximum Likelihood Estimator

Conditional maximum estimators to define robust estimators were proposed by Gervini and Yohai (2002), Marazzi and Yohai (2004) and Cuesta-Albertos et al. (2008).

We define the *CML estimator* $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$ given $z \in [a, b]$ by

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta; \tilde{S}),$$

where $\tilde{S}$ denotes the subsample $\tilde{S} = \{(x, y)/z \in [a, b]\}$
and $\mathcal{L}(\theta; \tilde{S})$ is the conditional likelihood of y given $x_1, \ldots, x_n$ and $(x, y) \in \tilde{S}$.

Step 3. Conditional Maximum Likelihood Estimator

Conditional maximum estimators to define robust estimators were proposed by Gervini and Yohai (2002), Marazzi and Yohai (2004) and Cuesta-Albertos et al. (2008).

We define the *CML estimator* $\hat{\theta} = (\hat{\beta}, \hat{\alpha})$ given $z \in [a, b]$ by

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta; \tilde{S}),$$

where $\tilde{S}$ denotes the subsample $\tilde{S} = \{(x, y)/z \in [a, b]\}$ and $\mathcal{L}(\theta; \tilde{S})$ is the conditional likelihood of y given $x_1, \ldots, x_n$ and $(x, y) \in \tilde{S}$.

The cutoff values are determined so that, in the case that S does not contain outliers, the conditioned sample $\tilde{S}$ approaches S when $n \to \infty$. In this case, the CML estimator tends to the ML estimator and is therefore fully efficient.

Monte Carlo Study

We first performed four experiments with uncontaminated samples of size $n = 100, 400, 1000, 2000$.

Monte Carlo Study

We first performed four experiments with uncontaminated samples of size $n = 100, 400, 1000, 2000$.

To measure the quality of the estimates we used the mean absolute error (MAE). The MAE of $\hat{\beta}$ and $\hat{\alpha}$ are

Monte Carlo Study

We first performed four experiments with uncontaminated samples of size $n = 100, 400, 1000, 2000$.
To measure the quality of the estimates we used the mean absolute error (MAE). The MAE of $\hat{\beta}$ and $\hat{\alpha}$ are

$$\text{MAE}(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\beta}_i - \beta_0\right\| \quad \text{MAEE}(\hat{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\|\hat{\alpha}_i - \alpha_0\|$$

Monte Carlo Study

We first performed four experiments with uncontaminated samples of size $n = 100, 400, 1000, 2000$.
To measure the quality of the estimates we used the mean absolute error (MAE). The MAE of $\hat{\beta}$ and $\hat{\alpha}$ are

$$\mathrm{MAE}(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\beta}_i - \beta_0\right\| \quad \mathrm{MAEE}(\hat{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\alpha}_i - \alpha_0\right\|$$

where $\hat{\beta}_i$ and $\hat{\alpha}_i$ are the estimates based on the i^{th} sample.

We simulate the following model.

$$x = (1, x^*), x \sim N_5(\mathbf{0}, \mathbf{I}),$$

We simulate the following model.

$$x = (1, x^*), x \sim N_5(\mathbf{0}, \mathbf{I}),$$

$$y|x \sim NB(exp(\beta_0^T x), \alpha_0)$$

We simulate the following model.

$$x = (1, x^*), x \sim N_5(\mathbf{0}, \mathbf{I}),$$

$$y|x \sim NB(exp(\beta_0^T x), \alpha_0)$$

$$\beta_0 = (1.5, 0.5, 0.25, 0, 0, 0)$$

We simulate the following model.

$$x = (1, x^*), x \sim N_5(\mathbf{0}, \mathbf{I}),$$
$$y|x \sim NB(exp(\beta_0^T x), \alpha_0)$$
$$\beta_0 = (1.5, 0.5, 0.25, 0, 0, 0)$$
$$\alpha_0 = 0.8.$$

We simulate the following model.

$$x = (1, x^*), x \sim N_5(\mathbf{0}, \mathbf{I}),$$

$$y|x \sim NB(exp(\beta_0^T x), \alpha_0)$$

$$\beta_0 = (1.5, 0.5, 0.25, 0, 0, 0)$$

$$\alpha_0 = 0.8.$$

Relative efficiency for clean samples

The tables report the empirical relative efficiencies measured as the ratios of the MAE of the robust estimates with respect to the corresponding MAE of the maximum likelihood estimates.

	100	400	1000	2000
Initial	0.51	0.49	0.49	0.49
Final	0.75	0.89	0.94	0.94
ACHE	0.71	0.75	0.78	0.75

Table: efficiency of regression estimator

	100	400	1000	2000
Initial	0.68	0.69	0.71	0.72
Final	0.80	0.84	0.91	0.93
ACHE	0.76	0.79	0.83	0.83

Table: efficiency of dispersion estimator

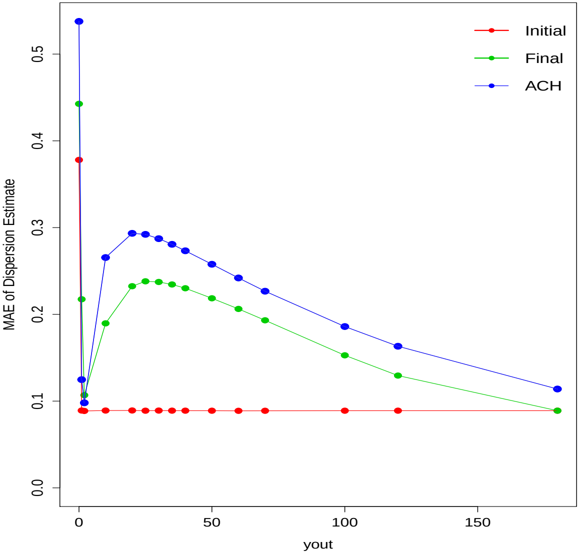
We consider contaminated samples were we replace 10% of the observations with outliers of the form (x_0, y_0) with

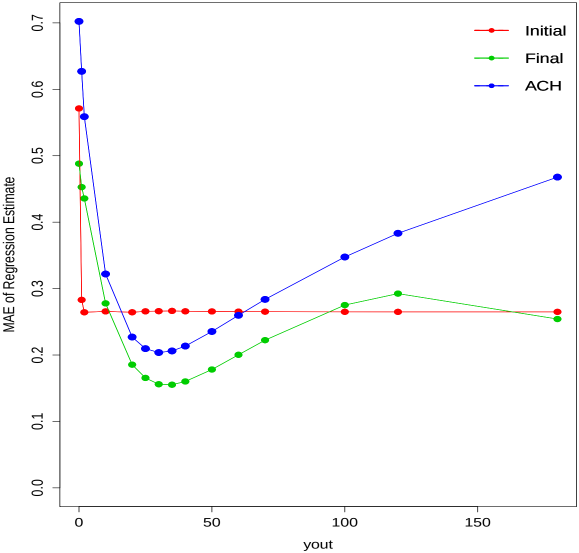
We consider contaminated samples were we replace 10% of the observations with outliers of the form (x_0, y_0) with

$$x_0 = (3, 2, 0, 0, 0)$$

We consider contaminated samples were we replace 10% of the observations with outliers of the form (x_0, y_0) with

$$x_0 = (3, 2, 0, 0, 0)$$
$$y_0 \in \{0, 1, 2, 10, 20, 25, 30, 35, 40, 50, 60, 70, 100, 120, 180\}.$$


Initial
Final
ACH
MAE of Dispersion Estimate
0.0
0.1
0.2
0.3
0.4
0.5
0
50
100
150
yout



Initial
Final
ACH
MAE of Regression Estimate
yout
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
50
100
150

Real Data Example

We consider a sample of 649 hospital stays (256 male and 393 female patients) for the "major diagnostic category" "Diseases and Disorders of the Endocrine, Nutritional And Metabolic System".

Real Data Example

We consider a sample of 649 hospital stays (256 male and 393 female patients) for the "major diagnostic category" "Diseases and Disorders of the Endocrine, Nutritional And Metabolic System". The data are shown in the following figure (two outliers with LOS = 84 and LOS = 122 fall beyond the upper limit of the figure).

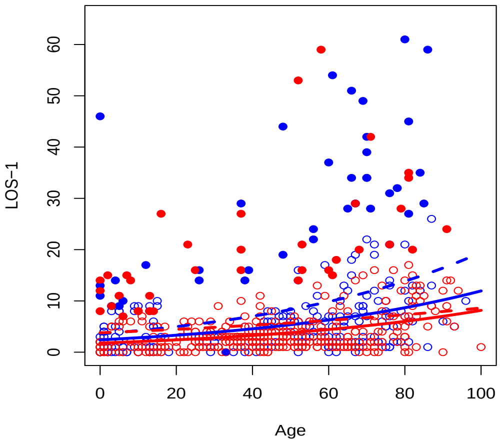


Figure: Data: LOS and Age of 649 patients. Blue circles are men, red circles are female, filled circles are outliers. Fitted models according to CML (solid lines) and ML (broken lines).

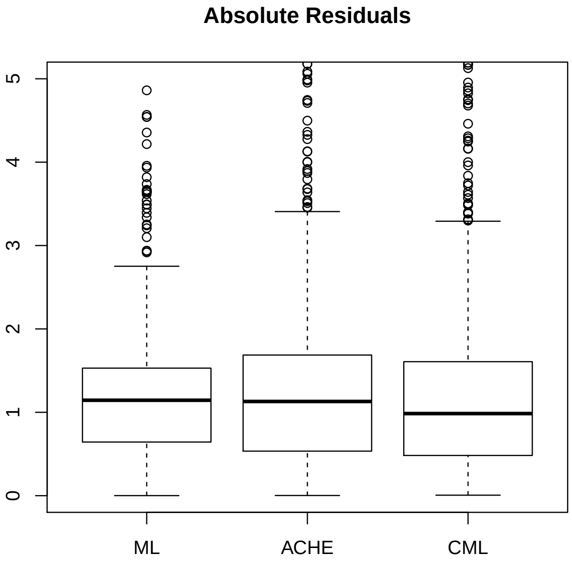
We study the relationship between LOS and two covariates: Age of the patient (x_1 in years) and Sex of the patient ($x_2 = 0$ for males and $x_2 = 1$ for females).

We study the relationship between LOS and two covariates: Age of the patient (x_1 in years) and Sex of the patient ($x_2 = 0$ for males and $x_2 = 1$ for females).
We consider a NB model with logarithmic link and linear predictor $\beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_1 x_3$.

We study the relationship between LOS and two covariates: Age of the patient (x_1 in years) and Sex of the patient ($x_2 = 0$ for males and $x_2 = 1$ for females).
We consider a NB model with logarithmic link and linear predictor $\beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_1 x_3$.

The CML identified 78 outliers (i.e. 12% contamination) shown by filled circles in the Figure. We also computed the ML estimator (ML*) after removal of these outliers.

	INTER.	AGE	SEX	AGE-SEX	α
ML	1.266	0.017	0.064	-0.009	1.067
ACH	1.257	0.010	-0.652	0.006	0.548
CML	0.950	0.015	-0.395	-0.000	0.574
ML*	0.848	0.015	-0.399	-0.000	0.460


Absolute Residuals
0
1
2
3
4
5
ML
ACHE
CML