13 Some Recent Advances in Measurement Error Models and Methods*

Hans Schneeweiß[1] and Thomas Augustin[2]

[1] Department of Statistics, Ludwig-Maximilians-Universität
schneew@stat.uni-muenchen.de
[2] Department of Statistics, Ludwig-Maximilians-Universität
augustin@stat.uni-muenchen.de

Summary: 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.

13.1 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, 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,

*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.

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 13.2 introduces the measurement error model. In Section 13.3 we discuss briefly the identification problem. Sections 13.4 to 13.6 deal with various estimation procedures, and Section 13.7 compares their efficiencies. Section 13.8 addresses survival models. A special type of measurement errors, viz., misclassification errors is dealt with in Section 13.9. Section 13.10 has some concluding remarks.

13.2 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 13.8.

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

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