

## A quadratic bootstrap method and improved estimation in logistic regression

Peer-reviewed author version

CLAESKENS, Gerda; AERTS, Marc & MOLENBERGHS, Geert (2002) A quadratic bootstrap method and improved estimation in logistic regression. In: Statistics and Probability Letters, 61(4). PII S0167-7152(02)00397-8. p. 383-394.

DOI: 10.1016/S0167-7152(02)00397-8

Handle: <http://hdl.handle.net/1942/420>

# A quadratic bootstrap method and improved estimation in logistic regression

Gerda Claeskens<sup>1</sup>, Marc Aerts<sup>2</sup> and Geert Molenberghs<sup>2</sup>

<sup>1</sup> Department of Statistics, Texas A&M University, College Station, TX 77843, USA, and

<sup>2</sup> Center for Statistics, Limburgs Universitair Centrum, B-3590 Diepenbeek, Belgium

## Abstract

This paper presents a quadratic one-step bootstrap method for binary response data. Rather than resampling from the original sample, the proposed method resamples summands appearing in the quadratic approximation of the estimates. It enjoys the same computational simplicity as its linear analogue while being more accurate. Moreover it allows the construction of a bias corrected estimator and improved confidence intervals. A small simulation study illustrates the improved finite sample behaviour for binary response data.

Key Words: Bias correction; Bootstrap confidence interval; One-step bootstrap; Logistic regression.

## 1 Introduction

It is well-known that the bootstrap can be very computer-intensive, especially if no analytic method can be used and simulation based approximations are required. In cases where the computation of the estimators requires iterations, the one-step linear bootstrap can be applied to save computer time. It is a simple and attractive method based on a linear representation of the estimators. Instead of calculating thousands of bootstrap estimates iteratively, the one-step approach uses only one step of the iterative process. The original idea is due to Schucany and Wang (1991), see also Shao and Tu (1995, Section 5.4.7) or Davison and Hinkley (1997).

Although the one-step linear method is asymptotically equivalent to the fully iterative one, our experience in the setting of binary response data showed to interpret its results with care. Compared to the normal based confidence intervals, the linear one-step bootstrap tends to produce shorter confidence intervals but simulations show an equivalent decrease in coverage probability. Also, by definition, the one-step approach is not able to detect the bias of the estimator. These shortcomings seem to be greatly eliminated by the one-step *quadratic* bootstrap, which allows the construction of a bias corrected estimator and improved confidence intervals.

---

<sup>1</sup>Corresponding author. Tel.:+1-979-8455650; fax:+1-979-8453144; e-mail: Gerda@stat.tamu.edu

The quadratic bootstrap is based on a quadratic approximation of the estimator and has been studied in Aerts and Claeskens (2001) for hypotheses tests. This paper focuses on the use of this improved one-step bootstrap approach to sharpen estimators in terms of bias and mean squared error and to improve interval estimation. The context of logistic regression is a particularly interesting and instructive application area for this method, even though we wish to stress that the scope of the one-step quadratic bootstrap is much broader than logistic modeling.

The paper is organized as follows. In the next section, we introduce the quadratic one-step bootstrap method and prove its consistency. In Section 3 we show how to use the quadratic bootstrap to define a bias corrected estimator, and how to implement a double bootstrap algorithm to estimate the variance of a bias corrected estimator. In Section 4 we construct improved confidence intervals, of which the finite sample performance is illustrated by a simulation study.

## 2 The quadratic one-step bootstrap

### 2.1 Estimation in logistic regression

Let  $Y_{i1}, \dots, Y_{in_i}$  be independent identically distributed Bernoulli random variables with probability function  $f_i(y) = \pi_i^y(1 - \pi_i)^{1-y}$ ,  $y \in \{0, 1\}$ ,  $i = 1, \dots, p$ . Associated populations in the context of logistic regression correspond to different covariate levels  $x_i$ . Let  $n_i$  be the number of replications at  $x_i$ , and let  $y_{ij}$  indicate whether the  $j$ th outcome in population  $i$  is a “success” or not. The number  $p$  of possibly different (associated) populations is considered as fixed whereas the number  $n_i$  of observations from the distinct populations become large as  $n = \sum_{i=1}^p n_i$  tends to infinity, according to  $n_i/n \rightarrow \lambda_i$  where  $\sum_{i=1}^p \lambda_i = 1$  with  $\lambda_i > 0$ . Logistic regression implies modeling the success probability  $\pi_i$  as a function of  $x_i$  by model parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ . For example, in a linear logit model ( $r = 1$ )

$$\ln(\pi_i/(1 - \pi_i)) = \theta_0 + \theta_1 x_i, \quad i = 1, \dots, p.$$

Maximum likelihood inference for associated populations is based on  $r$  dimensional *score* functions  $\boldsymbol{\psi}_i(y, \mathbf{t}) = \frac{\partial}{\partial \mathbf{t}} \log f_i(y, \mathbf{t})$ , where the “true” parameter  $\boldsymbol{\theta}$  is defined as the solution to

$$\sum_{i=1}^p \lambda_i E[\boldsymbol{\psi}_i(Y_{i1}; \boldsymbol{\theta})] = \mathbf{0}. \quad (1)$$

Solving the system of equations

$$\sum_{i=1}^p \sum_{j=1}^{n_i} \boldsymbol{\psi}_i(Y_{ij}; \boldsymbol{\theta}) = \mathbf{0}, \quad (2)$$

leads to the estimator  $\hat{\boldsymbol{\theta}}_n$  for  $\boldsymbol{\theta}$ .

More specifically, the score vector for linear logistic regression is given as follows:

$$\begin{aligned} \psi_i(y; \theta) = & \left( \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} - 1/(1 + \exp\{-(\theta_0 + \theta_1 x_i)\}), \right. \\ & \left. \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} x_i - x_i/(1 + \exp\{-(\theta_0 + \theta_1 x_i)\}) \right)^t. \end{aligned}$$

For logistic regression, bootstrap procedures based on Pearson's residuals have been studied in Simonoff and Tsai (1988), Moulton and Zeger (1989,1991) and Lee (1990). Here, we propose to resample the score and the differentiated score values.

## 2.2 One-step bootstrap estimation

Based on a linear approximation, we define a bootstrap replicate of  $\hat{\theta}_n$  as

$$\hat{\theta}_n^* = \hat{\theta}_n - \left( \sum_{i=1}^p \sum_{j=1}^{n_i} \dot{\psi}_{ij}^*(\hat{\theta}_n) \right)^{-1} \sum_{i=1}^p \sum_{j=1}^{n_i} \psi_{ij}^*(\hat{\theta}_n) \quad (3)$$

where, for each  $i = 1, \dots, p$ ,  $(\psi_{ij}^*(\hat{\theta}_n), \dot{\psi}_{ij}^*(\hat{\theta}_n)), j = 1, \dots, n_i$  is a sample with replacement from the set

$$\left\{ \left( \psi_i(Y_{ij}, \hat{\theta}_n), (\partial/\partial \theta) \psi_i(Y_{ij}, \hat{\theta}_n) \right), \quad j = 1, \dots, n_i \right\}.$$

Note that  $\psi_i(Y_{ij}, \hat{\theta}_n)$  is a  $r \times 1$  vector and  $(\partial/\partial \theta) \psi_i(Y_{ij}, \hat{\theta}_n)$  is a  $r \times r$  matrix. A similar linearization idea is used in simulation approaches for the bootstrap, as the linear bootstrap (Davison, Hinkley and Schechtman, 1986) and the one-step bootstrap (Schucany and Wang, 1991). For linear models where  $Y = \mathbf{X}\beta + \varepsilon$ , the idea of resampling scores has also been proposed by Hu and Zidek (1995).

The rationale behind the semiparametric bootstrap resampling scheme is as follows. The first term  $\hat{\theta}_n$  at the right-hand side of (3) reflects the “true” parameter in the bootstrap world and the second term represents the random fluctuation of the bootstrap replicate  $\hat{\theta}_n^*$  around this value.

Definition (3) follows from a linear approximation of the score equations. One can improve on this by including quadratic and higher order terms. A possible approach is suggested by the second and third order efficient approximations as discussed in, e.g., Ghosh (1994). We focus attention to the following second order approximation (simplified to one population),

$$\begin{aligned} \mathbf{0} = & \sum_{j=1}^n \psi(Y_j, \theta) + \sum_{k=1}^r \sum_{j=1}^n \frac{\partial}{\partial \theta_k} \psi(Y_j, \theta) (\hat{\theta}_{nk} - \theta_k) \\ & + \sum_{k=1}^r \sum_{\ell=1}^r \sum_{j=1}^n \frac{\partial^2}{\partial \theta_k \partial \theta_\ell} \psi(Y_j, \theta) (\hat{\theta}_{nk} - \theta_k) (\hat{\theta}_{n\ell} - \theta_\ell) + O_P(n^{-1/2}). \end{aligned} \quad (4)$$

By calculations similar to those of Ghosh (1994), the expansion (4) suggests the following one-step quadratic estimator

$$\hat{\boldsymbol{\theta}}_n^* = \hat{\boldsymbol{\theta}}_n + \mathbf{U}_n^* - \frac{1}{2} \left( \sum_{j=1}^n \dot{\boldsymbol{\psi}}_j^*(\hat{\boldsymbol{\theta}}_n) \right)^{-1} \sum_{k=1}^r \sum_{\ell=1}^r \sum_{j=1}^n \ddot{\boldsymbol{\psi}}_j^*(\hat{\boldsymbol{\theta}}_n)_{k,\ell} \mathbf{U}_{nk}^* \mathbf{U}_{n\ell}^* \quad (5)$$

with

$$\mathbf{U}_n^* = - \left( \sum_{i=1}^p \sum_{j=1}^{n_i} \dot{\boldsymbol{\psi}}_{ij}^*(\hat{\boldsymbol{\theta}}_n) \right)^{-1} \sum_{i=1}^p \sum_{j=1}^{n_i} \dot{\boldsymbol{\psi}}_{ij}^*(\hat{\boldsymbol{\theta}}_n).$$

This bootstrap estimator is based on the values  $(\boldsymbol{\psi}_j^*(\hat{\boldsymbol{\theta}}_n), \dot{\boldsymbol{\psi}}_j^*(\hat{\boldsymbol{\theta}}_n), \ddot{\boldsymbol{\psi}}_j^*(\hat{\boldsymbol{\theta}}_n))$ ,  $j = 1, \dots, n$  taken with replacement from the set

$$\left\{ \left( \boldsymbol{\psi}(Y_j, \hat{\boldsymbol{\theta}}_n), (\partial/\partial \boldsymbol{\theta}) \boldsymbol{\psi}(Y_j, \hat{\boldsymbol{\theta}}_n), (\partial^2/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T) \boldsymbol{\psi}(Y_j, \hat{\boldsymbol{\theta}}_n) \right), \quad j = 1, \dots, n \right\}.$$

It is expected that the last term at the right-hand side of (5) improves the representation of the random variation about estimate  $\hat{\boldsymbol{\theta}}_n$ . This is confirmed in the simulation study.

Both bootstrap procedures (linear and quadratic) lead to consistent estimators for the distribution of  $\hat{\boldsymbol{\theta}}_n$ . We first list the assumed regularity conditions. For  $k = 1, \dots, p$ ,

(A1) The parameter space  $\Theta$  is an open subset of  $\mathbb{R}^r$ . Second order partial derivatives of  $\boldsymbol{\psi}_k(y, \mathbf{t})$  w.r.t.  $\mathbf{t}$  exist, are continuous in  $y$  and are integrable.

(A2) There exists a function  $H_1$  such that  $E[H_1(Y)^2] < \infty$  and for each  $j = 1, \dots, r$ , for each  $k = 1, \dots, p$ ,  $(\partial/\partial t_j) \boldsymbol{\psi}_k(y; \mathbf{t})$  is bounded in absolute value by  $H_1(y)$  uniformly in some neighborhood of  $\boldsymbol{\theta}$ .

(A3) The  $r \times r$  matrices

$$\mathbf{J}(\boldsymbol{\theta}) = - \sum_{k=1}^p \lambda_k E\{(\partial/\partial \theta_k) \boldsymbol{\psi}_k(y; \boldsymbol{\theta})\}$$

and

$$\mathbf{K}(\boldsymbol{\theta}) = \sum_{k=1}^p \lambda_k E\{\boldsymbol{\psi}_k(y; \boldsymbol{\theta}) \boldsymbol{\psi}_k(y; \boldsymbol{\theta})^t\}$$

exist, and  $\mathbf{K}(\cdot)$  is positive definite in  $\boldsymbol{\theta}$ .

(A4) There exists a  $\delta > 0$  and a function  $H_2$  such that  $E[H_2(Y)] < \infty$ , and for each  $k$ ,  $|\boldsymbol{\psi}_k(y; \mathbf{t})|^{2+\delta}$  is bounded by  $H_2(y)$ , uniformly in some neighborhood of  $\boldsymbol{\theta}$ .

(A5) For each  $k$  and  $\ell$ ,

$$E \left[ \sup_{\|\mathbf{h}\| \leq d} |\boldsymbol{\psi}_k(Y; \boldsymbol{\theta} + \mathbf{h}) \boldsymbol{\psi}_\ell(Y; \boldsymbol{\theta} + \mathbf{h}) - \boldsymbol{\psi}_k(Y; \boldsymbol{\theta}) \boldsymbol{\psi}_\ell(Y; \boldsymbol{\theta})| \right]$$

and

$$E \left[ \sup_{\|\mathbf{h}\| \leq d} \left| \frac{\partial}{\partial t_k} \boldsymbol{\psi}_\ell(Y; \boldsymbol{\theta} + \mathbf{h}) - \frac{\partial}{\partial t_k} \boldsymbol{\psi}_\ell(Y; \boldsymbol{\theta}) \right| \right]$$

both tend to zero as  $d \rightarrow 0$ .

### Theorem

If the above regularity conditions hold, then, for  $\hat{\boldsymbol{\theta}}_n^*$  defined by (3) or (5) and for almost all sample paths  $(Y_1, Y_2, \dots)$ , as  $n \rightarrow \infty$

- (i) The bootstrap estimator  $\hat{\boldsymbol{\theta}}_n^*$  converges in bootstrap probability to  $\boldsymbol{\theta}$ .
- (ii) For any continuous function  $g : \mathbb{R}^r \rightarrow \mathbb{R}^{r'}$  ( $r' \leq r$ ), such that the distribution function of  $g(\mathbf{Z})$ , with  $\mathbf{Z}$  a  $r'$ -dimensional normal distributed random variable, is continuous, we have that

$$\sup_{t \in \mathbb{R}^{r'}} \left| P^* \{g(n^{1/2}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n)) \leq \mathbf{t}\} - P \{g(n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})) \leq \mathbf{t}\} \right| = o(1).$$

**Proof.** Statement (i) follows from (ii). To prove (ii), we need the strong consistency of the estimator  $\hat{\boldsymbol{\theta}}_n$  as defined in (2). By classical maximum likelihood theory (as e.g. in Ferguson, 1996), it can be shown that conditions (A1) and (A2) guarantee the strong consistency. Moreover, if in addition (A3) holds, we have that

$$\left( \mathbf{J}_n^{-1}(\boldsymbol{\theta}) \mathbf{K}_n(\boldsymbol{\theta}) \mathbf{J}_n^{-1}(\boldsymbol{\theta}) \right)^{-1/2} n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_r(\mathbf{0}, \mathbf{I}_r), \quad (6)$$

where in the above expression, the  $r \times r$  matrix

$$\mathbf{J}_n(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} \frac{\partial}{\partial \boldsymbol{\theta}} \psi_i(Y_{ij}, \boldsymbol{\theta})$$

and

$$\mathbf{K}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} \psi_i(Y_{ij}, \boldsymbol{\theta}) \psi_i(Y_{ij}, \boldsymbol{\theta})^t.$$

The next statements concerning bootstrap random variables hold conditionally on  $Y_1, \dots, Y_n$ , for almost all sample paths  $Y_1, Y_2, \dots$ . The notation  $E^*$ ,  $\text{Var}^*$  and  $\mathcal{D}^*$  stand for, respectively, the bootstrap expectation, variance and convergence in distribution, conditionally on  $Y_1, \dots, Y_n$ .

Choosing  $\delta > 0$  according to condition (A4), the strong consistency of  $\hat{\boldsymbol{\theta}}_n$  and (A4) imply that for any  $r$ -dimensional vector  $\mathbf{v}$ , having norm equal to 1,

$$\sum_{i=1}^n E^* [ |n^{-1/2} \mathbf{v}^T \boldsymbol{\psi}_i^*(\hat{\boldsymbol{\theta}}_n)|^{2+\delta} ] = O_P(n^{-\delta/2}). \quad (7)$$

The semiparametric resampling scheme is such that  $E^*[\mathbf{K}_n^*(\hat{\boldsymbol{\theta}}_n)] = \mathbf{K}_n(\hat{\boldsymbol{\theta}}_n)$ . By an application of Theorem 2.9 of Iverson and Randles (1989) to each of the  $r^2$  components of the matrix  $\mathbf{K}_n(\hat{\boldsymbol{\theta}}_n)$ , conditions (A3), (A5) and the strong consistency of  $\hat{\boldsymbol{\theta}}_n$  imply that  $E^*[\mathbf{K}_n^*(\hat{\boldsymbol{\theta}}_n)]$  converges to  $\mathbf{K}(\boldsymbol{\theta})$  almost surely as  $n$  tends to infinity. Together with (7) this implies (Liapunov's condition)

$$\frac{\sum_{i=1}^n E^* [ |n^{-1/2} \mathbf{v}^T \boldsymbol{\psi}_i^*(\hat{\boldsymbol{\theta}}_n)|^{2+\delta} ]}{\left( \sum_{i=1}^n E^* [ (n^{-1/2} \mathbf{v}^T \boldsymbol{\psi}_i^*(\hat{\boldsymbol{\theta}}_n))^2 ] \right)^{1+\delta/2}} \rightarrow 0.$$

By the Cramér-Wold theorem it then follows that

$$n^{-1/2} \sum_{i=1}^n \psi_i^*(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\mathcal{D}^*} \mathcal{N}(\mathbf{0}, \mathbf{K}(\boldsymbol{\theta})).$$

Applying Theorem 2.9 of Iverson and Randles once more, it can be shown that  $E^*[\mathbf{J}_n^*(\hat{\boldsymbol{\theta}}_n)] \rightarrow \mathbf{J}(\boldsymbol{\theta})$  and  $\text{Var}^*[\mathbf{J}_n^*(\hat{\boldsymbol{\theta}}_n)] = O_P(n^{-1})$  such that  $\mathbf{J}_n^*(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P^*} \mathbf{J}(\boldsymbol{\theta})$ .

Combining the above results, an application of Slutsky's theorem leads to

$$n^{1/2} \mathbf{U}_n^* \xrightarrow{\mathcal{D}^*} \mathcal{N}(\mathbf{0}, \mathbf{J}(\boldsymbol{\theta})^{-1} \mathbf{K}(\boldsymbol{\theta}) \mathbf{J}(\boldsymbol{\theta})^{-1}). \quad (8)$$

Statement (ii) follows from the asymptotic normality results (6) and (8) and Pólya's theorem. Note that for the one-step quadratic estimator, conditional on the original sample, the order of the quadratic term is  $O_{P^*}(n^{-1})$  a.s. ■

### 3 Improved estimators

Although the estimator  $\hat{\boldsymbol{\theta}}_n$  obtained from (2) is asymptotically unbiased, we will show how the bootstrap procedure using the quadratic one-step bootstrap can be used for finite sample bias correction.

#### 3.1 Bias corrected estimation

Finite sample bias correction is often obtained by application of bootstrap methods. For some recent literature about this subject, we refer to Kim and Singh (1998) and MacKinnon and Smith (1998). In practical applications a large number, say  $B$ , resamples are taken, resulting in a set of  $B$  bootstrap estimators  $\hat{\boldsymbol{\theta}}_n^{*1}, \dots, \hat{\boldsymbol{\theta}}_n^{*B}$ . From this set a bias corrected estimator is defined as

$$\hat{\boldsymbol{\theta}}_n^{bc} = 2\hat{\boldsymbol{\theta}}_n - \frac{1}{B} \sum_{i=1}^B \hat{\boldsymbol{\theta}}_n^{*i}. \quad (9)$$

For bias estimation, the second order approximation turns out to be very useful, which is not completely unexpected since the bias is a second order aspect of the asymptotic properties of the estimator. The intuition behind equation (9) should be clear. We subtract from the estimator  $\hat{\boldsymbol{\theta}}_n$ , the estimated bias based on the  $B$  bootstrap replicates. The estimated bias is calculated as  $\sum_{i=1}^B \hat{\boldsymbol{\theta}}_n^{*i} / B - \hat{\boldsymbol{\theta}}_n$ .

Table 1 shows that the quadratic one-step bootstrap estimator is quite able to estimate the finite sample bias.

The settings in this simulation are as follows. We generated 2000 data sets of size  $n = 10$  and  $n = 25$ , for each value of  $x$ , from a logistic regression model

$$\text{logit}\{P(Y = 1)\} = \beta_0 + \beta_1 x,$$

with  $(\beta_0, \beta_1)$  equal to  $(-1, -1)$ ,  $(-2.5, 1)$  or  $(-2.5, 2)$ , and  $x = 0, 0.25, 0.5$  and  $1$ . For each of these 2000 data sets we constructed 1000 one-step quadratic bootstrap replicates, the latter were used to obtain the bias corrected estimates  $(\hat{\beta}_0^{bc}, \hat{\beta}_1^{bc})$ . There were some numerical problems with obtaining the original estimates  $(\hat{\beta}_0, \hat{\beta}_1)$ ; the number of generated sets of data without these problems is indicated in Table 1 (conv).

An important observation is that the bias correction even *decreases* the variance, as the simulated standard deviation  $\sigma(\hat{\beta}_0^{bc})$  and  $\sigma(\hat{\beta}_1^{bc})$  are, for all settings in this study, smaller than the corresponding simulated values of  $\sigma(\hat{\beta}_0)$  and  $\sigma(\hat{\beta}_1)$  for the non-bias corrected estimators, respectively. In fact, the reduction in Mean Squared Error (MSE) is quite spectacular. Note that this reduction is much larger for the smallest sample size  $n=10$  than for the setting where  $n=25$ . The explanation for this is the finite sample bias which is less severe for larger samples, hence in those samples, there is less need for bias correction.

	$\beta_0 = -1$		$\beta_0 = -2.5$		$\beta_0 = -2.5$	
	$\beta_1 = -1$		$\beta_1 = 1$		$\beta_1 = 2$	
	$n = 10$	$n = 25$	$n = 10$	$n = 25$	$n = 10$	$n = 25$
conv.	1992	2000	1895	1997	1967	2000
$E(\hat{\beta}_0)$	-1.019	-1.020	-2.604	-2.582	-2.695	-2.599
$E(\hat{\beta}_0^{bc})$	-0.978	-1.006	-2.360	-2.486	-2.459	-2.513
$\sigma(\hat{\beta}_0)$	0.645	0.375	0.889	0.611	0.914	0.559
$\sigma(\hat{\beta}_0^{bc})$	0.588	0.362	0.724	0.554	0.758	0.519
$\frac{MSE(\hat{\beta}_0^{bc})}{MSE(\hat{\beta}_0)}$	0.833	0.933	0.679	0.807	0.660	0.835
$E(\hat{\beta}_1)$	-1.265	-1.070	0.889	1.027	2.152	2.096
$E(\hat{\beta}_1^{bc})$	-1.043	-0.988	0.852	1.001	1.959	2.022
$\sigma(\hat{\beta}_1)$	1.511	0.795	1.555	0.908	1.303	0.779
$\sigma(\hat{\beta}_1^{bc})$	1.314	0.747	1.297	0.836	1.119	0.735
$\frac{MSE(\hat{\beta}_1^{bc})}{MSE(\hat{\beta}_1)}$	0.734	0.875	0.701	0.846	0.728	0.878
$\frac{MSE(\hat{\beta}^{bc})}{MSE(\hat{\beta})}$	0.749	0.885	0.696	0.834	0.705	0.863

Table 1: Simulated mean, standard deviation and mean squared error values of original  $(\hat{\beta}_0, \hat{\beta}_1)$  and bias corrected  $(\hat{\beta}_0^{bc}, \hat{\beta}_1^{bc})$  estimators.



### 3.2 Double bootstrap and variance estimation

Since the bias corrected estimator seems to have interesting properties, we also study its distribution. This estimator is already based on a bootstrap resampling scheme, therefore we will need a second bootstrap stage to obtain this extra amount of information. The double bootstrap procedure reads as follows.

1. Using a nonparametric resampling scheme, that is, resample the pairs of data directly, we construct a set of bootstrap estimators. In the same way as described before, using the one-step quadratic bootstrap, perform a bias correction, and next, construct the bias corrected estimator.
2. To study the distribution of the resulting estimator, we construct from each of the resampled data new bootstrap estimators via the one-step quadratic bootstrap. These values can now be used to compute the estimator's variance, quantiles, etc.

Table 2 illustrates how a double bootstrap method is used to get estimators for the variance of the bias corrected estimator. We restrict attention to the slope parameter and to samples of size 10. It seems that all variance estimators somewhat underestimate the true variability. The double bootstrap variance estimator however is clearly less variable which leads to a substantial reduction in mean squared error.

We now list some details on the results of the simulation study presented in Table 2. The design of the covariate  $x$  is the same as in the previous setting. The number of simulated data sets equals 500, and for each simulated data set we used 500 replicates for the outer and 250 replicates for the inner bootstrap loop. This table contains information about the variability of the variance estimators. The simulated mean of all three estimators for the standard deviation is everywhere smaller than the simulated standard deviation  $\sigma(\hat{\beta}_1)$ . In this simulation we observe the following ordering:

$$\sigma(\hat{\beta}_1) \geq E[\hat{\sigma}(\hat{\beta}_1)] \geq E[\hat{\sigma}(\hat{\beta}_1^{bc})].$$

## 4 Bootstrap confidence intervals

### 4.1 Construction of the intervals

Confidence intervals for the parameter  $\theta$  can be derived from the asymptotic normality result by using the Wald statistic as a pivot. Next to this classical approach, the appropriate quantiles can also be selected from the bootstrap approximation to the asymptotic distribution. In this section we

	$\beta_0 = -1$	$\beta_0 = -2.5$	$\beta_0 = -2.5$
$n = 10$	$\beta_1 = -1$	$\beta_1 = 1$	$\beta_1 = 2$
conv.	499	480	493
$\sigma(\hat{\beta}_1)$	1.511	1.555	1.303
$E(\hat{\sigma}(\hat{\beta}_1))$	1.320	1.514	1.258
$\text{Var}(\hat{\sigma}(\hat{\beta}_1))$	0.171	0.203	0.082
$MSE(\hat{\sigma}(\hat{\beta}_1))$	0.207	0.205	0.084
$\sigma(\hat{\beta}_1^{bc})$	1.314	1.297	1.119
$E(\hat{\sigma}(\hat{\beta}_1^{bc}))$	1.247	1.103	1.067
$\text{Var}(\hat{\sigma}(\hat{\beta}_1^{bc}))$	0.136	0.101	0.022
$MSE(\hat{\sigma}(\hat{\beta}_1^{bc}))$	0.141	0.138	0.025

Table 2: Simulated mean, variance and mean squared error values of double bootstrap variance estimators of  $\hat{\beta}_1$  and of the bias corrected estimator  $\hat{\beta}_1^{bc}$ .

construct bootstrap confidence intervals from the so-called hybrid bootstrap (see, e.g., Shao and Tu, 1995, Sections 4.1 and 4.2). A  $100(1 - \alpha)\%$  confidence interval for the parameter  $\theta$  is defined as

$$\{\theta : z_L^* \leq \sqrt{n}(\hat{\theta}_n - \theta) \leq z_R^*\}$$

where  $z_L^*$  is the  $100\alpha/2\%$  and  $z_R^*$  is the  $100(1 - \alpha/2)\%$  quantile of the distribution of  $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ . In the following section, confidence intervals based on the quadratic one-step bootstrap estimator are compared to the normal-based intervals.

## 4.2 Simulation results

Data are generated from the following logistic model:

$$\text{logit}\{P(Y = 1)\} = \theta_0 + \theta_1 x,$$

where  $x$  belongs to  $\{0, .25, .5, 1\}$ ,  $\theta_0$  is -1.5 or -2.5;  $\theta_1$  takes values in  $\{1, 2\}$  and the sample size varies between 10 and 25 observations per covariate level. For each simulated data set, 1000 bootstrap estimators are constructed.

For 1000 simulated data sets, Table 3 shows the mean lengths of the confidence intervals for the slope parameter for both the normal and the bootstrap procedures. Intervals are constructed at the 90%, 95% and 99% confidence level. Since it turned out that the bootstrap intervals were everywhere shorter than their classical counterparts, the table also shows the percentage of reduction in length.

The most important observation is that the bootstrap intervals all have significantly smaller length for all the situations considered in this simulation setting. Moreover, as shown in Table 4, also their coverage probability is usually somewhat higher.

Except for the smallest sample size and  $\theta_0 = -2.5$  where the coverage probabilities are too small for the quadratic bootstrap, the quadratic bootstrap performs extremely well. Note that in those cases the reduction in length of the confidence intervals is enormous: even up to 16%. We should interpret these results with care, since exactly for these extreme cases, the number of convergences is smaller. The reason for this is not directly clear. Possibly the number of events was very small in those cases, which caused some extreme situations to lead to divergent estimators.

	Normal approx.			Quadratic one-step Bootstrap			Linear one-step Bootstrap		
$n$	90%	95%	99%	90%	95%	99%	90%	95%	99%
$\theta_0 = -2.5, \theta_1=1$									
10	4.38	5.22	6.87	4.03 (8.1%)	4.70 (10%)	5.79 (16%)	4.17 (5.0%)	5.10 (2.4%)	6.65 (3.1%)
15	3.76	4.48	5.88	3.53 (6.0%)	4.16 (7.0%)	5.30 (10%)	3.67 (2.4%)	4.43 (9.9%)	5.84 (0.7%)
25	2.80	3.34	4.39	2.71 (3.5%)	3.22 (3.7%)	4.20 (4.5%)	2.78 (0.9%)	3.32 (0.6%)	4.38 (0.1%)
$\theta_0 = -2.5, \theta_1=2$									
10	3.97	4.73	6.21	3.76 (5.3%)	4.40 (6.9%)	5.61 (9.7%)	3.87 (2.4%)	4.62 (2.2%)	6.06 (2.5%)
15	3.20	3.81	5.00	3.07 (4.0%)	3.64 (4.3%)	4.69 (6.2%)	3.15 (1.5%)	3.76 (1.3%)	4.93 (1.4%)
25	2.45	2.92	3.84	2.39 (2.5%)	2.84 (2.8%)	3.70 (3.6%)	2.43 (0.9%)	2.90 (0.8%)	3.81 (0.7%)
$\theta_0 = -1.5, \theta_1=1$									
10	3.40	4.05	5.32	3.27 (3.7%)	3.87 (4.4%)	4.98 (6.4%)	3.34 (1.8%)	3.97 (2.0%)	5.77 (2.8%)
15	2.70	3.21	4.23	2.63 (2.3%)	3.13 (2.6%)	4.07 (3.6%)	2.67 (1.2%)	3.18 (1.2%)	4.15 (1.7%)
25	2.05	2.44	3.21	2.03 (1.2%)	2.41 (1.4%)	3.16 (1.7%)	2.04 (0.6%)	2.43 (0.7%)	3.19 (0.8%)
$\theta_0 = -1.5, \theta_1=2$									
10	3.40	4.05	5.32	3.25 (4.2%)	3.82 (5.7%)	4.82 (9.5%)	3.34 (1.8%)	3.95 (2.4%)	5.13 (3.7%)
15	2.69	3.20	4.21	2.62 (2.6%)	3.09 (3.4%)	3.96 (5.7%)	2.65 (1.3%)	3.15 (1.6%)	4.11 (2.2%)
25	2.04	2.44	3.20	2.01 (1.4%)	2.39 (2.0%)	3.10 (3.2%)	2.03 (0.6%)	2.41 (0.9%)	3.16 (1.2%)

Table 3: Simulated mean length (reduction in length) of confidence intervals for the slope parameter in a linear logistic regression model.

Table 3 also shows the reduction in length of confidence intervals for the slope parameter in a linear logistic regression model when the *linear* one-step bootstrap is applied. It clearly demonstrates the need of the quadratic approximation, for which the results are everywhere superior to those of

		Normal approx.			Quadr. Bootstrap			Linear Bootstrap		
$n$	conv	90%	95%	99%	90%	95%	99%	90%	95%	99%
$\theta_0 = -2.5, \theta_1=1$										
10	( 944)	84.32*	91.00*	97.24	87.92	91.10*	94.81*	80.08*	88.77*	93.96*
15	( 993)	88.82	94.26	98.79	92.15	96.07	98.59	85.60*	90.74*	96.88*
25	( 997)	88.16	93.58	98.60	91.57	96.19	99.20	86.46*	92.68*	97.89*
$\theta_0 = -2.5, \theta_1=2$										
10	( 985)	91.47	96.14	99.29	93.81•	96.55	98.78	86.70*	92.49*	98.29
15	( 997)	88.67	94.88	98.80	90.97	96.19	99.70	86.36*	93.48	97.89*
25	(1000)	90.60	95.20	99.30	92.20	96.00	99.70	89.20	94.10	98.50
$\theta_0 = -1.5, \theta_1=2$										
10	(1000)	89.30	94.40	99.30	93.00•	97.10•	99.70	87.10*	92.50*	97.00*
15	(1000)	88.20	94.00	98.90	91.10	95.90	99.60	86.70*	92.70*	97.80*
25	(1000)	87.90	94.50	99.40	90.50	96.40	99.60	87.50*	93.50	98.80
$\theta_0 = -1.5, \theta_1=2$										
10	(1000)	89.10	94.70	99.50	93.00•	96.40	99.50	85.70*	91.20*	97.00*
15	(1000)	91.20	95.90	99.50	93.00•	96.90•	99.50	89.50	95.00	98.80
25	(1000)	89.00	94.60	99.30	89.10	95.30	99.30	88.60	94.40	98.50

Table 4: Simulated coverage probabilities (as %) of confidence intervals for the slope parameter in linear logistic regression model. A \* indicates a too small, and • a too large number, according to the 0.01 level.

the linear approximation. There still is a reduction in length of the confidence intervals when the linear one-step bootstrap is applied, but in Table 4 we see that the coverage probabilities decrease too! For most of the settings in this simulation study, the coverage probabilities are much too small for the linear bootstrap, while they are within the range of allowable values (at 1%) for the normal approximation. As a summary, in these settings, the linear bootstrap performs worse than the much simpler normal approximation, this in contrast to the quadratic bootstrap, which outperforms both.

## 5 Discussion

It should be clear that the applications of the one-step bootstrap method are not restricted to those shown here. There are various situations in which this resampling technique can be used. In its full generality, the quadratic bootstrap is applicable in any setting where the estimating functions  $\psi(\cdot, \boldsymbol{\theta})$  satisfy the regularity conditions, in particular, they need to possess at least two continuous derivatives with respect to  $\boldsymbol{\theta}$ , this to ensure the validity of the quadratic expansion.

Although presented here for a full likelihood model, the one-step quadratic bootstrap can be applied for estimation by means of generalized estimating equations (GEE), which can, for example, easily include ways to deal with overdispersion in the data.

Non- and semiparametric models such as for example local polynomial estimation and regression spline smoothing are another area of application. After specification of the suitable estimating function  $\psi$ , the bootstrap method proceeds as shown above. For an application of a one-step linear bootstrap method for local polynomial estimation, we refer to Claeskens and Aerts (2000).

## References

- Aerts, M. and Claeskens, G. (2001). Bootstrap tests for misspecified models, with application to clustered binary data, *Journal of Computational Statistics and Data Analysis* **36**, 383–401.
- Claeskens, G. and Aerts, M. (2000). Bootstrapping local polynomial estimators in likelihood-based models, *Journal of Statistical Planning and Inference* **86**, 63–80.
- Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
- Davison, A.C., Hinkley, D.V. and Schechtman, E. (1986). Efficient bootstrap simulations, *Biometrika* **73**, 555–566.
- Ferguson, T.S. (1996) *A course in large sample theory*. Chapman & Hall, London.

- Ghosh, J.K. (1994). *Higher Order Asymptotics*. NSF-CBMS regional conference series in probability and statistics, Vol. 4.
- Hu, F. and Zidek, J. (1995). A bootstrap based on the estimation equations of the linear model, *Biometrika* **82**, 263–275.
- Iverson, H.K., and Randles, R.H. (1989). The effects on convergence of substituting parameter estimates into U-statistics and other families of statistics, *Probability Theory and Related Fields* **81**, 453–471.
- Kim, Y. and Singh, K. (1998). Sharpening estimators using resampling, *Journal of Statistical Planning and Inference* **66**, 121–146.
- Lee, K.-W. (1990). Bootstrapping logistic regression models with random regressors, *Comm. Stat. Th. Meth.* **19**, 2527–2539.
- Liang, K.-Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**, 13–22.
- MacKinnon, J.G. and Smith, A.A. (1998). “Approximate bias correction in econometrics”, *Journal of Econometrics* **85**, 205–230.
- Moulton, L.H., and Zeger, S.L. (1989). Analyzing repeated measures on generalized linear models via the bootstrap, *Biometrics* **45**, 381–394.
- Moulton, L.H. and Zeger, S.L. (1991). Bootstrapping generalized linear models, *Computational Statistics and Data Analysis* **11**, 53–63.
- Schucany W.R. and Wang S. (1991). One-step bootstrapping for smooth iterative procedures, *Journal of the Royal Statistical Society, Series B* **53**, 587–596.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer, New York.
- Simonoff, J.S. and Tsai, C.-L. (1988). Jackknifing and bootstrapping quasi-likelihood estimators, *Journal of Statistical Computation and Simulation* **30**, 213–232.