Qual Quant (2011) 45:1505–1518
DOI 10.1007/s11135-010-9359-z

RESEARCH NOTE

Linear indices in nonlinear structural equation models: best fitting proper indices and other composites

Theo K. Dijkstra · Jörg Henseler

Published online: 11 August 2011

Abstract The recent advent of nonlinear structural equation models with indices poses a new challenge to the measurement of scientific constructs. We discuss, exemplify and add to a family of statistical methods aimed at creating linear indices, and compare their suitability in a complex path model with linear and moderating relationships. The composites used include principal components, generalized canonical variables, partial least squares, factor extraction ('LISREL'), and a newly developed method: best fitting proper indices. The latter involves the construction of linear combinations of indicators that maximize the fit of (non-)linear structural equations in terms of these indices; the weights as well as the loadings of the indicators are sign restricted so that each indicator contributes to as well as reflects its own index in a predefined way. We use cross-validation to evaluate the methods employed, and analyze the most general situation with a complete interaction specification using the bootstrap. The methods are exemplified using an empirical data set. An additional novel feature is the use of simulations to delineate the range of the possible parameter estimates.

Keywords Best fitting proper indices · Generalized canonical variables · Partial least squares · Latent factor scores · Indices · Interaction · Flat maximum

The authors thank Wynne W. Chin and Vincenzo Esposito Vinzi for helpful comments on a previous version of this paper.

T. K. Dijkstra (✉)
Department of Economics and Econometrics, University of Groningen,
Landleven 5, 9747 AD Groningen, The Netherlands
e-mail: t.k.dijkstra@rug.nl

J. Henseler
Institute for Management Research, Radboud University Nijmegen,
P.O. Box 9108, 6500 HK Nijmegen, The Netherlands
e-mail: j.henseler@fm.ru.nl

1 Introduction

High dimensional concepts linked to each other in a web of linear relationships have been a popular vehicle for research in the social sciences, as substantiated by the popularity of 'Structural Equation Modeling'. The software designed in the sixties and the seventies of the previous century, like LISREL, and its later extensions has been instrumental in resolving the numerical issues associated with fitting and testing latent factor models. In the last decade a new development has gained momentum: the incorporation of indices as an alternative to the traditional psychometric measurement model (Diamantopoulos and Winklhofer 2001). Indices encompass that a construct is built by a linear combination of constituting variables or components (Bollen and Lennox 1991). While there is a rich scientific discussion on the conceptual foundations of index measurement (c. f. Howell et al. 2007a,b; Bollen 2007; Bagozzi 2007), the discussion has largely ignored the statistical aspect of which method should be used to obtain the index weights.

In this paper we would like to draw attention to a family of methods that is able to provide index weights. The methods involved can be characterised roughly as variations and extensions of principal components and canonical variables analysis. An outstanding difference with the latent factor modeling approach is that the alternatives do not rely on distributional assumptions. Typically, there are no unobserved variables lurking behind the observables, allowing the high-dimensional theoretical covariance matrix of the latter to be described as the value of a nonlinear function of parameters from a low-dimensional space. They are best seen as *prescriptions* for dimension reduction in terms of the observables, or for the construction of indices, i. e. adequate linear combinations of the observables. The aim is generally to condense the data in a way that reflects their salient features adequately, often related to prediction purposes. All of them can handle nonlinearities in a structural model with little or no effort.

We will use for demonstrational purposes a dataset that has been put together and analyzed by Chin et al. (2003).[1] It is relatively small (it consists of 12-dimensional observations on each of 250 individuals) and shows evidence of interaction. The observations are values of indicator variables, in fact coded answers to questions. They can be partitioned into three groups, each group intending to measure one of three constructs, namely the *perceived-usefulness*, the *enjoyment* and the *intention-to-use* a particular IT-application. The observations were collected to find out whether there is a negative interaction between *usefulness* and *enjoyment*, in the explanation of *intention-to-use*. The issue is tackled by constructing *indices*, linear compounds of the indicators, and by measuring their relationships. Chin et al. (2003) build on the Technology Acceptance Model of Davis (1989). Section 2 describes the data briefly.

The next six sections describe a number of approaches to extract information from the data, in the form of linear compounds of the indicators. We give the weights, and the standardized loadings, i. e. the correlations between indicators and indices. We also report the coefficients of the non-linear regressions with the *intention-to-use* index as dependent variable, and the indices for *usefulness* and *enjoyment* as well as their product, and on one occasion also their squared values, as regressors. As a rule we impose *properness* on the indices, meaning that both the weights and the loadings are nonnegative. So each indicator contributes in a positive way to its own index, and reflects it in a positive way as well. In practice variables will be selected as indicators with these conditions in mind. It has been shown elsewhere (Dijkstra 2008) that no matter what covariances the indicators happen to have, proper indices

[1] We thank Wynne W. Chin for providing us with the data.

always exist (the set of admissible coefficient vectors is non-empty, convex and compact). We report measures of fit, and give an indication of the predictive quality of each prescription by cross-validation of the predictions for the *intention-to-use* indicators. We do not provide classical statistical inference measures, except for the prescription that adds a full quadratic term to the regression; here we employ the bootstrap to estimate the distribution function of the estimators, assuming the set of individuals is a random sample from some population.

The prescriptions gradually increase in 'complexity', directly related to the measure and nature of the mutual dependencies between the indicators. First we apply stand-alone constructions, where the coefficients or weights for each index are determined only by the correlations between its own, direct indicators. Section 3 reports the results for a principal component analysis. But we also include the results for an even simpler approach: equally weighted averaging of the indicators per group (or block). In spite of its lack of sophistication, or maybe just because of it, this prescription will turn out to be quite acceptable. In Sect. 4 we turn to indices that capitalize on the mutual linear dependencies between the indicators: generalized canonical variables, as discussed by Kettenring (1971). Of the many possible choices we use the MAXVAR-prescription. Sections 5 and 6 incorporate explicitly the interaction in constructing indices. In Sect. 5 we present PLS, three version in fact: Chin et al's product indicator approach, the Chin et al/Henseler and Fassott two-stage approach, and essentially Wold's original prescription. The second approach sidesteps interaction in the construction phase, but is included because it fits in naturally here. Then we follow in Sect. 6 a suggestion by Dijkstra (2008) in constructing indices that maximize the (internal) fit of a nonlinear regression: best-fitting-proper-indices (BFPI). We use the full quadratic term here, and as mentioned above, employ the bootstrap to calculate standard errors, possible bias, confidence intervals and to test hypotheses. In the penultimate section we use a latent factor model to construct indices, following a suggestion by Jöreskog (2000), based on an idea by Anderson and Rubin (1956). We also investigate whether a nonlinear factor model can ever lead to the simple prescription BFPI. Section 8 concludes by comments on the relative acceptable performance of the method of averages, the related *flat maximum* phenomenon.[2] It also puts all of the methods used into perspective, by delineating the ranges of possible parameter estimates via simulation.

2 Data description

The dataset consists of 250 observations on manifest variables, $\{z_1, z_2, \ldots, z_{250}\}$. In what follows all vectors are column vectors, and we usually stack vectors a and b by $[a; b]$, as in MATLAB. Each vector z representing one individual can be partitioned as $[x; m; y]$, where x reflects the answers to six questions measuring the *perceived-usefulness* of a particular IT-application, m reflects the answers to three questions about the measure of *enjoyment* one expects to obtain from using the application, and y finally reflects the answers to three questions concerning the *intention-to-use* the application. All answers are coded, either on a seven or an eleven point scale. We have recoded one of the indicators ('ENJ2') for *enjoyment*, in order to align the scores with the emotional value, consistently with the other indicators (score i is replaced by $8-i$). In this paper we will analyze the relationships between the standardized variables (manifest as well as indices). Our adjustment yields a 12 by 12 correlation matrix of observed variables with positive entries only (see Table 1).

[2] All computations, except for PLS, were done with MATLAB (Version R2006b, The MathWorks, Inc. 2007), with `fmincon` as the general purpose vehicle for optimization. The PLS-based approaches were implemented in R (2.5.1; R Development Core Team 2007) based on the algorithm description by Tenenhaus et al. (2005).

Table 1 Correlation matrix of the data of Chin et al. (2003)

	x_1	x_2	x_3	x_4	x_5	x_6	m_1	m_2	m_3	y_1	y_2	y_3
x_1	1	0.81	0.81	0.81	0.77	0.77	0.39	0.39	0.19	0.50	0.56	0.58
x_2		1	0.90	0.89	0.86	0.76	0.40	0.27	0.20	0.49	0.52	0.55
x_3			1	0.92	0.88	0.80	0.41	0.32	0.20	0.53	0.55	0.60
x_4				1	0.89	0.83	0.48	0.36	0.25	0.54	0.58	0.59
x_5					1	0.85	0.46	0.32	0.25	0.52	0.57	0.58
x_6						1	0.53	0.39	0.28	0.57	0.63	0.63
m_1							1	0.65	0.52	0.39	0.46	0.45
m_2								1	0.32	0.42	0.51	0.43
m_3									1	0.19	0.23	0.26
y_1										1	0.85	0.85
y_2											1	0.85
y_3												1

We note that, roughly, the x-indicators as well as the y-indicators show both a higher within-correlation and a higher between-correlation than the m-indicators. In the following sections we will check whether *indices*, carefully constructed linear compounds of the x-, m- and y-indicators, can make a negative interaction relationship visible and whether they can be of value for predictive purposes.

3 Stand-alone constructions: simple averages and principal components

Since all correlations are positive, a very simple way to construct indices is by just taking the average of each block, of each set of scores on the indicators per concept. The loading of an indicator on the index of its block is then proportional to the average correlation of the indicator with all indicators within its block. Relationships between the indices are given by straightforward (non-)linear regressions or otherwise. However, a simple average is not necessarily the best one-dimensional extraction of the high-dimensional information contained in the distribution of the indicators. For a variety of quadratic performance measures, the best way to capture the information is by means of principal components (see e. g. Rao 1973). Since all correlations are positive, for each block Perron-Frobenius' theorem guarantees that the corresponding first principal component has positive components, and that the indicators have positive loadings. So both approaches will generate *proper indices* for this data set. For the correlation matrix at hand they may be expected to yield similar results, which is indeed borne out by the computations (see first two columns of Table 2).

Note that the weights are standardized so that the indices have unit variance (we divide by N (=250) and not by 249, since we generally follow the rule that a functional of a distribution is estimated by the same functional of the empirical distribution). Regression of the y-index on the other indices together with their product, in deviation of its mean, also yields the same picture for both approaches. So for the method of averages:

$$\begin{aligned}\text{intention-to-use} = \text{constant} + 0.46 \cdot \text{perceived-usefulness} + 0.20 \cdot \text{enjoyment} \\ -0.14 \cdot (\text{perceived-usefulness} \times \text{enjoyment}) + \text{residual}. \qquad (1)\end{aligned}$$

In what follows, β_x, β_m, and β_{xm} will stand for regression coefficients for constructed indices, not 'population parameters'. The interaction term has a negative coefficient β_{xm},

Table 2 Coefficients per method

Coefficient	Average	PC	MAXVAR	PLS-PI	PLS-2S	PLS-Wold	BFPI
Weights							
x_1	0.18	0.17	0.03	0.18	0.27	0.27	0.33
x_2	0.18	0.18	0	0.17	−0.14	−0.14	0.02
x_3	0.18	0.18	0	0.18	0.13	0.14	0.10
x_4	0.18	0.18	0.27	0.18	0.16	0.16	0.04
x_5	0.18	0.18	0	0.18	0.00	−0.01	0
x_6	0.18	0.17	0.74	0.20	0.63	0.63	0.59
m_1	0.41	0.45	0.69	0.46	0.47	0.47	0.10
m_2	0.41	0.40	0.40	0.49	0.61	0.61	0.94
m_3	0.41	0.36	0.02	0.24	0.04	0.04	0
y_1	0.35	0.35	0	0.33	−0.10	−0.08	0.11
y_2	0.35	0.35	0.59	0.36	0.61	0.59	0.46
y_3	0.35	0.35	0.45	0.36	0.64	0.53	0.47
Loadings							
x_1	0.89	0.89	0.82	0.89	0.88	0.88	0.91
x_2	0.94	0.94	0.82	0.93	0.82	0.82	0.85
x_3	0.95	0.95	0.87	0.95	0.88	0.88	0.89
x_4	0.96	0.96	0.91	0.96	0.90	0.90	0.90
x_5	0.94	0.94	0.89	0.94	0.89	0.89	0.89
x_6	0.90	0.90	0.79	0.90	0.97	0.97	0.96
m_1	0.89	0.90	0.95	0.91	0.89	0.89	0.70
m_2	0.81	0.81	0.85	0.87	0.93	0.93	1.00
m_3	0.76	0.73	0.51	0.65	0.48	0.48	0.35
y_1	0.95	0.95	0.88	0.95	0.86	0.86	0.91
y_2	0.95	0.95	0.97	0.95	0.97	0.96	0.96
y_3	0.95	0.95	0.95	0.95	0.95	0.96	0.96
β_x	0.46	0.48	0.46	0.45	0.45	0.46	0.46	0.48
β_m	0.20	0.20	0.21	0.21	0.23	0.24	0.24	0.27
β_{x^2}		0.02						0.01
β_{m^2}		0.08						0.06
β_{xm}	−0.14	−0.20	−0.14	−0.13	−0.16	−0.14	−0.14	−0.19
R^2	0.49	0.50	0.49	0.51	0.50	0.54	0.54	0.55

Note: The PLS product indicator approach yields an additional index (which is not reported here), representing the interaction, with 18 weights for $\{x_1 \times m_1, \ldots, x_6 \times m_3\}$, ranging from 0.05 to 0.08

meaning that the higher the perceived-usefulness of the IT-application, the lower the impact of enjoyment on the intention-to-use. In this section we will not assess the stability/significance of the interaction coefficient, but check instead whether inclusion of the product term in the regression helps to improve the prediction of the *standardized* y-indicators. We will employ cross-validation for this purpose. This means that for each individual i, for each vector z_i, we will construct the indices and calculate the coefficients on the dataset *without* individual i. We emphasize that the construction and calculation procedure is *identical* for both the reduced and the full data set (so in particular, if we drop one of the individuals, the data on the remaining individuals will be standardized anew). We then predict the y-index for individual i with its x- and m-index, using the coefficients determined on the reduced data set. The vector y is predicted by multiplying its predicted y-index by the estimated loadings. This produces 250 'honest' prediction error vectors, 'honest' in the sense that z_i plays no role whatsoever in the construction of the prediction for y_i, as opposed to 'fitting' errors or 'reproduction' errors. We will report the results for the method of averages only, since the principal components method gives the same picture.

For linear terms only the errors are on the average slightly positive, so the predictions are a wee bit too low, and the eigenvalues of the covariance matrix of prediction errors are (1.51, 0.15, 0.15). If we include the product term we again get slightly improved prediction errors; the eigenvalues of their second order moments matrix equal (1.42, 0.15, 0.15). The difference between the matrices is small but positive (elementwise), it is almost positive semi-definite. Inclusion of interaction is therefore modestly helpful for prediction. We emphasize that this is not a gratuitous conclusion. While adding a term will always increase the fit, i. e. the R^2, prediction improvement is *no free lunch*.

4 Mutually informative constructions with linear relationships: generalized canonical variables

In the previous section we discussed indices whose construction ignored their embedding in a wider setting. Whatever the contents of the other blocks in the analysis, the constructions would be the same. The principal components method e. g. extracts the 'maximum amount of information' per block but completely ignores the between block relationships in its definition of the indices. One could argue that these relationships are the very reason for studying the indicators together. Generalized canonical variables present themselves as logical alternatives: they are 'maximally correlated indices', tied together, linearwise, as closely as possible. 'Maximally correlated' is ambiguous for three or more indices. Kettenring (1971) has defined a number of possible operationalizations, each based on a different distance measure between the correlation matrix of the indices and the identity matrix of the same size. Typically, a function of the eigenvalues of the correlation matrix of the indices, as a function of the weights, is optimized. Examples are minimization of the smallest eigenvalue, or the product of the eigenvalues; alternatively one could maximize the largest eigenvalue, or the sum of the squared eigenvalues. A simple approach, also covered, is to maximize the sum of the correlations (SUMCOR). Kettenring (1971) discusses in depth the motivation of the various methods and their interrelationships, their extensions to higher stages, as well as the design of algorithms, and their convergence properties (see also Gnanadesikan 1997).

We will report some calculation results for one of the methods studied, the maximization of the largest eigenvalue, called MAXVAR, *subject* to the demand of properness. MAXVAR produces indices for the x-, m- and y-block that are explained best by *their* first principal component. The third columns of Table 2 collects the weights and the loadings for MAXVAR. We note that MAXVAR could live without x_2, x_3, x_5 as well as y_1 and perhaps also m_3. The weights differ substantially from the weights obtained by the method of averages. Nevertheless, the difference between the regression coefficients is very small. The results (not displayed) for cross-validation are also very similar; the prediction errors are of about the same size, MAXVAR being slightly worse than the method of averages. It confirms that interaction helps to improve the predictions somewhat.

5 Mutually informative constructions with interaction: PLS

PLS is a family of alternating least squares algorithms,'prescriptions', that extend principal component and canonical variables analysis. It was designed by Wold (1966, 1982) for the analysis of high dimensional data in a low-structure environment. It has undergone various extensions and modifications, see Lohmöller (1989); Tenenhaus et al. (2005), or the Handbook of Computational Statistics volume II (2008) for an overview, discussion and

applications. PLS' potential for handling non-linear relationships was already recognized by Wold in 1982.

In the literature on Structural Equation Modeling (SEM), interactions between latent factors were first studied by Busemeyer and Jones (1983) and Kenny and Judd (1984). They suggested to extend the list of indicators by the products of the exogenous and mediator indicators. For our problem it would mean stacking $x_i \times m_j$ for all combinations of i and j. The theoretical covariance matrix of the extended vector of indicators, with a much higher dimension and a large number of additional nonlinearities, is then fitted to the sample covariance matrix. There are many variations, all aiming for a reduction in complexity and a higher stability of estimates, but we will not cover them here (see e. g. Schumacker and Marcoulides 1998).

Chin et al. (2003) transferred this idea to PLS path modeling. They added to the x- and the m-block a new block xm, say, with typical element $x_i \times m_j$. This new block is treated as any other in the PLS-algorithms, and estimation or construction is relatively straightforward. We will describe briefly one of the so-called 'mode A'-versions of PLS, that focus on indicator prediction; they can be seen as generalized principal components. There is another mode, mode B, which is really a generalization of canonical variables, related to SUMCOR, see Dijkstra (1981, 2008).

The algorithm is essentially a sequence of regressions in terms of weightvectors. The weightvectors obtained at convergence satisfy *fixed point equations* (see Dijkstra (1981, 2008) for a general analysis of such equations and ensuing convergence issues). If we denote by a, b, c and d the limiting *standardized* weightvectors for the x-, m-, y- and xm-block respectively we have:

- a is proportional to the regression of x on $c^\top y$;
- b is proportional to the regression of m on $c^\top y$;
- d is proportional to the regression of xm on $c^\top y$;
- c is proportional to the regression of y on $s_{a,c}a^\top x + s_{b,c}b^\top m + s_{d,c}d^\top(xm - \overline{xm})$, where the scalar $s_{a,c}$ is the sign of the correlation between $a^\top x$ and $c^\top y$, and the other s-scalars are defined similarly. The bar in $\overline{xm}$ signifies the mean.

The algorithm takes arbitrary starting vectors, and then basically follows the sequence of regressions above, each time inserting updates when available. Loadings and inner regression coefficients are calculated in a straightforward way, given the constructed indices.

The approach outlined takes its direction from the SEM-literature, by incorporating the products of the x- and m-indicators. Simple and appealing as this may be, one could feel that there is conceptually also a bit of a drawback, since the interaction index will never equal the product of the indices. It has been suggested by Chin et al. (2003) and elaborated by Henseler and Fassott (2008), to first employ the PLS-approach *ignoring* the interaction, so in effect deleting all reference to xm in the regressions above, and then define the interaction index by the product of the x- and m-indices so obtained. This two-stage prescription is not unlike another approach in SEM, see the section on latent factor modeling. Wold (1982) has proposed a way to deal with nonlinearities that is straightforward to adapt to the interaction problem. It simply amounts to replacing $d^\top xm$ by $(a^\top x) \times \left(b^\top m\right)$, adjusting the sign scalars accordingly. Variations in PLS are defined by different choices for the scalars, with different functions of the correlations, but we stick to the original approach. We note that PLS-indices are not automatically *proper*, in the sense that weights and loadings are nonnegative. It is not evident how to impose this property in general.

We report the numerical results for the *indicator product*-, the *two-stage*- and the *Wold*-approach. First we look at the weights and the loadings as listed in Table 2. The loadings

are similar, for all prescriptions involved, with the main exception of m_3. *Product indicator* and *average* produce weights of the same order, again with m_3 as the main exception. But *two-stage* and *Wold* choose weights rather different from the former, different in both size and sign. Particularly the negative signs impose a critical issue. If manifest variables contribute to the respective index in a different way, i. e. they deliver a negative instead of a positive contribution, the meaning of the index may change. Even though the weights are different, the regression coefficients are comparable again, with little for *product indicator* to distinguish itself from *average*.

We do get some improvement though, compared to *average* i. e., in terms of cross-validation. Differences become apparent when one compares the eigenvalues: (1.40, 0.15, 0.15) for the PLS product indicator approach, (1.39, 0.15, 0.15) for the PLS two-stage approach and (1.39, 0.15, 0.15) for Wold's approach, compared with (1.42, 0.15, 0.15) for the average.

6 Mutually informative constructions with interaction: best-fitting-proper-indices

The way PLS incorporates the nonlinearity in the mutual dependencies is rather indirect. It can also be done explicitly. We will exemplify this for the extension of the linear specification with a *full* quadratic expression (i. e. the product of the indices plus the respective squared terms): we will construct indices of the form $a^\top x$, $b^\top m$ and $c^\top y$ such that the regression of $c^\top y$ on $\left\{1, a^\top x, b^\top m, \left(a^\top x\right)^2, \left(b^\top m\right)^2, \left(a^\top x\right) \times \left(b^\top m\right)\right\}$ has maximal R^2 subject to the usual standardization (i. e., the sample variances of the indices equal 1) and to *properness* (i. e., weights as well as loadings are non-negative). We believe this type of prescription to be new. The numerical results in Table 2 are indicated by *BFPI*. There are substantial differences between the weights, but the loadings are not too far apart, save for the moderator indicators. Also, the values of the nonlinear regression are similar.

The coefficients of the squared indices seem to be rather small, indeed one may wonder whether they contribute anything at all. One way to assess that is by means of the *bootstrap*. For the *BFPI*-approach we generated 10,000 bootstrap samples, and calculated for each sample all 31 quantities (the weights, the loadings, the regression coefficients including the constant, and the R^2). This yields (an approximation of) 'the nonparametric Maximum Likelihood estimate of the distribution of the estimators', see e. g. Efron (1982) and Davison and Hinkley (2003). The 95% shortest empirical confidence intervals for β_{x^2} and β_{m^2} are $[-0.19, +0.12]$ and $[-0.11, +0.16]$, respectively. The squared indices could therefore be deleted. This is confirmed by refitting with exclusion of the squares: the results, not reported here, are rather similar to what we reported above, with an almost *identical* value of R^2. The coefficients β_x and β_m are clearly significant though, with 95% shortest confidence intervals equal to [0.37, 0.59] and [0.18, 0.38], respectively. The interaction term has an estimated coefficient β_{xm} equal to -0.19. Bootstrapping reveals that it is biased downward; a better estimate is -0.15. Its significance is not overwhelming: the bootstrap values for β_{xm} have a standard deviation of about 0.08, and its 95% shortest confidence interval is equal to $[-0.30, +0.02]$. Boootstrapping shows that the R^2 is biased upwards; a better estimate is 0.52. The loading for m_3 is relatively small, its bias-corrected value even smaller, 0.28; its 95% shortest confidence interval does not include zero: [0.03, 0.53].

We have also used cross-validation to assess three different *BFPI*-prescriptions: 250 'honest' predictions were calculated for the y-indicators, using the *linear*, the *linear plus interaction*, and the *linear plus interaction plus squares* regression. Note that each prescription determines the indices in a different way. Somewhat remarkably, all three second order

moments around zero of the prediction errors had the same second and third eigenvalues, they differed only in their first (i.e. largest) eigenvalue. The best prescription (i.e. with the smallest largest eigenvalue) is based on *linear plus interaction*, followed by *linear*, and then by the more extensive specification *linear plus interaction plus squares*. This exemplifies that more predictors lead not necessarily to better predictive performance, although the fit will by necessity improve. The differences are not large, but, if anything, they confirm that interaction helps to predict, as opposed to inclusion of the squared indices in addition to the interaction term. For completeness' sake we report the eigenvalues of the second order moments matrix of the prediction errors for the *linear plus interaction* prescription as (1.39, 0.15, 0.14). The largest eigenvalues of the two contenders here are 1.46 for the linear case, and 1.49 for the full specification. We note that our *BFPI linear plus interaction* prescription is 'even' better than the method of averages.

7 Latent factor modeling

7.1 A latent factor prescription

In this section we fit a structured correlation matrix to the observed correlation matrix of the indicators, and use the results to construct indices. Until now no structure was assumed or postulated for the joint distribution of the indicators, in fact not even the existence of a joint distribution was necessary (although the employment of the bootstrap and the interpretation of the results in terms of 'significance', 'confidence interval', and 'bias', are more easily justified when the individuals, the z_i's, can be seen as a random sample from some stable population). We just calculated all sorts of aspects of the data in an attempt to reduce its 'complexity' effectively. In the same vein we will use the notion of a factor model *instrumentally*, without imparting an existential meaning to factors or parameters.

So consider:

$$\begin{bmatrix} x \\ m \\ y \end{bmatrix} = \begin{bmatrix} \lambda_x & & \\ & \lambda_m & \\ & & \lambda_y \end{bmatrix} \begin{bmatrix} \eta_x \\ \eta_m \\ \eta_y \end{bmatrix} + \begin{bmatrix} \delta_x \\ \delta_m \\ \delta_y \end{bmatrix} \tag{2}$$

Empty cells contain zeros. Here η_x, η_m and η_y are scalar latent factors corresponding with the x-, m- and y-block, respectively, and x loads on η_x with λ_x etc. All vectors have zero mean, their second moments exist, $\eta \equiv [\eta_x; \eta_m; \eta_y]$ is independent of $\delta \equiv [\delta_x; \delta_m; \delta_y]$, and all components of the latter (12 by one) vector are mutually independent. It is evident that none of these assumptions can be confirmed. Their function is to generate possibly useful restrictions. In this case they imply that

$$\Sigma = \Lambda \Phi \Lambda^\top + \Theta \tag{3}$$

where Σ is the theoretical correlation matrix of the indicators, Φ stands for the correlation matrix of η, Λ is the matrix with the factor loadings, and Θ is the diagonal matrix containing the unique variances of the elements of δ (the $diag$ of Θ is the $diag$ of $(I - \Lambda \Phi \Lambda^\top)$).[3] There is an unlimited supply of fitting functions, each of which, when minimized, will produce a numerical value for Σ that is 'as close as possible' to S, the correlation matrix of the indicators. The results will vary with the fitting function chosen, the variation being smaller

[3] In general, Σ and S are variance/covariance matrices. Here, we regard them as correlation matrices here, because we assume standardized manifest variables.

Table 3 Factor loadings

Indicator	$[\lambda_x; \lambda_m; \lambda_y]^\top$	$diag(\Theta)^\top$
x_1	0.86	0.26
x_2	0.90	0.20
x_3	0.93	0.13
x_4	0.96	0.08
x_5	0.93	0.14
x_6	0.92	0.16
m_1	0.91	0.17
m_2	0.74	0.45
m_3	0.49	0.76
y_1	0.87	0.24
y_2	0.94	0.12
y_3	0.95	0.10

the better the fit. We will use trace $\left[(S-\Sigma)^2\right]$, possibly the conceptually simplest fitting function. One of many alternatives is trace $\left[\left((S-\Sigma)\,S^{-1}\right)^2\right]$, whose choice is usually justified by an appeal to 'normality' and the desideratum of 'asymptotic efficiency'. We tend to see these justifications as heuristic devices, helpful in generating fitting functions which may be reasonable and useful from a number of perspectives. At any rate, we will not adhere to the mistake of imputing a belief in normality to anyone who uses this last function; nor does our choice of the simple Euclidean distance warrant any supposition regarding a belief in any distribution whatsoever (c. f. also John Tukey (1980): "methodologies have no assumptions").

Fitting Σ to S by minimizing trace $\left[(S-\Sigma)^2\right]$ yields the results as shown in Table 3 for the factor loadings (not to be confused with the loadings of the previous sections, which are correlations between *observables*) and unique variances.

The fit is not bad: the mean absolute error is about 0.03, the largest discrepancies are for the correlation between x_6 and m_1 (the fitted value is 0.1 smaller than the observed value of 0.53) and for the correlation between m_2 and y_2 (the fitted value is 0.12 smaller than the observed value of 0.51).

Now the latent factors cannot be observed, but one could follow Anderson and Rubin (1956) and minimize with respect to the η_i's:

$$\sum_{i=1}^{N} (z_i - \Lambda\eta_i)^\top \Theta^{-1} (z_i - \Lambda\eta_i) \tag{4}$$

with Λ and Θ as fitted to S, subject to

$$\frac{1}{N}\sum_{i=1}^{N} \eta_i \eta_i^\top = \Phi. \tag{5}$$

So the constructed η will be linear in z, and since they share the correlation matrix with the latent factors, they will satisfy the *same* regression equations; see also Jöreskog (2000) for further discussion and an application. From this paper one obtains that the constructed η satisfies:

$$\eta_{\text{constructed}} = UD^{\frac{1}{2}} \left(VL^{\frac{1}{2}}V^\top\right)^{-1} D^{\frac{1}{2}} U^\top \Lambda^\top \Theta^{-1} z \tag{6}$$

where

$$[U, D] = eig(\Phi), \tag{7}$$

Table 4 Factor weights and loadings (latent factor modeling)

Indicators	Weights			Loadings
	$\eta_x^{\text{constructed}}$	$\eta_m^{\text{constructed}}$	$\eta_y^{\text{constructed}}$	
x_1	0.09	0.00	0.00	0.87
x_2	0.12	0.00	0.00	0.93
x_3	0.19	0.00	0.00	0.96
x_4	0.30	0.00	0.01	0.97
x_5	0.19	0.00	0.00	0.95
x_6	0.16	0.00	0.00	0.90
m_1	0.00	0.76	0.02	0.98
m_2	0.00	0.23	0.01	0.78
m_3	0.00	0.09	0.00	0.58
y_1	0.00	0.01	0.18	0.92
y_2	0.00	0.02	0.37	0.95
y_3	0.01	0.03	0.47	0.95

so U is the matrix of eigenvectors of Φ and D is the diagonal matrix of corresponding eigenvalues, and

$$[V, L] = eig\left(D^{\frac{1}{2}} U^{\top} \Lambda^{\top} \Theta^{-1} S \Theta^{-1} \Lambda U D^{\frac{1}{2}}\right). \tag{8}$$

Inserting the numbers found leads to the indices as shown in Table 4.

It is gratifying to see that the indices are predominantly determined by their *own* indicators. A small bootstrap experiment indicated that this is generally true for data similar to the data at hand, except for the coefficients of the y-indicators for $\eta_m^{\text{constructed}}$ which are not negligible, so we refrained from doing a cross-validation. The tendency to include indicators from other blocks was much more outspoken for the fitting function trace $\left[\left((S-\Sigma)\,S^{-1}\right)^2\right]$, whose fitted Σ was also less close to S in terms of signs and sizes of the elements of $S-\Sigma$ (not reported here).

When we regress the y-index, $\eta_y^{\text{constructed}}$, on a constant and the other indices together with their product we get values of 0.44, 0.26, and -0.14 for the coefficients β_x, β_m, and β_{xm}, respectively. The coefficient of determination reaches 0.54. These results are rather not unlike the other approaches. The last column of Table 4 exhibits the loadings, again defined as the correlation between an (observable) index and its indicators.

7.2 Factor models with interaction and simple indices?

The Anderson and Rubin approach is not guaranteed to yield indices that are driven solely by their own indicators. One could fit a one-factor-structure per block, estimate the latent factors by means of regressions and proceed as before. For this dataset and the Euclidean fitting function, it leads to a drop in R^2 to 0.45 and the index for the m-block collapses into m_1, to mention two of the main differences (there are also similarities). One might wonder whether the measurement model of the previous subsection can ever be consistent with a conditional expectation $E\,(y \mid x, m)$ that is linear in $\left\{1, a^{\top}x, b^{\top}m, \left(a^{\top}x\right)^2, \left(b^{\top}m\right)^2, \left(a^{\top}x\right) \times \left(b^{\top}m\right)\right\}$ for appropriate choices of a and b, if in addition

$$\eta_y = \pi_0 + \pi_x \eta_x + \pi_m \eta_m + \pi_{xm} \eta_x \eta_m + \nu \tag{9}$$

where ν is a zero-mean residual independent of $[\eta_x; \eta_m]$, and the π's are regression coefficients. It appears natural to assume even more and take ν independent of $[x; m]$, since this only

requires that v is independent of the measurement errors of η_x and η_m. So $E\,(v \mid x, m) = 0$. We need the conditional expectation of $[\eta_x; \eta_m]$ to be linear in $[x, m]$. This will be true under joint normality of $[\eta_x; \eta_m; x; m]$, a condition which seems virtually necessary also granted the other assumptions made so far, in particular with regard to the measurement relations (see Kagan et al. 1973; Dijkstra 1981, 2008). So with joint normality:

$$\begin{bmatrix} \eta_x \\ \eta_m \end{bmatrix} = \begin{bmatrix} B_{xx} & B_{xm} \\ B_{mx} & B_{mm} \end{bmatrix} \begin{bmatrix} x \\ m \end{bmatrix} + \begin{bmatrix} \varepsilon_x \\ \varepsilon_m \end{bmatrix} \tag{10}$$

with an obvious notation, where ε is independent of $[x; m]$ and $E\,(\varepsilon \mid x, m) = 0$. Inserting this into the interaction-expression for η_y quickly yields that consistency is present if and only if both $[B_{xx}; B_{mx}]$ and $[B_{xm}; B_{mm}]$ have rank 1. If at least one of the x- and m-blocks have at least two indicators the rank condition will be satisfied only on sets of measure zero. In other words, this non-linear factor model will 'never' lead exactly to one of the simple prescriptions we have used before. It remains to be determined what kind of prescriptions, say for prediction or index construction, approaches produce that build on latent factors with nonlinear dependencies, and what quality they have.

8 Conclusion

We have reported numerical results for the dataset of Chin et al. (2003), for various prescriptions of how to construct linear indices of the manifest variables. Each prescription was motivated by another 'philosophy', one more 'sophisticated' than the other. We had the 'no-brainer', no offence intended, that averaged the indicators per block. We looked at principal components which capitalize on the information within each block separately. The generalized canonical variables on their turn exploited the linear mutual dependencies between the blocks. Methods like PLS and best-fitting-proper-indices took in addition non-linearities between the blocks into account. And we ventured into the mystic realm of the unobservable, occupied by latent factors. The no-brainer *average* proved to be a worthy 'contestant/opponent', extracting a comparable amount of information from the blocks, with not a bad internal fit, and quite a good prediction (cross-validation) performance (all in *relative* terms). The other methods are useful for the assessment of the relative importance of the indicators, which *average*, or any approach with *a priori* fixed weights, cannot by its very nature do at all. The others also showed a somewhat higher internal fit, with the exception of the blockwise latent factor approach.

The relatively acceptable performance, at little or no cost, of the simple average is a manifestation of the *flat maximum* phenomenon. Its ubiquity is well recognized in the forecasting and decision making (expert systems) literature, see e.g. Lovie and Lovie (1986). The latter argue that instead of bemoaning the flat maximum one should welcome it: instead of focusing on weighting systems one can invest most of the effort on the design and choice of the (predictor) variables. This is not to say there is no use for more elaborate weighting systems. They will be of assistance to assess the relative importance of the observables. And as we would like to emphasize, they *are* called for when correlations between the indicators per block or concept do not have the same sign and—unlike the case we analyzed here—cannot be made to have the same sign by simultaneous transformations.

Finally, as promised, we put all of the methods used into perspective by delineating the ranges of all possible parameter estimates. We generated randomly for each block weighting vectors, independently within and between blocks: 10,000 independent drawings from the unit simplexes of $\mathbb{R}^6$ and $\mathbb{R}^3$ (twice). For each triplet of weighting vec-

Table 5 Confidence intervals of regression estimates

Parameter	Mean	Std. dev.	Maximum	Minimum	95% Interval
p_x	0.46	0.04	0.59	0.36	[0.39, 0.54]
p_m	0.19	0.06	0.35	0.03	[0.09, 0.30]
p_{xm}	−0.14	0.01	−0.11	−0.18	[−0.16, −0.12]
R^2	0.47	0.03	0.55	0.34	[0.40, 0.53]
λ_{m_3}	0.72	0.17	1.00	0.34	[0.45 , 1.00]

tors (a, b, c) we calculated all values of interest for the case where $c^\top y$ is regressed on $\{1, a^\top x, b^\top m, (a^\top x) \times (b^\top m)\}$. Table 5 collects some of the results. The last column gives the shortest intervals that contain 95% of the values. We included the loading for m_3 because the inward oriented prescriptions tended to assign this indicator a rather lower loading than the others. The high values of λ_{m_3} for arbitrary weighting vectors affirms the special position of m_3: indices that somehow capitalize on the information in the data have little use for it. It is also clear that the R^2's found are rather high, in particular the *BFPI*−value is unsurpassed (which was to be expected). The R^2 of the method of averages is not as large, but still lies in the upper quadrant of the distribution: about 74% of the randomly chosen weighting vectors give a *smaller* R^2. We see that for β_{xm} the playing field is smallish: there is little room to distinguish oneself. If the object was to find indices that support the hypothesis of negative interaction, *any* three indices would do! But recall that the bootstrap analysis for the *BFPI*-method indicated that other samples of similar size could well yield values closer to zero, in fact the value 0 is contained in the shortest 95% confidence interval for *its* interaction coefficient.

References

Anderson, T.W., Rubin, H.: Statistical inference in factor analysis. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 150 (1956)

Bagozzi, R.: On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007). Psychol. Methods **12**(2), 229–237 (2007)

Bollen, K.A.: Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). Psychol. Methods **12**(2), 219–228 (2007)

Bollen, K.A., Lennox, R.: Conventional wisdom on measurement, a structural equation perspective. Psychol. Bull. **110**(2), 305–314 (1991)

Busemeyer, J.B., Jones, L.E.: Analysis of multiplicative combination rules when causal variables are measured with error. Psychol. Bull. **93**(3), 549–562 (1983)

Chin, W.W., Marcolin, B.L., Newsted, P.R.: A partial least squares latent variable modeling approach for measuring interaction effects. Results from a monte carlo simulation study and an electronic-mail emotionadopion study. Inf. Syst. Res. **14**(2), 189–217 (2003)

Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)

Davison, A.C., Hinkley, D.V.: Bootstrap methods and their application. 2nd edn. Cambridge University Press, Cambridge (2003)

Diamantopoulos, A., Winklhofer, H.M.: Index construction with formative indicators: an alternative to scale development. J. Mark. Res. **38**, 269–277 (2001)

Dijkstra, T.K.: Latent variables in linear stochastic models: reflections on maximum likelihood and partial least squares methods. PhD thesis, Groningen, 2nd edn. Published in 1985 by Sociometric Research Foundation (1981)

Dijkstra, T.K.: Latent variables and indices: Herman Wold's basic design and partial least squares. In: Vinzi, V.E., Chin, W.W., Henseler, J., Wang, H. (eds.) Handbook of Partial Least Squares: Concepts, Methods, and Applications, Computational Statistics, vol. 2. Springer, Heidelberg (2008, in print)

Efron, B.: The Jackknife, the Bootstrap, and Other Resampling Plans. Vol. 92, Society for Industrial and Applied Mathematics, Philadelphia (1982)
Gnanadesikan, R.: Methods for Statistical Data Analysis of Multivariate Observations. 2nd edn. Wiley series in Probability and Mathematical Statistics Wiley, New York (1997)
Henseler, J., Fassott, G.: Testing moderating effects in PLS path models: An illustration of available procedures. In: Vinzi, V.E., Chin, W.W., Henseler, J., Wang, H. (eds.) Handbook of Partial Least Squares: Concepts, Methods, and Applications, Computational Statistics, vol. 2. Springer, Heidelberg (2008, in print)
Howell, R.D., Breivik, E., Wilcox, J.B.: Is formative measurement really measurement? reply to Bollen (2007) and Bagozzi (2007). Psychol. Methods **12**(2), 238–245 (2007a)
Howell, R.D., Breivik, E., Wilcox, J.B.: Reconsidering formative measurement. Psychol. Methods **12**(2), 205–218 (2007b)
Jöreskog, K.G.: Latent variable scores and their uses. Scientific Software International, http://www.ssicentral.com/lisrel (2000)
Kagan, A.M., Linnik, Y.V., Rao, C.R.: Characterization problems in mathematical statistics. Probability and Mathematical Statistics Wiley, New York (1973)
Kenny, D.A., Judd, C.M.: Estimating the nonlinear and interactive effects of latent variables. Psychol. Bull. **96**(1), 201–210 (1984)
Kettenring, J.R.: Canonical analysis of several sets of variables. Biometrika **58**(3), 433–451 (1971)
Lohmöller, J.B.: Latent Variable Path Modeling with Partial Least Squares. Physica, Heidelberg (1989)
Lovie, A.D., Lovie, P.: The flat maximum effect and linear scoring models for prediction. J. Forecast. **5**(3), 159–168 (1986)
R Development Core Team: R: a language and environment for statistical computing. http://www.R-project.org (2007)
Rao, C.R.: Linear Statistical Inference and Its Applications. Wiley, New York (1973)
Schumacker, R.E., Marcoulides, G.A. (eds.): Interaction and Nonlinear Effects in Structural Equation Modeling. Lawrence Erlbaum Associates, Mahwah, NJ (1998)
Tenenhaus, M., Vinzi, V.E., Chatelin, Y.M., Lauro, C.: PLS path modeling. Comput. Stat. Data Anal. **48**(1), 159–205 (2005)
The MathWorks, Inc: MATLAB: the language of technical computing (2007)
Tukey, J.W.: Methodological comments focused on opportunities. In: Brillinger, D.R., Tukey, J.W. (eds.) The Collected Works of John W. Tukey (Reprint 1986) vol. 4, pp. 819–867. Philosophy, Wadsworth Advanced Books & Software, chap 23 (1980)
Wold, H.O.A.: Non-linear estimation by iterative least squares procedures. In: David, F.N. (ed.) Research papers in statistics, pp. 411–444. Wiley, London (1966)
Wold, H.O.A.: Soft modelling: the basic design and some extensions. In: Jöreskog, K.G., Wold, H.O.A. (eds.) Systems under Indirect Observation. Causality, Structure, Prediction, pp. 1–54. North-Holland, Amsterdam (1982)