Robust estimation in structural equation models using Bregman divergences

S. Penev[1] T. Prvan[2]

(Received 29 October 2012; revised 5 August 2013)

Abstract

Structural equation models seek to find causal relationships between latent variables by analysing the mean and the covariance matrix of some observable indicators of the latent variables. Under a multivariate normality assumption on the distribution of the latent variables and of the errors, maximum likelihood estimators are asymptotically efficient. The estimators are significantly influenced by violation of the normality assumption and hence there is a need to robustify the inference procedures. We propose to minimise the Bregman divergence or its variant, the total Bregman divergence, between a robust estimator of the covariance matrix and the model covariance matrix, with respect to the parameters of interest. Our approach to robustification is different from the standard approaches in that we propose to achieve

http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/6306 gives this article, © Austral. Mathematical Soc. 2013. Published September 23, 2013, as part of the Proceedings of the 16th Biennial Computational Techniques and Applications Conference. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

the robustification on two levels: firstly, choosing a robust estimator of the covariance matrix; and secondly, using a robust divergence measure between the model covariance matrix and its robust estimator. We focus on the (total) von Neumann divergence, a particular Bregman divergence, to estimate the parameters of the structural equation model. Our approach is tested in a simulation study and shows significant advantages in estimating the model parameters in contaminated data sets and seems to perform better than other well known robust inference approaches in structural equation models.
Subject class: 62G05, 62G35
Keywords: Total Bregman distance, Structural equation/model

Contents

1 Introduction **C575**

2 Robust estimation using Bregman divergences **C577**

3 Simulation study **C580**

4 Conclusion **C583**

References **C588**

1 Introduction

Structural equation modelling is a popular technique used to quantify the causal relationships between hidden (latent) and observable variables. A typical structural equation model (SEM) has a structural equation, the equation relating the latent exogenous (independent) and endogenous (dependent) variables, and a set of linear equations which form a matrix equation, relating the latent variables to both the exogenous and endogenous observed variables.

Within both the structural and matrix equations there are parameters that are unknown and must be estimated to obtain some quantitative relationship. The method used to estimate these parameters cannot rely on a process such as least squares estimation (as used in regression modelling) since the values of the latent variables are not known for any given observable output vector. In our new method, we minimise the distance between the model covariance matrix $\Sigma(\theta)$ and a robust estimated covariance matrix of the data U to obtain estimates of the model parameters of interest.

Under multivariate normality assumptions on the endogenous variables, inference procedures and their properties are well known [2, e.g.]. The traditional maximum likelihood (ML) divergence (discrepancy function) used in covariance structure modelling is (up to an additive constant)

$$F(\theta) = \log|\Sigma(\theta)| + \operatorname{tr}\left[S\Sigma^{-1}(\theta)\right] - \log|S|\,, \tag{1}$$

where tr is the trace, the estimated covariance matrix of the observed data is the sample covariance matrix S, the model covariance matrix is $\Sigma(\theta)$, and $|\Sigma|$ is the determinant of Σ.

In recent years, attention focused on non-normal data. It was found that such data often appears in the behavioural sciences. In addition, the SEM methodology found applications in areas such as risk and insurance [1, e.g.]. In these applications, non-normal data is often the rule rather than an exception. Non-normality, if not properly taken into account, may have a profound effect on inference.

Yuan, Bentler and others [7, 8, 9, and references therein] studied the drastic effects of non-normal data on the ML fit in terms of biasedness and inefficiency. The problem was revisited by Zhong and Yuan [10] who performed a thorough simulation study of the effects of outliers, leverage points (and the combination of both) on the structural equation fit.

All the robust procedures suggested until now have the logical disadvantage that the data is weighted depending on how likely the data is to be 'from the model' rather than 'contaminated'. They present fits to a robust variant of

covariance matrices but do not sufficiently modify the divergence criterion once the robust covariance matrix is chosen. In our view, the robust fit should necessarily be a two step procedure. In the first step, there is a need to find a robust estimator of the covariance matrix of the endogenous variables. Then, in the second step, a robust divergence measure must be chosen when accomplishing the minimum distance fitting procedure.

The two step philosophy was also advocated by Yuan and Bentler [7]. However, they proceeded with a Wishart-based likelihood after the robust covariance matrix was substituted. In contrast, we propose an alternative and intrinsically more robust divergence function in the second step for cases where we believe the multivariate normality assumption is violated. This violation takes many different forms. For illustrative purposes, we choose a mixture of two multivariate normal distributions.

We present results from a simulation study for a standard SEM with two latent exogenous variables, one latent endogenous variable and six observable variables with different levels of contamination in the data. For the two types of Bregman distances considered, our approach works best for high contamination, whereas the traditional approach referred to above only performs well for data that is not contaminated.

2 Robust estimation using Bregman divergences

Our new divergence measure, which is a particular Bregman divergence, is now outlined. We minimise the von Neumann divergence (vND) between the matrices U and $\Sigma(\theta)$ over the set of parameter values θ, where U is a robust estimator of the true covariance matrix Σ. We also consider a variant of the procedure where we minimise the (symmetrised) total von Neumann divergence (TvND) between U and $\Sigma(\theta)$. Applying either of the two approaches delivers very similar outcomes.

Dhillon and Tropp [3] provide a thorough discussion of the use of Bregman divergences between matrices. Given a differentiable, strictly convex function ϕ defined on the set of symmetric positive definite matrices, we define the Bregman divergence of the matrix X from the matrix Y as

$$\tilde{d}(X,Y) = \phi(X) - \phi(Y) - \operatorname{tr}\left[\nabla\phi(Y)^T(X-Y)\right].$$

The vND is a particular Bregman divergence which arises when we choose $\phi(X) = \operatorname{tr}(X\log X - X)$. Here $\log X = U(\operatorname{diag}\log D)U^T$ where $X = U(\operatorname{diag} D)U^T$ is the spectral decomposition of X. Since $\nabla\phi(X) = \log X$, we arrive at the 'ordinary' vND functional

$$\tilde{d}(X,Y) = \operatorname{tr}\left[X(\log X - \log Y) - X + Y\right]. \tag{2}$$

The divergence functional is non-symmetric and therefore the relative nearness of the matrices differs depending on their order as arguments in the functional. Inspired by the definition (2), we define the TvND associated with real symmetric positive definite matrices X and Y, as

$$d(X,Y) = \frac{\operatorname{tr}\left[X(\log X - \log Y) - X + Y\right]}{\sqrt{1 + \operatorname{tr}(\log Y \log Y)}}. \tag{3}$$

The normalisation in the denominator causes the difference between the usual vND and the TvND. Vemuri et al. [5] suggested similar normalisations, based on heuristic discussions, for vector divergence measures. There is some advantage in using TvND rather than vND in the sense that the divergence (3) is invariant under rotations of the coordinate system whereas (2) is not [5]. The minimisation of vND and TvND will deliver equivalent results if the matrix U is the Y entry in $d(X,Y)$ and $\tilde{d}(X,Y)$, and the matrix $\Sigma(\theta)$ is the X entry. However, if the matrix U is the X entry, the outcomes of the two minimisations processes will be different. Nielsen and Boltz [4] reviewed several ways to symmetrise the TvND. One simple and attractive way is to take the average $(1/2)[\tilde{d}(X,Y) + \tilde{d}(Y,X)]$. We call this the symmetrised TvND (STvND).

Our aim is to demonstrate the robustness of estimators that arise from the minimisation of these divergence measures in applications in the context of SEM fitting. Ultimately, the theoretical challenge is to show that the resulting influence function is bounded for the Bregman divergence-based estimators, unlike the unbounded maximum likelihood estimators. Vemuri et al. [5] showed such robustness properties for a specific measure of centre (the so-called 't-center') when the TvND is used as a divergence measure. However, their approach cannot be used in the context of SEM, and we have not yet provided a general theoretical result. This article presents a set of systematic simulation results that are indicative of the robustness of the Bregman divergence-based estimators.

The sample covariance matrix S is the maximum likelihood estimator of the observed data under the assumption the errors are independently and identically normally distributed. To fit a robust covariance matrix to the data, we suggest the minimum covariance determinant method (MCD) as a highly robust procedure, with well known asymptotic properties. The MCD method utilises the subset which is $h \times 100\%$ of the n observations, with $h > 0.5$, and which has the sample covariance matrix with the smallest possible determinant. The MCD scale estimate is the average of these observations and the MCD estimate of scatter is their empirical covariance matrix, multiplied by a consistency factor.

Our simulation study demonstrates the advantages of the fits obtained using the two stage procedure. The minimisation of (2), (3) or $(1/2)[\tilde{d}(X, Y) + \tilde{d}(Y, X)]$ delivers the minimum divergence within the space of functional values, between the variable matrix $\Sigma(\theta)$ and the robust estimator U of the 'true' covariance matrix Σ, with constraints on the entries of the $\Sigma(\theta)$ matrix. This is opposed to the maximum likelihood estimation procedure where the goal is to minimise the functional in (1) to determine the maximisation of the likelihood function.

Vemuri et al. [5] demonstrated that the Bregman divergences tend to show robustification in minimisation problems involving matrices, and thus these

divergence measures are desirable for this study.

3 Simulation study

For this study we applied the Matlab function `fmincon` (constrained functional minimum) to the variance-covariance matrix and the population matrix within the three important functionals: the maximum likelihood functional, the standard vND and the TvND. Random data was generated for the model and parameters estimated using each functional. The contamination level ranged from 0% up to 30%.

The structural equation model considered is a standard model with two latent exogenous variables ξ_1 and ξ_2, one latent endogenous variable η_1, and six observable variables Y_i for $i = 1, 2, \ldots, 6$. In terms of linear equations the model is

$$\begin{aligned} \eta_1 &= \gamma_1\xi_1 + \gamma_2\xi_2 + \zeta_1\,, \quad Y_1 = \eta_1 + \epsilon_1\,, \quad Y_2 = \lambda_{21}\eta_1 + \epsilon_2\,, \quad Y_3 = \xi_1 + \epsilon_3\,, \\ Y_4 &= \lambda_{42}\xi_1 + \epsilon_4\,, \quad Y_5 = \xi_2 + \epsilon_5\,, \quad Y_6 = \lambda_{63} + \epsilon_6\,, \end{aligned} \tag{4}$$

where γ_1, γ_2, λ_{21}, λ_{42} and λ_{63} are model parameters,

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \phi_{12} \\ \phi_{12} & 1 \end{pmatrix}\right],$$

$\zeta_1 \sim N(0, \psi)$ and $\epsilon_i \sim N(0, 1)$ for $i = 1, 2, \ldots, 6$, with $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$. Two restrictions arise from this model: $|\phi_{12}| \leqslant 1$ and $\psi \geqslant 0$. In a condensed form, equation (4) is

$$\mathbf{y} = \Lambda \begin{pmatrix} \eta_1 \\ \xi_1 \\ \xi_2 \end{pmatrix} + \boldsymbol{\epsilon}\,, \tag{5}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_6)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_6)^T$ and

$$\Lambda = \begin{pmatrix} 1 & \lambda_{21} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_{42} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_{63} \end{pmatrix}^T .$$

The elements λ_{11}, λ_{32} and λ_{53} of the matrix Λ were set to 1 as a standard approach to guarantee identification of the model. The variance-covariance matrix of this model is

$$\Sigma(\theta) = \Lambda \Phi \Lambda^T + I_6 ,$$

where Φ is the variance-covariance matrix of the vector of latent variables $(\eta_1, \xi_1, \xi_2)^T$. Simple algebra gives

$$\Phi = \begin{pmatrix} \gamma_1^2 + \gamma_2^2 + 2\gamma_1\gamma_2\phi_{12} + \psi & \gamma_1 + \gamma_2\phi_{12} & \gamma_2 + \gamma_1\phi_{12} \\ \gamma_1 + \gamma_2\phi_{12} & 1 & \phi_{12} \\ \gamma_2 + \gamma_1\phi_{12} & \phi_{12} & 1 \end{pmatrix} . \qquad (6)$$

The unknown parameters are $\theta = (\lambda_{21}, \lambda_{42}, \lambda_{63}, \gamma_1, \gamma_2, \phi_{12}, \psi)^T$. We doubled the sample sizes from $n = 100$ to $n = 200$ and then to $n = 400$ and performed 100 simulations for a particular level of contamination, ranging from none (0%) to 30%, for the given values of θ.

We contaminated the data by changing the variance of the vector $\boldsymbol{\epsilon}$ from 1 to 7^2 for a certain percentage of data. These variances are equivalent to a normal mixture of $N(0, 1)$ or $N(0, 49)$, respectively, for each of the six independent components of the vector $\boldsymbol{\epsilon}$. This allows for a more comfortable interpretation of the $\boldsymbol{\epsilon}$ components as noise since the mean is preserved at zero but the variance is inflated. We experimented with other types of contamination, such as applying a mean shift to the assumed values of $\boldsymbol{\epsilon}$ for a certain percentage of data. The results were similar for the mean shift model. In each trial run `fmincon` requires a starting input value for minimisation. We used the assumed parameter values as the starting values to remove any chance of convergence to another local minimum of the vND and TvND functionals. We ran 100 trials for the contamination levels 0%, 5%, 10%, 20% and 30%.

The Matlab function MCD [11] was used to obtain robust estimates of the covariance matrix. We used the default of $h = 0.75$ for the proportion of observations used to obtain an estimate of the covariance matrix. Verboven and Hubert [6] provide details about the implementation of the Matlab function MCD.

The random data was generated from parameters $\theta = (2, 3, 3, 2, 0.8, -0.3, 1.5)^T$. We used

$$\sum_{i=1}^{100} (\theta_j - \hat{\theta}_{ij})^2 / 100 \,,$$

as a measure of the model fit for each parameter ($j = 1, \ldots, 7$) where $\hat{\theta}_{ij}$ denotes the estimate of θ_j at the ith iteration. Tables 1–5 present the fit results from:

- MLE*, the standard MLE (ML estimate) which minimises the functional (1);
- MLE, implementing minimisation of the functional in the form (1) but with S replaced by the robust MCD-based estimator U with $h = 0.75$;
- TvND, implementing minimisation of the functional in (3) with first argument U and second argument $\Sigma(\theta)$ (so results of vND and TvND fit are different);
- STvND fit;
- vND fit with first argument $\Sigma(\theta)$ and second argument U.

When there was no contamination the estimators obtained from the traditional ML divergence using S performed best in our simulations. When there was contamination, even when sample sizes were large, the traditional ML divergence using S usually performed worst. For contaminated data, the quality of the estimators obtained with the traditional ML divergence using S started deteriorating at a contamination of only 5%, with drastic deterioration at higher levels of contamination. Most of the time, the traditional ML divergence using U was competitive with vND and STvND up to 20% contamination.

Table 1: Error estimates for 0% contamination

	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ
$n = 400$							
MLE	0.0036	0.016	0.018	0.011	0.0087	0.0034	0.053
MLE*	0.0027	0.010	0.009	0.009	0.0064	0.0025	0.029
TVND	0.0036	0.024	0.026	0.016	0.0084	0.0032	0.050
STVND	0.0036	0.024	0.026	0.016	0.0085	0.0032	0.051
VND	0.0036	0.025	0.027	0.016	0.0085	0.0032	0.052
$n = 200$							
MLE	0.0062	0.030	0.028	0.030	0.024	0.0061	0.082
MLE*	0.0051	0.018	0.019	0.023	0.016	0.0048	0.052
TVND	0.0063	0.041	0.042	0.035	0.023	0.0059	0.077
STVND	0.0063	0.042	0.043	0.034	0.023	0.0060	0.080
VND	0.0063	0.043	0.044	0.034	0.023	0.0060	0.083
$n = 100$							
MLE	0.015	0.074	0.072	0.069	0.041	0.014	0.19
MLE*	0.011	0.052	0.037	0.050	0.033	0.009	0.11
TVND	0.016	0.087	0.089	0.071	0.043	0.013	0.18
STVND	0.016	0.089	0.092	0.070	0.043	0.013	0.19
VND	0.016	0.092	0.095	0.070	0.043	0.013	0.20

However, for 30% contamination VND usually performed best, very closely followed by STVND. The TVND did not perform as well as expected.

4 Conclusion

When there is contamination in the data, using the traditional ML divergence measure to estimate the parameters does not deliver a good fit. This is not surprising as this measure is particularly tailored for high efficiency under the

Table 2: Error estimates for 5% contamination

	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ
$n = 400$							
MLE	0.0031	0.017	0.020	0.011	0.0090	0.0029	0.052
MLE*	0.0178	0.102	0.127	0.128	0.0656	0.0081	1.840
TvND	0.0032	0.022	0.026	0.014	0.0092	0.0029	0.051
STvND	0.0032	0.022	0.027	0.014	0.0092	0.0029	0.052
vND	0.0032	0.022	0.027	0.014	0.0092	0.0029	0.053
n=200							
MLE	0.0087	0.027	0.029	0.034	0.026	0.0075	0.079
MLE*	0.0352	0.181	0.164	0.189	0.094	0.0152	2.303
TvND	0.0088	0.036	0.038	0.036	0.026	0.0077	0.075
STvND	0.0088	0.037	0.038	0.036	0.026	0.0077	0.077
vND	0.0089	0.038	0.039	0.036	0.026	0.0077	0.080
n=100							
MLE	0.014	0.066	0.065	0.061	0.048	0.013	0.18
MLE*	0.066	0.537	0.310	0.238	0.167	0.018	2.77
TvND	0.013	0.076	0.080	0.068	0.051	0.012	0.17
STvND	0.013	0.077	0.082	0.067	0.051	0.012	0.17
vND	0.013	0.079	0.084	0.067	0.051	0.012	0.18

ideal non-contaminated model. Hence it trades off robustness for achieving high efficiency when the endogenous variables are multivariate normal distributed with zero mean. The simulated data represents a *mixture* of two normal distributions and as such violates the normality assumption. However, this violation does not have a detrimental effect on our proposed robust estimators. When replacing S with the robust estimator U of the covariance matrix in the divergence measure, a better performance was achieved. Yet this new divergence measure was not the best divergence measure when more than 20% of the data was contaminated. The 'ordinary' vND, which delivers equivalent results to the TvND if U is the second entry in the func-

Table 3: Error estimates for 10% contamination

	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ
$n = 400$							
MLE	0.0034	0.013	0.013	0.012	0.0080	0.0029	0.044
MLE*	0.0346	0.412	0.416	0.285	0.1423	0.0170	6.121
TvND	0.0034	0.016	0.018	0.014	0.0078	0.0029	0.044
STvND	0.0034	0.016	0.019	0.014	0.0078	0.0029	0.044
vND	0.0034	0.016	0.019	0.014	0.0079	0.0029	0.045
$n = 200$							
MLE	0.0067	0.029	0.029	0.030	0.022	0.0057	0.060
MLE*	0.0775	0.426	0.515	0.332	0.157	0.0240	5.956
TvND	0.0067	0.035	0.034	0.031	0.021	0.0056	0.058
STvND	0.0067	0.036	0.035	0.031	0.021	0.0056	0.059
vND	0.0067	0.037	0.035	0.031	0.022	0.0056	0.061
$n = 100$							
MLE	0.011	0.069	0.047	0.062	0.039	0.015	0.16
MLE*	0.130	0.760	0.759	0.397	0.252	0.024	8.53
TvND	0.011	0.081	0.058	0.071	0.042	0.014	0.16
STvND	0.011	0.082	0.060	0.071	0.042	0.015	0.17
vND	0.011	0.083	0.061	0.071	0.043	0.015	0.17

tional, gradually improved its performance with increasing contamination and performed best compared to other divergences at 30% contamination. The TvND minimisation also performed exceptionally well and almost matched the performance of the vND. It is difficult to pinpoint a clear winner between these two methods. For a contamination of 30%, it seems that the standard vND is slightly better, with no such advantage demonstrated at lower levels of contamination.

The use of Bregman divergences for robust estimation in structural equation models is in its infancy and there are many avenues left unstudied and questions unanswered that deserve future attention. For example, other robust

Table 4: Error estimates for 20% contamination

	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ
$n = 400$							
MLE	0.0036	0.013	0.017	0.014	0.0077	0.0030	0.045
MLE*	0.0797	0.958	1.068	0.696	0.2418	0.0307	17.45
TvND	0.0037	0.015	0.019	0.015	0.0083	0.0028	0.047
STvND	0.0037	0.015	0.019	0.015	0.0082	0.0028	0.047
vND	0.0037	0.016	0.020	0.015	0.0082	0.0028	0.046
$n = 200$							
MLE	0.0061	0.028	0.025	0.033	0.023	0.0059	0.074
MLE*	0.1970	1.753	1.445	0.687	0.257	0.0360	19.39
TvND	0.0061	0.031	0.031	0.032	0.026	0.0059	0.074
STvND	0.0061	0.031	0.032	0.032	0.026	0.0059	0.074
vND	0.0061	0.031	0.032	0.032	0.026	0.0060	0.075
$n = 100$							
MLE	0.015	0.058	0.050	0.052	0.047	0.015	0.16
MLE*	0.778	2.695	2.391	0.756	0.347	0.051	17.3
TvND	0.016	0.078	0.066	0.065	0.048	0.015	0.16
STvND	0.016	0.079	0.066	0.065	0.048	0.015	0.16
vND	0.016	0.079	0.067	0.065	0.048	0.015	0.17

divergences could be used instead of the vND. Burg's divergence, a Bregman divergence with $\phi(X) = -\log\det(X)$, is another promising alternative. Dhillon and Tropp [3] listed several other divergences.

The robust approach presented here can be extended to more general models. We are very optimistic about its applicability to robust inference about mean and covariance structure models, which are extensions of the standard covariance structure models. The difference between the models is that the mean vector of the observations is assumed to be a nuisance parameter in covariance structure models but is modelled in mean and covariance structure models. Yuan and Bentler [8] discuss robust inference in mean and covariance

Table 5: Error estimates for 30% contamination

	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ
$n = 400$							
MLE	0.019	0.23	0.25	0.16	0.062	0.010	3.7
MLE*	0.130	2.43	2.08	0.94	0.305	0.038	27
TvND	0.019	0.40	0.54	0.13	0.054	0.0084	4.3
STvND	0.019	0.37	0.50	0.12	0.052	0.0082	3.9
vND	0.019	0.33	0.45	0.11	0.050	0.0081	3.5
$n = 200$							
MLE	0.028	0.26	0.25	0.15	0.063	0.015	3.0
MLE*	0.306	2.84	2.75	0.89	0.313	0.042	40
TvND	0.026	0.40	0.41	0.11	0.058	0.013	3.4
STvND	0.025	0.36	0.37	0.11	0.057	0.013	3.0
vND	0.024	0.32	0.34	0.10	0.056	0.013	2.6
$n = 100$							
MLE	0.074	0.33	0.30	0.21	0.15	0.021	1.9
MLE*	4.196	4.70	5.04	1.24	0.49	0.048	38
TvND	0.070	0.32	0.32	0.18	0.14	0.019	2.3
STvND	0.067	0.28	0.28	0.17	0.13	0.019	1.9
vND	0.064	0.24	0.25	0.15	0.13	0.018	1.5

structure models and suggest a two stage approach but do not apply Bregman divergences in the estimation at any stage.

Finally, more should be said about the asymptotic properties of our estimators via investigation of their influence function. This is a future avenue of research.

References

[1] E. G. Baranoff, S. Papadopoulos and T. W. Seger. Capital and Risk Revisited: A Structural Equation Model Approach for Life Insurers. *Journal of Risk and Insurance,* 74:653–681, 2007. doi:10.1111/j.1539-6975.2007.00229.x C576

[2] K. Bollen. *Structural Equations with Latent Variables*. Wiley, New York, 1989. C576

[3] I. S. Dhillon and J. A. Tropp. Matrix Nearness Problems with Bregman Divergences. *SIAM Journal on Matrix Analysis and Applications*, 29:1120–1146, 2008. doi:10.1137/060649021 C578, C586

[4] F. Nielsen and S. Boltz. The Burbea–Rao and Bhattacharyya Centroids. *IEEE transactions in information theory*, 57:5455–5466, 2011. doi:10.1109/TIT.2011.2159046 C578

[5] B. C. Vemuri, M. Liu, S.-I. Amari and F. Nielsen. Total Bregman Divergence and Its Applications to DTI Analysis. *IEEE Transactions on medical imaging*, 30:475–483, 2011. doi:10.1109/TMI.2010.2086464 C578, C579

[6] S. Verboven and M. Hubert. LIBRA: a MATLAB library for robust analysis, *Chemometrics and intelligent laboratory systems*, 75:127–136, 2005. doi:10.1016/j.chemolab.2004.06.003 C582

[7] K.-H. Yuan and P. M. Bentler. Structural equation modeling with robust covariances. *Sociological Methodology*, 28:363–396, 1998. doi:10.1111/0081-1750.00052 C576, C577

[8] K.-H. Yuan and P. M. Bentler. Robust mean and covariance structure analysis, *British Journal of Mathematical and Statistical Psychology*, 51:63–88, 1998. doi:10.1111/j.2044-8317.1998.tb00667.x C576, C586

[9] K.-H. Yuan, P. M. Bentler and W. Chan. Structural Equation Modeling with Heavy Tailed Distributions, *Psychometrika*, 69:421–436, 2004. doi:10.1007/BF02295644 C576

[10] X. Zhong and K.-H. Yuan. Bias and Efficiency in Structural Equation Modeling: Maximum Likelihood Versus Robust Methods. *Multivariate Behavioral Research,* 46:229–265, 2011. doi:10.1080/00273171.2011.558736 C576

[11] LIBRA: a Matlab Library for Robust Analysis. http://wis.kuleuven.be/stat/robust/LIBRA/LIBRA-home C582

Author addresses

1. **S. Penev**, School of Mathematics and Statistics, UNSW, Sydney NSW 2052, Australia.
mailto:s.penev@unsw.edu.au

2. **T. Prvan**, Department of Statistics, Macquarie University, Sydney, NSW 2109, Australia