Correcting Measurement Error Bias in Interaction Models with Small Samples

Josep Bisbe[1], Germà Coenders[2], Willem E. Saris[1], and
Joan Manuel Batista-Foguet[1]

Abstract

Several methods have been suggested to estimate non-linear models with
interaction terms in the presence of measurement error. Structural equation
models eliminate measurement error bias, but require large samples.
Ordinary least squares regression on summated scales, regression on factor
scores and partial least squares are appropriate for small samples but do not
correct measurement error bias. Two stage least squares regression does
correct measurement error bias but the results strongly depend on the
instrumental variable choice. This article discusses the old disattenuated
regression method as an alternative for correcting measurement error in
small samples. The method is extended to the case of interaction terms and
is illustrated on a model that examines the interaction effect of innovation
and style of use of budgets on business performance. Alternative reliability
estimates that can be used to disattenuate the estimates are discussed. A
comparison is made with the alternative methods. Methods that do not
correct for measurement error bias perform very similarly and considerably
worse than disattenuated regression.

1   Introduction

When an interaction effect between two continuous variables is present, the
effect of one of the variables on the dependent one is different for different values
of the other variable. If both variables are measured without error, a particular
form of ordinary least squares (OLS) regression (i.e. moderated regression
analysis, a regression analysis including the product of the variables that interact
as an additional variable) can be used to estimate and test interaction effects (Irwin
and McClelland 2001; Jaccard et al. 1990).

---

[1] ESADE, Universitat Ramon Llull, Spain; josep.bisbe@esade.edu
[2] Corresponding author. Department of Economics, University of Girona. Postal address:
Faculty of Economics and Business. Campus Montilivi. 17071 Girona, Spain;
germa.coenders@udg.es

Interaction effects are relevant to a number of research problems in disciplines that use questionnaire data that are subject to measurement error. If variables are measured with error, OLS leads to biased estimates. This bias is usually negative and is hence referred to as measurement error attenuation bias. In this case it is common practice to use summated rating scales (SRS, Spector 1992) by averaging multiple indicators of each variable of interest, and then using these SRS as variables in the OLS regression model. However, attenuation bias is not eliminated completely because the SRS are not perfectly reliable, even if they are more reliable than single items. Thus, OLS on SRS is only useful for predictive purposes, not for theory building or testing.

This distinction between prediction and testing will be very relevant throughout this article. For estimating the parameters of a model according to a theory, absence of bias is a key requirement. In this case, measurement error correction is necessary because the parameters that are theoretically of interest are those that relate error-free variables. However, when the aim of the researcher is merely to predict the observed dependent variable from the available (i.e. observed with error) explanatory variables, then the relevant relation is that among variables containing errors. In this case it turns out that the biased parameter estimates yield optimal predictions in the sense of being unbiased estimates of the conditional expectation of the dependent variable given the observed (i.e. with error) values of the explanatory variables.

Many methods for modelling variables measured with error have been suggested in the literature. Among the most widely used are structural equation models (SEM), two-stage least squares regression (TSLS), partial least squares (PLS) and ordinary least squares regression on factor scores (FSR). However, these methods either rely on large samples, depend on an arbitrary choice of instrumental variables, or do not properly correct measurement error bias.

SEM can in principle be used to eliminate attenuation bias (Bollen, 1989; Goldberger and Duncan, 1973; Raykov and Marcoulides, 2000). In recent years, different developments have been proposed to examine interaction effects using this approach following a seminal paper by Kenny and Judd (1984). Kenny and Judd's approach requires each latent variable to relate to at least two indicators and implies the formation of one or more interaction indicators based on the products of the main effect indicators (Batista-Foguet et al., 2004a, 2004b; González et al., in press; Jackard and Wan, 1995; Jöreskog, 2000; Jöreskog and Yang, 1996; Marsh et al., 2004; Moulder and Algina, 2002; Ping, 1995; Saris et al., in press). For instance, if item1 and item2 are indicators of the first variable and item3 and item4 of the second, the possible interaction indicators would be item1×item3, item1×item4, item2×item3 and item2×item4. Except those of Batista-Foguet et al. (2004a) and Marsh et al. (2004), SEM approaches have a great degree of complexity, requiring non-linear constraints. More importantly, with sample sizes below 100 or 200, SEM are hardly applicable because the properties of estimators and tests are

asymptotic (Cortina et al., 2001; Schumacker and Marcoulides, 1998; Boomsma and Hoogland, 2001).

The requirement of large samples is particularly relevant when cases are not individuals but organisations. For instance, in management research, sampling units may be firms and respondents may be top managers who are limited in number. Consequently, and despite the cautions regarding minimum sample sizes (Medsker et al., 1994), often business studies (Abernethy and Vagnoni, 2004; Bisbe and Otley, 2004; Chalos and Poon, 2000; Chenhall, 2004; Chong and Chong, 2002) report findings based on small samples (i.e. less than 100 data points).

TSLS regression (Koopmans and Hood, 1953; see Johnston, 1972 for an introduction; Bollen, 1996 for its application to measurement error correction and Bollen and Paxton, 1998 for its use in interaction models) also eliminates the measurement error bias. In this method, one indicator is selected as regressor and the others as instrumental variables. Instrumental variables may be any variables that are highly correlated with the regressor measured with error but uncorrelated with the disturbance term. The other indicators of the explanatory variable can be used as instrumental variables if their measurement error is uncorrelated with the measurement error of the regressor and with the disturbance of the dependent variable. In the case of product indicators, this will be fulfilled only if instrumental product indicators do not include products of the same items that are multiplied in the indicator that is in the regression equation. In the above example, if item1, item3 and item1×item3 are used as indicators, the valid instruments are item2, item4 and item2×item4. A drawback of the TSLS method and of any limited information method is that the estimates will usually change depending on which indicators are chosen as regressors and which as instruments. If all model assumptions hold and the sample size is large, differences should be small, but not otherwise (Bollen and Paxton, 1998; Jöreskog and Sörbom, 1989). In particular, the assumption of uncorrelatedness between the measurement errors of the instruments and the disturbance is crucial in this respect (Saris et al., in press).

PLS (Chin, 1998; Chin and Newsted, 1999; Fornell and Cha, 1994; Wold, 1975) does not rely on large samples but fails to correct measurement error bias properly. PLS constitutes a rather complex procedure that is in fact not so far from OLS on SRS, from which it differs by the fact that the weights of the indicators are not equal but computed from the optimization of certain criteria. On the positive side, PLS shares with OLS regression the property of providing optimal predictions and can successfully be applied for predictive purposes or whenever the aim of the analysis is exploratory, the theory is weak, or the number of variables is too large for formal modelling (Jöreskog and Wold, 1982). Wold (1982) introduced the term "soft modelling" to refer to these situations. However, on the negative side, and in the same way as OLS on SRS, PLS has the limitation that it does not eliminate measurement error bias, as it is consistent only under perfect reliability or with an infinite number of items per dimension (Dijkstra, 1983; Fornell and Cha, 1994; Hulland, 1999; O'Loughlin and Coenders, 2004; Wold, 1982).

Therefore, and despite its potential usefulness for prediction with small samples, PLS is not adequate when theory is well developed and the researcher's purpose is the estimation of population parameters related to the theory, for which the lack of bias is a key requirement.

In FSR (e.g. Hair et al., 1998), separate factor analysis models (e.g. Kim and Mueller, 1978a, 1978b) are fitted for each dimension, factor scores are saved and used as variables in an OLS regression. It is similar to PLS in that it computes composite scores from a weighted sum of the indicators and in that it does not correct for measurement error bias. The weights of the items used to compute the factor scores do differ from PLS, but not to a large extent. Anyway, changing the weights of the indicators is reported to have a minor impact (McDonald, 1996). The simpler uniform weights that are used in OLS on SRS have the advantage over both FSR and PLS that they are fixed instead of being sample dependent, and thus are more suitable for comparative research.

In this article we concentrate on a very simple method for correcting the effects of measurement error in the estimation of grounded theoretical models including interaction terms with small sample sizes. The long known disattenuated regression (DR) on SRS (Lord and Novick, 1968) is discussed in this article as an alternative to OLS, SEM, TSLS, PLS and FSR that is unbiased while being very little demanding, both in terms of statistical expertise and of sample size. In short, this method estimates the reliability of the SRS, uses this information to compute the variances of the SRS that would have been obtained in the absence of measurement error, and substitutes these variances in the covariance matrix from which OLS estimates are computed. An extension of DR is developed for the case of interaction effects. Obtaining a good reliability estimate is of course a crucial issue. Two reliability estimates are discussed: α (Cronbach, 1951) and Ω (Heise and Bohrnstedt, 1970). α can be obtained from the covariances among items while Ω requires estimating a factor analysis model, but Ω is based on less stringent assumptions. The estimation of the reliability of interaction terms cannot be done by just using the α and Ω formulae, but a suitable method is developed in this article.

The DR procedure will be illustrated on an example of Bisbe and Otley (2004), who estimated and tested the moderating effects of an interactive use of budgets on the impact of innovation on performance using OLS on SRS. The results obtained under OLS and DR regression will be compared to those obtained with TSLS, FSR and PLS. The results of applying SEM on the same data, although not very reliable due to the small sample size, can be found in Batista-Foguet et al. (2004b).

This article is structured as follows. First we present the DR method. Then we summarize the model of Bisbe and Otley (2004) and the data collection mode and measurement instruments used. Then, the DR method is extended to the interaction effect case. Next we provide the results of the different analyses and sizeable differences emerge between DR and the methods that do not correct attenuation bias (PLS, OLS and FSR), and sizeable differences appear within the TSLS method depending on the instrumental variable choice. Based on these results, some recommendations are given to applied researchers facing small samples.

2   Disattenuated regression on summated rating scales

SRS are often used when an unobservable concept, assumed to be unidimensional, is measured by multiple indicators. A SRS is computed as either the sum or the average of these indicators (in this article we assume, without loss of generality, that they are averaged; for summated items Equations 2.1 and 2.2 will change somewhat). This has a threefold purpose: 1) properly defining a composite construct by combining observable variables; 2) increasing measurement reliability by averaging out random errors of measurement from single indicators and 3) increasing parsimony as only equations relating the composites (of which there are fewer than variables) are needed.

Unfortunately, if OLS regression is estimated on the SRS, measurement error correction is not complete. It has long been known that a sum or an average of several measures is more reliable than just one measure (Simpson, 1755). However, this average is only perfectly reliable when the number of items approaches infinity or the reliability of all items approaches one. As a result, the OLS estimates of regression coefficients will still suffer from attenuation bias.

DR is a simple method for solving measurement error bias in the estimation of well-grounded theoretical models when sample sizes are small. DR is a method whose foundations were laid long ago (Spearman, 1904) and which, for large samples, became superseded by more modern alternatives such as 2SLS and SEM and thus stopped being used.

The first step in a DR is to estimate the reliability of the SRS. Reliability is defined as 1 minus the percentage of variance of the SRS that corresponds to random measurement error. So, the product of the total variance of the SRS and reliability yields the so-called true variance. A disattenuated regression proceeds as an OLS regression in which true variances are substituted for total variances. Any OLS regression software that accepts covariance matrices as means of data input as well as any SEM software can thus perform a disattenuated regression.

Reliability of an SRS is usually computed as Cronbach's α (Cronbach, 1951) on the assumption that items are at least tau-equivalent (e.g. Bollen, 1989: 215-216). This assumption implies that all items are an unweighed sum of the true score plus a random error term. These random error terms are assumed not to contain any systematic component (the items thus measure the true score and only one true score), and to be mutually uncorrelated. An observed consequence of tau-equivalence is that all covariances among all pairs of items are equal (the opposite does not hold, i.e. covariances may be equal and yet items may not be tau-equivalent). Cronbach's α is a very popular measure and its computation is performed by most commercial software packages as:

$$\alpha = \frac{p}{p-1}\left(1 - \frac{\sum_{j=1}^{p} \mathrm{var}(item_j)}{p^2 \times \mathrm{var}(SRS)}\right) \qquad (2.1)$$

where p is the number of items, $\mathrm{var}(item_j)$ the variance of the jth item and $\mathrm{var}(SRS)$ the variance of the SRS constructed as the average of the items.

If the tau-equivalence assumption is fulfilled, then the disattenuated regression estimates obtained in this way are consistent. Otherwise, α is biased (Cortina, 1993; Novick and Lewis, 1967; Raykov, 1997), and, unlike what is usually understood, this bias can be both positive and negative in sign (Raykov, 2001). Unfortunately, empirical studies do not usually perform any test of the tau-equivalence assumption when applying α.

A myriad of alternative estimates of reliability that are based on more relaxed assumptions is available. Among them, one of the simplest is Heise and Bohrnstedt's Ω (Heise and Bohrnstedt, 1970). In order to estimate Ω, a unidimensional factor analysis model must be fitted to the items in each dimension. The estimates of the model will include the so-called communalities or percentages of true score variance in each item. Reliability is estimated as:

$$\Omega = 1 - \frac{\sum_{j=1}^{p}\left[\mathrm{var}(item_j) \times (1 - h_j)\right]}{p^2 \times \mathrm{var}(SRS)} \qquad (2.2)$$

where h_j is the communality of the jth item. A factor analysis model requires p (number of items) to be equal or larger than three.

This measure assumes that items are congeneric. This assumption implies that all items are a *weighted* sum of the true score plus a random error term, which makes it possible for the contribution of the true score to the different items to be different. As before, these random error terms are assumed not to contain any systematic component and to be mutually uncorrelated. An observed consequence of congeneric measurement when the number of items is equal to or larger than four is that the unidimensional factor analysis model fits the inter-item correlations well (the opposite does not hold, i.e. the one-factor model may perfectly fit the correlations and yet items may not be congeneric; See Coenders et al., 2006 and Saris et al., in press for large sample tests of congeneric measurement). If the model is estimated by maximum likelihood, most commercial software packages will produce a χ^2 test of the fit of the model to the correlations. Otherwise, the residual correlations may be examined one by one to check that they are all small.

In this article we use the two simple approaches described to estimate reliability of an SRS in order to perform a DR, and we compare their results to OLS, TSLS, PLS and FSR using Bisbe and Otley's (2004) data. First we present Bisbe and Otley's conceptual framework and specify a regression model with interaction effects and errors in the variables and show how α and Ω have to be used to estimate a DR in the presence of interaction effects.

3   Conceptual framework

The management literature has long considered innovation to be one of the major determinants of long-term organizational performance in contemporary environments (e.g. Clark and Fujimoto, 1991; Kanter, 2001). Most empirical studies (e.g. Capon et al. 1992; de Moerloose, 2000) have shown a positive relationship between innovation and performance. Another determinant of business performance is the use of management control systems (MCS). Simons (1991, 1995) states that, when used interactively, MCS focus on strategic uncertainties and become a recurring forum and agenda for a continuous and challenging debate in which top managers are involved.

Following Simons' framework, it can be expected that, by orientating the contents and the adequacy of the innovation initiatives, an interactive use of MCS will positively influence the success of innovation initiatives and, consequently, will enhance their impact on performance.   Thus, the relationship between innovation and performance will be affected by the extent to which MCS are used interactively, suggesting a moderated causal relationship in which an *interaction effect* is present (See Bisbe and Otley, 2004, for details).
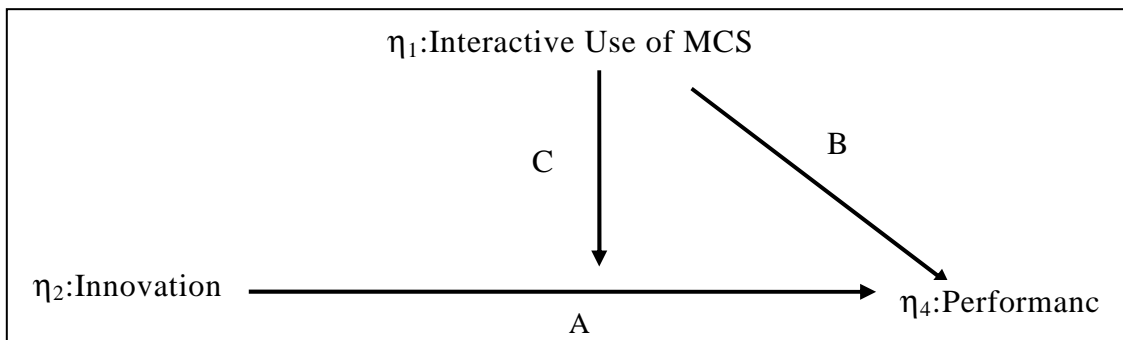
Figure 1: Conceptual framework.

Drawing on Simons' framework (Simons 1990, 1991, 1995), Bisbe and Otley (2004) estimate and test the model in Figure 1 that includes the aforementioned interaction effect (arrow C). Besides, Figure 1 includes the *main effects* of both

innovation and interactive use of MCS on performance (arrows A and B). Following Irwin and McClelland (2001), all main effects of the variables that interact must be introduced in the interaction model (even if they are expected to be theoretically irrelevant) in order to obtain meaningful estimates of the interaction effect.

4   Data collection and instruments

Data were gathered through the administration of a written questionnaire to a sample of Chief Executive Officers of medium-sized, mature manufacturing firms with headquarters located in Catalonia, Spain (Bisbe and Otley, 2004). Mature medium-sized firms were defined as those with an annual turnover of between 18 and 180 million euro, with between 200 and 2000 employees and founded at least ten years before the survey was administered. Exploitation of the Dun and Bradstreet/CIDEM 2000 database (referring to 1998) resulted in 120 firms fulfilling the screening criteria.

   Once revised after pilot tests, questionnaires were distributed and traced following Dillmann's (2000) guidelines. Out of the 120 questionnaires distributed, 58 were returned, all of which were complete. This compares well with the response rate of similar studies. Cases where the executives reported not to have been in their current position for at least three years (n=18) were excluded. The resulting useable sample size was n=40. The measurement instruments used for each of the variables in Figure 1 are described next.

   The *interactive use of MCS* (η_1) was defined based on Simons´ (1990, 1995, 2000) work in terms of the patterns of attention posed by top managers and considering budgets as type of MCS. Using Simons´ framework, and developing the instruments suggested by Abernethy and Brownell (1999) and Davila (2000), Bisbe and Otley (2004) developed a 3-item instrument in a 1 to 7 scale containing:

   item 1)   Degree to which information from the control system is discussed face-to-face merely on an exception basis.

   item 2)   Extent to which it demands frequent and regular attention from the top manager.

   item 3)   Extent to which it demands frequent and regular attention from operating managers at all levels of the organization.

Product innovation (η_2) was understood from an output perspective and was defined as the development and launching of products that are in some respect unique and distinctive from existing products. It referred to innovative behaviour during the last three years in relative terms, in comparison with the industry average. The measure of product innovation used in Bisbe and Otley (2004) was

drawn from instruments used by Capon et al. (1992), Thomson and Abernethy (1998) and Scott and Tiesen (1999) and contained 3 items on a 1 to 7 scale:

    item 4)   Rate of introduction of new products.

    item 5)   Tendency of firms to pioneer.

    item 6)   Part of the product portfolio corresponding to recently launched products.

Performance (η_4) was defined as the degree of goal attainment along several dimensions, including both financial and non-financial aspects, during the three years previous to the administration of the questionnaire. An instrument for the evaluation of strategic business unit effectiveness (Chenhall and Langfield-Smith, 1998; Chong and Chong, 1997; Govindarajan 1984, 1988; Govindarajan and Gupta, 1985; Gupta and Govindarajan, 1984) was adapted by Bisbe and Otley (2004), who used eight questions related to both financial (sales growth rate, revenue growth rate, return on investment, profit/sales ratio) and customer perspectives (customer satisfaction, customer retention, customer acquisition and increase in market share). A single item (item7) was constructed from the assessments of the firm's performance on those eight aspects weighted according to their perceived importance by respondents themselves.

5   The regression model with interaction effects and measurement errors

The model is composed of two parts, one relating the variables, called structural part, and one containing measurement error, called measurement part. The relationships of the structural part in Figure 1 are formalized in Equation 5.1. For simplicity we do not use a different notation for endogenous and exogenous variables:

$$\eta_4 = \beta_{41}\eta_1 + \beta_{42}\eta_2 + \beta_{43}\eta_3 + \zeta_4 \qquad (5.1)$$

where:

- η_1 is the interactive use of budgets corrected for measurement error, centred with zero mean.
- η_2 is innovation corrected for measurement error, centred with zero mean.
- $\eta_3 = \eta_1\eta_2$ is the interaction term corrected for measurement error, referred to the mean-centred η_1 and η_2 centred again after computing the product. As in moderated regression analysis (Irwin and Mclelland, 2001; Jaccard et al., 1990) the interaction is constructed as the product of both variables that interact.

- η_4 is performance corrected for measurement error, centred with zero mean.
- ζ_4 is a disturbance term, with zero expectation and uncorrelated with η_1, and η_2.

The β_{43} interaction term implies that the effect of the interactive use of budgets on performance depends on the innovation level or that the effect of innovation on performance depends on the level of interactive budget use. For instance, the effect of η_1 of η_4 conditional on a given value of η_2 is $\beta_{41}+\beta_{43}\eta_2$. Thus, the interpretation of the main effects β_{41} and β_{42} is that occurring when the value of the other variable is zero (if η_1 and η_2 are mean centred, for the mean value of the other variable). Standardization of the η variables cannot be performed as it would prevent η_3 from being equal to the product of η_1 and η_2. This implies that standardized parameter estimates have no interpretation (Jaccard et al., 1990).

As regards the measurement part, for each variable the same measurement equation is assumed:

$$y_j = \eta_j + e_j \quad \text{where } j=1,...,4 \tag{5.2}$$

where:
- y_1 and y_2 are not items but centred SRS of the interactive use of budgets and innovation.
- $y_3=y_1y_2$ is the centred product of centred SRS used for the interaction term.
- y_4=item7, the single weighted item measuring performance.
- e_j are measurement errors with zero expectation.
- e_1, e_2, e_4, η_1, η_2 and ζ_4 are assumed to be mutually independent, not just uncorrelated.

The decomposition of y_3 as a function of the error-free component η_3 and the error term e_3 is more complicated as it involves the product of two variables. Following a similar procedure as Jöreskog and Yang (1996), if the above assumptions of mean centring and independence hold, the interaction term can be decomposed as:

$$y_3=y_1y_2 =(\eta_1+e_1) \times (\eta_2+e_2) = \eta_1\eta_2+(\eta_1e_2+\eta_2e_1+e_1e_2)=\eta_3 + e_3 \tag{5.3}$$

with $\text{cov}(e_3,e_1)=\text{cov}(e_3,\eta_1)=\text{cov}(e_3,\eta_2)=\text{cov}(e_3,e_2)=0.$

6 Extension of DR to interaction effects. Steps of the analysis

6.1 Computation of reliabilities and true variances of main effects

Table 1 first shows the necessary information to calculate α and Ω for y_1 and y_2: the variances and covariances of the two sets of items, the variances of the SRS and the maximum likelihood communality estimates in two separate factor analysis models, as provided by a standard package such as SPSS 10.1. For instance, for Ω and the interactive use of budgets we have:

$$\Omega_1 = 1 - \frac{3.259 \times (1-0.755) + 2.743 \times (1-0.778) + 2.615 \times (1-0.366)}{9 \times 2.125} = 0.840 \qquad (6.1)$$

The items measuring interactive use of budgets have markedly different covariances (the ratio of the smallest over the largest covariance is 1.6) and thus the application of Cronbach's α is dubious in this particular case. This results in α and Ω being somewhat different for this dimension. All α's and Ω's are substantially different from 1, much too different for OLS to be appropriate.

Table 1: Computation of α and Ω.

Statistics for items in y_1 (budget use)				Statistics for items in y_2 (Innovation)			
Variances and covariances				Variances and covariances			
	item1	item2	item3		item4	item5	item6
item1	3.259			item4	2.984		
item2	2.291	2.743		item5	1.994	3.208	
item3	1.535	1.429	2.615	item6	1.577	1.682	1.887
var(y_1)= var(SRS$_1$)	2.125			var(y_2)= var(SRS$_2$)	2.065		
α_1	0.824			α_2	0.848		
	item1	item2	item3		item4	item5	item6
communalities (h_j)	0.755	0.778	0.366	communalities (h_j)	0.626	0.663	0.705
Ω_1	0.840			Ω_2	0.852		

To estimate the reliability of the single item y_4=item7, a direct question about overall performance was also included in the questionnaire. This question might play the role of an external criterion. If both measurements (this single question and item7) are valid for performance (measure performance and only performance), their correlation is the geometric mean of the reliabilities of both.

The high value of this correlation at 0.73 makes validity a reasonable assumption. We take this figure as the reliability of the dependent variable y_4.

Once the reliability of y_j is obtained, its product by the total variance of y_j is the true score variance of y_j or the variance of y_j corrected for measurement error, that is, the estimate of the variance of η_j. For instance, for the interactive use of budgets and using Ω we would have:

$$\text{var}(\eta_1)=\Omega_1\times\text{var}(y_1)=0.840\times2.125=1.784 \tag{6.2}$$

Table 2 shows the covariance matrix of y_1 to y_4, the variances corrected for measurement error (covariances do not change), and the error variances.

Table 2: Raw and corrected variances and covariances among dimensions.

	raw variances and covariances among dimensions			
	y_1	Y_2	y_3	y_4
y_1	2.125			
y_2	0.464	2.065		
y_3	-0.286	-0.242	4.283	
y_4	0.078	-0.604	-0.593	1.003
	true variances (corrected for measurement error)			
	$\text{var}(\eta_1)$	$\text{var}(\eta_2)$	$\text{var}(\eta_3)$	$\text{var}(\eta_4)$
α	1.751	1.751	2.961	0.732
Ω	1.784	1.759	3.035	0.732
	error variances			
	$\text{var}(e_1)$	$\text{var}(e_2)$	$\text{var}(e_3)$	$\text{var}(e_4)$
α	0.374	0.314	1.322	0.271
Ω	0.340	0.306	1.248	0.271

6.2   Computation of the true variance of the interaction effect

The presence of interaction terms makes measurement error correction somewhat more complicated. One might think that the computation of the reliability of y_3 could be performed by selecting the 9 products of items item1×item4, item1×item5,..., item3×item6 as tau-equivalent or congeneric indicators of y_3 and then estimating α or Ω in the usual way. However, these pairs of items overlap and thus can lead to correlated measurement errors. As an alternative, 3 non-overlapping pairs could be used (e.g. item1×item4, item2×item5 and item3×item6) but then y_3 would fail to be equal to y_1y_2, η_3 would fail to be equal to $\eta_1\eta_2$ and thus could not be interpreted as a proper interaction term any more.

By using Equation 5.3 and some properties of variances of products of independent variables we can obtain an appropriate true variance for $\eta_3=\eta_1\eta_2$:

$$\text{Var}(\eta_3)=\text{var}(y_3)-\text{var}(e_3)=$$
$$=\text{var}(y_3)-\text{var}(\eta_1)\text{var}(e_2)-\text{var}(\eta_2)\text{var}(e_1)-\text{var}(e_1)\text{var}(e_2) \qquad (6.3)$$

where $\text{var}(e_j)$ can be obtained as $\text{var}(y_j)-\text{var}(\eta_j)$. For instance, using the Ω approach, $\text{var}(\eta_3)$ in Table 2 is computed as:

$$\text{var}(\eta_3)=4.283-\text{var}(e_3)= 4.283-1.784\times0.306-1.759\times0.340-0.340\times0.306=3.035 \quad (6.4)$$

```
DR USING OMEGA
OBSERVED VARIABLES
Y1 Y2 Y3 Y4
COVARIANCES
 1.784
 0.464  1.759
-0.286 -0.242  3.035
 0.078 -0.604 -0.593 0.732
SAMPLE SIZE: 40
RELATIONSHIPS
Y4 = Y1 Y2 Y3
END OF PROBLEM
```

Figure 2: Sample SIMPLIS input file.

7 Results

The estimates and t-values for Equation 5.1 are displayed in Table 3 for OLS, DR using α, DR using Ω, PLS (using PLS-PC 1.8, with 30 jackknife resamples, the path weighting scheme, the outward measurement model, and item1×item4, item2×item5 and item3×item6 as interaction indicators), FSR (using SPSS 10.1, with maximum likelihood estimation and regression factor score computation) and TSLS (using SPSS 10.1). Even if raw estimates are the only ones that can be interpreted for models with interaction terms, PLS and FSR only compute standardized estimates and thus the latter have also been computed for OLS and DR for comparative purposes only. Since more than one TSLS estimate is possible depending on the instrumental variable choice, the maximum and minimum estimates across all possible choices are reported.

The R^2 obtained with methods that do not correct for attenuation (OLS, PLS and FSR) are very similar and much lower than those under DR. Both disattenuated regression variants provide similar R^2. The TSLS R^2 is strictly non comparable because this method corrects for measurement error in the explanatory variables only. What is most striking about this method is the huge difference in

the R^2 that can be obtained depending on the instrument choice. The results obtained by Batista-Foguet et al. (2004b) using SEM on the same data closely resemble those of DR, in spite of the small sample size.

The regression coefficients follow the same pattern. They are erratic for TSLS; lower and roughly similar for OLS, FSR and PLS; higher and roughly similar for both methods correcting for attenuation, though slightly higher for DR using α. Though in this case α and Ω have yielded similar estimates, this does not need to be the case under all circumstances, especially when there are strong departures from the tau-equivalence assumption.

Table 3: Estimates under 6 approaches.

parameter	OLS on SRS			DR on SRS (α)			DR on SRS (Ω)		
	esti-mate	t-value	stand. estim.	esti-mate	t-value	stand. estim.	esti-mate	t-value	stand. estim.
β_{41}	.088	0.9	.129	0.116	1.5	.179	0.114	1.5	.178
β_{42}	.330	3.4	.473	0.406	5.4	.629	0.403	5.3	.625
β_{43}	.151	2.3	.312	0.222	4.0	.447	0.217	3.8	.441
R^2	.295			0.528			0.520		

parameter	PLS			FSR			TSLS	
	esti-mate	t-value	stand. estim.	esti-mate	t-value	stand. estim.	estimate range	t-value range
β_{41}	-	1.1	.207	-	1.1	.159	-.013/.289	-0.1/1.8
β_{42}	-	3.1	.498	-	3.2	.474	.226/.372	2.0/2.8
β_{43}	-	1.4	.269	-	1.9	.267	.051/.253	0.6/2.3
R^2	.309			.285			.011/.209	

The significance at the 5% level also varies across approaches in the case of the β_{43} parameter, which fails to be significant under some of the approaches that do not correct for measurement error attenuation and for some instrumental variable choices in the case of TSLS. In the analysis using disattenuated regression with the Ω reliability estimate (which is correct under the mildest assumptions), we find support for a positive interaction effect between interactive use of budgets and innovation on performance. Overall we can conclude that large biases can be encountered when correction for attenuation is omitted.

8   Discussion

As expected, in this article we have found substantial differences in estimates, t-values and R^2 depending on whether measurement error correction is applied or not, even if SRS are used. For large samples, the researcher can use efficient methods like SEM for performing this correction. For small samples this is not the

case, but researchers can resort to the classic DR technique, which can be extended to the interaction case. Other techniques that are sometimes believed to correct for measurement error attenuation (PLS and FSR) in fact do not.

When DR has been applied in the past, it has traditionally been carried out with Cronbach's α. The use of Heise and Bohrnstedt's Ω is not substantially more complex and relies on milder assumptions and should thus be preferred on a general basis, even if, for our particular data set, results obtained for α and Ω were quite similar.

The DR standard errors (and thus the t-values) can tend to slightly underestimate uncertainty because α and Ω are treated as known, thus leading to the possibility of true null hypotheses being rejected. There have been many attempts to produce exact standard errors of disattenuated estimates (e.g. of correlations) but they rely either on complex simulations or on large-sample methods (Charles, 2005). Exact standard errors can be obtained with SEM and TSLS but only for large samples. PLS has been claimed to produce correct standard errors with resampling methods like the jackknife or the bootstrap. However, even if PLS itself does not require large sample sizes, resampling methods do (e.g. Efron and Tibshirani, 1993; LePage and Billard, 1992; Shao and Tu, 1995). Besides, since PLS does not correct measurement error bias, one may wonder what the use is for a correct standard error around a biased point estimate.

PLS yielded point estimates and R^2 which were very similar to those of the much simpler OLS. This finding is in accordance with that of McDonald (1996), who shows that merely changing the weights of the items used to compose the scale has a minor impact on the results. FSR results were also very similar to those obtained under PLS and OLS. In fact, PLS scores are very close to being principal components, which are considered by some as a special case of factor scores. In any case, in the context of models with interaction terms there is one respect in which OLS is superior to FSR and one respect in which FSR is superior to PLS. OLS makes it possible to compute unstandardized estimates which are the ones that can be interpreted for such models. PLS selects weights of the interaction indicators independently of those of the main effect indicators and thus η_3 fails to be equal to $\eta_1\eta_2$, which compromises its interpretation as a proper interaction term.

Nevertheless, it must be admitted that there has been growing interest in PLS (even if the technique dates back to the 1970's, half of its applications reported in the Social Sciences Citation Index in November 2005 were published in 2001 or later). This is probably due to a mystification of the "soft modelling" term (McDonald, 1996). In fact, many of these applications of PLS are for non-predictive purposes, for which the presence of bias is a fundamental drawback. Our particular application could not be further from a soft-modelling situation: our aim is parameter estimation, the number of variables is small, and theory is well grounded.

Summarizing, all these considerations enable us to make some recommendations for applied researchers. If the aim of the analysis is parameter estimation or theory testing, then lack of bias is a key requirement. In this situation, if sample size is large enough, SEM and TSLS will be the approaches of choice. For small sample sizes, disattenuated regression should be used. Reliability should best be estimated by Ω, as it relies on milder assumptions than α. Disattenuated regression can also be applied for regression models with only main effects (i.e. without interactions). The procedure is essentially the same but becomes much simpler as the computation of the true variance of interaction effects (section 6.2) is not needed.

If the aim of the analysis is prediction, then the appropriate technique would be OLS regression using SRS, no matter what the sample size is. The resulting equations will yield optimal predictions of the dependent composite scores conditional on the explanatory composite scores, but parameter estimates will not reflect any population characteristic or relationship.

Of course, a researcher may be interested in both estimation and prediction and perform two appropriate analyses on the same data.

References

[1] Abernethy, M.A. and Brownell, P. (1999): The role of budgets in organizations facing strategic change: an exploratory study. *Accounting, Organizations and Society*, **24**, 189-204.

[2] Abernethy, M.A. and Vagnoni, E. (2004): Power, organization design and managerial behaviour. *Accounting, Organizations and Society*, **29**, 207-226.

[3] Batista-Foguet, J.M., Coenders, G., and Saris, W.E. (2004a): A parsimonious approach to interaction effects in structural equation models: an application to consumer behaviour. *Working Papers of ESADE*, **183**, 1-28.

[4] Batista-Foguet, J.M., Coenders, G., Saris, W.E., and Bisbe, J. (2004b): Simultaneous estimation of indirect and interaction effects using structural equation models. *Metodološki Zvezki,* **1**, 163-184.

[5] Bisbe, J. and Otley, D.T. (2004): The effects of the interactive use of management control systems on product innovation. *Accounting, Organizations and Society*, **29**, 709-737.

[6] Bollen, K.A. (1989): *Structural Equations with Latent Variables.* New York: Wiley.

[7] Bollen, K.A. (1996): An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika,* **35**, 89-110.

[8] Bollen, K.A. and Paxton, P. (1998): Interactions of latent variables in structural equation models. *Structural Equation Modeling*, **5**, 267-293.

[9] Boomsma, A. and Hoogland, J. (2001): The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, S., and D. Sörbom (Eds): *Structural Equation Modeling: Present and Future,* 139-168. Lincolnwood, II: Scientific Software International.

[10] Capon, N., Farley, J.U., Lehmann, D.R., and Hulbert, J.M. (1992): Profiles of product innovators among large US manufacturers. *Management Science,* **38**, 157-169.

[11] Chalos, P. and Poon, M.C. (2000): Participation and performance in capital budgeting teams. *Behavioral Research in Accounting,* **12***,* 199-229.

[12] Charles, E.P. (2005): The correction for attenuation due to measurement error. Clarifying concepts and creating confidence sets. *Psychological Methods*, **10**, 206-226.

[13] Chenhall, R.H. (2004): The role of cognitive and affective conflict in early implementation of activity-based cost management. *Behavioral Research in Accounting*, **16**, 19-44.

[14] Chenhall, R.H. and Langfield-Smith, K. (1998): The relationship between strategic priorities, management techniques and management accounting: an empirical investigation using a systems approach. *Accounting, Organizations and Society,* **23***,* 243-264.

[15] Chin, W.W. (1998): The partial least squares approach for structural equation modelling. In G.A. Marcoulides (Ed): *Modern Methods for Business Research*, 237-246. Mahway NJ: Lawrence Erlbaum.

[16] Chin, W.W. and Newsted, P.R. (1999): Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed): *Statistical Strategies for Small Sample Research*, 307-341. Thousand Oaks: Sage.

[17] Chong, V.K. and Chong, K.M. (1997): Strategic choices, environmental uncertainty and SBU performance: a note on the intervening role of management accounting systems. *Accounting and Business Research,* **27**, 268-276.

[18] Chong, V.K. and Chong, K.M. (2002): Budget goal commitment and informational effects of budget participation on performance: a structural equation modeling approach. *Behavioral Research in Accounting*, **14***,* 65-86.

[19] Clark, K.B. and Fujimoto, T. (1991): *Product Development Performance*. Boston: Harvard Business School Press.

[20] Coenders, G., Batista-Foguet, J.M., and Saris, W.E. (2006): *Temas Avanzados en Modelos de Ecuaciones Estructurales [Advanced Topics in Structural Equation Models]*. Madrid: La Muralla.

[21] Cortina, J.M. (1993): What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* **78**, 98-104.

[22] Cortina, J.M., Chen, G., and Dunlap, W.P. (2001): Testing interaction effects in Lisrel: examination and illustration of available procedures. *Organizational Research Methods,* **4**, 324-360.

[23] Cronbach, L.J. (1951): Coefficient alpha and the internal structure of tests. *Psychometrika,* **16**, 297-334.

[24] Davila, A. (2000): An empirical study on the drivers of management control systems design in new product development. *Accounting, Organizations and Society,* **25***,* 383-410.

[25] Dijkstra, T. (1983): Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics,* **22**, 67-90.

[26] Dillman, D.A. (2000): *Mail and Internet Surveys*. New York: Wiley.

[27] Efron, B. and Tibshirani, R.J. (1993): *An Introduction to the Bootstrap*. New York: Chapman and Hall.

[28] Fornell, C. and Cha, J. (1994): Partial least squares. In R.P. Bagozzi (Ed): *Advanced Methods in Marketing Research*, 52-78. Cambridge: Blackwell.

[29] Goldberger, A.S. and Duncan, O.D. (1973): *Structural Equation Models in the Social Sciences*. New York: Academic Press.

[30] González, M., Casas, F., and Coenders, G. (in press): A complexity approach to psychological well-being in adolescence: major strengths and methodological issues. *Social Indicators Research.*

[31] Govindarajan, V. (1984): Appropriateness of accounting data in performance evaluation: an empirical examination of environmental uncertainty as an intervening variable. *Accounting, Organizations and Society,* **9**, 33-48.

[32] Govindarajan, V. (1988): A contingency approach to strategy implementation at the business-unit level: integrating administrative mechanisms with strategy. *Academy of Management Journal,* **31**, 828-853.

[33] Govindarajan, V. and Gupta, A.K. (1985): Linking control systems to business unit strategy: impact on performance. *Accounting, Organizations and Society,* **10***,* 51-66.

[34] Gupta, A.K. and Govindarajan, V. (1984): Business unit strategy, managerial characteristics and business unit effectiveness at strategy implementation. *Academy of Management Journal,* **27**, 24-41.

[35] Hair, J.F., Anderson, R.E. Tatham, R.L., and Black, W.C. (1998): *Multivariate Data Analysis (5th edition).* Englewood Cliffs, NJ: Prentice Hall.

[36] Heise, D.R. and Bohrnstedt, G.W. (1970): Validity, invalidity and reliability. In E.F. Borgatta and G.W. Bohrnstedt (Eds): *Sociological Methodology*, 104-129. San Francisco: Jossey-Bass.

[37] Hulland, J. (1999): Use of Partial Least Squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal,* **20**, 195-224.

[38] Irwin, J.R. and McClelland, G. (2001): Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research,* **38**, 100-109.

[39] Jaccard, J., Turrisi, R., and Wan, C.K. (1990): *Interaction Effects in Multiple Regression.* Newbury Park: Sage.

[40] Jaccard, J. and Wan, C.K. (1995): Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: multiple indicator and structural equation approaches. *Psychological Bulletin,* **116**, 348-357.

[41] Johnston, J. (1972): *Econometric Methods (2nd edition).* New York: McGraw Hill.

[42] Jöreskog, K.G. (2000): *Latent Variable Scores and Their Uses.* Lincolnwood, IL: Scientific Software International
http://www.ssicentral.com/lisrel/techdocs/lvscores.pdf.

[43] Jöreskog, K.G. and Sörbom, D. (1989): *LISREL 7. A Guide to the Program and Applications (2nd edition).* Chicago, IL: SPSS Inc.

[44] Jöreskog, K.G. and Sörbom, D. (1993): *LISREL8: Structural Equation Modeling with the SIMPLIS Command Language.* Chicago: Scientific Software International.

[45] Jöreskog, K.G. and Wold, H. (1982): The ML and PLS techniques for modeling with latent variables: historical and comparative aspects. In H. Wold and K.G. Jöreskog (Eds): *Systems Under Indirect Observation: Causality, Structure, Prediction, Part 1*, 263-270. Amsterdam: North Holland.

[46] Jöreskog, K.G. and Yang, F. (1996): Nonlinear structural equation models: the Kenny-Judd model with interaction effects. In G.A. Marcoulides and R.E. Schumacker (Eds): *Advanced Structural Equation Modeling*, 57-88. Mahwah, NJ:Lawrence Erlbaum.

[47] Kanter, R.M. (2001): *Evolve! Succeeding in the Digital Culture of Tomorrow.* Boston: Harvard Business School Press.

[48] Kenny, D.A. and Judd C.M. (1984): Estimating the non-linear and interactive effects of latent variables. *Psychological Bulletin,* **96**, 201-210,

[49] Kim J.O. and Mueller, C.W. (1978a): *Introduction to Factor Analysis. What It Is and How to Do It.* Thousand Oaks: Sage.

[50] Kim J.O. and Mueller, C.W. (1978b): *Factor Analysis. Statistical Methods and Practical Issues.* Thousand Oaks: Sage.

[51] Koopmans T.C. and Hood W.C. (1953): The estimation of simultaneous linear economic relationships. In W.C. Hood and T.C. Koopmans (Eds): *Studies in Econometric Method*, 112-199. New Haven: Yale University Press.

[52] LePage, R. and Billard, L. (1992): *Exploring the Limits of Bootstrap.* New York: Wiley.

[53] Lord, F.M. and Novick, M.R. (1968): *Statistical Theories of Mental Test Scores.* Reading: Addison-Wesley.

[54] Marsh, H.W., Wen, Z., and Hau, K.T. (2004): Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods,* **9**, 275-300.

[55] McDonald, R.P. (1996): Path analysis with composite variables. *Multivariate Behavioral Research,* **31**, 239-270.

[56] Medsker, G.J., Williams, L.J., and Holahan, P.J. (1994): A review of current practices for evaluating causal models in organizational behaviour and human resources management research. *Journal of Management,* **20**, 439-464.

[57] Moerloose, Ch. de (2000): Turning innovation into success. *European Business Forum*, **1**, 29-35.

[58] Moulder, B.C. and Algina, J. (2002): Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling,* **9**, 1-19.

[59] Novick, M.R. and Lewis, C. (1967): Coefficient Alpha and the reliability of composite measurement. *Psychometrika,* **32**, 1-13.

[60] O'Loughlin, C. and Coenders, G. (2004): Estimation of the European Customer Satisfaction Index: maximum likelihood versus partial least squares. Application to postal services. *Total Quality Management and Business Excellence,* **15**, 1231-1255.

[61] Ping, R.A. (1995): A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research,* **32**, 336-347.

[62] Raykov, T. (1997): Scale reliability, Cronbach's coefficient Alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research,* **32**, 329-353.

[63] Raykov, T. (2001): Bias of Cronbach's coefficient Alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement,* **25**, 69-76.

[64] Raykov, T. and Marcoulides, G.A. (2000): *A First Course in Structural Equation Modeling.* Mahwah, NJ: Lawrence Erlbaum.

[65] Saris, W.E., Batista-Foguet, J.M., and Coenders, G (in press): Selection of indicators for the interaction term in structural equation models with interaction. *Quality & Quantity.*

[66] Schumacker, R.E. and Marcoulides, G.A. (1998): *Interaction and Non-Linear Effects in Structural Equations.* Mahwah, NJ: Lawrence Erlbaum.

[67] Scott, T.W. and Tiessen, P. (1999): Performance measurement and managerial teams. *Accounting, Organizations and Society,* **24**, 263-285.

[68] Shao, J. and Tu, D. (1995): *The Jackknife and Bootstrap.* New York: Springer.

[69] Simons, R. (1990): The role of management control systems in creating competitive advantage: new perspectives. *Accounting, Organizations and Society*, **15**, 127-143.

[70] Simons, R. (1991): Strategic orientation and top management attention to control systems. *Strategic Management Journal,* **12**, 49-62.

[71] Simons, R. (1995): *Levers of Control.* Boston: Harvard Business School Press.

[72] Simons, R. (2000): *Performance Measurement and Control Systems for Implementing Strategies.* Upper Saddle River: Prentice Hall.

[73] Simpson, T. (1755): A letter to the right Honorable George Earl of Marclesfield, president of the Royal Society on the advantage of taking the mean of a number of observations in practical astronomy, *Philosophical Transactions of the Royal Society,* **49***, 82-93.

[74] Spearman, C. (1904): The proof and measurement of association between two things. *American Journal of Psychology,* **15***, 72-101.

[75] Spector, P.E. (1992): *Summated Rating Scale Construction. An Introduction.* Newbury Park: Sage.

[76] Thomson, G. and Abernethy, M.A. (1998): Product innovation and management control system design: the concept of fit. *Paper presented at the 4th International Management Control Systems Research Conference, Reading (UK), July 1998.*

[77] Wold, H. (1975): Path models with latent variables: the NIPALS approach. In H.M. Blalock, A. Aganbegian, F.M. Borodkin, R. Boudon, and V. Cappecchi (Eds): *Quantitative Sociology. International Perspectives on Mathematical and Statistical Modeling*, 307-357. New York: Academic Press.

[78] Wold, H. (1982): Soft modeling: the basic design and some extensions. In K.G. Jöreskog and H. Wold (Eds): *Systems under Indirect Observation. Causality, Structure, Prediction, Vol. 2*, 1-54. Amsterdam: North Holland.