RT-MAE 9705

ESTIMATION IN WEIBULL REGRESSION MODEL WITH MEASUREMENT ERROR

by

Patricia Gimenez, Heleno Bolfarine
and
Enrico A. Colosimo

Palavras-Chave: consistent estimator; corrected score; error-in-variables; naive estimator.

Classificação AMS: 62F05, 62J05
(AMS Classification)

- Abril de 1997 -

ESTIMATION IN WEIBULL REGRESSION MODEL WITH MEASUREMENT ERROR

Patricia Gimenez[a] Heleno Bolfarine[a]

Enrico A. Colosimo[a,b]

[a]Departamento de Estatística, IME/USP, Caixa Postal 20570, 01452-990, São Paulo-SP, Brazil

[b]Departamento de Estatística, ICEx/UFMG, Caixa Postal 702, 30161-970, Belo Horizonte-MG, Brazil

Key Words: consistent estimator; corrected score; error-in-variables; naive estimator.

ABSTRACT

A new estimator is proposed for the Weibull regression model with measurement error. The naive estimator obtained by ignoring measurement error is asymptotically biased. This new estimator is established based on the fact that asymptotically the naive estimator converges to the true parameter times an attenuation factor. A brief review of the bias-corrected estimators literature is presented as well as an exhaustive simulation comparing the estimators properties.

1 Introduction

Failure time data is common in many research areas. By failure, we mean the occurence of a event of interest such as death in a clinical trial or product failure in a industrial accelerated life test. By failure time, we mean the period of time taken for the event to occur. An important component of failure time data is the possible presence of censoring. This refers to the circumstance where some subjects or itens under study are event free due to either early withdrawal or termination of the study. The parametric Weibull model plays a special position among the methods for the analysis of failure time data.

Most studies in life sciences, biology, engineering, demography and economics dealing with failure time data involve covariates that can not be recorded exactly. Errors arise, most notably as measurement errors. Examples include a follow-up study of A-bomb survivors where the variable radiation received is measured with error (Okajima, Mine and Nakamura, 1985, Pierce et al., 1992), amount of nitrogen in the soil in a study related to the yield of a certain grain (Fuller, 1987), biologic covariates, such as sistolic blood pressure, daily intake of saturated fat in the famous Framingham Heart prospective study dealing with cardiovascular disease (Gordon and Kannel, 1968).

When predictors are measured with error, the naive estimator obtained by ignoring measurement error is asymptotically biased. Several methods have been developed in order to obtain reduced bias estimators. Some approaches considered replacing contamined covariates by an estimate of them before parameter estimation. Whittemore (1989) proposed Stein estimates and Carroll and Stefanski (1990) proposed an estimation of the conditional expectation of the true covariates given the observed values. These two estimators are very similar when the conditional expectation is linear in the observed values. Stefanski (1985, 1989a) showed how to reduce the bias of the naive estimator such that second-order unbiased estimators were obtained. These estimators are easily computed but yield only approximately consistent estimates. Consistent parameter estimators for nonlinear regression models are difficult to obtain as established by Stefanski (1985), Stefanski and Carroll (1985), Carroll et. al. (1984), Armstrong (1985) and Whittemore and Keller (1988) among others. Nakamura (1990) proposed an estimation method based on a corrected score function that generates consistent and asymptotically normal estimates. A new estimator is proposed by removing the asymptotic bias of the naive estimator. This new estimator is established based on the fact that the naive estimator converges to the true parameter times an attenuation factor.

There are very limited results in literature with respect to small samples properties of these estimators. Zhao and Lee (1996) in a recent paper examine the small sample properties of some bias-corrected estimators for generalized linear measurement error models. However, they do not considered censored observations and replications for the estimation of the error variance. The purpose of this paper is to propose a new estimator and compare its properties with those of some important bias-corrected estimators for the Weibull regression model. This model is described in Section 2, a brief review of the bias-corrected estimators literature as well as a new estimator are presented in Section 3. The simulation study and its results are presented in Section 4.

2 Model Formulation and Notation

The Weibull regression model for the ith failure time T_i of a sample of n observations is usually written in terms of its logaritm as

$$\log T_i = \mathbf{z}_i'\beta + \sigma\epsilon_i \tag{1}$$

where $\mathbf{z}_i' = (1, z_{i1}, \ldots, z_{ip})$ is a $p+1$ vector of covariates, $\beta' = (\beta_0, \ldots, \beta_p)$ is a $p+1$ vector of parameters, σ is a scale parameter and ϵ_i is a random error which has a standard extreme value distribution, $f(\epsilon_i) = e^{\epsilon_i}\exp(e^{\epsilon_i})$.

Consider the classic random censorship model in which a vector of censoring times $(C_1, \ldots, C_n)'$ is assumed independent of the failure times T_is. The observed data are the $min(T_i, C_i)$ and δ_i, that is the failure indicator, such as $y_i = min(logT_i, logC_i)$, $i = 1, \ldots, n$.

Estimation of $\theta = (\beta, \sigma)$ is based on the loglikelihood function which in the presence of censoring observations is written for the model (1), taking $\mathbf{z}_i$ as fixed and assuming that the distribution of the censoring times does not depend on parameters, as

$$\begin{aligned} l(\theta; \mathbf{y}, \mathbf{Z}) &= \sum_{i=1}^{n} l_i(\theta; y_i, \mathbf{z}_i) \\ &= \sum_{i=1}^{n} \delta_i \left\{ \frac{(y_i - \mathbf{z}_i'\beta)}{\sigma} - \log\sigma \right\} - exp((y_i - \mathbf{z}_i'\beta)/\sigma) \end{aligned} \tag{2}$$

where $\mathbf{Z}$ is the design matrix of dimension n by $p+1$ and $\mathbf{y}' = (y_1, \ldots, y_n)$. Estimates of θ are obtained by maximizing (2), which is equivalent to finding the roots defined by the score vector $\mathbf{U}(\theta; \mathbf{y}, \mathbf{Z}) = (\mathbf{U}_1(\theta; \mathbf{y}, \mathbf{Z}), \mathbf{U}_2(\theta; \mathbf{y}, \mathbf{Z}), \mathbf{U}_3(\theta; \mathbf{y}, \mathbf{Z}))'$, where

$$\begin{aligned} U_1(\theta; \mathbf{y}, \mathbf{Z}) &= \frac{1}{\sigma}\sum_{i=1}^{n}\left[\exp((y_i - \mathbf{z}_i'\beta)/\sigma) - \delta_i\right], \\ U_2(\theta; \mathbf{y}, \mathbf{Z}) &= \frac{1}{\sigma}\sum_{i=1}^{n}\left[\exp((y_i - \mathbf{z}_i'\beta)/\sigma)\mathbf{z}_i - \delta_i\mathbf{z}_i\right], \\ U_3(\theta; \mathbf{y}, \mathbf{Z}) &= \sum_{i=1}^{n}\left\{-\delta_i\left[\frac{1}{\sigma} + \frac{(y_i - \mathbf{z}_i'\beta)}{\sigma^2}\right] + \exp((y_i - \mathbf{z}_i'\beta)/\sigma)\frac{(y_i - \mathbf{z}_i'\beta)}{\sigma^2}\right\} \end{aligned} \tag{3}$$

We are concerned with the situation that **Z** can not be recorded directly, but instead we observe a surrogate **X** having measurement error. Considering an additive error model

$$\mathbf{x}_i = \mathbf{z}_i + \mathbf{u}_i, \quad i = 1, \ldots, n$$

where the random error measurement $\mathbf{u}_1, \ldots, \mathbf{u}_n$, is independent of **Z** and **y**, has zero mean and covariance matrix $\Sigma_{\mathbf{u}}$. This covariance matrix may be assumed known or estimated from replications of the $\mathbf{x}_i$s. If the covariates **Z** are fixed constants, a functional model is obtained; if **Z** are independent and identically distributed random vectors from some distribution, a structural model is defined. Although the model (1) exhibits the functional form, the majority of the estimation methods are appropriate for both functional and structural models.

3 Parameter Estimation

The naive estimator $\hat{\theta}_N$ obtained by solving the equations $\mathbf{U}(\hat{\theta}_N; \mathbf{y}, \mathbf{X}) = 0$, such that, the true covariates **Z** are replaced by the observed ones **X** is known to be inconsistent. We review next some typical methods used to improve parameter estimation and propose a new estimator based on the asymptotic properties of the naive estimator. In order to simplify the notation, the presentation consider the measurement error model with just one covariate ($p = 1$). For a general form of the estimators refer to the original papers and the generalization of the new estimator is straightforward. Initially let's consider that $\Sigma_{\mathbf{u}} = \sigma_u^2$ is known. By the end of this section we will establish the estimator for σ_u^2 based on replications for the situation where σ_u^2 is unknown.

3.1 Z estimation approach

Whittemore (1989) and Carroll and Stefanski (1990), based just in the observed values of the covariates, proposed different estimators for **Z**. The methods consist basically of estimating the unobserved covariates **Z** and then estimating of the parameters of interest θ in the usual way. That is, finding θ such that $\mathbf{U}(\theta; \mathbf{y}, \hat{\mathbf{Z}}) = 0$.

Whittemore (1989) considering a functional model with independent normal measurement errors, that is, $u_i \sim N(0, \sigma_u^2)$; $i = 1, \ldots, n$, proposed the James-Stein estimate

$$\hat{z}_i = \hat{B}\bar{x} + (1 - \hat{B})x_i \tag{4}$$

where $\bar{x} = 1/n \sum_{i=1}^{n} x_i$, $\hat{B} = \frac{\sigma_u^2(n-3)}{\hat{S}}$ and $\hat{S} = \sum_{i=1}^{n}(x_i - \bar{x})^2$. This estimator has a justification in terms of the sum of the mean square error loss and a interpretation in terms of the empirical Bayes estimator.

Carroll and Stefanski (1990) considering a structural model proposed the conditional expectation of z_i given x_i estimate, that is, $\hat{z}_i = \hat{E}(z_i/x_i)$. If $u_i \sim N(0, \sigma_u^2)$ and $z_i \sim N(\mu, \nu^2)$, $E(z_i/x_i)$ is linear and can be estimated as

$$\hat{E}(z_i/x_i) = \hat{\mu} + \frac{\hat{\sigma}_x^2 - \sigma_u^2}{\hat{\sigma}_x^2}(x_i - \hat{\mu})$$

where $\hat{\mu} = \bar{x}$ and $\hat{\sigma}_x^2 = \hat{S}/(n-1)$ such that

$$\hat{E}(z_i/x_i) = \hat{B}'\bar{x} + (1 - \hat{B}')x_i \qquad (5)$$

where $\hat{B}' = \frac{\sigma_u^2(n-1)}{\hat{S}}$. Observe that these estimators (4) and (5) have the same form and just a minor difference between $\hat{B}$ and $\hat{B}'$.

3.2 Naive correction approach

Stefanski (1985) derived first order bias of the naive estimator and proposed the following bias-corrected estimator

$$\hat{\theta}_S = \hat{\theta}_N + \frac{\sigma_u^2}{2}\left\{\sum_{i=1}^{n}\frac{\partial \mathbf{U}_i(\hat{\theta}_N; y_i, x_i)}{\partial \theta}\right\}^{-1}\left\{\sum_{i=1}^{n}\frac{\partial^2 \mathbf{U}_i(\hat{\theta}_N; y_i, x_i)}{\partial x_i^2}\right\}. \qquad (6)$$

The method is applicable to both functional and structural cases. Stefanski's (1985) approach is valid only when the covariate measurement error is small.

3.3 Corrected Score Approach

Nakamura (1990) considers a correction for score functions. The main idea is finding a corrected log-likelihood, that is, a function $l^*(\theta; \mathbf{y}, \mathbf{X})$ such that

$$E[l^*(\theta; \mathbf{y}, \mathbf{X})/\mathbf{y}, \mathbf{Z}] = l(\theta; \mathbf{y}, \mathbf{Z})$$

and then consider the corrected score function

$$\mathbf{U}^*(\theta; \mathbf{y}, \mathbf{X}) = \frac{\partial l^*(\theta; \mathbf{y}, \mathbf{X})}{\partial \theta}$$

where, $E[./\mathbf{y}, \mathbf{Z}]$ and $\frac{\partial}{\partial \theta}$ being interchangeable we have that

$$E[\mathbf{U}^*(\theta; \mathbf{y}, \mathbf{X})/\mathbf{y}, \mathbf{Z}] = \mathbf{U}(\theta; \mathbf{y}, \mathbf{Z}). \qquad (7)$$

The estimador θ^* obtained by solving $\mathbf{U}^*(\theta; \mathbf{y}, \mathbf{X}) = 0$ such that $-\frac{\partial \mathbf{U}^*(\theta; \mathbf{y}, \mathbf{X})}{\partial \theta}$ is positive definite, is called the corrected estimador for θ. As noted by Nakamura (1990) $E[\mathbf{U}^*(\theta^0; \mathbf{y}, \mathbf{X})] = 0$, where θ^0 is the true parameter value. Under some regularity conditions and this property, consistency and asymptotic normality of the estimator θ^* can be established. The corrected score method depends critically on the assumed normality of the measurement error. Corrected score functions satisfying (7) may not exist and finding them may not be an easy task. These issues are studied in details in Stefanski (1989b). The corrected score function $\mathbf{U}^*(\theta; \mathbf{y}, \mathbf{X}) = (U_1^*(\theta; \mathbf{y}, \mathbf{X}), U_2^*(\theta; \mathbf{y}, \mathbf{X}), U_3^*(\theta; \mathbf{y}, \mathbf{X}))'$ takes the following form for the Weibull regression model:

$$
\begin{aligned}
U_1^*(\theta; \mathbf{y}, \mathbf{X}) &= \frac{1}{\sigma}\sum_{i=1}^{n}[\exp((y_i - \beta_0 - \beta_1 x_i)/\sigma - \xi) - \delta_i], \\
U_2^*(\theta; \mathbf{y}, \mathbf{X}) &= \frac{1}{\sigma}\sum_{i=1}^{n}\left[\exp((y_i - \beta_0 - \beta_1 x_i)/\sigma - \xi)(x_i + \frac{\beta_1\sigma_u^2}{\sigma}) - \delta_i x_i\right], \\
U_3^*(\theta; \mathbf{y}, \mathbf{X}) &= \sum_{i=1}^{n} -\delta_i\left[\frac{1}{\sigma} + \frac{(y_i - \beta_0 - \beta_1 x_i)}{\sigma^2}\right] + \exp((y_i - \beta_0 - \beta_1 x_i)/\sigma \\
&\quad -\xi)\left[\frac{(y_i - \beta_0 - \beta_1 x_i)}{\sigma^2} - \frac{\beta_1\sigma_u^2}{\sigma^3})\right], \qquad (8)
\end{aligned}
$$

where $\xi = \frac{\beta_1^2\sigma_u^2}{2\sigma^2}$.

3.4 New Estimator

The naive estimator $\hat{\theta}_N$ converges not to the true parameter θ^0 but rather to θ which satisfies

$$E[\mathbf{U}(\theta; \mathbf{y}, \mathbf{X})] = 0. \qquad (9)$$

This expectation is taken over θ^0 and generally $\theta \neq \theta^0$ (Stefanski, 1985).

Evaluating the expression (9) for the non-censored structural Weibull regression model (1), that is, taking $z \sim N(\mu, \nu^2)$, results in

$$\frac{1}{\sigma}E\left[\exp\left(\frac{y - \beta_0 - \beta_1 x}{\sigma}\right) - 1\right] = 0, \qquad (10)$$

$$\frac{1}{\sigma}E\left[\exp\left(\frac{y - \beta_0 - \beta_1 x}{\sigma}\right)x - x\right] = 0, \qquad (11)$$

$$E\left[-\frac{1}{\sigma} - \frac{(y - \beta_0 - \beta_1 x)}{\sigma^2} + \exp\left(\frac{y - \beta_0 - \beta_1 x}{\sigma}\right)\frac{(y - \beta_0 - \beta_1 x)}{\sigma^2}\right] = 0.$$

From (10) and (11) we get the following system of equations

$$\exp\left(\frac{\beta_0^0 - \beta_0 + (\beta_1^0 - \beta_1)\mu}{\sigma} + \xi + \gamma\right)\Gamma(\sigma^0/\sigma + 1) - 1 = 0, \qquad (12)$$

$$\exp\left(\frac{\beta_0^0-\beta_0+(\beta_1^0-\beta_1)\mu}{\sigma}+\xi+\gamma\right)\Gamma(\sigma^0/\sigma+1))\left[\mu+\frac{\nu^2}{\sigma}(\beta_1^0-\beta_1)-\frac{\sigma_u^2}{\sigma}\beta_1\right]$$
$$-\mu=0, \tag{13}$$

where $\xi=\frac{\beta_1^2\sigma_u^2}{2\sigma^2}$ and $\gamma=\frac{\nu^2(\beta_1^0-\beta_1)^2}{2\sigma^2}$. From (12) and (13), we can find

$$\beta_1=k_\nu\beta_1^0$$

where $k_\nu=\frac{\nu^2}{\nu^2+\sigma_u^2}$. It can be observed that k_ν is the same attenuation factor from the ordinary least squares with measurement error. It is usual called reliability ratio (Fuller, 1987).

In this way a consistent estimator for β_1 can be obtained as

$$\tilde{\beta}_1=\hat{k}_\nu^{-1}\hat{\beta}_{1N}$$

where $\hat{k}_\nu=\frac{\hat{\sigma}_x^2-\sigma_u^2}{\hat{\sigma}_x^2}$ with $\hat{\sigma}_x^2$ as in (3.1).

3.5 Estimation of σ_u^2

It was assumed so far that the measurement error variance σ_u^2 is known. The parameter σ_u^2 can be estimated provided we have repeated measurements of x_i. That means, suppose that x_{ij}; $j=1,\ldots,k$ represent k independent and identically distributed replicates such that

$$x_{ij}=z_i+u_{ij},$$

$i=1,\ldots,n$. Replications enable us to estimate σ_u^2 by the usual components of variance analysis, as follows

$$\hat{\sigma}_u^2=\frac{\sum_{i=1}^n\sum_{j=1}^k(x_{ij}-\bar{x}_{i.})^2}{n(k-1)} \tag{14}$$

where $\bar{x}_{i.}=1/k\sum_{j=1}^k x_{ij}$. Replacing x_i by $\bar{x}_{i.}$, σ_u^2 by $\hat{\sigma}_u^2/k$ and $\bar{x}$ by $\bar{x}_{..}=1/n\sum_{i=1}^n\bar{x}_{i.}$ in the expressions of the estimates in (3.1)-(3.4) above we can estimate θ without assume that σ_u^2 is known.

4 Simulation Study

In this section we perform Monte Carlo simulation for comparing the performance of the new estimator and the others described in the previous section. The simulation study is based on a Weibull regression model with shape parameter equal to 2.

Two independent sets of independent random variables $\mathbf{T}'=(T_1,\ldots,T_n)$ and $\mathbf{C}'=(C_1,\ldots,C_n)$ are generated for each repetition and the lifetime

$min(T_i, C_i)$ and δ_i are recorded. T_i is a realization of a Weibull(2, $exp(z_i'\beta)$) and C_i, corresponding to the random censoring mechanism, is uniform on $(0, \tau)$. The true covariate is generated once as a standard normal and it is maintained the same in all repetitions. The error variable is generated as a normal with mean 0 and variance σ_u^2. The parameters β_0 and β_1 are set equal to zero and one respectively and 1000 replications are run for each simulation.

The simulations are performed for several combinations varying the sample sizes ($n = 50, 100, 300$), the variance of the error measurement ($\sigma_u^2 = 0.1, 0.3, 0.5$) and the proportion of censoring ($F = 0, 25, 50\%$) in the sample. The proportion of censoring, $P(C_i < T_i)$, is obtained by controlling the value of the parameter τ. The simulations were run in a CRAY computer using the IMSL software at CCE/USP. Tables I, II and III display the simulations sample means and the root of the mean square error (RMSE) of the various estimators of β_1. The error-free estimate presented in the first line of these tables is the maximum likelihood estimator based on the unobserved data $(y_1, z_1), \ldots, (y_n, z_n)$, that is solution for the equation $\mathbf{U}(\theta; \mathbf{y}, \mathbf{Z}) = 0$. Tables IV, V and VI display the simulations using an estimated value for σ_u^2 based on two replications for each value of z.

5 Discussion

Some conclusions can be drawn from the simulation studies presented in the previous section.

1. The sample size and the proportion of censoring appear that do not affect the bias magnitude of the estimates. As it is known their effects is on the standard deviation of the estimates.

2. As expected the naive estimate attenuates to zero as the error measurement increases. The same seems to happen with the Stefanski's estimate but in a much slower speedy. The others estimators perform well in terms of reducing the bias

3. As expected all estimates are more variable than the naive and error-free ones, and their variability increases as the value of σ_u^2 increases.

4. Comparisons of the estimators using the mean square error have to take into account the magnitude of σ_u^2 and the amount of information provided by the sample (measured by the sample size and the proportion of censoring). Essentially for a small measurement error ($\sigma_u^2 = 0.1$) the naive estimator is so good as the others one. In situations with little information such as $n = 50$ and $F = 50\%$, the standard deviation dominates the bias and even in a bad situation such as $\sigma_u^2 = 0.5$, the naive estimate

has the smaller mean square error. In the others situations it seems that the corrected estimators are really effective and the Whittemore and the new estimators are behaving a little better than the others.

5. Stefanski's estimate has the biggest values for the bias for large values of σ_u^2 since the naive correction only works well for small values of this term. On the other hand, the others estimators do not impose such kind of constraints.

6. The estimate (14) for the error variance σ_u^2 in Tables IV, V and VI is very accurate even for situations with little information such as $n = 50$ and $F = 50\%$. In this way the results obtained in these tables are compared with the formers where σ_u^2 is known and replaced by $\hat{\sigma}_u^2/2$.

In general, it seems that all corrected-bias estimators work well for moderate to large values of σ_u^2 except the Stefanski's estimate that does not work well for $\sigma_u^2 = 0.5$. However, the new estimator is simple and easier to calculate than the others.

Table I - Estimators Performance (no censoring)

Estimate	$n = 50$			$n = 100$			$n = 300$		
	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
Error-free	1.00	0.30	0.30	1.00	0.20	0.20	1.00	0.12	0.12
$\sigma_u^2 = 0.1$									
Naive	0.90	0.28	0.30	0.90	0.19	0.22	0.91	0.12	0.15
New	0.99	0.31	0.31	0.99	0.21	0.21	1.01	0.13	0.13
Nakamura	0.99	0.31	0.31	0.99	0.21	0.21	1.01	0.13	0.13
Stefanski	0.98	0.31	0.31	0.98	0.21	0.21	1.00	0.13	0.13
Whittemore	0.98	0.31	0.31	0.98	0.21	0.21	1.01	0.13	0.13
Carroll-Stefanski	0.99	0.31	0.31	0.99	0.21	0.21	1.01	0.13	0.13
$\sigma_u^2 = 0.3$									
Naive	0.75	0.26	0.36	0.75	0.18	0.31	0.76	0.11	0.27
New	0.96	0.33	0.34	0.98	0.23	0.23	1.00	0.14	0.14
Nakamura	0.97	0.33	0.33	0.98	0.24	0.24	1.00	0.15	0.15
Stefanski	0.92	0.33	0.33	0.93	0.22	0.24	0.94	0.14	0.15
Whittemore	0.95	0.33	0.33	0.97	0.23	0.23	1.00	0.14	0.14
Carroll-Stefanski	0.96	0.33	0.33	0.97	0.23	0.23	1.00	0.14	0.14
$\sigma_u^2 = 0.5$									
Naive	0.64	0.24	0.43	0.65	0.17	0.39	0.65	0.10	0.37
New	0.94	0.35	0.37	0.97	0.25	0.25	0.99	0.16	0.16
Nakamura	0.94	0.36	0.36	0.96	0.26	0.26	0.99	0.16	0.16
Stefanski	0.86	0.33	0.36	0.86	0.23	0.26	0.87	0.14	0.19
Whittemore	0.92	0.34	0.35	0.96	0.25	0.25	0.98	0.16	0.16
Carroll-Stefanski	0.94	0.35	0.36	0.97	0.25	0.25	0.99	0.16	0.16

Table II - Estimators Performance (25% censoring)

Estimate	$n = 50$			$n = 100$			$n = 300$		
	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
Error-free	1.04	0.37	0.37	1.01	0.26	0.26	1.01	0.15	0.15
$\sigma_u^2 = 0.1$									
Naive	0.95	0.35	0.36	0.92	0.25	0.26	0.91	0.14	0.16
New	1.01	0.38	0.38	0.99	0.27	0.27	1.01	0.16	0.16
Nakamura	1.05	0.40	0.40	1.02	0.28	0.28	1.01	0.15	0.15
Stefanski	1.01	0.39	0.39	0.99	0.27	0.27	1.00	0.16	0.16
Whittemore	1.01	0.38	0.38	0.99	0.27	0.27	1.01	0.16	0.16
Carroll-Stefanski	1.01	0.38	0.38	0.99	0.27	0.27	1.00	0.15	0.15
$\sigma_u^2 = 0.3$									
Naive	0.81	0.33	0.38	0.77	0.23	0.32	0.76	0.13	0.27
New	0.98	0.41	0.41	0.97	0.28	0.29	1.00	0.17	0.17
Nakamura	1.07	0.45	0.46	1.04	0.33	0.33	1.01	0.18	0.18
Stefanski	0.95	0.40	0.41	0.93	0.28	0.29	0.95	0.17	0.17
Whittemore	0.97	0.40	0.40	0.96	0.28	0.28	1.00	0.17	0.17
Carroll-Stefanski	0.98	0.41	0.41	0.97	0.28	0.28	1.00	0.17	0.17
$\sigma_u^2 = 0.5$									
Naive	0.66	0.29	0.45	0.67	0.22	0.39	0.66	0.12	0.36
New	0.96	0.43	0.43	0.96	0.30	0.30	0.99	0.18	0.18
Nakamura	0.95	0.43	0.43	1.05	0.37	0.37	1.02	0.21	0.21
Stefanski	0.88	0.40	0.42	0.86	0.28	0.31	0.88	0.17	0.21
Whittemore	0.94	0.42	0.42	0.95	0.30	0.30	0.99	0.18	0.18
Carroll-Stefanski	0.96	0.43	0.43	0.96	0.30	0.30	0.99	0.18	0.18

Table III - Estimators Performance (50% censoring)

	$n = 50$			$n = 100$			$n = 300$		
Estimate	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
Error-free	1.06	0.47	0.47	1.03	0.33	0.33	1.01	0.19	0.19
$\sigma_u^2 = 0.1$									
Naive	0.94	0.42	0.43	0.91	0.31	0.32	0.91	0.18	0.20
New	1.02	0.48	0.48	1.00	0.34	0.34	1.00	0.19	0.19
Nakamura	1.03	0.48	0.48	1.01	0.34	0.34	1.01	0.20	0.20
Stefanski	1.03	0.49	0.49	1.01	0.35	0.35	1.01	0.20	0.20
Whittemore	1.02	0.47	0.47	1.00	0.34	0.34	1.00	0.19	0.19
Carroll-Stefanski	1.02	0.48	0.48	1.00	0.34	0.28	1.00	0.19	0.19
$\sigma_u^2 = 0.3$									
Naive	0.78	0.38	0.43	0.75	0.26	0.36	0.75	0.16	0.30
New	0.98	0.50	0.50	0.97	0.35	0.35	0.99	0.21	0.21
Nakamura	0.99	0.50	0.50	0.98	0.36	0.36	1.00	0.22	0.22
Stefanski	0.96	0.51	0.51	0.94	0.35	0.36	0.95	0.21	0.22
Whittemore	0.97	0.50	0.50	0.97	0.35	0.35	0.99	0.21	0.21
Carroll-Stefanski	0.98	0.50	0.50	0.97	0.35	0.35	0.99	0.21	0.21
$\sigma_u^2 = 0.5$									
Naive	0.68	0.34	0.47	0.65	0.24	0.43	0.64	0.14	0.39
New	0.95	0.53	0.53	0.96	0.37	0.37	0.98	0.22	0.22
Nakamura	0.95	0.51	0.51	0.95	0.38	0.38	0.99	0.23	0.23
Stefanski	0.89	0.52	0.53	0.87	0.35	0.37	0.88	0.20	0.23
Whittemore	0.94	0.52	0.52	0.95	0.36	0.36	0.98	0.22	0.22
Carroll-Stefanski	0.95	0.53	0.53	0.95	0.37	0.37	0.98	0.22	0.22

Table IV - Estimators Performance (no censoring) and two replications for each z

Estimate	$n = 50$			$n = 100$			$n = 300$		
	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
Error-free	1.00	0.30	0.30	1.00	0.20	0.20	1.00	0.12	0.12
$\sigma_u^2 = 0.1$									
$\hat{\sigma}_u^2$	0.10	0.02	0.02	0.10	0.01	0.01	0.10	0.01	0.01
Naive	0.96	0.26	0.26	0.96	0.20	0.20	0.96	0.11	0.12
New	1.00	0.27	0.27	1.00	0.20	0.20	1.00	0.12	0.12
Nakamura	1.01	0.28	0.28	1.01	0.21	0.21	1.00	0.12	0.12
Stefanski	1.00	0.27	0.27	1.00	0.20	0.20	1.00	0.12	0.12
Whittemore	1.00	0.27	0.27	1.00	0.20	0.20	1.00	0.12	0.12
Carroll-Stefanski	1.00	0.27	0.27	1.00	0.20	0.20	1.00	0.12	0.12
$\sigma_u^2 = 0.3$									
$\hat{\sigma}_u^2$	0.30	0.06	0.06	0.30	0.04	0.04	0.30	0.03	0.03
Naive	0.90	0.26	0.28	0.89	0.19	0.22	0.88	0.11	0.16
New	1.01	0.29	0.29	1.01	0.22	0.22	1.01	0.13	0.13
Nakamura	1.01	0.30	0.30	1.01	0.23	0.23	1.01	0.13	0.13
Stefanski	1.00	0.30	0.30	1.00	0.22	0.22	0.99	0.12	0.12
Whittemore	1.00	0.29	0.29	1.00	0.22	0.22	1.00	0.13	0.13
Carroll-Stefanski	1.00	0.29	0.29	1.01	0.22	0.22	1.00	0.13	0.13
$\sigma_u^2 = 0.5$									
$\hat{\sigma}_u^2$	0.50	0.10	0.10	0.50	0.07	0.07	0.50	0.04	0.04
Naive	0.84	0.25	0.30	0.82	0.19	0.26	0.80	0.12	0.23
New	1.01	0.32	0.32	1.01	0.24	0.24	1.01	0.14	0.14
Nakamura	1.02	0.33	0.33	1.02	0.25	0.25	1.01	0.16	0.16
Stefanski	0.99	0.31	0.31	0.98	0.21	0.21	0.97	0.13	0.13
Whittemore	1.00	0.31	0.31	1.01	0.24	0.24	1.01	0.14	0.14
Carroll-Stefanski	1.00	0.31	0.31	1.01	0.24	0.24	1.01	0.14	0.14

Table V - Estimators Performance (25% censoring) and two replications for each z

	$n = 50$			$n = 100$			$n = 300$		
Estimate	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
Error-free	1.04	0.37	0.37	1.01	0.26	0.26	1.01	0.15	0.15
$\sigma_u^2 = 0.1$									
$\hat{\sigma}_u^2$	0.10	0.02	0.02	0.10	0.01	0.01	0.10	0.01	0.01
Naive	0.98	0.32	0.32	0.97	0.24	0.24	0.95	0.13	0.14
New	1.02	0.34	0.34	1.01	0.23	0.23	1.00	0.14	0.14
Nakamura	1.02	0.34	0.34	1.02	0.26	0.26	1.00	0.14	0.14
Stefanski	1.02	0.34	0.34	1.02	0.26	0.26	1.00	0.14	0.14
Whittemore	1.02	0.34	0.34	1.01	0.26	0.26	1.00	0.14	0.14
Carroll-Stefanski	1.02	0.34	0.34	1.01	0.26	0.26	1.00	0.14	0.14
$\sigma_u^2 = 0.3$									
$\hat{\sigma}_u^2$	0.30	0.06	0.06	0.30	0.04	0.04	0.30	0.03	0.03
Naive	0.91	0.32	0.33	0.89	0.24	0.26	0.87	0.13	0.18
New	1.02	0.36	0.36	1.01	0.27	0.27	1.00	0.15	0.15
Nakamura	1.03	0.37	0.37	1.02	0.28	0.28	1.01	0.15	0.15
Stefanski	1.02	0.37	0.37	1.01	0.28	0.28	0.99	0.15	0.15
Whittemore	1.01	0.36	0.36	1.01	0.27	0.27	1.00	0.15	0.15
Carroll-Stefanski	1.02	0.36	0.36	1.01	0.27	0.27	1.00	0.15	0.15
$\sigma_u^2 = 0.5$									
$\hat{\sigma}_u^2$	0.50	0.10	0.10	0.50	0.07	0.07	0.50	0.04	0.04
Naive	0.85	0.31	0.34	0.82	0.23	0.29	0.80	0.12	0.23
New	1.02	0.38	0.38	1.01	0.28	0.28	1.00	0.16	0.16
Nakamura	1.04	0.40	0.40	1.01	0.30	0.30	1.01	0.16	0.16
Stefanski	1.02	0.39	0.39	0.99	0.29	0.29	0.97	0.15	0.16
Whittemore	1.02	0.37	0.37	1.00	0.28	0.28	0.99	0.16	0.16
Carroll-Stefanski	1.02	0.38	0.38	1.01	0.28	0.28	1.00	0.16	0.16

Table VI - Estimators Performance (50% censoring) and two replications for each z

	$n = 50$			$n = 100$			$n = 300$		
Estimate	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
Error-free	1.06	0.47	0.47	1.03	0.33	0.33	1.01	0.19	0.19
$\sigma_u^2 = 0.1$									
$\hat{\sigma}_u^2$	0.10	0.02	0.02	0.10	0.01	0.01	0.10	0.01	0.01
Naive	0.99	0.41	0.41	0.98	0.31	0.31	0.96	0.17	0.17
New	1.04	0.43	0.43	1.02	0.32	0.32	1.01	0.18	0.18
Nakamura	1.04	0.43	0.43	1.03	0.32	0.32	1.01	0.19	0.19
Stefanski	1.05	0.44	0.44	1.03	0.33	0.33	1.01	0.18	0.18
Whittemore	1.04	0.43	0.43	1.02	0.32	0.32	1.01	0.18	0.18
Carroll-Stefanski	1.04	0.43	0.43	1.02	0.32	0.32	1.01	0.18	0.18
$\sigma_u^2 = 0.3$									
$\hat{\sigma}_u^2$	0.30	0.06	0.06	0.30	0.04	0.04	0.30	0.03	0.03
Naive	0.93	0.40	0.40	0.90	0.29	0.31	0.88	0.16	0.20
New	1.04	0.45	0.45	1.02	0.33	0.33	1.00	0.19	0.19
Nakamura	1.05	0.48	0.48	1.03	0.35	0.35	1.01	0.19	0.19
Stefanski	1.06	0.48	0.48	1.03	0.35	0.35	1.01	0.19	0.19
Whittemore	1.04	0.45	0.45	1.02	0.33	0.33	1.00	0.19	0.19
Carroll-Stefanski	1.04	0.45	0.36	1.02	0.33	0.33	1.00	0.19	0.19
$\sigma_u^2 = 0.5$									
$\hat{\sigma}_u^2$	0.50	0.10	0.10	0.50	0.07	0.07	0.50	0.04	0.04
Naive	0.87	0.38	0.41	0.83	0.28	0.33	0.81	0.16	0.25
New	1.05	0.47	0.47	1.02	0.35	0.35	1.00	0.20	0.20
Nakamura	1.05	0.50	0.50	1.04	0.37	0.37	1.02	0.20	0.20
Stefanski	1.06	0.50	0.50	1.02	0.36	0.36	0.99	0.20	0.20
Whittemore	1.04	0.47	0.47	1.02	0.35	0.35	1.00	0.19	0.19
Carroll-Stefanski	1.04	0.47	0.47	1.02	0.35	0.35	1.00	0.20	0.20

BIBLIOGRAPHY

Armstrong,B. (1985). Measurement error in the generalized linear model. *Communications in Statistics - Simulation and Computation*, **14**, 529-544.

Carroll,R.J., Spiegelman,C.H.,Lan,K.K.,Bailey,K.T. and Abbott,R.D. (1984). On errors-in-variables for binary regression models. *Biometrika*, **71**, 19-25.

Carroll,R.J. and Stefanski,L.A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Assoaciation*, **85**, 652-663.

Fuller,W.A. (1987). *Measurement Error Models*. J. Wiley, New York.

Gordon,T. and Kannel,W.E. (1968). *Introduction and general background in the Framinghan study - The Framinghan study, sections 1 and 2*. National Heart, Lung and Blood Institute , Betheesda, MD.

Nakamura,T. (1990). Corrected score function for errors in variables models: Methodology and application to generalized linear models. *Biometrika*, **77**, 127-137.

Okajima,S., Mine,M. and Nakamura,T. (1985). Mortality of registered A-bomb survivors in Nagasaki, Japan, 1970-1984. *Radiation Research*, **103**, 419-431.

Pierce,D.A., Stram,D.O.,Vaeth,M., Schafer,D.(1992). Some insights into the errors in variables problem provided by consideration of radiation dose-response analysis for the A-bomb survivors. *Journal of the American Statistical Assoaciation*, **87**, 351-359.

Stefanski,L.A. (1985). The effects of measurement error on parameter estimation. *Biometrika*, **72**, 583-592.

Stefanski,L.A. (1989a). Correcting data for measurement error in generalized linear models. *Communications in Statistics - Theory and Methods*, **18**, 1715-1733.

Stefanski,L.A. (1989b). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Communications in Statistics - Theory and Methods*, **18**, 4335-4358.

Stefanski,L.A. and Carroll,R.J. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics*, **13**, 1335-1351.

Whittemore, A. S. (1989). Error-in-variables regression using Stein estimates. *The American Statistician*, **43**, 226-228.

Whittemore,A.S. and Keller, (1988). Approximations for regression with covariate measurement error. *Journal of the American Statistical Assoaciation*, **83**, 1057-1066.

Zhao, Y., Lee, A.H. (1996). A simulation study of estimators for generalized linear measurement error models. *Journal of Statistical Computation and Simulation*, **54**, 55-74.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9701 - BOFARINE, H.; ARELLANO-VALE, R.B. Weak nondifferential measurement error models. IME-USP, 1997. 12p. (RT-MAE-9701)

9702 - FERRARI, S.L.P.; CORDEIRO, G. M.; CRIBARI-NETO, F. Higher Order Asymptotic Refinements for Score Tests in Proper Dispersion Models. IME-USP, 1997. 14p. (RT-MAE-9702)

9703 - DOREA, C.; GALVES, A.; KIRA, E.; ALENCAR, A. P. Markovian modeling of the stress contours of Brazilian and European Portuguese. IME-USP, 1997. 11p. (RT-MAE-9703)

9704 - FONTES, L.R.G.; ISOPI, M.; KOHAYAKAWA, Y.; PICCO, P. The Spectral Gap of the REM under Metropolis Dynamics. IME-USP, 1997. 24p. (RT-MAE-9704)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil