Schneeweiss, Hans; Augustin, Thomas

Working Paper

Some recent advances in measurement error models and methods

Discussion Paper, No. 452

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Schneeweiss, Hans; Augustin, Thomas (2005) : Some recent advances in measurement error models and methods, Discussion Paper, No. 452, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München,
https://doi.org/10.5282/ubm/epub.1821

This Version is available at:
https://hdl.handle.net/10419/31008

Some Recent Advances in Measurement Error Models and Methods

Hans Schneeweiß Thomas Augustin*

Abstract

A measurement error model is a regression model with (substantial) measurement errors in the variables. Disregarding these measurement errors in estimating the regression parameters results in asymptotically biased estimators. Several methods have been proposed to eliminate, or at least to reduce, this bias, and the relative efficiency and robustness of these methods have been compared. The paper gives an account of these endeavors. In another context, when data are of a categorical nature, classification errors play a similar role as measurement errors in continuous data. The paper also reviews some recent advances in this field.

Keywords Measurement errors, error in variables, misclassification, efficiency comparison, survival analysis, JEL C13, C20, C24, C25

*This work was supported by the Deutsche Forschungsgemeinschaft (DFG) within the frame of the Sonderforschungsbereich SFB 386. We thank two anonymous referees for their helpful comments.

1 Introduction

A measurement error model is a, linear or non-linear, regression model with (substantial) measurement error in the variables, above all in the regressor variable. Disregarding these measurement errors in estimating the regression parameters (naive estimation) results in asymptotically biased, i.e. inconsistent, estimators. This is the motivation for investigating measurement error models. Measurement errors are found in almost all fields of application. A classical example in econometrics is Friedman's (1957) "permanent income hypothesis". Another example is the measurement of schooling as a predictor of wage earnings (Card, 2001). In epidemiology, various studies may be cited where the impact of an exposure to noxious substances on the health status of people is studied (e.g., Heid *et al.*, 2002). In engineering, the calibration of measuring instruments deals with measurement errors by definition (Brown, 1982). Many more examples can be found in the literature, in particular in the monographs by Schneeweiss and Mittag (1986), Fuller (1987), Carroll *et al.* (1995), Cheng and Van Ness (1999), Wansbeek and Meijer (2000). Recently measurement error methods have been applied in the masking of data to assure their anonymity (Brand, 2002). The data are artificially distorted in various ways including through the addition of random errors.
Several estimation methods have been proposed to eliminate, or at least to reduce, the bias of the naive estimation method. The present paper reviews some of these methods and compares their efficiencies.
Section 2 introduces the measurement error model. In Section 3 we discus briefly the identification problem. Section 4 to 6 deal with various estimation procedures, and Section 7 compares their efficiencies. Section 8 addresses survival models. A special type of measurement errors, viz., misclassification errors is dealt with in Section 9. Section 10 has some concluding remarks.

2 Measurement error models

A measurement error model consists of three parts:

1. A *regression model* relating an unobservable (generally vector-valued, but here for simplicity scalar) regressor variable ξ to a response variable y given by a conditional distribution $f(y|\xi;\theta)$, where θ is an unknown parameter vector. Quite often only the conditional mean function $\mathbb{E}(y|\xi) = m^*(\xi,\beta)$, the regression in the narrower sense, is given, supplemented by a conditional variance function $\mathbb{V}(y|\xi) = v^*(\xi,\beta,\varphi)$, where θ comprises β and φ plus possibly other parameters describing the distribution of y.
Two major examples, that we will often refer to, are the polynomial model

(for a survey see Cheng and Schneeweiss, 2002),

$$y = \beta_0 + \beta_1 \xi + \cdots + \beta_k \xi^k + \epsilon$$

with $m^*(\xi, \beta) = \beta_0 + \beta_1 \xi + \cdots + \beta_k \xi^k$ and $v^* = \sigma_\epsilon^2$, and the log-linear Poisson model

$$y|\xi \sim Po(\lambda), \quad \lambda = \exp(\beta_0 + \beta_1 \xi)$$

with $m^*(\xi, \beta) = v^*(\xi, \beta) = \lambda$. Survival models are considered separately in Section 8.

2. A *measurement model* that relates the unobservable ξ to an observable surrogate variable x, given by a conditional distribution $g(x|\xi; \alpha)$. The so-called non-differentiality property requires that $f(y|\xi, x) = f(y|\xi)$. The classical measurement model assumes an additive random error δ with mean zero, which is independent of ξ and (by non-differentiality) of y:

$$x = \xi + \delta.$$

An alternative is the so-called Berkson model, where δ is independent of x instead of being independent of ξ (e.g., Küchenhoff *et al.*, 2003). Here we shall only consider the classical model. Typically δ is assumed to be normally distributed: $\delta \sim \mathrm{N}(0, \sigma_\delta^2)$.

3. A *distribution of the latent regressor variable* ξ. The distribution may be specified by a density $h(\xi; \gamma)$ with an unknown parameter vector γ. We then have the *structural variant* of the model. Another possibility is that ξ is not considered a random variable but rather an unknown parameter pertaining to the observation x. In this case, which is called the *functional variant*, the number of parameters ξ grows with the sample size. We do not deal with this case here (but see Cheng and Van Ness, 1999). Instead, following Caroll *et al.* (1995), we distinguish between structural and functional estimation methods. The former use the distribution of ξ, the latter do not, even if such a distribution exists. Estimation of β is based on an i.i.d. sample of data (x_i, y_i), $i = 1, \cdots, n$. For an example of estimation in the context of time series see Nowak (1993).

3 Identifiability

Since ξ is latent, the parameters of the model may not be identified. This is the case in the linear model and in the probit model both with normally distributed regressor and error variables. In such cases additional pieces of

information are necessary in order to be able to construct consistent estimators for β, for more details see Cheng and Van Ness (1999). But even if the model is identified (as is often the case in non-linear models - for the logistic model see Küchenhoff, 1995; for the quadratic regression model, see Huang and Huwang, 2001), additional information may be of great help to enhance the efficiency of estimation. The most prominent pieces of extra information are knowledge of the error process, in particular the measurement error variance σ_δ^2, and knowledge of instrumental variables. Here we will only deal with the first type of information (for the second see, e.g., Schneeweiss and Mittag, 1986, and Wansbeek and Meijer, 2000). Knowledge of σ_δ^2 may come from repeated measurements or from a validation subsample. For an example where knowledge of $h(\xi)$ is used see Hu and Ridder (2005).

4 Naive estimation and bias correction

Suppose a consistent estimator $\hat{\beta}$ for the original, error-free model is available. Simply replacing ξ with x in this estimator gives rise to the so-called *naive* estimator $\hat{\beta}_N$. Simple as it is, this estimator is almost always not consistent. As an example consider the linear model $y = \alpha + \beta\xi + \epsilon$. The naive estimator of β is the LS estimator $\hat{\beta}_N = \frac{s_{xy}}{s_x^2}$, which has the bias $-\frac{\sigma_\delta^2}{\sigma_x^2}\beta$. Note that $|\beta|$ is systematically underestimated by $|\hat{\beta}_N|$ (attenuation effect). This has the undesirable consequence that a strong effect of the covariate ξ on y may not be detectable anymore once the covariate has been corrupted by measurement errors. In the multiple linear model, measurement errors have a more complicated effect (see, e.g., Schneeweiss and Mittag, 1986). In the quadratic model the attenuation effect is expressed as a flattening of the curvature at the peak of the parabola (Kuha and Temple, 2003). A segmented linear regression shows a smooth curve connecting the two segments instead of the sharp kink of the error-free model (Küchenhoff and Caroll, 1997).
When the (asymptotic) bias $B = \text{plim}\hat{\beta}_N - \beta$ can be evaluated (typically as a function of β and possibly other parameters), it is sometimes possible to correct the naive estimator such that a consistent estimator results. For instance, the bias of $\hat{\beta}_N$ in the linear model can be easily corrected if σ_δ^2 is known:

$$\hat{\beta}_C = \frac{s_x^2}{s_x^2 - \sigma_\delta^2}\,\hat{\beta}_N = \frac{s_{xy}}{s_x^2 - \sigma_\delta^2}$$

is a consistent estimator of β. Another example is the bias correction of the naive ML estimator in a logistic model (Küchenhoff, 1992).

5 Functional estimation methods

Functional estimators do not use the distribution of ξ. They are therefore immune against possible misspecifications of $h(\xi)$ and they are also valid when ξ is nonstochastic. In this latter case the problem of estimating the incidental parameters ξ_i arises (Cheng and Van Ness, 1999). However, one can circumvent this problem and can directly find estimators for the parameter of interest β. We present two such estimators: CS and SIMEX.

5.1 Corrected score (CS) estimator

Suppose we have a (vector-valued) unbiased estimating (or simply: score) function $\psi(y, \xi; b)$ such that $b = \beta$ is the only solution to the equation $\mathbb{E}[\psi(y, \xi; b)|\xi] = 0$. Then the solution $\hat{\beta}$ of $\sum_{i=1}^{n} \psi(y_i, \xi_i; \hat{\beta}) = 0$, assuming it exists uniquely, is (under general regularity conditions) a consistent estimator of β. However, as ξ is unobservable, this estimator is not feasible. Therefore, one may try to find a so-called corrected score function $\psi_{CS}(y, x; b)$ such that

$$\mathbb{E}[\psi_{CS}(y, x; b)|y, \xi] = \psi(y, \xi; b)$$

(Nakamura, 1990). With the help of the iterative expectation principle ψ_{CS} can be seen to be an unbiased estimating function, and so, under mild regularity conditions, $\hat{\beta}_{CS}$ solving

$$\sum_{i=1}^{n} \psi_{CS}(y_i, x_i; \hat{\beta}_{CS}) = 0$$

is a consistent and asymptotically normal estimator (the CS estimator). Its asymptotic covariance matrix is given by the sandwich formula

$$\Sigma_{CS} = \frac{1}{n} A_{CS}^{-1} B_{CS} A_{CS}^{-\top}, \qquad \text{with } A_{CS} = -\mathbb{E}\left(\frac{\partial \psi_{CS}}{\partial \beta}\right),\ B_{CS} = \mathbb{E}(\psi_{CS}\psi_{CS}^{T}),$$

where $\psi_{CS} = \psi_{CS}(y, x; \beta)$. A common score function of the error-free model is

$$\psi(y, \xi; b) = [y - m^*(\xi, b)]\, v^{*-1} \frac{\partial m^*(\xi, b)}{\partial b}.$$

We then need to find functions f_1 and f_2 such that

$$\mathbb{E}[f_1(x, b)|\xi] = v^{*-1} m_b^*, \qquad \mathbb{E}[f_2(x, b)|\xi] = m^* v^{*-1} m_b^*,$$

where m_b^* is short for $\partial m^*(\xi, b)/\partial b$. Stefanski (1989) gives conditions for the existence of such functions. If they exist, then $\psi_{CS} = yf_1 - f_2$.
In the polynomial model one can construct polynomials $t_r(x)$ of degree r such that $\mathbb{E}[t_r(x)|\xi] = \xi^r$ (Cheng and Schneeweiss, 1998, and Cheng *et al.*, 2000). The corrected score function is then given by

$$\psi_{CS}(y, x; b) = H(x)b - yt(x),$$

where $t(x) = (t_0(x), \cdots, t_k(x))^\top$ and $H(x)$ is a $(k+1) \times (k+1)$ matrix with $H_{rs}(x) = t_{r+s}(x)$, $r, s = 0, \cdots, k$, from which the CS estimator is found as $\hat{\beta}_{CS} = \bar{H}^{-1}\overline{yt}$, where the bar denotes averaging over the sample values (x_i, y_i).
In the Poisson model (see Shklyar and Schneeweiss, 2005), the corrected score function is given by

$$\psi_{CS}(y, x; b_0, b_1) = \left(y - \lambda e^{-\frac{1}{2}b_1^2\sigma_\delta^2}\right)(1, x)^\top + b_1\sigma_\delta^2 e^{-\frac{1}{2}b_1^2\sigma_\delta^2}(0, 1)^\top.$$

5.2 Simulation-extrapolation (SIMEX) estimator

One cannot subtract the measurement error, but one can add a random error to the x_i and thereby study the effect of measurement errors on the estimate of β. This idea gives rise to the following method (Cook and Stefanski, 1994):

1. Compute the naive estimate $\hat{\beta}_N =: \hat{\beta}_{(0)}$.
2. Add random noise to the x_i: $x_i'(a) = x_i + \delta_i'(a)$, $\delta_i' \sim N(0, a\sigma_\delta^2)$ and compute the naive estimate with these artificial data (y_i, x_i').
3. Repeat this step m times with a fixed a and average the m naive estimates to get an estimate $\hat{\beta}_{(a)}$.
4. Do this for a series of a's, e.g. $a = 0.1, 0.2, \cdots, 2$. One may plot the resulting points $(a, \hat{\beta}_{(a)})$.
5. Fit a curve through these points by least squares using some convenient function, e.g., a quadratic one.
6. Extrapolate this curve to $a = -1$, which corresponds to the situation of no measurement error. Then $\hat{\beta}_{SIMEX} = \hat{\beta}_{(-1)}$.

This procedure is easy to apply, as it uses only the naive estimation method given from the original error-free model. It is, however, very computer intensive and it only gives a consistent estimator if the correct extrapolation curve has been used (see Carroll *et al.*, 1996). The quadratic curve may be convenient, but it is rarely the correct curve. SIMEX estimators are therefore often biased, but the bias is typically greatly reduced as compared to the bias of the naive estimator (Wolf, 2004).

6 Structural estimation methods

Structural estimation methods use the information given in the distribution of the regressor variable. Note, however, that this distribution $h(\xi;\gamma)$ contains the unknown (nuisance) parameter vector γ. Typically γ can be estimated from the data x_i alone without recourse to the regression model. For instance, if $\xi \sim N(\mu_\xi, \sigma_\xi^2)$ the nuisance parameters μ_ξ and σ_ξ^2 can be estimated by $\bar{x}$ and $s_x^2 - \sigma_\delta^2$, respectively. For how to estimate γ in a distribution which is a mixture of normals see Thamerus (2003). Replacing γ with a consistent estimate $\hat{\gamma}$ does not alter the consistency property of $\hat{\beta}$, though it does have an effect on the asymptotic variance (cf. Caroll *et al.*, 1995). For simplicity, let us assume in the sequel that γ is known. We will consider three estimators: ML, QS, and RC.

6.1 Maximum likelihood (ML) estimator

The joint density of x and y is given by

$$q(x, y; \theta, \alpha, \gamma) = \int f(y|\xi;\theta) \cdot g(x|\xi;\alpha) \cdot h(\xi;\gamma)\, d\xi$$

Maximizing it with respect to θ, α, γ gives the ML estimator. Though being the most efficient estimator, it has two drawbacks: it relies on the complete joint distribution of x and y and is therefore sensitive to any kind of misspecification and, due to the integral, it is in most cases extremely difficult to compute, not the least because all the parameters have to be estimated simultaneously. Although the computational burden can be greatly alleviated by using simulation methods (simulated ML, simulated LS, see, e.g., Wansbeek and Meijer, 2000, Li, 2000, or Hsiao and Wang, 2000), there is still demand for simpler, and more robust, estimation methods. Two of these, QS and RC, will now be discussed.

6.2 The quasi score (QS) estimator

The *(structural) quasi score* (QS) estimator is constructed by means of the conditional mean and variance function of y given x:

$$\mathbb{E}(y|x) = m(x;\beta), \qquad \mathbb{V}(y|x) = v(x;\beta,\varphi).$$

These are computed starting from the original mean and variance functions given ξ:

$$m(x;\beta) = \mathbb{E}\left[m^*(\xi;\beta)|x\right], \quad v(x;\beta,\varphi) = \mathbb{V}\left[m^*(\xi;\beta)|x\right] + \mathbb{E}\left[v^*(\xi;\beta,\varphi)|x\right].$$

For these computations we need the conditional distribution of ξ given x. In some cases this distribution may be found directly from validation data. In most other cases it is computed from $g(x|\xi;\alpha)$ and $h(\xi;\gamma)$. Therefore m and v do not only depend on β (and φ) but also on α and γ. Here we assume that α and γ are given. In the classical measurement error model with $\delta \sim \mathrm{N}(0, \sigma_\delta^2)$ and $\xi \sim \mathrm{N}(\mu_\xi, \sigma_\xi^2)$ the conditional distribution of ξ given x is simply given by

$$\xi|x \sim \mathrm{N}(\mu(x), \tau^2) \text{ with } \mu(x) = \mu_x + \left(1 - \frac{\sigma_\delta^2}{\sigma_x^2}\right)(x - \mu_x)\,,\ \tau^2 = \sigma_\delta^2\left(1 - \frac{\sigma_\delta^2}{\sigma_x^2}\right).$$

The quasi score function for β then is

$$\psi_{QS}(y, x; b, \varphi) = [y - m(x; b)]\, v^{-1}(x; b, \varphi) m_b(x, b).$$

This should be supplemented by a quasi score function for φ, which we have suppressed for ease of presentation. Given φ, the QS estimator is found as the solution to

$$\sum_{1=1}^{n} \psi_{QS}(y_i, x_i; \hat{\beta}_{QS}, \varphi) = 0.$$

As ψ_{QS} is an unbiased estimating function, $\hat{\beta}_{QS}$ is, under appropriate regularity conditions, a consistent, asymptotically normal estimator with an asymptotic covariance matrix that is again given by a sandwich formula (Kukush and Schneeweiß, 2005).
For the polynomial model first construct $\mathbb{E}(\xi^r|x) = \mu_r(x)$, which is a polynomial of degree r. The QS estimator is then found from the heteroscedastic regression equation

$$\begin{aligned} y &= \beta_0 + \beta_1\mu_1(x) + \cdots + \beta_k\mu_k(x) + u \\ \sigma_u^2 &= \sigma_\epsilon^2 + \sum_{i=0}^{k}\sum_{i=0}^{k} (\mu_{rs}(x) - \mu_r(x)\mu_s(x))\, \beta_r\beta_s \end{aligned}$$

by applying an iteratively reweighted least squares procedures (Kukush *et al.*, 2001).
For the Poisson model (see Shklyar and Schneeweiss, 2005),

$$\begin{aligned} m(x;\beta) &= \exp(\beta_0 + \beta_1\mu(x) + \frac{1}{2}\beta_1^2\tau^2) \\ v(x;\beta) &= m(x;\beta) + \left[\exp(\beta_1^2\tau^2) - 1\right] m^2(x;\beta). \end{aligned}$$

6.3 The regression calibration (RC) estimator

The regression calibration estimator is even simpler to compute than the QS estimator (see Carroll *et al.*, 1995). One replaces the variable x in the naive estimator by $\mu(x)$, which is the best linear predictor of ξ given x (Gleser, 1990).
Thus in the polynomial model, the RC estimator is the LS estimator of the regression

$$y = \beta_0 + \beta_1\mu(x) + \cdots + \beta_k\mu(x)^k + \epsilon.$$

In the Poisson model, the RC estimator is the ML estimator of a Poisson model with $\lambda = \exp\{\beta_0 + \beta_1\mu(x)\}$.
Unfortunately, the RC estimator is inconsistent in general, an exception being the linear model, where RC=QS=CS. But in most cases the bias is greatly reduced as compared to the naive estimator and often negligible (Wolf, 2004).

7 Efficiency comparison

In this section we compare CS and QS with respect to their relative efficiencies. Various results that have been found in the last years will be summarized (Kukush and Schneeweiß 2005, Shklyar and Schneeweiss, 2005, Schneeweiss and Cheng, 2005, Shyklar *et al.*, 2005).
We assume that $\delta \sim N(0, \sigma_\delta^2)$ and $\xi \sim N(\mu_\xi, \sigma_\xi^2)$. Thus we are in the structural case. In addition, a very general regression model of the exponential family is assumed:

$$f(y|\xi) = \exp\left(\frac{y\lambda - c(\lambda)}{\varphi} + a(y,\varphi)\right), \quad \text{with} \quad \lambda = \lambda(\xi,\beta).$$

This model comprises the polynomial and the Poisson model as well as other generalized linear models. Note that in this model $m^* = c'(\lambda)$ and $v^* = \varphi c''(\lambda)$, which will be the basis for constructing the CS estimator, see Section 5.1. Clearly, the ML estimator is the most efficient one. One might speculate that QS is more efficient than CS, as the latter ignores the information inherent in the distribution of ξ. However, this is not at all clear, as QS is not ML. Nevertheless one can, indeed, prove that the presumption is correct, i.e.: $\Sigma_{ML} \leq \Sigma_{QS} \leq \Sigma_{CS}$, at least as long as the nuisance parameters μ_ξ and σ_ξ^2 are given and need not be estimated. Thus if ML is avoided because of its complexity, QS seems to be the estimator of ones choice.
But QS depends on the distribution $f(\xi)$ of the latent regressor. If this distribution is misspecified, then $\hat{\beta}_{QS}$ will typically be biased. Suppose the true

distribution is a finite mixture of normals which cluster around the single normal, erroneously assumed to be the true distribution, and suppose the average distance ϑ of the modes (and the variances) of the mixture components is small and tends to zero, then the misspecification bias of $\hat{\beta}_{QS}$ is of the order ϑ^2. Therefore, in most cases, the bias is practically negligible. There are, however, other forms of misspecification which are not that benign. In any case, misspecification of the regressor distribution is a serious problem with QS.
¿From that point of view, one might prefer CS as the more robust estimator. Even more so, as for small measurement errors, QS and CS and also ML become almost equally efficient anyway. More precisely:

$$\Sigma_{CS} = \Sigma_{ML} + O(\sigma_\delta^4), \quad \Sigma_{QS} = \Sigma_{ML} + O(\sigma_\delta^4).$$

One can also compare CS and QS to the naive method (N). Of course, N is biased. But according to a general rule of thumb one might surmise that the bias of N is compensated by a smaller covariance matrix. Most often this is true, but there are cases where $\Sigma_{CS} - \Sigma_N$ is indefinite or where $\Sigma_{QS} < \Sigma_N$.

8 Survival Analysis

In survival analysis the time until a certain event occurs ('survival time') is considered. The characteristic issue making survival analysis a separate area of research is the problem of censoring: Typically not all survival times T_i, $i = 1, \cdots, n$, can be observed completely; for a subset of the units it is only known that unit i is still alive at some censoring time C_i.

8.1 Measurement Error in Cox-type Models

Mainly two classes of regression models have been studied. The first one, which is to due Cox (1972), relates the individual hazard rate $\lambda(t|\xi)$ to the covariates ξ and the regression parameter β according to the relationship $\lambda(t|\xi) = \lambda_0(t) \cdot \exp(\beta\xi)$. The so-called baseline hazard rate $\lambda_0(t)$ characterizes the dynamic development of risk over time, and is assumed not to depend on i, i.e., the hazards are proportional to each other. Most often $\lambda_0(t)$ is seen as an unspecified nuisance function making the model semiparametric. In particular in econometrics, also parametric versions are of interest (e.g., Flinn and Heckman, 1982).

There are two classical papers on measurement errors in Cox-type models, namely the work by Prentice (1982) and Nakamura (1992), both providing -

to some extent - negative results. Prentice (1982), who relies on the structural case, has shown that a simple likelihood-based correction along the lines of Section 6.1 is not possible (see also Augustin and Schwarz, 2002): The resulting induced relative risk has the form

$$\lambda(t|x) = \lambda_0(t) \cdot \mathbb{E}(\exp(\beta\xi)|x, \{T \geq t\}) \,. \tag{1}$$

Via the event $\{T \geq t\}$ appearing in the conditional expectation, the second factor depends on the previous history of the process, and so the characteristic multiplicative form of the Cox model is lost. As a consequence partial likelihood maximization, i.e., the usual estimation method for the Cox model, can not be directly applied anymore.

However, as Prentice also argued, the effect of this time dependence can be expected to be small if the failure intensity is very low. Under this so-called *rare disease assumption* the condition $\{T \geq t\}$ is almost always satisfied, and so (1) can be solved analytically for normal measurement errors. Then the resulting estimator for β coincides with that obtained from regression calibration, which moreover turns out to be the same as the naive estimator multiplied by the simple deattanuation factor known from linear regression (cf. Section 4). Pepe *et al.* (1989) discuss the accuracy of this approximation (see also Hughes, 1993) and derive further results on handling (1) directly.

Further structural approaches are provided by Hu *et al.* (1998). In general, structural approaches appear promising for dealing with Berkson errors, which, for instance, occurs in cohort studies on exposure to risk factors (Bender *et al.*, 2005; Küchenhoff, *et al.*, 2003).

The classical paper from the functional point of view is Nakamura (1992), who tries to apply his general method of corrected score function (Nakamura, 1990; see also Section 5.1) to partial likelihood estimation. However, the partial likelihood has a singularity in the complex plain, and so - according to a general result from Stefanski (1989) - a corrected score function can not exist. Nakumara (1992) therefore proposes to correct first and second order approximations, instead. The resulting estimators behave not only well in simulation studies, but, surprisingly, the estimator based on first order correction even turned out to be consistent (Kong and Gu, 1999). Moreover, Kong *et al.* (1998) derive a corresponding correction of the cumulative baseline hazard rate $\Lambda_0(t) := \int_0^t \lambda_0(u)du$. Both results are extended in Kong and Gu (1999) to the case of non-normal measurement error. Huang and Wang (2000) suggest a nonparametric variant based on replication data.

A different justification of Nakamura's method for the Cox model and related work is provided by Augustin (2004). He shows that these seemingly approximate corrections are exact corrections, indeed, arising in a straightforward

manner when Nakamura's original concept of corrected score function is applied to the so-called Breslow likelihood instead of partial likelihood. This approach immediately extends to those proportional hazards models where the baseline hazard rate is parameterized and to almost arbitrary measurement error distributions.
Alternative functional correction methods include Buzas' (1998) approach and applications of the so-called conditional score principle in longitudinal Cox models (see, in particular, Tsiatis and Davidian, 2004).

8.2 Accelerated failure time models

The second class of survival models assumes a linear relationship between the log-survival time and the linear predictor: $\ln T = \beta_0 + \beta\xi + \sigma\epsilon$. This model provides a superstructure upon the common parametric duration models like the Weibull, log-logistic, log-normal and gamma model, which are obtained by appropriate specification of ϵ. Recently, also the non-parametric variant, where the distribution of ϵ is left unspecified, has experienced a renaissance. Correction methods for the Weibull model under covariate measurement error have been presented and compared by Gimenez, Bolfarine and Colosimo (1999). Skinner and Humphreys (1999), Wolff and Augustin (2003) and Augustin and Wolff (2004) discuss Weibull regression under error-prone or heaped lifetimes.
The simple linear structure in the logarithm of T also suggests to use mean and variance function models. Augustin (2002, (Chapter 5f.)) derives the corresponding corrected estimating equations to adjust for measurement errors, both from the structural as well as from the functional point of view. The methods obtained allow for a unified treatment of all the commonly used parametric duration models and are the first to handle measurement errors in the covariates and lifetimes simultaneously. Censoring, however, needs additional attention (cf. Augustin, 2002, Theorem 6.2.2), since the estimation equations do not rely on the likelihood anymore.

9 Misclassification

Misclassification of categorical variables is another type of measurement error. As an example, consider a generalized linear model (GLM) for a dichotomous response variable y taking values 0 and 1 with

$$\mathbb{P}(y = 1|x) = G(\kappa), \quad \kappa = x\beta$$

and suppose the response y is occasionally misclassified as y^*. Then using y^* instead of the unknown y in estimating β will produce a bias.
Define the misclassification probabilities

$$\pi_{ij} := \mathbb{P}(y^* = i|y = j, x) = \mathbb{P}(y^* = i|y = j),$$

where the second equality is a consequence of the nondifferentiality postulate. If the π_{ij} are known (as, e.g., when misclassification is used as a masquing device to anonymize data, see Ronning, 2005), or if they can be estimated, (e.g., through a validation study, see Schuster, 1998), then consistent estimators can be constructed. Just observe that

$$\mathbb{P}(y^* = 1|x) = \pi_{11}G(\kappa) + \pi_{10}(1 - G(\kappa)) =: H(\kappa)$$

is again a GLM and can be estimated by conventional methods. For further details see Hausman *et al.* (1998).
Recently Küchenhoff *et al.* (2005) developed a variant of the SIMEX method (see Section 5.2) to be applied to models of the above kind and to more complicated ones. By artificially contorting the data y^* through further misclassification and estimating the resulting models in a naive way, i.e., as if the data were not misclassified, one gets an idea of the amount of bias due to misclassification. One can then extrapolate to the state of no misclassification.

10 Concluding remarks

In this survey we restricted our presentation to parametric regression models in explicit form. We should like to mention a few other approaches.
Functional relations between variables ξ_1 and ξ_2, say, can also be given in the implicit form $f(\xi_1, \xi_2; \beta) = 0$. If instead of ξ_1 and ξ_2 we observe surrogates x_1 and x_2 with additive measurement errors: $x_i = \xi_i + \delta_i$, $i = 1, 2$, and if the error variances are known to be equal, then orthogonal, or total, least squares (TLS) is the method of choice. TLS works nicely in linear models (Cheng and Van Ness, 1999), but leads to biased estimation in nonlinear models. But there is an asymptotic small-σ_δ theory (Fuller, 1987; Amemiya and Fuller, 1988). For the quadratic model, consistent estimators exist (Kukush *et al.*, 2004).
We mentioned that masquing of data can be seen as a method of adding artificial measurement errors to the data. However these measurement errors are often of a quite different type than those considered in this paper. In particular, microaggregation is such a method, which may lead to biased

regression estimators. In order to deal with this bias new methods have been developed (Schmid *et al.*, 2005a,b). A related field, deserving further attention, is the analysis of rounding and heaping errors (e.g., Wolff and Augustin, 2003).

Bibliography

Amemiya, Y., Fuller, W. (1988). Estimation for the nonlinear functional relationship. *Annals of Statistics* **16** 147-160.

Augustin, T. (2002). *Survival Analysis under Measurement Error.* Habilitation (post-doctoral thesis). University of Munich.

Augustin, T. (2004). An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scandinavian Journal of Statistics* **31** 43-50.

Augustin, T., Schwarz, R. (2002). Cox's proportional hazards model under covariate measurement error – A review and comparison of methods. In *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications* (S. Van Huffel, P. Lemmerling eds.), 175-184. Kluwer, Dordrecht.

Augustin, T., Wolff, J. (2004). A bias analysis of Weibull models under heaped data. *Statistical Papers* **45** 211-229.

Bender, R., Augustin, T., Bletter, M. (2005). Simulating survival times for Cox regression models. *Statistics in Medicine* **24** 1713-1723.

Brown, P. J. (1982). Multivariate calibration. *Journal of the Royal Statistical Society* **B 44** 287-321.

Brand, R. (2002). Microdata protection through noise addition. In *Inference Control in Statistical Databases – From Theory to Practice.* (J. Doningo-Ferrer ed.), Lecture Notes in Computer Science 2316. Springer, Berlin.

Buzas, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *Journal of Statistical Planning and Inference* **67** 247-257.

Card, D. (2001). Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica* **69** 1127-1160.

Carroll, R. J., Ruppert, D., Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

Carroll, R. J., Küchenhoff H., Lombard, F., Stefanski, L.A. (1996). Asymptotics for the simex estimator in structural measurement error models. *Journal of the American Statistical Association* **91** 242-250.

Cheng, C.-L., Schneeweiss, H. (1998). Polynomial regression with errors in the variables. *Journal of the Royal Statistical Society* **B 60** 189-199.

Cheng, C.-L., Schneeweiss, H., Thamerus, M. (2000). A small sample estimator for a polynomial regression with errors in the variables. *Journal of the Royal Statistical Society* **B 62** 699-709.

Cheng, C.-L., Schneeweiss, H. (2002). On the polynomial measurement error model. In *Total Least Squares and Errors-in-Variables Modeling* (S. van Huffel, P. Lemmerling eds.), 131-143. Kluwer, Dordrecht.

Cheng, C.-L., Van Ness, J.W. (1999). *Statistical Regression with Measurement Error.* Arnold, London.

Cook, J., Stefanski, L. A. (1994). Simulation-extrapolation estimation for parametric measurement error models. *Journal of the American Statistical Association* **89** 1314-1328.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* **B 34** 187-220.

Flinn, C. J., Heckman, J. J. (1982). Models for the analysis of labor force dynamics. *Advances in Econometrics* **1** 35-95.

Friedman, M. (1957). *A Theory of the Consumption Function.* Princeton University Press. NJ.

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

Gimenez, P., Bolfarine, H., Colosimo, E. A. (1999). Estimation in Weibull regression model with measurement error. *Communications in Statistics – Theory and Methods* **28** 495-510.

Gleser, L. J. (1990). Improvement of the naive estimation in nonlinear errors-in-variables regression. In *Statistical Analysis of Measurement Error Models and Application* (P. J. Brown, W. A. Fuller eds.), *Contemporary Mathematics* **112** 99-114.

Hausman, J. A., Abrevaya, J., Scott-Morton, F. M. (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* **87** 239-269.

Heid, I., Küchenhoff, H., Wellmann, J., Gerken, M., Kreienbrock, L. (2002). On the potential of measurement error to induce differential bias on odds ratio estimates: an example from radon epidemiology. *Statistics in Medicine* **21** 3261-3278.

Hsiao, C., Wang, Q. K. (2000). Estimation of structural nonlinear errors-in-variables models by simulated least-squares method. *International Economic Review* **41** 523-542.

Hu, P., Tsiatis A. A., Davidian M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54** 1407-1419.

Hu, Y., Ridder, G. (2005). *Estimating a nonlinear model with measurement error using marginal information.* http://www-rcf.usc.edu/ ridder/Wpapers/EIV-marg-final.pdf.

Huang, H. S., Huwang, L. (2001). On the polynomial structural relationship. *The Canadian Journal of Statistics* **29** 493-511.

Huang, Y., Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *Journal of the American Statistical Association* **95** 1209-1219 (Correction: **98** 779).

Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics* **49** 1056-1066.

Kong, F. H., Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica* **9** 953-969.

Kong, F. H., Huang, W., Li, X. (1998). Estimating survival curves under proportional hazards model with covariate measurement errors. *Scandinavian Journal of Statistics* **25** 573-587.

Küchenhoff, H., (1992). Estimation in generalized linear models with covariate measurement error using the theory of misspecified models. In *Statistical Modelling* (P. van der Heijden, W. Jansen, B. Francis, G. Seeber, eds.), 185-193. Elsevier Science, Publishers.

Küchenhoff, H. (1995). The identification of logistic regression models with errors in the variables. *Statistical Papers* **36** 41-48.

Küchenhoff, H., Bender, R., Langer, I., Lenz-Tönjes, R. (2003). Effect of Berkson measurement error on parameter estimates in Cox regression models. Discission Paper 346/SFB 386, Universität München.

Küchenhoff, H., Carroll, R. J. (1997). Segmented regression with errors in predictors: semiparametric and parametric methods. *Statistics in Medicine* **16** 169-188.

Küchenhoff, H., Mwalili, S., Lesaffre, E. (2005). A general method for dealing with misclassification in regression: The Misclassification SIMEX. To appear in *Biometrics*.

Kuha, J. T., Temple, J. (2003). Covariate measurement error in quadratic regression. *International Statistical Review* **71** 131-150.

Kukush, A., Markovsky, I., Van Huffel, S. (2004). Consistent estimation in an implicit quadratic measurement error model. *Computational Statistics & Data Analysis* **47** 123-147.

Kukush, A., Schneeweiss, H., Wolf, R. (2001). Comparison of three estimators in a polynomial regression with measurement errors. Discussion Paper 233, Sonderforschungsbereich 386, University of Munich.

Kukush, A., Schneeweiss, H. (2005). Comparing different estimators in a nonlinear measurement error model. I and II. *Mathematical Methods of Statistics* **14** 53-79 and 203-223.

Li, T. (2000). Estimation of nonlinear errors-in-variables models: a simulated minimum distance estimator. *Statistics and Probability Letters* **47** 243-248.

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

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48** 829-838.

Nowak, E. (1993). The identification of multivariate linear dynamic error-in-variables models. *Journal of Econometrics* **59** 213-227.

Pepe, M. S., Self, M. S., Prentice, R. L. (1989). Further results in covariate measurement errors in cohort studies with time to response data. *Statistics in Medicine* **8** 1167-1178.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69** 331-342.

Ronning, G. (2005). Randomized response and the binary probit model. *Economics Letters* **86** 221-228.

Schmid, M., Schneeweiss, H., Küchenhoff, H. (2005a). Consistent estimation of a simple linear model under microaggregation. Discussion Paper 415/SFB 386, Universität München.

Schmid, M., Schneeweiss, H., Küchenhoff, H. (2005b). Statistical inference in a simple linear model under microaggregation. Discussion Paper 416/SFB 386, Universität München.

Schneeweiss, H., Cheng, C.-L. (2005). Bias of the quasi score estimator of a measurement error model under misspecification of the regressor distribution. To appear in *Journal of Multivariate Analysis*.

Schneeweiss, H., Mittag, H. J. (1986). *Lineare Modelle mit fehlerbehafteten Daten.* Physica-Verlag, Heidelberg.

Schuster, G. (1998). ML estimation from binomial data with misclassifications - a comparison: internal validation versus repeated measurements. In *Econometrics in Theory and Practice* (R. Galata, H. Küchenhoff, eds.), 45-58. Physika, Heidelberg.

Shklyar, S., Schneeweiss, H. (2005). A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors. *Journal of Multivariate Analysis* **94** 250-270.

Shklyar, S., Schneeweiss, H., Kukush, A. (2005). Quasi Score is more efficient than Corrected Score in a polynomial measurement error model. Discussion Paper 445/SFB 386, Universität München.

Skinner, C. J., Humphreys, K. (1999). Weibull regression for lifetimes measured with error. *Lifetime Data Analysis* **5** 23-37.

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

Thamerus, M. (2003). Fitting a mixture distribution to a variable subject to heteroscedastic measurement errors. *Computional Statistics* **18** 1-17.

Tsiatis A., Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* **14** 809-834.

Wansbeek, T., Meijer, E. (2000). *Measurement Error and Latent Variables in Econometrics.* Elsevier, Amsterdam.

Wolf, R. (2004). *Vergleich von funktionalen und strukturellen Messfehlerverfahren.* Logos Verlag, Berlin.

Wolff, J., Augustin, T. (2003). Heaping and its consequences for duration analysis - a simulation study. *Allgemeines Statistisches Archiv – Journal of the German Statistical Association* **87** 1-28.