Bootstrap confidence intervals for trilinear partial least squares regression

Sven Serneels

Pierre J. Van Espen*

Department of Chemistry, University of Antwerp, Belgium

February 4, 2005

revised version

*Correspondence to S. Serneels, Departement Scheikunde, Universiteit Antwerpen, Universiteitsplein 1, 2610 Antwerpen (Belgium) E-mail: sven.serneels@ua.ac.be. . Tel.: +32/3/8202378; Fax.: +32/3/8202376; url: http://www.chemometrix.ua.ac.be.

Bootstrap confidence intervals for trilinear partial least squares regression

Abstract

The boostrap is a successful technique to obtain confidence limits for estimates where it is theoretically impossible to establish an exact expression thereunto. Trilinear partial least squares regression (tri-PLS) is an estimator for which this is the case; in the current paper we thus propose to apply the bootstrap in order to obtain confidence intervals for the predictions made by tri-PLS. By dint of an extensive simulation study, we show that bootstrap confidence intervals have a desirable coverage. Finally, we apply the method to an identification problem of micro-organisms and show that from the bootstrap confidence intervals, the organisms can (up to a misclassification probability of 3.5 %) correctly be identified.

Keywords: bootstrap, confidence interval, uncertainty, trilinear partial least squares regression, tri-PLS, prediction.

1 Introduction

For the purpose of multivariate calibration, most frequently regression techniques are applied which should optimally describe the relation between

a predictor and a predictand. The design of these regression techniques strongly depends on the dimensionality and structure of the predictor. For univariate predictors, least squares regression is optimal at the normal model and hence the most frequently applied technique. For multivariate predictors, several regression techniques are used, depending on the structure of the data. Often the dimensionality of the predictor exceeds the number of observations at hand, or some of the predictor variables may be prone to correlation. In these cases, one tries to summarize the information in the predictor data into a few uncorrelated components, whereupon OLS regression is performed. Partial least squares regression [1] has been particularly successful for this type of data.

Other estimators need to be applied for three-way predictor data matrices. These data matrices consist of n observations of a matrix variable. Three-way data arise naturally in many fields of analytical chemistry, if one has at hand e.g. two different spectra for each observation, or a response at several time or pressure points. Early proposals to treat this type of data consisted of unfolding the three-way array by horizontally aligning the slabs of the $n \times p \times q$-array into a $n \times pq$-matrix and henceforth applying two-way calibration methods such as PLS to these unfolded data, thus disregarding the three-way nature of the data. It should be clear that this order of proceeding can never lead to as good results as when one would apply a

truly trilinear technique which has especially been designed for such data. Hence, several trilinear regression methods have been proposed. As is the case with PLS, trilinear partial least squares regression (tri-PLS) [2] is particularly attractive because it computes uncorrelated components (also called *latent variables*) which respect the three-way structure of the data, and have a maximal covariance with the predictand vector. Hence these latent variables mathematically summarize the information in the predictor which is relevant for prediction of the predictand, without disregarding the original three-way structure of the data. Indeed tri-PLS has been shown to out-perform PLS applied to the unfolded data matrix in terms of being more parsimonious and less prone to predictor noise [3].

The goal of multivariate calibration is to be able to predict a response for new samples. Once the regression coefficients have been determined, prediction itself consists of a single matrix multiplication. An important question to address, however, is how to assess the uncertainty of the predictions made. In the case of trilinear partial least squares, an exact expression for the prediction uncertainty cannot be derived. Hence, Faber and Bro propose a roughly approximate estimate, based on a local linearization of the estimator [4]. However, the correctness of this estimate depends on whether several assumptions hold, which is often hard to verify in practice. Nonetheless, the Faber and Bro estimate is very user-

friendly and has been shown to be successful in several applications (see e.g. [5, 6]). As an exact theoretical derivation of the standard errors is also impossible in the case of partial least squares, data driven techniques such as the jackknife and the bootstrap have proven to be a viable alternatives. The jackknife can be seen as an approximation to the bootstrap [7], which has very recently been successfully applied to yield confidence intervals for related trilinear techniques such as PARAFAC and Tucker3 analysis [8]. Hence, in the current work, we decided to investigate the applicability of the bootstrap in order to obtain confidence intervals for the predictions made by univariate trilinear partial least squares regression (also denoted *tri-PLS1*).

It will be illustrated how to construct these bootstrap confidence intervals. The fact that they may be interpreted as true confidence intervals in the statistical sense, will be shown by dint of an extensive simulation study. Finally, as an illustration of the method, two practical examples will be given.

2 Trilinear partial least squares regression

Let $\mathbf{X}$ and $\mathbf{y}$ denote the calibration data matrices. Let $\mathbf{X} \in I\!R^{n \times p \times q}$ and $\mathbf{y} \in I\!R^{n \times 1}$, respectively, where n is the number of samples at hand and $\mathbf{y}$

is the vector to be predicted. Furthermore, $X \in I\!R^{n \times pq}$ denotes the so-called *unfolded* data matrix, where each of the slabs of $\mathbf{X}$ have been aligned next to each other. Matrices will always be denoted by upper-case letters. The columns of a matrix will be denoted by the corresponding lower-case bold-face letter. Three-way matrices will always be denoted by bold-face upper-case letters. Let vec$(\cdot)$ denote the vectorization operator, which vertically stacks the columns of its argument underneath each other. Hence if A is a $p \times q$ matrix, vec(A) will be a $pq \times 1$-vector. Let $\text{vec}^{-1}_{p,q}$ denote the operator which re-shapes a pq-vector into a $p \times q$ matrix, such that $\text{vec}^{-1}_{p,q}(\text{vec}(A)) = A$. By $\otimes$ we denote the Kronecker product and by $\boxtimes$ we denote the Khatri-Rao (or columnwise Kronecker) product.

Tri-PLS1 models the latent structure in the data according to the following model ($i \in [1, \min(n, pq)]$):

$$X = T_i (W_i^q \boxtimes W_i^p)^T + E_{X,i} \tag{1a}$$

$$\mathbf{y} = T_i \mathbf{b}_i + \varepsilon_{\mathbf{y},i} \tag{1b}$$

In the previous model E_X and $\varepsilon_{\mathbf{y}}$ are random error matrices.

Several algorithms have been proposed to compute estimates of all parameters in the above model [2, 9, 10]. However, de Jong [10] reported one algorithm to out-perform all other variants in terms of computational

properties. Hence this algorithm was used throughout the current paper. For the sake of completeness, it will briefly be repeated here. Let $\mathbf{e}_0 = \mathbf{y}$, then estimates based on i latent variables for all model parameters can be inductively obtained as follows:

$$Z_i = \mathrm{vec}_{p,q}^{-1}(X^T \mathbf{e}_{i-1}) \tag{2a}$$

$$\mathbf{w}_i^p, \mathbf{w}_i^q = \text{dominant singular vectors of } Z_i \tag{2b}$$

$$\mathbf{w}_i = \mathbf{w}_i^q \otimes \mathbf{w}_i^p \tag{2c}$$

$$T_i = XW_i \tag{2d}$$

$$\mathbf{b}_i = \left(T_i^T T_i\right)^{-1} T_i^T \mathbf{y} \tag{2e}$$

$$e_i = \left[I_n - T_i \left(T_i^T T_i\right)^{-1} T_i^T\right] \mathbf{y} \tag{2f}$$

An important quantity to estimate are the regression coefficients relating X and $\mathbf{y}$. From the observation $\mathbf{y} = T_i \mathbf{b}_i = XW_i \mathbf{b}_i$, it follows that:

$$\beta_i = W_i \mathbf{b}_i \tag{3}$$

A topic which cannot be disregarded in practice, is how to determine the optimal value of i. It can be expected that, as is the case in PLS regression, an optimal model dimension $i = h$ exists, such that any choice of $i < h$ leads to an increase in bias, whereas a choice of $i > h$ leads to an increase in prediction variance. Hence it is important to obtain a correct estimate of h before passing to prediction of new samples. The most currently applied technique thereunto is cross validation. In cross validation, an arbitrary number of samples are left out of the calibration matrix. The model parameters are estimated based on the remaining samples and these estimates are then used to predict the response for the samples that have been left out. This process is repeated throughout an arbitrary number of iterations, whereafter a root mean squared error of the deviations between predictand and predicted value is computed. The choice of i which yields the lowest value in RMSEP is considered to be an estimate of h. The type of cross-validation described here is usually being referred to as *random* cross-validation. Simpler types of cross-validation exist such as *leave one out* cross-validation, but these have been reported to over-estimate the optimal model dimensionality [11].

3 Bootstrapping tri-PLS1

As heeded in the Introduction, a three-way $n \times p \times q$-data matrix corresponds to n observations of a $p \times q$-dimensional stochastical variable. These n observations are drawn from a population from which, at least in theory, an infinite number of samples could have been drawn. Based on these n samples at hand, estimates are obtained for all model parameters.

The idea of the bootstrap is to assess the variance of these estimates by re-sampling from the set of samples at hand. In fact, one thus tries to mimic the process which has in the statistical sense led to the generation of the data. This procedure of re-sampling is repeated a vast number of times (usually $m = 500$, $m = 1000$ or $m = 2000$ times). The variation observed among these re-sampled data sets is assumed to be representative how the samples may vary when being drawn from the population. This is the only assumption being made when applying the bootstrap and has been proven to be viable in many an application (see e.g. [7, 8]).

In practice, the following steps lead to a tri-PLS1 bootstrap confidence interval:

1. compute the required estimate $\hat{\theta}^{\text{sample}}$ from the original data matrices;

2. select a random number $k \leq n$ of samples from the augmented data

matrix $\begin{pmatrix} X & \mathbf{y} \end{pmatrix}$ and replace them with k randomly chosen samples from the same data matrix;

3. repeat the previous process m times, hence constructing m bootstrap data matrices X_i^{boot} and $\mathbf{y}_i^{\text{boot}}$;

4. compute m boostrap estimates of θ from these bootstrap data matrices;

5. the $100\frac{1-\alpha}{2}$th and $100\frac{\alpha}{2}$th percentiles of the empirical distribution of the bootstrap estimates $\hat{\theta}_i^{\text{boot}}$ are considered to be an $100\alpha\%$ confidence interval for the original estimate $\hat{\theta}^{\text{sample}}$.

The quality of the bootstrap estimates depends on the number of bootstrap samples that are being constructed. It is considered good practice to use m = 2000 bootstrap samples if one envisages the construction of confidence intervals [7]. Throughout this paper, the results we will report will be based on m = 2000 bootstrap samples.

4 Simulation study

In the previous section we described how to construct bootstrap confidence intervals from a given sample. The basic question which arises is

whether these confidence intervals really behave as expected, i.e. whether they cover the true value in 100α% of the cases occurring. We investigated this aspect in a simulation study, for the special case where α = .95. In this case we expect the true values of the predictand to lie within the bootstrap confidence limits for 95% of the predictions made. The simulation study was set up as follows: at first a population was created consisting of 2000 samples, which follow the model:

$$X_0 = T_3(W_3^q \boxtimes W_3^p)^T \tag{4a}$$

$$\mathbf{y}_0 = T_3\mathbf{b}_3 \tag{4b}$$

Afterwards noise was added to these data: $X = X_0 + E_X$, $\mathbf{y} = \mathbf{y}_0 + \varepsilon_\mathbf{y}$ where both error terms are taken from a standard normal with zero mean and variance equal to $\sigma^2 I$ with I being an identity matrix of appropriate dimensions.

In the simulation study concerning the bootstrap applied to PARAFAC [8], the bootstrap confidence intervals for the model parameters were investigated. In this case this would imply that we investigate the confidence intervals for T, W^q, W^p, $\mathbf{b}$ and eventually β. However, practical applications of tri-PLS1 differ somewhat from applications of PARAFAC in the sense that for tri-PLS1, the emphasis is most frequently not on estimation

of these model parameters, but on prediction. Hence, in order to set up the simulation study as realistically as possible, also here we investigated the coverage performance of the bootstrap confidence intervals for the predicted response. As a result of the same considerations, it was observed that in practical applications the number of samples available is almost always moderate, the availability of 100 samples already seldomly occurring. Envisaging to stay as close to practice as possible, we selected calibration data matrices of 40 observations from these 2000 observations, for which we computed the boostrap confidence intervals for the predicted response. We compared the confidence intervals with the known true values of **y**; if in 2 cases out of 40 this value was not covered by the respective interval, a coverage of 95% could be concluded. We repeated this process for 500 times, after which we report the average of the 500 coverage percentages observed. It is also interesting to know the "best" and "worst" case scenarios, i.e. the minimum of all observed coverage percentages as well as the maximum. It is expected that the average coverage percentages will all be close to 95% and that the maximum deviation from this number will not be excessively big. However, as a result of the small calibration matrix sizes, an undercoverage of three observations already corresponds to a coverage of 87.5%. Hence, these ranges should be seen as *indicative* of what might happen in practice, but may in fact be a somewhat pessimistic

Table I: Average coverage percentages of bootsrap confidence intervals for predicted responses of simulated data at different noise levels. The relevant dimensions of the data matrices were $p = 11$, $q = 8$ and $h = 3$.

	$\sigma_X = 0$	$\sigma_X = 0.01$	$\sigma_X = 0.1$	$\sigma_X = 0.3$	$\sigma_X = 0.8$
$\sigma_{\mathbf{y}} = 0$	94.45	94.69	94.75	94.49	94.39
$\sigma_{\mathbf{y}} = 0.01$	94.37	94.84	94.35	94.34	94.52
$\sigma_{\mathbf{y}} = 0.1$	94.03	94.3	94.32	94.12	94.97
$\sigma_{\mathbf{y}} = 0.3$	94.15	94.09	94.55	94.93	94.33
$\sigma_{\mathbf{y}} = 0.8$	94.57	94.41	94.45	94.16	94.01

estimate. The same holds for overcoverage: an overcoverage of 2 samples corresponds to 100% coverage. The average coverage percentages, on the contrary, correspond to 2000 analyzed samples and are hence a realistic estimate of the confidence intervals' performance.

Different noise levels lead to a somewhat different behaviour of the tri-PLS estimator. Hence it was decided to repeat the same simulation study for several values of σ_X and $\sigma_{\mathbf{y}}$. The results of the simulation study are summarized in Table I. From Table I we see that indeed the average coverage percentages are in all cases very close to the ideal 95 %, corroborating the assumption that the boostrap intervals can be seen as confidence intervals. The results show the same trends as observed in [8]

for PARAFAC. However, the fact that none of the observed values exceeds 95% can be interpreted as a slight downward bias. Efron and Tibshirani report a "better" bootstrap method, called the BC_a method [7], which is proposed as a bias-corrected alternative. However, in the current simulation study, the BC_a method did not alter the picture that is observed from table I but did drastically increase the bootstrap computation times and hence it was decided not to pursue this path. If we should conclude that we indeed observe some biasedness in the average coverage percentages for the bootstrap confidence intervals, we note that this bias is very small (never exceeds 1%) and that hence, the bootstrap confidence intervals can indeed be interpreted as such.

Another important aspect of a confidence interval is its length. Unacceptably long confidence intervals may also have good coverage properties (although a strong overcoverage may be expected in that case). We will not include the lengths of the individual confidence intervals into this section on the simulation study, but it will be evident from the Example in the next Section that indeed the boostrap confidence intervals for tri-PLS1 are not unacceptably long.

5 Example: application to the identification of micro-organisms by dint of the electronic nose

In the current section, we would like to illustrate the applicability of the methodology described in the previous sections to experimental data.

In a recent paper we described the identification of micro-organisms based on measurements by the electronic nose and tri-PLS1 calibration [6]. The uncertainty of the predictions plays an important rôle throughout the paper: it enables to eliminate erroneous predictions. The estimate of uncertainty used in this paper is the Faber and Bro estimate [4].

The experiment comprised the identification of 10 types of micro-organisms, whose gaseous excrements can be considered to be a "fingerprint". These excrements were led over ten sensors; the measurement was repeated at sixty time intervals of a second. Hence, the data have an implicit three-way structure, their dimensionality being $n \times 60 \times 10$. Apart from the cultures of the ten micro-organisms in question, also measurements were included which correspond to no micro-organism at all. For each of the eleven classes of micro-organisms (the eleventh corresponding to the absence of micro-organisms), a binary response had to be predicted: if the predicted value was close to one, it was concluded that the micro-organism corre-

sponding to that class was present in the sample. It shall be clear that the uncertainty of the predictions made significantly influences which conclusion to infer from these results. The same design as in Reference [6] has been used for the calibration data matrix, such that it contained 105 samples, approximately corresponding to 10 samples of each class (less samples were available for class 9 due to experimental considerations).

A separate, unknown, validation set, containing for each class two samples which correspond to that class, had to be predicted separately. The percentage of samples misclassified in this validation set was considered to be a measure of the performance of tri-PLS1 regression for this type of application; a misclassification exceeding 5% is considered unacceptable in clinical practice. In the current example, we will show the bootstrap confidence intervals for the predicted values corresponding to classes 1 and 5.

For class one, an estimate of h was obtained by means of random cross validation (200 iterations) from which it was decided that the optimal model dimensionality was equal to nine. In Table II, the predicted values for class 1 based on 9 latent variables are given, as well as the 95% Faber and Bro confidence intervals thereof and the respective 95% bootstrap confidence intervals. From Table II it can be seen that the bootstrap confidence intervals are similar to those obtained by a local linearization. In particular the

Table II: Predicted values for class 1, pseudo confidence intervals based on a local linearization (3rd column) and 95% bootstrap confidence intervals (4th column). The true classes to which the observations belong are given in comparison in the first column. In the last row, the average lengths of both intervals over these 22 cases are given.

True class	$\hat{\mathbf{y}}_{\text{val},9}^{\text{class 1}}$	PI (LL)	PI (Bootstrap)
7	0.11	[0.00,0.21]	[0.03,0.16]
2	-0.17	[-0.28,-0.06]	[-0.23,-0.02]
4	0.17	[0.07,0.28]	[0.04,0.28]
3	-0.01	[-0.14,0.12]	[-0.22,0.37]
11	-0.01	[-0.11,0.10]	[-0.03,0.04]
10	0.21	[0.11,0.32]	[0.09,0.30]
6	-0.15	[-0.26,-0.04]	[-0.20,-0.05]
8	-0.02	[-0.13,0.08]	[-0.06,0.04]
5	0.07	[-0.03,0.18]	[-0.05,0.16]
1	1.37	[1.25,1.49]	[1.09,1.66]
9	0.01	[-0.10,0.12]	[-0.03,0.04]
4	0.31	[0.20,0.42]	[0.11,0.49]
6	-0.05	[-0.16,0.06]	[-0.12,0.07]
2	-0.11	[-0.22,0.00]	[-0.19,0.02]
1	1.00	[0.89,1.11]	[0.77,1.14]
7	0.05	[-0.05,0.16]	[-0.02,0.11]
3	0.02	[-0.10,0.14]	[-0.16,0.22]
11	-0.01	[-0.12,0.10]	[-0.04,0.03]
5	-0.03	[-0.14,0.07]	[-0.13,0.07]
9	0.04	[-0.07,0.15]	[0.00, 0.07]
10	0.19	[0.09,0.30]	[0.08,0.29]
8	-0.01	[-0.12,0.10]	[-0.07,0.07]
Average length		0.22	0.22

average lengths of the intervals over all 22 intervals for the validation set, are identical. The performance of the Faber and Bro intervals is striking: in [5] the authors explicitly write that the intervals obtained by their approach cannot be seen in true confidence intervals in the statistical sense, but that they yield “almost superimposed” confidence bounds. Nevertheless, the Faber and Bro pseudo confidence intervals seem to perform as well as the 95% bootstrap confidence intervals do.

Howbeit, some differences between both methods can be observed. The bootstrap prediction intervals show a higher degree of specificity to the sample. In the cases where the predicted value is close to zero, the bootstrap confidence intervals are almost without exception smaller than the estimates based on a local linearization (see e.g. both predictions of samples belonging to class 11). This is easily understandable if one considers the fact that the sole factor being specific to the sample in the computation of the Faber and Bro estimate is the sample leverage, which is also a (non-robust) indication of outlyingness of the sample. One can expect these sample leverages to be high when a sample is present which is significantly outlying with respect to the model estimated by tri-PLS1. The latter will be the case in the analysis for class 5, where based on the high prediction error an erroneous prediction can be eliminated. However, the Faber and Bro estimate is insensitive to small differences between ob-

servations. Nontheless, e.g. the smaller prediction error for the samples corresponding to class 11 can be understood, as the measurements of blank nutrients are less prone to experimental variability than measurements of samples containing micro-organisms, albeit this explication does not hold for all the samples where a discrepancy is observed between both measures of prediction error. Due to the similarity of both types of confidence intervals, both confidence intervals lead to the conclusion that observations 10 and 15 belong to class 1, which is correct.

Let us now pass on to the predicted responses for class 5. For this class, the data can less well be modelled by tri-PLS 1, which already reflects in the estimate of $\hat{h} = 14$. In Table III the predicted responses for class 5 based on 14 latent variables are given, together with their 95% confidence intervals based on a local linearization and the bootstrap.

As is explained in [6], at first sight one would conclude from these predicted responses that three observations belong to this class. The confidence interval for observation 4, however, includes both 0 and 1, from which we can infer that this observation does not belong to class 5 and that the only observations belonging to this class are observations 9 and 20, which is correct.

Again the confidence intervals obtained by the bootstrap are very similar to those obtained by a local linearization of the tri-PLS1 estimator. The

Table III: Predicted values for class 5, pseudo confidence intervals based on a local linearization (3rd column) and 95 % bootstrap confidence intervals (4th column). The true classes to which these observations belong are given in comprison in the first column. In the last row, the average lengths of both intervals over these 22 cases are given.

True class	$\hat{\mathbf{y}}_{val,14}^{class\ 5}$	PI (LL)	PI (Bootstrap)
7	0.21	[0.02,0.41]	[-0.04,0.35]
2	0.11	[-0.09,0.31]	[-0.17,0.20]
4	0.16	[-0.10,0.42]	[-0.01,0.42]
3	0.58	[-0.33,1.49]	[-0.31,2.49]
11	-0.07	[-0.24,0.09]	[-0.10,0.03]
10	0.26	[0.05,0.47]	[0.01,0.49]
6	0.39	[0.16,0.62]	[0.22,0.70]
8	0.18	[0.01,0.35]	[-0.06,0.24]
5	0.77	[0.54,0.99]	[0.66,1.08]
1	0.10	[-0.30 ,0.49]	[-0.62,0.77]
9	0.12	[-0.04,0.28]	[-0.26,0.16]
4	0.17	[-0.05,0.40]	[0.04,0.62]
6	0.21	[-0.01,0.42]	[0.02,0.50]
2	0.13	[-0.08,0.34]	[-0.05,0.29]
1	-0.40	[-0.67,-0.12]	[-0.70,0.00]
7	0.18	[0.00,0.37]	[0.05,0.35]
3	0.27	[-0.39,0.92]	[-0.21,1.28]
11	-0.06	[-0.22,0.10]	[-0.09,0.04]
5	0.65	[0.41,0.89]	[0.35,1.01]
9	0.14	[-0.03,0.30]	[0.03,0.18]
10	0.22	[-0.01,0.46]	[0.04,0.47]
8	0.17	[-0.01,0.35]	[-0.06,0.26]
Average lengths		0.53	0.59

bootstrap confidence intervals show one remarkable feature. The very high prediction error for observation 4, when computed by the Faber and Bro estimate, is caused by the fact that observation 4 has a high leverage with respect to the model. As the leverages are an important factor in the Faber and Bro estimate (the other being the mean-squared error of calibration), it can be expected that high leverage points have a high prediction error. However, for the bootstrap no direct “factor” which accounts for leverage can be detected in the process of its computation. Nevertheless, the bootstrap confidence intervals are also very long for this observation, indicating that the bootstrap *implicitly* takes into account these leverages. This can be easily understood as the quality of the prediction for these high leverage points significantly deteriorates if in the bootstrap data matrices the few samples which are closest to these leverage points, are replaced by other samples. Hence one can expect that prediction of high leverage points based on bootstrap data matrices will have a much higher variability than prediction of normal observations. Notwithstanding that small differences between both methods can be observed, both estimates lead to the same conclusions and hence to the same misclassification probability.

6 Conclusions

In the current paper, it has been investigated whether the bootstrap can be applied in order to obtain confidence intervals for the predictions made by tri-PLS1. At first, from a theoretical point of view, there is no reason to believe why this should not be the case. Secondly, an extensive simulation study has shown that the bootstrap confidence intervals indeed have desirable coverage properties, i.e. that an $100\alpha\%$ bootstrap confidence interval indeed covers the true value in about 95 % of all cases. Moreover, the bootstrap is also user-friendly, as thanks to the optimal tri-PLS1 algorithm [10], bootstrap confidence intervals can be obtained in a short time span. For example, the computation of the bootstrap confidence intervals from the example, i.e. for $X \in I\!R^{105\times600}$, took 33 seconds on a Pentium 1.4 GHz processor.

In the example, bootstrap confidence intervals have been compared to their counterpart obtained by a user-friendly expression based on a local linearization [4]. For this example, it had already been illustrated that linear approximation was successful [6]. From the results given in the Example, it can be seen that the bootstrap confidence intervals perform in no sense inferior to their counterparts obtained by a local linearization and that indeed bootsrap confidence intervals can be used for this application as well.

Moreover, a positive property of the bootstrap confidence intervals is their higher degree of specificity to the sample; it has been observed that an interpretation can be given for the lengths of some bootstrap confidence intervals based on experimental reflections.

To conclude, note that the bootstrap confidence intervals as proposed here are prone to very few model assumptions (the use of percentile intervals gives the method some non-parametric flavour), so that they can be used for a wide range of applications.

Acknowledgements

Research financed by a PhD grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The authors would also like to thank Maarten Moens and Frank Blockhuys for the data in the Example section, as well as an anonymous referee for some very helpful comments.

References

[1] H. Wold, in: P.R. Krishnaiah (ed.), Multivariate Analysis III, Academic Press, New York (1973), 383-407.

[2] L. Ståhle, Chemometr. Intell. Lab. Syst., 7 (1989), 95-100.

[3] R. Bro, Multiway analysis in the food industry, PhD Thesis, University of Amsterdam, Amsterdam, The Netherlands, 1996.

[4] N.M. Faber and R. Bro, Chemometr. Intell. Lab. Syst., 61 (2003), 133-149.

[5] R. Bro, Å. Rinnan and N.M. Faber, Chemometr. Intell. Lab. Syst., 75 (2005), 69-76.

[6] S. Serneels, M. Moens, P.J. Van Espen, F. Blockhuys, Anal. Chim. Acta, 516 (2004), 1-5.

[7] B. Efron, R.J. Tibshirani, An introduction to the bootstrap, Chapman and Hall, New York, 1993.

[8] H.A.L. Kiers, J. Chemometr., 18 (2004), 22-36.

[9] R. Bro, J. Chemometr., 10 (1996), 47-61.

[10] S. de Jong, J. Chemometr., 12 (1998), 77-81.

[11] K. Baumann, H. Albert and M. von Korff, J. Chemometr. 16 (2002), 339-350.