Effects of measurement errors in predictor selection of linear regression model [1]

Kimmo Vehkalahti [a,*], Simo Puntanen [b], Lauri Tarkkonen [a]

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

[b] *Department of Mathematics, Statistics & Philosophy, FI-33014 University of Tampere, Finland*

Abstract

We study the effects of random measurement errors in the predictor selection of the linear regression model using a measurement framework. The variances of the measurement errors are estimated instead of the usual procedures where they are assumed to be known. By specifying a measurement model, we solve the problem of measurement in a reduced *true score space*, and then create various measurement scales to be used as predictors in the regression model. We examine the stability of the predictor selection and the predicted validity and reliability of the prediction scales by extensive Monte Carlo simulations. Varying the magnitude of the measurement error variance we compare four sets of predictors: 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.

Key words: Measurement error, Regression, Reliability, Validity, Factor analysis
AMS 2000 subject classification: 62J05, 62H25, 62H20, 91C05

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

[*] Corresponding author.
Email address: Kimmo.Vehkalahti@helsinki.fi (Kimmo Vehkalahti).

[1] Partially based on a contributed paper "Linear regression model with measurement framework" presented by the first author in the 55th Session of the International Statistical Institute (Sydney, Australia, April 2005).

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 can not 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.

A classic treatment of measurement errors in regression models is provided by the *errors-in-variables* regression models [1]. The fundamental assumption of those models is that each observed variable has its own true value, disturbed by a random measurement error. This assumption may lead to problems in the model identification, since there are simply too many parameters to be estimated. The usual procedure is to assume that the measurement error variances—or the reliabilities of the observed variables—are known (see, e.g., [2,3,1]). This may be a reasonable assumption in the physical sciences and engineering (see, e.g., [3, 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., [3, pp. 698–699] or [1, p. 106]) does not either provide a satisfactory solution to the problem of measurement in the above-mentioned fields.

In the most general form, the errors-in-variables regression models combine the regression model with the factor analysis model [1, Sec. 4.3]. Another method combining these two models has been called *factor analysis regression* [4–6]. 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* [7,8], 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 [9–11], 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 the estimation of 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*, i.e., 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 [9–11]. 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. The dimension reduction 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 $\boldsymbol{x} = (x_1, \ldots, x_p)'$ measure k $(< p)$ unobservable true scores $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)'$ with unobservable measurement errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$, and assume that $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ and $\mathrm{cov}(\boldsymbol{\tau}, \boldsymbol{\varepsilon}) = \boldsymbol{0}$. The structure of the measurement is specified by the measurement model

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \tag{2.1}$$

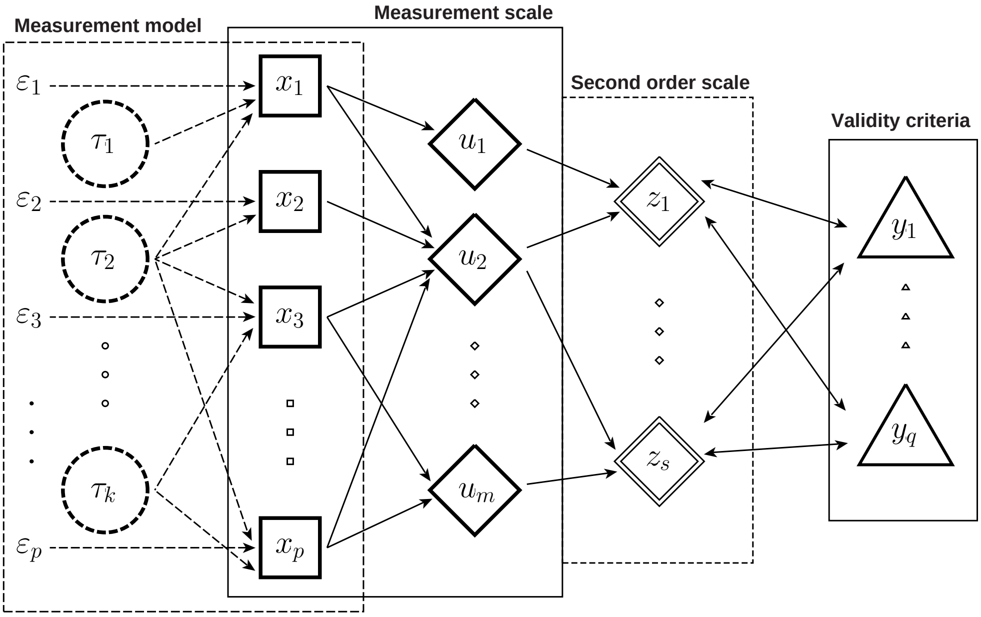

Fig. 1. Elements of the measurement framework.

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ is defined as the expectation of $\boldsymbol{x}$, and the matrix $\boldsymbol{B} \in \mathbb{R}^{p \times k}$ specifies the relationship between $\boldsymbol{x}$ and $\boldsymbol{\tau}$ [9, p. 176]. In Fig. 1, the measurement model is represented by the first frame on the left. The true scores appear as circles and the observed variables as squares. The arrows between them indicate their relationship, corresponding to the elements of the matrix $\boldsymbol{B}$. There is a random measurement error ε_i related to each x_i.

It is often practical to assume that $\text{cov}(\boldsymbol{\tau}) = \boldsymbol{I}_k$, an identity matrix of order k, and $\text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}_d = \text{diag}(\psi_1^2, \ldots, \psi_p^2)$. (We note that throughout this paper we may use the subscript d to indicate a diagonal matrix.) With the above assumptions the measurement model (2.1) conforms with the orthogonal factor analysis model [9, p. 177] and it follows that

$$\text{cov}(\boldsymbol{x}) = \boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{\Psi}_d, \tag{2.2}$$

where the true score variation is separated from the measurement error variation. We assume that $\boldsymbol{B}$ has full column rank, and that $\boldsymbol{\Sigma}$ is positive definite.

2.2 Measurement scale

The variables $\boldsymbol{x}$ are used in further analyses by creating multivariate measurement scales $\boldsymbol{u} = \boldsymbol{A}'\boldsymbol{x}$, where $\boldsymbol{A} \in \mathbb{R}^{p \times m}$ is a matrix of the weights, e.g., factor score coefficients or predetermined values according to a theory [9, pp. 177–178]. In Fig. 1, the measurement scale is represented by the second frame from the left. The distinct scales appear as diamonds, the arrows pointing from the observed variables to the scales symbolizing the weights of the scales. The observed variables are surrounded in Fig. 1 by two frames, as their task is to connect the measurement model and the measurement scale.

The conceptions regarding the measurement errors in the measurement scales are based on the measurement model (2.1) and its assumptions. According to Eq. (2.2) we have

$$\operatorname{cov}(\boldsymbol{u}) = \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A} + \boldsymbol{A}'\boldsymbol{\Psi}_d\boldsymbol{A}, \tag{2.3}$$

which gives the separate contributions to the variances and covariances of the scales by the true scores and the measurement errors. The assumptions made about the measurement model imply that the variances of the scales can be estimated without any extra assumptions.

The measurement scale $\boldsymbol{u}$, being a linear combination of the observed variables, may be called a *first order scale*. Similarly, $\boldsymbol{z} = \boldsymbol{W}'\boldsymbol{u}$ where $\boldsymbol{W} \in \mathbb{R}^{m\times s}$ is a weight matrix, can be called a *second order scale* [9, p. 178]. It is represented in Fig. 1 by the second frame from the right, the distinct scales appearing as double diamonds. By (2.3) we have similarly

$$\operatorname{cov}(\boldsymbol{z}) = \boldsymbol{W}'\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{W} = \boldsymbol{W}'\boldsymbol{A}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}\boldsymbol{W} + \boldsymbol{W}'\boldsymbol{A}'\boldsymbol{\Psi}_d\boldsymbol{A}\boldsymbol{W}. \tag{2.4}$$

The reliabilities of the scales are obtained as ratios of the variances, i.e., the diagonal elements of the matrices in (2.3) and (2.4). Vehkalahti et al. [12] have suggested this general estimator of reliability to be called *Tarkkonen's rho*. In the case of the measurement scales $\boldsymbol{u}$ and $\boldsymbol{z}$, Tarkkonen's rho is a diagonal matrix which can be written in the forms [9, pp. 179–180]

$$\boldsymbol{\rho}_{\boldsymbol{u}} = \{\boldsymbol{I}_m + (\boldsymbol{A}'\boldsymbol{\Psi}_d\boldsymbol{A})_d \times [(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A})_d]^{-1}\}^{-1}, \tag{2.5}$$

and

$$\boldsymbol{\rho}_{\boldsymbol{z}} = \{\boldsymbol{I}_s + (\boldsymbol{W}'\boldsymbol{A}'\boldsymbol{\Psi}_d\boldsymbol{A}\boldsymbol{W})_d \times [(\boldsymbol{W}'\boldsymbol{A}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}\boldsymbol{W})_d]^{-1}\}^{-1}, \tag{2.6}$$

respectively.

2.3 Linear regression model with measurement framework

Our aim here is to take advantage of the measurement framework by creating different measurement scales to be used as predictors in the linear regression model. In general, we have m scales $\boldsymbol{u} = (u_1, \ldots, u_m)'$ which are obtained basically by

$$\boldsymbol{u} = \boldsymbol{A}'\boldsymbol{x} = \boldsymbol{A}'\boldsymbol{\mu} + \boldsymbol{A}'\boldsymbol{B}\boldsymbol{\tau} + \boldsymbol{A}'\boldsymbol{\varepsilon}, \tag{2.7}$$

but we recall that $\boldsymbol{\tau}$ and $\boldsymbol{\varepsilon}$ are unobservable. However, by Eq. (2.3) we can assess their contributions to the scales as soon as we have the estimates of the measurement model parameters, that is, the elements of $\boldsymbol{B}$ and the diagonal elements of $\boldsymbol{\Psi}_d$. The weight matrix $\boldsymbol{A}$ will vary depending on the measurement scale to be applied. In the simulation experiments, we will consider four different sets of predictors, and hence four different weight matrices.

We can now write the linear regression model in the form

$$y = \beta_0 + \boldsymbol{\beta}'\boldsymbol{u} + \delta, \tag{2.8}$$

where y is the response variable, β_0 is the intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$ is the vector of the regression coefficients, $\boldsymbol{u}$ is the vector of the predictors, and δ is a model error, an unobservable random error in the assumed linear relationship between y and $\boldsymbol{u}$.

2.4 Prediction scale and predictive validity

In view of the measurement framework, the prediction scale $z = \boldsymbol{\beta}'\boldsymbol{u}$ is a second order measurement scale, the variance of which is $\mathrm{var}(z) = \boldsymbol{\beta}' \,\mathrm{cov}(\boldsymbol{u})\boldsymbol{\beta}$. Thus the reliability of z can be estimated by

$$\rho_{zz} = \left(1 + \frac{\boldsymbol{\beta}'\boldsymbol{A}'\boldsymbol{\Psi}_d\boldsymbol{A}\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{A}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}\boldsymbol{\beta}}\right)^{-1}, \tag{2.9}$$

which is a special case of the Eq. (2.6), when $\boldsymbol{W} = \boldsymbol{\beta}$ and $s = 1$.

The response variable y represents an external criterion for the *predictive validity* of z. The predictive validity of a measurement scale is assessed by the correlation between the scale and the criterion, and it is denoted by ρ_{zy} [9, p. 182]. In the case of the regression model, ρ_{zy} is equal to the multiple correlation coefficient. In Fig. 1, predictive validity is indicated by the arrows between the second order scales and the validity criteria.

In this study, we will not consider the measurement errors in y, since our focus is on the predictors. In practice, however, y could also be a measurement scale—perhaps based on a different measurement model. We could also have several distinct criteria for different prediction purposes. In the most general case, we would simultaneously have a multidimensional criterion $\boldsymbol{y} = (y_1, \ldots, y_q)'$ and multiple (second order) scales $\boldsymbol{z} = (z_1, \ldots, z_s)'$, which would lead, e.g., to the canonical correlations of $\boldsymbol{y}$ and $\boldsymbol{z}$. However, this study is restricted to the case of the linear regression model, i.e., using a single criterion y and one-dimensional prediction scales z.

The measurement errors in the predictors cause a reduction in all correlations, including ρ_{zy}. If we could eliminate the measurement errors from $z = \boldsymbol{\beta}'\boldsymbol{u}$, we could calculate the true value of z, denoted by ζ, simply by

$$\zeta = \boldsymbol{\beta}'\boldsymbol{u} - \boldsymbol{\beta}'\boldsymbol{A}'\boldsymbol{\varepsilon} = \boldsymbol{\beta}'\boldsymbol{A}'(\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{\tau}),$$

but it is not possible, since we will not have the estimates of $\boldsymbol{\tau}$ or $\boldsymbol{\varepsilon}$. However, we will obtain an estimate of the correlation between ζ and y by applying the *correction for attenuation* (see [9, p. 182]) in the form

$$\rho_{\zeta y} = \frac{\rho_{zy}}{\sqrt{\rho_{zz}}}, \tag{2.10}$$

where the estimate of the reliability of z obtained by Eq. (2.9) is essential. In psychometrics, the square root of the reliability in Eq. (2.10) is sometimes called the *reliability index*.

3 Simulation studies

We conduct simulation studies, because the complexity of the procedures introduced in the previous section makes it impossible to study the effects of the measurement errors analytically. The simulation studies are based on specifying the design of a "true" measurement structure, which is repeatedly used to generate random samples and estimate the parameters of interest. We study four different sets of predictors with a varying magnitude of artificial measurement error variance.

3.1 Design of the measurement structure

Without losing generality, we assume that the observed variables are standardized, that is, $\mathrm{E}(\boldsymbol{x}) = \boldsymbol{0}$ and $\mathrm{cov}(\boldsymbol{x}) = \mathrm{cor}(\boldsymbol{x}) = \boldsymbol{\Sigma}$. Then it is obvious that $\boldsymbol{\Sigma}$ is known as soon as $\boldsymbol{B}$ is known. Hence we can specify the design of the "true" measurement structure by choosing the elements of the matrix $\boldsymbol{B}$ so that $\boldsymbol{B}$ has full column rank and $\boldsymbol{\Sigma}$ is positive definite.

The chosen measurement structure consists of $k = 3$ true scores and $p = 13$ variables. Table 1 presents the matrix $\boldsymbol{B}$ together with the rowwise and columnwise sums of squares of its elements. The variables $x_1, \ldots, x_9$ are the primary contributors to the true scores τ_1, τ_2, and τ_3. That part of the matrix $\boldsymbol{B}$ has a pure simple structure (see, e.g., [13, p. 573]). Without the rest of the variables, which act as a sort of confounders, the structure would be overly simple. In practice, the variables $x_{10}, \ldots, x_{13}$ could represent some background information affecting the traits measured by the other variables.

Table 1
Matrix $\boldsymbol{B}$ with the sums of squares of the elements (zeros omitted).

	True score			
Variable	τ_1	τ_2	τ_3	Sumsqr
x_1	0.9			0.81
x_2	0.8			0.64
x_3	0.7			0.49
x_4	0.6			0.36
x_5		0.8		0.64
x_6		0.7		0.49
x_7		0.6		0.36
x_8			0.7	0.49
x_9			0.6	0.36
x_{10}	0.5	0.5	0.5	0.75
x_{11}	0.6	−0.6		0.72
x_{12}		0.6	−0.6	0.72
x_{13}	−0.6		0.6	0.72
Sumsqr	3.27	2.46	1.82	

The true scores appear in the order of the columnwise sums of squares of the elements. Without the confounders τ_3 would be quite weak, with x_8 and x_9 as its only indicators. The rowwise sums of the squares (communalities in factor analysis) vary from 0.36 to 0.81. According to the measurement model, they indicate the variance generated by the true scores for each variable. Similarly, the rest of the variance is generated by the measurement errors. The variables x_4, x_7, and x_9 are the weakest ones, that is, most of their variance is due to the measurement errors. The confounders are among the best ones in this respect. To conclude, the design of the chosen measurement structure should be general enough to form a reasonable basis for the simulation studies.

3.2 Random samples

The random samples are repeatedly generated by creating p independent, normally distributed, standardized variables $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$ and computing $\boldsymbol{x} = \boldsymbol{C}\boldsymbol{\eta}$, where $\boldsymbol{C}$ is obtained from the spectral decomposition $\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}'$. Because $\boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{I}_p)$, it follows that $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$. We recall that $\boldsymbol{\Sigma}$ is known. The sample variation is established by using each time different seeds for a combined Tausworthe pseudo random number generator [14]. In order to estimate the pk elements of $\boldsymbol{B}$ and the p diagonal elements of $\boldsymbol{\Psi}_d$, that is, the required $13 \cdot 3 + 13 = 52$ parameters of the measurement model, it is reasonable to use at least $n = 100$ observations. To control for the sample size, the simulations are conducted with $n = 100, 300, 500, 1000$.

The response variable y is computed as a plain sum $y = \mathbf{1}'\boldsymbol{x} + \delta$, where $\mathbf{1}$ is the vector of ones. The model error $\delta \sim N(0, \sigma^2)$, where σ^2 brings some additional sampling variation in y, in order to make the relationship between y and the predictors a bit more complicated. We recall that in this study we ignore the measurement errors in y, including those implied by the design (see below), and interpret all the variation in y to be caused by sampling. To control for the sampling variation, the simulations are conducted with $\sigma^2 = 4$ and $\sigma^2 = 9$.

The design of the measurement structure implies that the basic magnitude of the measurement error variance in the $\boldsymbol{x}$ variables is given by $\boldsymbol{\Psi}_d^* = \boldsymbol{\Sigma} - \boldsymbol{B}\boldsymbol{B}'$. After the response variable y has been computed, artificial measurement error $\tilde{\boldsymbol{\varepsilon}}$ is added in $\boldsymbol{x}$ assuming that $\tilde{\boldsymbol{\varepsilon}} \sim N(\mathbf{0}, \theta^2 \boldsymbol{I}_p)$, where θ is given fixed values. The total variance of the measurement error to be estimated is therefore given by $\operatorname{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}_d = \boldsymbol{\Psi}_d^* + \theta^2 \boldsymbol{I}_p$. When $\theta = 0$, we have simply $\boldsymbol{\Psi}_d = \boldsymbol{\Psi}_d^*$. To study the effect of the measurement errors, the simulations are conducted increasing θ from 0 up to 2 in increments of 0.5.

3.3 Parameter estimation

For each sample we have to estimate the parameters of the measurement model (2.1) and the parameters of the regression model (2.8) using different sets of predictors. We will denote the sample correlation matrix by $\hat{\boldsymbol{\Sigma}}$, and the estimated factor matrix by $\hat{\boldsymbol{B}}_0$. The latter will be transformed according to the design, and then denoted by $\hat{\boldsymbol{B}}$. The estimates of the measurement error variances will be denoted by $\hat{\boldsymbol{\Psi}}_d$.

3.3.1 Parameters of the measurement model

Due to the assumptions we have made earlier, the parameters of the measurement model (2.1) can be estimated from $\hat{\boldsymbol{\Sigma}}$ by the maximum likelihood factor analysis. Because of this connection, we prefer the term “factor” to “true score” from now on. According to the design, the number of factors to be extracted is fixed at $k = 3$. By each choice of n, the sample size is sufficient to ensure that the parameter estimates are consistently found. However, it is possible that an element of $\hat{\boldsymbol{\Psi}}_d$ becomes negative (so called *Heywood case*, see, e.g., [15, p. 217]). These cases are excluded from the analyses.

Because of the usual rotational indeterminacy of the factor model, $\hat{\boldsymbol{B}}_0$ is not unique. In particular, it will not necessarily correspond to $\boldsymbol{B}$, despite the same number of the factors. Instead of using typical factor rotation methods, we have to find the particular transformation that gives the "best match" between $\hat{\boldsymbol{B}}_0$ and $\boldsymbol{B}$.

We need the following result, related to *Procrustes rotation* [16], but also to *transformation analysis* [17], where more emphasis is given on the deviations (or residuals) after finding the "best match". The optimal method of solution for orthogonal factors was found independently by several authors in 1966 [18–20]. Our formulation of the problem is adapted from Proposition 15.5 in [21, p. 162], and our proof is based on [22, pp. 95–98] and [15, pp. 253–255].

Lemma 1 *Let $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ be given $p \times k$ matrices $(k < p)$ and let $\mathscr{O}_{k\times k}$ be the set of orthogonal $k \times k$ matrices. Then*

$$\min_{\boldsymbol{Z}\in\mathscr{O}_{k\times k}} \|\boldsymbol{B}_1\boldsymbol{Z} - \boldsymbol{B}_2\|^2 = \|\boldsymbol{B}_1\boldsymbol{L} - \boldsymbol{B}_2\|^2, \tag{3.1}$$

where $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{V}'$ is obtained from the singular value decomposition $\boldsymbol{B}_1'\boldsymbol{B}_2 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$.

PROOF. To prove (3.1), we first write

$$\begin{aligned}\|\boldsymbol{B}_1\boldsymbol{Z} - \boldsymbol{B}_2\|^2 &= \operatorname{tr}[(\boldsymbol{B}_1\boldsymbol{Z} - \boldsymbol{B}_2)(\boldsymbol{B}_1\boldsymbol{Z} - \boldsymbol{B}_2)'] \\ &= \operatorname{tr}(\boldsymbol{B}_1\boldsymbol{B}_1') + \operatorname{tr}(\boldsymbol{B}_2\boldsymbol{B}_2') - 2\operatorname{tr}[\boldsymbol{Z}(\boldsymbol{B}_1'\boldsymbol{B}_2)'],\end{aligned} \tag{3.2}$$

and observe that minimizing (3.2) is equivalent to maximizing $\operatorname{tr}[\boldsymbol{Z}(\boldsymbol{B}_1'\boldsymbol{B}_2)']$. Using the the singular value decomposition $\boldsymbol{B}_1'\boldsymbol{B}_2 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$, where $\boldsymbol{U}, \boldsymbol{V} \in \mathscr{O}_{k\times k}$, we obtain

$$\operatorname{tr}[\boldsymbol{Z}(\boldsymbol{B}_1'\boldsymbol{B}_2)'] = \operatorname{tr}[\boldsymbol{Z}(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}')'] = \operatorname{tr}(\boldsymbol{U}'\boldsymbol{Z}\boldsymbol{V}\boldsymbol{D}) = \operatorname{tr}(\boldsymbol{R}\boldsymbol{D}), \tag{3.3}$$

where $\boldsymbol{R} = \boldsymbol{U}'\boldsymbol{Z}\boldsymbol{V} \in \mathscr{O}_{k\times k}$, being the product of orthogonal matrices. Since the elements of $\boldsymbol{R}$ can not exceed 1, we have

$$\operatorname{tr}(\boldsymbol{R}\boldsymbol{D}) = \sum_{j=1}^{k} r_{jj}d_j \le \sum_{j=1}^{k} d_j = \operatorname{tr}(\boldsymbol{D}),$$

and the maximum of (3.3) and hence the minimum of (3.2) is clearly attained when $\boldsymbol{R} = \boldsymbol{I}_k$. Our claim (3.1) follows by selecting $\boldsymbol{Z} = \boldsymbol{L} = \boldsymbol{U}\boldsymbol{V}'$, as then $\boldsymbol{R} = \boldsymbol{U}'\boldsymbol{U}\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_k$. □

We note that the orthogonality of $\boldsymbol{L}$ implies a symmetric solution, i.e., the roles of $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ can be interchanged and thus the results do not depend on the direction of the comparison. Therefore this form of the method has been called *symmetric transformation analysis* [20].

In our study, $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are given by $\hat{\boldsymbol{B}}_0$ and $\boldsymbol{B}$, respectively. Using Lemma 1 the required transformation matrix is hence obtained from the singular value decomposition $\hat{\boldsymbol{B}}_0'\boldsymbol{B} = \boldsymbol{UDV}'$ in the form $\boldsymbol{L} = \boldsymbol{UV}'$, and the "best match" with $\boldsymbol{B}$, denoted by $\hat{\boldsymbol{B}}$, is provided by the transformation $\hat{\boldsymbol{B}} = \hat{\boldsymbol{B}}_0\boldsymbol{L}$.

3.3.2 Parameters of the regression model

We now turn to the parameter estimation of the regression model (2.8). Our focus is on the predictors, and how they are affected by the measurement errors. We consider four different sets of predictors, denoted by P_1, P_2, P_3, and P_4. According to Eq. (2.7), the predictor sets are measurement scales $\boldsymbol{u} = \boldsymbol{A}'\boldsymbol{x}$ created with different weight matrices $\boldsymbol{A}$. They are summarized in Table 2 and defined as follows:

- P_1 is a trivial set in which each observed variable forms its own scale. This is formally denoted by $P_1 : \boldsymbol{u} = \boldsymbol{x} = \boldsymbol{I}_p\boldsymbol{x}$, that is, P_1 gives the predictors of the full regression model.
- P_2 involves a subset of r ($\leq p$) variables selected by a stepwise regression algorithm. Let $h_i = 1$, if variable x_i is selected, and $h_i = 0$ otherwise. The complete set of p variables is denoted by $P_2 : \boldsymbol{u} = \boldsymbol{h} = \boldsymbol{H}_d\boldsymbol{x}$, where $\boldsymbol{H}_d = \mathrm{diag}(h_1, \ldots, h_p)$. However, the regression models are estimated using a reduced set $P_2^* : \boldsymbol{u} = \boldsymbol{h}^* = \boldsymbol{I}_r\boldsymbol{x}^*$, where $\boldsymbol{x}^*$ includes only those variables with $h_i = 1$.
- P_3 is a set of k ($< p$) sums of the variables, with the weights either 0 or 1, depending on the elements of the true factor matrix $\boldsymbol{B} = (b_{ij})$ given in Table 1. Let $\boldsymbol{G} = (g_{ij}) \in \mathbb{R}^{p\times k}$, where $g_{ij} = 1$, if $b_{ij} > 0$ and $g_{ij} = 0$ otherwise. This set we are calling the factor sums is denoted by $P_3 : \boldsymbol{u} = \boldsymbol{g} = \boldsymbol{G}'\boldsymbol{x}$.
- P_4 is a set of k ($< p$) factor scores by regression method (see, e.g., [9, p. 178]). The set of the factor scores is denoted by $P_4 : \boldsymbol{u} = \boldsymbol{s} = \hat{\boldsymbol{B}}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{x}$.

The parameters of the regression model (2.8) are estimated from the simulated samples for each predictor set. The models include the constant term β_0, although it is not of any special interest in this study. Because of the $p - r$ variables omitted by the selection algorithm, the dimension of the predictor set P_2^* is incompatible with the measurement model (unless $r = p$, i.e., all the p variables are selected). Hence we must expand the vector of the estimated regression coefficients $\hat{\boldsymbol{\beta}}_{\boldsymbol{h}^*} = (\hat{\beta}_{h_1^*}, \ldots, \hat{\beta}_{h_r^*})'$ with $p - r$ zeroes to obtain $\hat{\boldsymbol{\beta}}_{\boldsymbol{h}} = (\hat{\beta}_{h_1}, \ldots, \hat{\beta}_{h_p})'$ for subsequent analyses.

Table 2
Predictor sets and prediction scales.

Set	Dim.	Weights	Predictors	Regression coefficients	Prediction scale
P_1	p	$\boldsymbol{I}_p$	$\boldsymbol{x}$	$\boldsymbol{\beta}_{\boldsymbol{x}} = (\beta_{x_1}, \ldots, \beta_{x_p})'$	$z_1 = \hat{\boldsymbol{\beta}}'_{\boldsymbol{x}}\boldsymbol{x}$
P_2^*	r	$\boldsymbol{I}_r$	$\boldsymbol{h}^* = \boldsymbol{x}^*$	$\boldsymbol{\beta}_{\boldsymbol{h}^*} = (\beta_{h_1^*}, \ldots, \beta_{h_r^*})'$	—
P_2	p	$\boldsymbol{H}_d$	$\boldsymbol{h} = \boldsymbol{H}_d\boldsymbol{x}$	—	$z_2 = \hat{\boldsymbol{\beta}}'_{\boldsymbol{h}}\boldsymbol{h}$
P_3	k	$\boldsymbol{G}$	$\boldsymbol{g} = \boldsymbol{G}'\boldsymbol{x}$	$\boldsymbol{\beta}_{\boldsymbol{g}} = (\beta_{g_1}, \ldots, \beta_{g_k})'$	$z_3 = \hat{\boldsymbol{\beta}}'_{\boldsymbol{g}}\boldsymbol{g}$
P_4	k	$\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{B}}$	$\boldsymbol{s} = \hat{\boldsymbol{B}}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{x}$	$\boldsymbol{\beta}_{\boldsymbol{s}} = (\beta_{s_1}, \ldots, \beta_{s_k})'$	$z_4 = \hat{\boldsymbol{\beta}}'_{\boldsymbol{s}}\boldsymbol{s}$

The vectors $\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{h}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{g}}$, and $\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}$ determine the prediction scales z_1, z_2, z_3, and z_4, respectively (see Table 2). The reliability of the prediction scales can be estimated with Eq. (2.9) by using the corresponding estimates of the matrices. Hence we have

$$\hat{\rho}_{z_1z_1} = \left(1 + \frac{\hat{\boldsymbol{\beta}}'_{\boldsymbol{x}}\hat{\boldsymbol{\Psi}}_d\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}}{\hat{\boldsymbol{\beta}}'_{\boldsymbol{x}}\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}'\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}}\right)^{-1}, \hat{\rho}_{z_2z_2} = \left(1 + \frac{\hat{\boldsymbol{\beta}}'_{\boldsymbol{h}}\boldsymbol{H}_d\hat{\boldsymbol{\Psi}}_d\boldsymbol{H}_d\hat{\boldsymbol{\beta}}_{\boldsymbol{h}}}{\hat{\boldsymbol{\beta}}'_{\boldsymbol{h}}\boldsymbol{H}_d\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}'\boldsymbol{H}_d\hat{\boldsymbol{\beta}}_{\boldsymbol{h}}}\right)^{-1},$$

$$\hat{\rho}_{z_3z_3} = \left(1 + \frac{\hat{\boldsymbol{\beta}}'_{\boldsymbol{g}}\boldsymbol{G}'\hat{\boldsymbol{\Psi}}_d\boldsymbol{G}\hat{\boldsymbol{\beta}}_{\boldsymbol{g}}}{\hat{\boldsymbol{\beta}}'_{\boldsymbol{g}}\boldsymbol{G}'\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}'\boldsymbol{G}\hat{\boldsymbol{\beta}}_{\boldsymbol{g}}}\right)^{-1}, \text{ and } \hat{\rho}_{z_4z_4} = \left(1 + \frac{\hat{\boldsymbol{\beta}}'_{\boldsymbol{s}}\hat{\boldsymbol{B}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Psi}}_d\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}}{\hat{\boldsymbol{\beta}}'_{\boldsymbol{s}}\hat{\boldsymbol{B}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{\beta}}_{\boldsymbol{s}}}\right)^{-1}.$$

These estimates are needed for correcting the predictive validity for attenuation. Following Eq. (2.10), we obtain

$$\hat{\rho}_{\zeta_i y} = \frac{\hat{\rho}_{z_i y}}{\sqrt{\hat{\rho}_{z_iz_i}}}, \quad i = 1, 2, 3, 4, \tag{3.4}$$

where y is the response and ζ_i is the true value of the prediction scale z_i.

3.4 Details of implementation

The simulations are conducted 1000 times for each parameter configuration (five choices of θ, four choices of n, and two choices of σ^2) making a total of $5 \cdot 4 \cdot 2 \cdot 1000 = 40\ 000$ simulations. In each one, a new random sample is generated and the following operations executed: a) maximum likelihood factor analysis, b) symmetric transformation analysis, c) stepwise selection of predictors P_2^*, d) regression analyses with the predictor sets P_1, P_2^*, P_3, and P_4, and e) computations of reliability and validity.

The simulations have been implemented in SURVO MM [23,24] using its matrix interpreter, statistical operations and sucros ("Survo macros"). The stepwise algorithm giving the set P_2^* (option STEP4 in the ready-made sucro STEPREG) uses a combination of a forward selection and a backward elimination. The former includes variables in the model if $|t| > 2$ separately for each predictor against the response, and the latter deletes them in order of the t-values until $|t| > 2$ for all predictors.

To speed up the simulations, a SURVO MM program module and a sucro have been programmed by the first author. The program module takes care of the regression analyses and computations of reliability and validity, while the sucro implements the general flow of the simulations and saving of the results in SURVO MM data files. (For general information on programming in Survo environment, see [24, pp. 399–443] or [25].) This paper has been composed and written using SURVO MM, its PostScript graphics and `PRINT` operation with LaTeX frontend, thus completing and documenting [26] the various tasks of the research process within the same environment.

4 Results

Results show that the highest predictive validity in the regression model is obtained by using the factor scores as predictors and applying the correction for attenuation. The factor scores also appear to be the most stable choice for the predictor selection, while the other alternatives tend to give biased results. In the following, we provide more detailed arguments to support our findings.

4.1 Predicted validity

Table 3 gives the minimum, mean, and maximum values of the estimates related to the predicted validity of the prediction scales z_i, $i = 1, 2, 3, 4$, namely, the predicted validity $\hat{\rho}_{z_i y}$ and the square root of the reliability $\hat{\rho}_{z_i z_i}$. Those estimates are needed in Eq. (3.4) to obtain the predicted validity corrected for attenuation, denoted by $\hat{\rho}_{\zeta_i y}$. The figures in Table 3 are based on 20 000 simulations of each value of the sample variation parameter σ^2, hence aggregating all the results for the values of θ and n.

As expected, the predicted validity of all scales is on average higher with $\sigma^2 = 4$ when compared to $\sigma^2 = 9$. The reliabilities, which do not depend on σ^2, are clearly lower for the prediction scales z_1, z_2, and z_3, which implies that the correction for attenuation gives dubious results for those scales, especially with $\sigma^2 = 4$. Indeed, the rightmost column of Table 3 gives the percentage of the cases where $\hat{\rho}_{\zeta_i y} > 1$. With the scale z_4 this may happen by chance, but for the other scales it is systematic, and the results could be anything, as

Table 3
Overall statistics of the estimates related to predicted validity.

		$\hat{\rho}_{z_i y}$			$\sqrt{\hat{\rho}_{z_i z_i}}$			$\hat{\rho}_{\zeta_i y}$			
σ^2	Scale	min	mean	max	min	mean	max	min	mean	max	> 1
4	z_1	0.418	0.793	0.971	0.096	0.765	0.937	0.709	1.041	6.897	80.4%
	z_2	0.246	0.780	0.964	0.070	0.765	0.996	0.247	1.025	6.828	72.6%
	z_3	0.292	0.764	0.958	0.123	0.772	0.953	0.575	0.991	4.778	40.6%
	z_4	0.077	0.741	0.947	0.557	0.822	1.000	0.102	0.899	1.086	1.7%
9	z_1	0.376	0.742	0.937	0.105	0.762	0.947	0.627	0.980	5.364	21.5%
	z_2	0.000	0.727	0.930	0.073	0.759	0.997	0.208	0.964	7.755	14.8%
	z_3	0.285	0.711	0.923	0.181	0.772	0.955	0.492	0.923	3.104	5.8%
	z_4	0.093	0.690	0.912	0.529	0.822	1.000	0.116	0.838	1.022	< 0.1%

the maximum values of $\hat{\rho}_{\zeta_i y}$ display. The factor scores is the only predictor set that utilizes the information from the measurement model, and therefore the correction for attenuation works for the corresponding prediction scale z_4.

Table 4 presents the predictive validity of the prediction scales z_i, $i = 1, 2, 3, 4$. The figures are now based on 1000 simulations of each parameter configuration where σ^2 represents the sample variation, θ represents the magnitude of the artificial measurement error, and n is the sample size. To save space, the results for $\theta = 0.5$, $\theta = 1.5$, and $n = 500$ are omitted. The values of the scale z_4 have been corrected for attenuation using Eq. (3.4).

From Table 4 we can infer that the highest predictive validity is obtained by using the factor scores as predictors and applying the correction for attenuation. The only exceptions are the cases where $\theta = 2$ and $n = 100$, i.e., the worst models. With larger n, the factor scores give consistent results for all values of θ, whereas the results of the other scales systematically decrease when θ is increased.

4.2 Stability

We consider the stability in the predictor selection as a sort of steadiness: the selection should occur systematically, not by chance. The stability is partially reflected by the results of the predictive validity, but we also find it useful to examine the estimated regression coefficients of each predictor set.

Table 4
Predictive validity of the scales z_i, the values of z_4 corrected for attenuation.

		$\sigma^2 = 4$			$\sigma^2 = 9$		
θ	Scale	$n = 100$	$n = 300$	$n = 1000$	$n = 100$	$n = 300$	$n = 1000$
0	z_1	0.942	0.937	0.936	0.885	0.874	0.870
	z_2	0.925	0.931	0.935	0.860	0.867	0.870
	z_3	0.923	0.922	0.923	0.859	0.858	0.857
	z_4	0.942	0.947	0.948	0.878	0.880	0.881
1	z_1	0.822	0.802	0.797	0.775	0.749	0.742
	z_2	0.790	0.798	0.796	0.735	0.743	0.741
	z_3	0.780	0.776	0.776	0.727	0.722	0.722
	z_4	0.891	0.946	0.962	0.830	0.881	0.896
2	z_1	0.658	0.616	0.604	0.627	0.580	0.563
	z_2	0.580	0.601	0.602	0.544	0.561	0.561
	z_3	0.577	0.571	0.571	0.542	0.534	0.532
	z_4	0.586	0.762	0.917	0.556	0.714	0.854

Figure 2 consists of four sub-figures A, B, C, and D, corresponding to the predictor sets P_1, P_2, P_3, and P_4, respectively. The sub-figures demonstrate how the regression coefficient of the first predictor of each set varies with different parameter configurations. Here, $\sigma^2 = 9$, and for each value of θ from 0 up to 2 in increments of 0.5 we have $n = 100, 300, 500, 1000$ from left to right. Hence, each sub-figure includes 20 000 data points of the estimated regression coefficients drawn against the observation number in the simulation data which have been sorted hierarchically by θ and n. The gray dots indicate that the predictor is statistically significant (on 0.05 level) whereas the black dots indicate non-significance. In Figure 2 B the significant predictors are those selected by the stepwise procedure, and the non-significant ones which have been deleted, are drawn as zeros and jittered vertically for a better visibility.

The overall result in Figure 2 is that the factor scores are stable predictors: although $\hat{\beta}_{s_1}$ has a quite large variance, its average stays on the same level when θ and n are varied (see Figure 2 D). The only exception is the case where $\theta = 2$ and $n = 100$. The other scales introduce a bias in their results, as the values of $\hat{\beta}_{x_1}$, $\hat{\beta}_{h_1^*}$, and $\hat{\beta}_{g_1}$ decrease systematically along the values of θ.

When $\theta = 0$, the variance of $\hat{\beta}_{s_1}$ (see Figure 2 D) is about the same as the variance of $\hat{\beta}_{x_1}$ (see Figure 2 A). When θ is increased, the variance of $\hat{\beta}_{s_1}$ also increases, unlike the variance of $\hat{\beta}_{x_1}$, which tends to decrease. Furthermore,

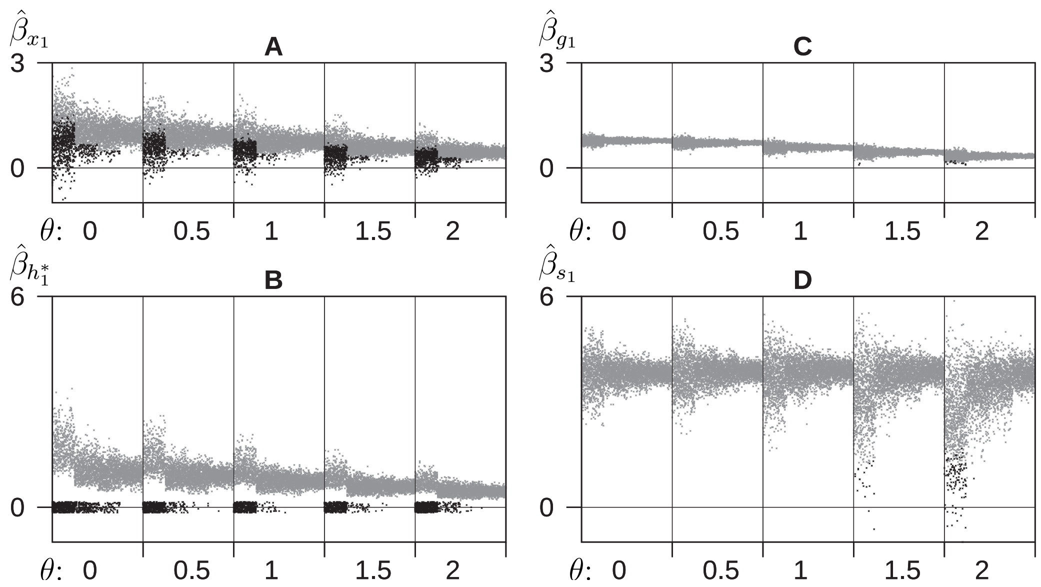

Fig. 2. Variation of the regression coefficient of the first predictor of each set.

$\hat{\beta}_{s_1}$ is always significant when $\theta \leq 1$, whereas this applies to $\hat{\beta}_{x_1}$ only when $n = 1000$. The stepwise selection (see Figure 2 B) appears to behave quite similarly compared to Figure 2 A, except that in the non-significant cases, the variable is deleted altogether. Not surprisingly, $\hat{\beta}_{g_1}$ has the smallest variance of all the coefficients (see Figure 2 C), because the weights used in creating the predictors of this set are always fixed at 0 or 1.

Table 5 presents statistics supporting Figure 2, namely the means and standard errors of the same regression coefficients as well as the percentage of the cases where they are significant. To save space, the results for $\theta = 0.5$, $\theta = 1.5$, and $n = 500$ are again omitted.

The bias mentioned above is easily detected from Table 5 as well. It is also evident that the factor scores is the only scale that works logically when θ is increased. All the other scales lead more or less to capitalizing on chance, because they can not separate the artificial measurement error variance from the true variance. This is clearly reflected in their regression coefficients.

5 Conclusions

Our results based on simulation studies suggest that if the linear regression model is applied within the measurement framework approach, then the factor scores should be the predictors of choice. Several findings justify this. Firstly, the factor scores take advantage of the information from the measurement model, which separates the true variance from the measurement error vari-

Table 5
Statistics of the regression coefficient of the first predictor of each set.

		$\hat{\beta}_{x_1}$			$\hat{\beta}_{g_1}$		
θ	Statistic	$n = 100$	$n = 300$	$n = 1000$	$n = 100$	$n = 300$	$n = 1000$
0	mean	1.573	1.086	0.990	0.782	0.784	0.783
	stderr	0.325	0.269	0.176	0.076	0.041	0.023
	sig	40.6%	87.8%	100%	100%	100%	100%
1	mean	1.027	0.764	0.748	0.589	0.589	0.589
	stderr	0.238	0.184	0.108	0.085	0.050	0.027
	sig	50.4%	96.2%	100%	100%	100%	100%
2	mean	0.673	0.478	0.450	0.347	0.344	0.343
	stderr	0.151	0.115	0.075	0.077	0.046	0.025
	sig	43.0%	91.3%	100%	98.8%	100%	100%
		$\hat{\beta}_{h_1^*}$			$\hat{\beta}_{s_1}$		
θ	Statistic	$n = 100$	$n = 300$	$n = 1000$	$n = 100$	$n = 300$	$n = 1000$
0	mean	1.707	1.106	0.986	3.804	3.823	3.825
	stderr	0.453	0.309	0.185	0.488	0.279	0.153
	sig	52.9%	81.9%	100%	100%	100%	100%
1	mean	1.081	0.774	0.742	3.680	3.832	3.893
	stderr	0.284	0.187	0.109	0.602	0.338	0.179
	sig	61.2%	96.8%	100%	100%	100%	100%
2	mean	0.697	0.490	0.455	2.808	3.376	3.810
	stderr	0.164	0.118	0.075	0.822	0.589	0.270
	sig	49.6%	92.5%	100%	89.9%	99.9%	100%

sig = % of cases (in 1000 replicates) where the coefficient was statistically significant

ance. Hence the factor scores is the scale of the highest reliability. Secondly, the prediction scale of the factor scores gives the highest predictive validity corrected for attenuation. The attenuation correction does not work for the other scales, because they can not separate the different sources of variation. Lastly, using the factor scores leads to stable regression coefficients, that is, on average the estimated coefficients stay on the same level, independently of the magnitude of the measurement error variance and the sample size. In addition, the coefficients of the factor score predictors are nearly always significant.

Reliability, validity and stability are quite important properties for predictors. If they are unacceptable, the coefficients and their interpretations will be easily affected by fluctuations of random measurement errors. Indeed, all the predictor sets in this study except the factor scores seem to lead to capitalizing on chance. Whether we use all the variables, a stepwise selection, or factor sums (i.e., variables weighted with 0 or 1 according to a factor structure), the results of the regression model become more or less unstable. To put it briefly, it is desirable to use the factor scores as predictors of the linear regression model, as in general, it leads to more reliable, more valid, and more stable results.

Acknowledgements

We would like to express our gratitude to professor *Seppo Mustonen* for his long-standing encouragement and support.

References

[1] W. A. Fuller, Measurement Error Models, Wiley, New York, 1987.

[2] H. M. Kim, A. K. M. E. Saleh, Improved estimation of regression parameters in measurement error models, Journal of Multivariate Analysis 95 (2005) 273–300.

[3] L. J. Gleser, The importance of assessing measurement reliability in multiple regression, Journal of the American Statistical Association 76 (419) (1992) 696–707.

[4] J. T. Scott, Factor analysis and regression, Econometrica 34 (1966) 552–562.

[5] D. N. Lawley, A. E. Maxwell, Regression and factor analysis, Biometrika 60 (1973) 331–338.

[6] Y. Isogawa, M. Okamoto, Linear prediction in the factor analysis model, Biometrika 67 (1980) 482–484.

[7] K. G. Jöreskog, A general method for analysis of covariance structures, Biometrika 57 (1970) 239–251.

[8] K. A. Bollen, Structural Equations with Latent Variables, Wiley, New York, 1989.

[9] L. Tarkkonen, K. Vehkalahti, Measurement errors in multivariate measurement scales, Journal of Multivariate Analysis 96 (2005) 172–189.

[10] K. Vehkalahti, Reliability of Measurement Scales, no. 17 in Statistical Research Reports, Finnish Statistical Society, Helsinki, Finland, 2000.

[11] L. Tarkkonen, On Reliability of Composite Scales, no. 7 in Statistical Studies, Finnish Statistical Society, Helsinki, Finland, 1987.

[12] K. Vehkalahti, S. Puntanen, L. Tarkkonen, Estimation of reliability: a better alternative for Cronbach's alpha, Reports on Mathematics 430, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland, 20 pp., `http://mathstat.helsinki.fi/reports/Preprint430.pdf` (2006).

[13] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, 3rd Edition, Wiley, New Jersey, 2003.

[14] S. Tezuka, P. L'Ecuyer, Efficient and portable combined Tausworthe random number generators, ACM Transactions on Modelling and Computer Simulation 1 (1991) 99–112.

[15] G. A. F. Seber, Multivariate Observations, Wiley, New York, 1984.

[16] J. R. Hurley, R. B. Cattell, Procrustes program: producing direct rotation to test a hypothesised factor structure, Behavioral Science 7 (1962) 258–262.

[17] Y. Ahmavaara, Transformation analysis of factorial data, Annales Academiæ Scientiarum Fennicæ, Series B 88 (1954).

[18] P. H. Schönemann, A generalized solution of the orthogonal Procrustes problem, Psychometrika 31 (1966) 1–10.

[19] N. Cliff, Orthogonal rotation to congruence, Psychometrika 31 (1966) 33–42.

[20] S. Mustonen, Symmetrinen transformaatioanalyysi [Symmetric transformation analysis, in Finnish], Report 24, Social Research Institute of Alcohol Studies, Helsinki, Finland (1966).

[21] J. Isotalo, S. Puntanen, G. P. H. Styan, Matrix tricks for linear statistical models: our personal Top Sixteen, Research report A 363, Dept. of Mathematics, Statistics & Philosophy, University of Tampere, Tampere, Finland (2005).

[22] S. Mustonen, Tilastolliset monimuuttujamenetelmät [Statistical Multivariate Methods, in Finnish], Survo Systems, Helsinki, Finland, 1995.

[23] S. Mustonen, SURVO MM: Computing environment for creative processing of text and numerical data, `http://www.survo.fi/mm/english.html` (2001).

[24] S. Mustonen, Survo, An Integrated Environment for Statistical Computing and Related Areas, Survo Systems, Helsinki, Finland, 1992.

[25] S. Mustonen, Programming SURVO 84 in C, SURVO 84C Contributions 3, Department of Statistics, University of Helsinki, Helsinki, Finland (1989).

[26] K. Vehkalahti, Leaving useful traces when working with matrices, Research Letters in the Information and Mathematical Sciences 8 (2005) 143–154, `http://iims.massey.ac.nz/research/letters/volume8/`.