

Effects of measurement errors in predictor selection of linear regression model

Kimmo Vehkalahti^{a,*}, Simo Puntanen^b, Lauri Tarkkonen^a

^a*Department of Mathematics and Statistics, University of Helsinki, P.O. Box 54, FI-00014 Helsinki, Finland*

^b*Department of Mathematics, Statistics and Philosophy, University of Tampere, FI-33014 Tampere, Finland*

Available online 13 May 2007

Abstract

Measurement errors may affect the predictor selection of the linear regression model. These effects are studied using a measurement framework, where the variances of the measurement errors can be estimated without setting too restrictive assumptions about the measurement model. In this approach, the problem of measurement is solved in a reduced *true score space*, where the latent true score is multidimensional, but its dimension is smaller than the number of the measurable variables. Various measurement scales are then created to be used as predictors in the regression model. The stability of the predictor selection as well as the estimated predicted validity and the reliability of the prediction scales is examined by Monte Carlo simulations. Varying the magnitude of the measurement error variance four sets of predictors are compared: all variables, a stepwise selection, factor sums, and factor scores. The results indicate that the factor scores offer a stable method for predictor selection, whereas the other alternatives tend to give biased results leading more or less to capitalizing on chance.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Measurement error; Regression; Reliability; Validity; Factor analysis

1. Introduction

The predictor selection of the linear regression model is affected, not only by the sampling variation, but also by the measurement errors. Let us assume that a predictor, say, x is measured with error. We can express this as $x = \tau + \varepsilon$, where τ is the true value of the predictor and ε is the random measurement error. It is reasonable to assume that ε is uncorrelated with τ , and hence we can write the variance of x as $\text{var}(x) = \text{var}(\tau) + \text{var}(\varepsilon)$, where $\text{var}(\tau)$ represents the sampling variation and $\text{var}(\varepsilon)$ represents the measurement error variation. Either of these may dominate in a given study. If the measurements are unreliable, we cannot improve the situation by increasing the sample size. Instead, we should have more accurate measurements. In many applications it would be preferable to reduce the effects of the measurement errors in the predictor selection, and hence make the models more stable. However, the measurement errors are often neglected in the statistical models, including perhaps the most widely applied one, the linear regression model.

* Corresponding author. Tel.: +358 9 191 24863; fax: +358 9 191 51400.
E-mail address: Kimmo.Vehkalahti@helsinki.fi (K. Vehkalahti).

A classic treatment of measurement errors in regression models is provided by the *errors-in-variables* regression models, also called measurement error models (Cheng and Van Ness, 1999; Fuller, 1987). The fundamental assumption of those models is that each observed variable has its own true value, which is disturbed by a random measurement error. This assumption may lead to problems in the model identification, since the number of parameters to be estimated is easily greater than the number of equations. Hence, additional assumptions are needed. The usual procedure is to assume that the measurement error variances—or the reliabilities of the observed variables—are known (see, e.g., Kim and Saleh, 2005; Cheng and Van Ness, 1999; Gleser, 1992; Fuller, 1987). This may be a reasonable assumption in the physical sciences and engineering (see, e.g., Gleser, 1992, pp. 698–699). However, in areas such as the social sciences or the behavioral sciences, it is usually unrealistic to assume that the reliabilities would be so well established that they could be treated as known. Taking independent replicate experiments to establish the magnitude of the measurement error (see, e.g., Gleser, 1992, pp. 698–699; Fuller, 1987, p. 106) does not provide a satisfactory solution in the above-mentioned fields, since the replicate measurements are seldom independent.

In the most general form, the errors-in-variables regression models combine the regression model with the factor analysis model (Fuller, 1987, Section 4.3). Another method combining these two models has been called *factor analysis regression* (Scott, 1966; Lawley and Maxwell, 1973; Isogawa and Okamoto, 1980). It allows any one of the variables in the factor model to be the dependent variable and uses the regression method to solve a set of simultaneous equations. A more general approach for combining factor models and regression models is provided by the *structural equation modeling* (Jöreskog, 1970; Bollen, 1989), which allows specifying and testing complicated models and relations that include measurement errors. The focus of these models is mainly on the structural relations, the connections between the latent variables.

Our approach for regression modeling with measurement errors is based on the measurement framework (Tarkkonen and Vehkalahti, 2005; Vehkalahti, 2000; Tarkkonen, 1987), where the fundamental assumption, and thus the main difference compared with the errors-in-variables regression models, is that the observed variables are measuring a latent structure, whose dimension is considerably smaller than the number of the variables. Instead of focusing on the single true values of each observed variable, the problem of measurement is solved in a reduced *true score space*. This approach allows us to estimate the measurement error variances without a need to make assumptions that might be unrealistic. In certain respects, our approach comes close to the structural equation modeling, since it also employs the factor model, but instead of the connections between the latent variables we stress the connections established by the *measurement scales*, that is, the linear combinations of the observed variables.

In this paper we take advantage of the measurement framework to study how the measurement errors affect the predictor selection of the linear regression model. We make Monte Carlo simulations based on a certain measurement structure using four different sets of predictors, which are measurement scales created within the measurement framework. Section 2 reviews the basic concepts of the measurement framework and establishes the connection with the linear regression model. Section 3 describes the settings of the simulation studies. Section 4 presents the results and Section 5 concludes.

2. Measurement framework

Our approach for regression modeling with measurement errors is based on the measurement framework (Tarkkonen and Vehkalahti, 2005; Vehkalahti, 2000; Tarkkonen, 1987). In this section we review the basic concepts of the framework and establish the connection between the framework and the linear regression model.

The measurement framework is illustrated in Fig. 1 (the details are explained in the text). The general aim of the framework is (1) to specify the structure of the measurement with a *measurement model*, (2) to estimate the parameters of the measurement model, including the measurement error variances, and (3) to create *measurement scales* for further use, e.g., for regression modeling. Solving the problem of measurement in a reduced true score space facilitates estimating the parameters of interest and assessing the validity and reliability of the measurement scales without extra assumptions.

2.1. Measurement model

Let p variables $\mathbf{x} = (x_1, \dots, x_p)'$ measure k ($< p$) unobservable true scores $\boldsymbol{\tau} = (\tau_1, \dots, \tau_k)'$ with unobservable measurement errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)'$, and assume that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\tau}, \boldsymbol{\varepsilon}) = \mathbf{0}$. The structure of the measurement

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات