

Robust estimation in structural equation models using Bregman and other divergences with t-centre approach to estimate the covariance matrix

S. Penev¹ T. Prvan²

(Received 27 February 2015; revised 20 January 2016)

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. Previous work minimized the von Neuman divergence or its variant the total von Neumann divergence to estimate the parameters,

<http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/9359> gives this article, © Austral. Mathematical Soc. 2016. Published February 8, 2016, as part of the Proceedings of the 17th 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.

with the minimum covariance determinant used as a robust estimator of the covariance matrix. We extend this approach by considering other divergences and by developing a robust estimate of the covariance matrix. The robust estimator of the covariance matrix developed is a t-centre like estimator based on several minimum covariance determinant estimators ranging from 0% contamination to 50% contamination. The simulation results are promising. The results can be used for robustifying the fit of structural equation models.

Contents

1 Introduction	C340
2 New two step procedure	C341
3 Simulation study	C345
4 Conclusion	C352
References	C353

1 Introduction

Structural equation modelling is a popular technique used to quantify relationships among variables where some of the variables are hidden (latent) and some are observed (measured). A typical structural equation model has a structural equation relating the exogenous (independent) and endogenous (dependent) variables, and a set of linear equations which are grouped together to form a matrix equation, relating the latent variables to both the exogenous and endogenous observed values. The unknown parameters within the equations must be estimated to obtain some quantitative relationship.

A process such as least squares estimation cannot be used to estimate these parameters because the values of the latent variables are not known for any given observable output vector. In our approach we minimise the distance between the model covariance matrix $\Sigma(\boldsymbol{\theta})$ and a robust estimated covariance matrix of the data \mathbf{U} as a means of obtaining estimates of the model parameters of interest. Moreover, we develop a t-centre like estimator of the covariance matrix based on several minimum covariance determinant estimators ranging from 0% contamination to 50%.

Under multivariate normality assumptions on the endogenous variables, inference procedures and their properties are well known [1, e.g.]. If the estimated covariance matrix of the observed data is the sample covariance matrix \mathbf{S} and the model covariance matrix is $\Sigma(\boldsymbol{\theta})$, then the traditional maximum likelihood (ML) divergence (discrepancy function) used in covariance structure modelling is (up to additive constant)

$$F(\boldsymbol{\theta}) = \log |\Sigma(\boldsymbol{\theta})| + \text{tr}[\mathbf{S}\Sigma^{-1}(\boldsymbol{\theta})] - \log |\mathbf{S}|, \quad (1)$$

where $|\Sigma|$ is the determinant of Σ .

Penev and Prvan [5] provide references for evidence of the drastic effects of non-normal data on the ML fit. They note that all robust procedures suggested until now put weights on the data, depending on how likely it is to be “from the model” rather than “contaminated”.

2 New two step procedure

Penev and Prvan [5] proposed a two step procedure to obtain a robust fit that does not use the ML divergence. The first step is to find a robust estimator of the covariance matrix of the endogenous variables and the second step is to choose a robust divergence measure and then to execute a minimum-distance type fitting to find an estimator of the model parameter vector $\boldsymbol{\theta}$.

Penev and Prvan [5] used the minimum covariance determinant estimator as the robust estimator of the covariance matrix. We now use a t-centre like estimator of the covariance matrix which has the advantage of not needing to commit to using a specified level of contamination. Penev and Prvan [5] developed the *total von Neumann divergence* (TvND) which was used to measure the distance between the robust estimator of the covariance matrix and the model covariance matrix. We now consider two different divergence measures and variants to measure the distance between the robust estimate of the covariance matrix and the model fitted covariance matrix. These divergence measures, as opposed to the ones we use previously [5], are specifically designed to robustly fit within the manifold of symmetric positive definite $\mathbf{d} \times \mathbf{d}$ matrices.

We start our discussion with Definition 1 for the logarithm of a matrix. Properties of such matrices were discussed by Higham [3]. Also, Definition 2 is for $\nabla_{\mathbf{X}}f(\mathbf{X})$ where $f(\mathbf{X})$ is a scalar function of the matrix \mathbf{X} .

Definition 1. *Assume \mathbf{Y} is a square $\mathbf{d} \times \mathbf{d}$ symmetric and positive definite matrix. Let $\mathbf{Y} = \mathbf{VDV}^T$ where \mathbf{V} is a matrix of eigenvectors of \mathbf{Y} and \mathbf{D} is a diagonal matrix whose diagonal elements are the corresponding eigenvalues of \mathbf{Y} . Then $\log \mathbf{Y} = \mathbf{VD}_L\mathbf{V}^T$ where \mathbf{D}_L is the diagonal matrix whose diagonal elements are the logs of the eigenvalues of \mathbf{Y} .*

Definition 2. *Let $\mathbf{X} = [x_{ij}]$ be a $\mathbf{p} \times \mathbf{p}$ matrix and let $f(\mathbf{X})$ be a scalar function of \mathbf{X} . The derivative of $f(\mathbf{X})$ with respect to the matrix \mathbf{X} is defined by the $\mathbf{p} \times \mathbf{p}$ matrix*

$$\nabla_{\mathbf{X}}f(\mathbf{X}) = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1p}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{p1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{pp}} \end{bmatrix}.$$

In our applications we use \mathbf{X} as the robust estimator and \mathbf{Y} as the model based estimator of the covariance matrix. Vemuri et al. [6] defined the *total squared Frobenius divergence* (FROBT) for matrices, $\mathbf{d}(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|^2 / \sqrt{1 + 4\|\mathbf{Y}\|_F^2}$. We call the average $\frac{1}{2}[\mathbf{d}(\mathbf{X}, \mathbf{Y}) + \mathbf{d}(\mathbf{Y}, \mathbf{X})]$ the symmetrised FROBT (SFROBT).

The squared Frobenius divergence (FROB) is the numerator $\|\mathbf{X} - \mathbf{Y}\|^2$. For $f(\mathbf{Y}) = \|\mathbf{Y}\|_{\mathbb{F}}^2$ we have $\|\nabla_{\mathbf{Y}}f(\mathbf{Y})\|_{\mathbb{F}}^2 = 4\|\mathbf{Y}\|_{\mathbb{F}}^2$.

Jayasumana et al. [4] suggested that the most appropriate way of measuring the squared distance in the manifold of symmetric positive definite matrices is $d(\mathbf{X}, \mathbf{Y}) = \|\log \mathbf{X} - \log \mathbf{Y}\|_{\mathbb{F}}^2$ which they called the Log-Euclidean distance and we call the log-Frobenius distance (FROBLOG). In our simulations we use FROBLOG, as well as a variant

$$d(\mathbf{X}, \mathbf{Y}) = \frac{\|\log \mathbf{X} - \log \mathbf{Y}\|_{\mathbb{F}}^2}{\sqrt{1 + \|\nabla_{\mathbf{Y}}f(\mathbf{Y})\|_{\mathbb{F}}^2}}, \quad (2)$$

where $f(\mathbf{Y}) = \|\log \mathbf{Y}\|_{\mathbb{F}}^2$ and thus $\|\nabla_{\mathbf{Y}}f(\mathbf{Y})\|_{\mathbb{F}}^2 = \|2\mathbf{V}\mathbf{D}^{-1}(\log(\mathbf{D}))\mathbf{V}^T\|_{\mathbb{F}}^2$. The measure (2), which we call total squared log Frobenius distance (FROBLOGT), is inspired by the total Bregman divergences used by Vemuri et al. [6].

We state explicitly that FROBLOGT is not strictly a total Bregman divergence. In order for it to be a total Bregman divergence the numerator in (2) has to be in the form $f(\mathbf{X}) - f(\mathbf{Y}) - \langle \mathbf{X} - \mathbf{Y}, \Delta f(\mathbf{Y}) \rangle$, with the scalar product that generates the Frobenius norm [6, compare eq. (2), p. 476], which is clearly not the case in (2). However, there is a strong heuristic thrust towards using FROBLOGT as defined in (2) in robust inference. The numerator of (2) is the squared log-Euclidean metric which is known to be the true geodesic distance introduced by the Riemannian metric on the Riemannian manifold of symmetric positive definite $d \times d$ matrices [4]. The denominator serves as a norming by the length of the projection of the infinitesimal covariance matrix estimator change on the tangent space to the manifold ‘at the model matrix’. A smaller projection length, being in the denominator of (2), implies big changes in the functional for the same value of the distance between the matrices, and this means non-robustness of the estimator. Hence, for robustness purposes, such changes should be minimized. This approach makes sure that small changes in the parameter value do not affect the model matrix estimator $\Sigma(\boldsymbol{\theta})$ drastically.

We use \mathbf{X} as the robust estimator and \mathbf{Y} as the model based estimator of

the covariance matrix as this might result in a significant difference between FROBLOGT and FROBLOG. This is the case we are most interested in. If the roles of X and Y are swapped, then FROBLOGT is only a known scalar multiple of FROBLOG so the resulting estimators will coincide.

Vemuri et al. [6] also defined a robust estimator of the covariance matrix based on the total squared Frobenius distance. Suppose we have n symmetric positive definite matrices S_1, \dots, S_n . The t-centre is $\sum_{i=1}^n w_i S_i$ where

$$w_i = \frac{1/\sqrt{1 + 4\|S_i\|_F^2}}{\sum_{j=1}^n 1/\sqrt{1 + 4\|S_j\|_F^2}}.$$

Our robust estimator of the covariance matrix inspired by the t-centre approach uses

$$w_i = \frac{1/\sqrt{1 + \|\nabla_{S_i} f(S_i)\|_F^2}}{\sum_{j=1}^n 1/\sqrt{1 + \|\nabla_{S_j} f(S_j)\|_F^2}},$$

with $f(S_i) = \|\log S_i\|_F^2$ where S_i with $i = 1, 2, \dots, 11$ are the minimum covariance determinant estimators of the covariance matrix based on 0%, 5%, ..., 50% contamination, respectively. Being estimates of the covariance matrix, the S_i are symmetric positive definite matrices. The minimum covariance determinant method is a highly robust procedure, with well known asymptotic properties [2, e.g.].

The minimization of any of the divergences considered above delivers the minimum distance 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 minimization is different from the maximum likelihood estimation procedure where the goal is to minimize (1) which gives rise to a result equivalent to the maximization of the likelihood function.

3 Simulation study

The standard structural equation model considered has two latent exogenous variables ξ_1 and ξ_2 , one latent endogenous variable η_1 , and six observable variables Y_i with $i = 1, 2, \dots, 6$. The structural equation model is

$$\begin{aligned} \eta_1 &= \gamma_1 \xi_1 + \gamma_2 \xi_2 + \zeta_1, & Y_1 &= \eta_1 + \epsilon_1, & Y_2 &= \lambda_{21} \eta_1 + \epsilon_2, \\ Y_3 &= \xi_1 + \epsilon_3, & Y_4 &= \lambda_{42} \xi_1 + \epsilon_4, & Y_5 &= \xi_2 + \epsilon_5, & Y_6 &= \lambda_{63} + \epsilon_6, \end{aligned} \quad (3)$$

where $\gamma_1, \gamma_2, \lambda_{21}, \lambda_{42}, \lambda_{63}$ are model parameters, $\zeta_1 \sim \mathbf{N}(0, \psi)$, $\epsilon_i \sim \mathbf{N}(0, 1)$ for $i = 1, 2, \dots, 6$ and $\text{cov}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$, and

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \sim \mathbf{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \phi_{12} \\ \phi_{12} & 1 \end{bmatrix} \right).$$

Two restrictions arise from this model: $|\phi_{12}| \leq 1$ and $\psi \geq 0$.

In a condensed form, equation (3) is

$$\mathbf{Y} = \Lambda \begin{bmatrix} \eta_1 \\ \xi_1 \\ \xi_2 \end{bmatrix} + \boldsymbol{\epsilon}, \quad (4)$$

where $\mathbf{Y} = (Y_1, Y_2, \dots, Y_6)$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_6)$ and

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

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

$$\Sigma = \Lambda \Phi \Lambda^T + \mathbf{I}_6,$$

where Φ is the covariance matrix of the vector of latent variables (η_1, ξ_1, ξ_2) . Simple algebra gives

$$\Phi = \begin{bmatrix} \gamma_1^2 + \gamma_2^2 + 2\gamma_1\gamma_2\phi_{12} + \psi & \gamma_1 + \gamma_2\phi_{12} & \gamma_2 + \gamma_1\phi_{12} \\ \gamma_1 + \gamma_2\phi_{12} & 1 & \phi_{12} \\ \gamma_2 + \gamma_1\phi_{12} & \phi_{12} & 1 \end{bmatrix}.$$

The unknown parameters are $\theta = (\lambda_{21}, \lambda_{42}, \lambda_{63}, \gamma_1, \gamma_2, \phi_{12}, \psi)$. For sample sizes $n = 100, 200, 400$ we performed 100 simulations for a particular level of contamination, ranging from none (0%) to 40%, for the given values of θ .

We contaminated the data by changing the variance of the vector of ϵ values from 1 to 7^2 for a certain percentage of the data, assuming normality which is equivalent to a normal mixture of $N(0, 1)$ up to $N(0, 49)$ for each of the six independent components of the vector ϵ . This allows for a more comfortable interpretation of the ϵ components as noise since the mean is preserved at zero but the variance is inflated. We also experimented with using other distributions for the contaminated data. Only a subset of the results for data simulated from a mixture of two normal distributions are given here.

The Matlab function MCD, which is a part of LIBRA: the Matlab Library for Robust Analysis¹ [7], was used to get the robust estimates of the covariance matrix. The implementation was straightforward. The Matlab function `fmincon` was used to minimize the divergence functionals. The Hessian matrix was not computed. We used the actual assumed parameter values as the starting value of the vector of parameters to remove any chance of convergence to another local minimum of the distance measures considered.

We use $\sum_{i=1}^{100} (\theta_j - \hat{\theta}_{ij})^2 / 100$ as a measure of how well the model fits for each parameter ($j = 1, \dots, 7$). The simulations presented are for $\theta = (2, 3, 3, 2, 0.8, -0.3, 1.5)$ with contamination 5%, 10%, 20%, 30%. For each parameter the relative efficacy compared to ML is calculated for each divergence measure, then the average of these is also calculated for each measure

¹<http://wis.kuleuven.be/stat/robust/LIBRA/LIBRA-home>

which we call the average relative efficacy (ARE). The last columns of Tables 1, 2, 3 and 4 report the ARE. An ARE under 100% means more efficient compared to the traditional ML divergence using the normal theory maximum likelihood estimator of the covariance matrix, and the lower value the better. The converse means less efficient than the traditional ML divergence using the normal theory maximum likelihood estimator of the covariance matrix. The ARE of a parameter is the measure of how well the model fits the parameter for the divergence measure divided by the measure of how well the model fits the parameter for the ML divergence, all multiplied by 100%. Hence the ARE for ML is always 100%.

In Tables 1, 2, 3 and 4 ML^* is the divergence given by (1) with S replaced by the robust t-centre like minimum covariance determinant based estimator U . Also, $STvND$ is the symmetrised version of $TvND$, and vND is the usual von Neumann divergence.

When there is no contamination, unsurprisingly, the traditional ML divergence using S (maximum likelihood estimate of the covariance matrix assuming multivariate normality) performs best; these results are not presented. When there is contamination, even when sample sizes are large, the traditional maximum likelihood divergence using S performs worst. For contaminated data, most of the time, the traditional ML divergence using U is competitive for below 20% contamination. For large contamination, the FROBLOGT divergence performs best. This is also true for 40% contamination results which are not presented here. There is no point considering contamination greater than or equal to 50% because this would mean that you do not trust the model. Increasing the number of simulations will not change the conclusions. Increasing the sample sizes above 400 should also not change the conclusions. Sample sizes much below 100 are not common in data applications.

Similar conclusions are reached when the contaminated data was simulated from a Chi-squared distribution centred about zero with 20 degrees of freedom, and also for a Chi-squared distribution centred about zero with 30 degrees of

Table 1: Error estimates for 5% contamination.

$n = 400$	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ	ARE
ML	0.018	0.10	0.13	0.13	0.065	0.0081	1.83	100%
ML*	0.0028	0.015	0.016	0.010	0.0085	0.0027	0.047	14.2%
TvND	0.0029	0.019	0.022	0.013	0.0087	0.0026	0.046	15.6%
STvND	0.0028	0.019	0.022	0.013	0.0087	0.0026	0.047	15.7%
vND	0.0028	0.019	0.023	0.013	0.0087	0.0026	0.048	15.7%
FROB	0.0029	0.019	0.025	0.016	0.0091	0.0026	0.051	16.5%
SFROBT	0.0029	0.019	0.025	0.016	0.0091	0.0026	0.051	16.5%
FROBT	0.0029	0.019	0.025	0.016	0.0091	0.0026	0.051	16.5%
FROBLOG	0.0027	0.017	0.018	0.011	0.0085	0.0027	0.049	14.8%
FROBLOGT	0.0027	0.017	0.018	0.011	0.0084	0.0027	0.048	14.8%
$n = 200$								
ML	0.035	0.18	0.16	0.19	0.094	0.015	2.30	100%
ML*	0.0075	0.025	0.023	0.031	0.023	0.0064	0.070	19.3%
TvND	0.0076	0.033	0.030	0.034	0.024	0.0065	0.066	21.1%
STvND	0.0076	0.034	0.031	0.034	0.024	0.0066	0.068	21.2%
vND	0.0077	0.034	0.031	0.034	0.024	0.0066	0.070	21.3%
FROB	0.0077	0.034	0.035	0.038	0.024	0.0067	0.067	22.1%
SFROBT	0.0077	0.034	0.035	0.038	0.024	0.0067	0.067	22.1%
FROBT	0.0077	0.034	0.035	0.038	0.024	0.0067	0.067	22.1%
FROBLOG	0.0074	0.029	0.025	0.031	0.023	0.0064	0.073	19.9%
FROBLOGT	0.0074	0.029	0.026	0.031	0.023	0.0064	0.071	19.9%
$n = 100$								
ML	0.066	0.54	0.31	0.24	0.17	0.018	2.77	100%
ML*	0.012	0.060	0.050	0.054	0.045	0.012	0.14	23.7%
TvND	0.012	0.070	0.060	0.059	0.047	0.011	0.13	24.2%
STvND	0.012	0.071	0.061	0.059	0.047	0.011	0.14	24.4%
vND	0.012	0.072	0.062	0.059	0.047	0.011	0.14	24.6%
FROB	0.012	0.071	0.069	0.068	0.048	0.011	0.14	25.3%
SFROBT	0.012	0.071	0.069	0.068	0.048	0.011	0.14	25.3%
FROBT	0.012	0.071	0.069	0.068	0.048	0.011	0.14	25.3%
FROBLOG	0.012	0.064	0.052	0.054	0.046	0.012	0.15	24.3%
FROBLOGT	0.012	0.064	0.053	0.054	0.046	0.012	0.14	24.3%

Table 2: Error estimates for 10% contamination

$n = 400$	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ	ARE
ML	0.035	0.41	0.42	0.28	0.14	0.017	6.12	100%
ML*	0.0031	0.011	0.011	0.013	0.0077	0.0025	0.038	5.61%
TvND	0.0031	0.013	0.015	0.015	0.0076	0.0025	0.039	5.93%
STvND	0.0031	0.013	0.015	0.015	0.0076	0.0025	0.038	5.93%
vND	0.0031	0.013	0.015	0.015	0.0076	0.0025	0.038	5.93%
FROB	0.0032	0.013	0.017	0.017	0.0076	0.0025	0.046	6.23%
SFROBT	0.0032	0.013	0.017	0.017	0.0076	0.0025	0.046	6.23%
FROBT	0.0032	0.013	0.017	0.017	0.0076	0.0025	0.046	6.23%
FROBLOG	0.0032	0.011	0.012	0.013	0.0076	0.0024	0.037	5.66%
FROBLOGT	0.0032	0.011	0.012	0.013	0.0076	0.0025	0.037	5.67%
$n = 200$								
ML	0.078	0.43	0.52	0.33	0.16	0.024	5.95	100%
ML*	0.0060	0.025	0.026	0.028	0.019	0.0058	0.052	9.18%
TvND	0.0060	0.030	0.028	0.030	0.019	0.0056	0.051	9.36%
STvND	0.0060	0.030	0.028	0.030	0.019	0.0056	0.050	9.37%
vND	0.0060	0.031	0.029	0.030	0.019	0.0056	0.050	9.39%
FROB	0.0061	0.030	0.031	0.033	0.018	0.0056	0.053	9.57%
SFROBT	0.0061	0.030	0.031	0.033	0.018	0.0056	0.054	9.57%
FROBT	0.0061	0.030	0.031	0.033	0.018	0.0056	0.054	9.56%
FROBLOG	0.0062	0.028	0.026	0.028	0.018	0.0057	0.053	9.20%
FROBLOGT	0.0062	0.028	0.026	0.028	0.018	0.0058	0.051	9.21%
$n = 100$								
ML	0.13	0.76	0.76	0.40	0.25	0.024	8.53	100%
ML*	0.010	0.057	0.042	0.053	0.036	0.013	0.17	14.6%
TvND	0.010	0.067	0.052	0.060	0.039	0.012	0.18	15.1%
STvND	0.0099	0.068	0.053	0.060	0.039	0.012	0.17	15.1%
vND	0.0098	0.068	0.054	0.059	0.039	0.012	0.17	15.2%
FROB	0.010	0.069	0.060	0.070	0.040	0.012	0.19	15.7%
SFROBT	0.010	0.069	0.060	0.070	0.040	0.012	0.19	15.7%
FROBT	0.010	0.069	0.060	0.070	0.040	0.012	0.19	15.7%
FROBLOG	0.010	0.058	0.045	0.053	0.038	0.013	0.17	14.8%
FROBLOGT	0.010	0.059	0.046	0.053	0.037	0.013	0.16	14.9%

Table 3: Error estimates for 20% contamination.

$n = 400$	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ	ARE
ML	0.080	0.96	1.07	0.70	0.24	0.031	17.46	100%
ML*	0.0052	0.029	0.035	0.047	0.024	0.0040	0.57	6.54%
TvND	0.0051	0.044	0.056	0.038	0.022	0.0035	0.61	6.57%
STvND	0.0051	0.042	0.053	0.037	0.022	0.0035	0.57	6.41%
vND	0.0051	0.040	0.050	0.036	0.022	0.0035	0.54	6.26%
FROB	0.0050	0.040	0.056	0.035	0.021	0.0034	0.50	6.15%
SFROBT	0.0050	0.041	0.057	0.035	0.021	0.0034	0.51	6.16%
FROBT	0.0050	0.041	0.057	0.034	0.020	0.0034	0.51	6.17%
FROBLOG	0.0057	0.036	0.041	0.041	0.023	0.0036	0.57	6.43%
FROBLOGT	0.0056	0.029	0.027	0.036	0.022	0.0037	0.42	5.93%
<hr/>								
$n = 200$								
ML	0.20	1.75	1.45	0.69	0.26	0.036	19.39	100%
ML*	0.0064	0.050	0.047	0.062	0.028	0.0069	0.51	7.31%
TvND	0.0063	0.055	0.060	0.050	0.028	0.0064	0.56	7.07%
STvND	0.0063	0.052	0.057	0.049	0.028	0.0064	0.51	6.92%
vND	0.0063	0.049	0.053	0.048	0.027	0.0064	0.46	6.77%
FROB	0.0064	0.051	0.061	0.048	0.027	0.0063	0.47	6.83%
SFROBT	0.0064	0.052	0.061	0.047	0.027	0.0063	0.47	6.84%
FROBT	0.0064	0.052	0.062	0.047	0.027	0.0063	0.47	6.85%
FROBLOG	0.0071	0.049	0.046	0.055	0.027	0.0065	0.48	6.95%
FROBLOGT	0.0071	0.043	0.036	0.051	0.027	0.0067	0.36	6.67%
<hr/>								
$n = 100$								
ML	0.78	2.69	2.39	0.76	0.35	0.051	17.32	100%
ML*	0.021	0.10	0.11	0.098	0.062	0.012	0.65	9.96%
TvND	0.020	0.11	0.13	0.097	0.061	0.012	0.74	10.05%
STvND	0.020	0.11	0.12	0.094	0.058	0.012	0.65	9.79%
vND	0.020	0.10	0.11	0.090	0.056	0.013	0.56	9.56%
FROB	0.021	0.11	0.12	0.10	0.058	0.012	0.59	9.88%
SFROBT	0.021	0.11	0.12	0.10	0.058	0.012	0.60	9.89%
FROBT	0.021	0.11	0.13	0.10	0.057	0.012	0.61	9.89%
FROBLOG	0.021	0.10	0.10	0.091	0.058	0.012	0.57	9.60%
FROBLOGT	0.021	0.091	0.082	0.086	0.058	0.013	0.41	9.32%

Table 4: Error estimates for 30% contamination.

$n = 400$	λ_{21}	λ_{42}	λ_{63}	γ_1	γ_2	ϕ_{12}	ψ	ARE
ML	0.13	2.43	2.08	0.94	0.30	0.038	28.00	100%
ML*	0.014	0.22	0.21	0.19	0.075	0.0095	3.65	16.2%
TvND	0.013	0.42	0.51	0.16	0.070	0.0082	4.81	18.7%
STvND	0.013	0.36	0.44	0.15	0.066	0.0077	4.15	16.9%
vND	0.013	0.30	0.37	0.14	0.061	0.0072	3.49	15.1%
FROB	0.013	0.31	0.40	0.13	0.059	0.0069	3.40	15.1%
SFROBT	0.013	0.32	0.41	0.13	0.059	0.0069	3.48	15.2%
FROBT	0.013	0.33	0.43	0.13	0.058	0.0069	3.55	15.4%
FROBLOG	0.016	0.28	0.30	0.15	0.066	0.0078	3.64	15.7%
FROBLOGT	0.015	0.10	0.082	0.078	0.046	0.0064	1.36	9.2%
<hr/>								
$n=200$								
ML	0.31	2.84	2.75	0.89	0.31	0.042	40.46	100%
ML*	0.025	0.26	0.24	0.18	0.072	0.013	3.70	15.6%
TvND	0.023	0.46	0.47	0.14	0.068	0.011	4.84	16.7%
STvND	0.023	0.40	0.40	0.13	0.064	0.011	4.11	15.3%
vND	0.022	0.33	0.33	0.12	0.059	0.010	3.38	13.8%
FROB	0.022	0.35	0.36	0.11	0.060	0.010	3.41	14.0%
SFROBT	0.022	0.36	0.38	0.11	0.059	0.010	3.49	14.1%
FROBT	0.022	0.37	0.39	0.11	0.059	0.010	3.57	14.2%
FROBLOG	0.029	0.30	0.28	0.14	0.064	0.011	3.61	14.6%
FROBLOGT	0.028	0.15	0.11	0.086	0.051	0.011	1.47	10.5%
<hr/>								
$n=100$								
ML	4.21	4.70	5.04	1.24	0.49	0.048	38.28	100%
ML*	0.067	0.40	0.38	0.30	0.16	0.017	4.01	17.0%
TvND	0.066	0.52	0.52	0.26	0.15	0.015	5.32	17.1%
STvND	0.064	0.43	0.44	0.24	0.14	0.015	4.45	15.7%
vND	0.063	0.35	0.36	0.22	0.13	0.015	3.61	14.5%
FROB	0.068	0.40	0.41	0.23	0.14	0.015	3.86	15.1%
SFROBT	0.068	0.41	0.42	0.23	0.14	0.015	3.98	15.2%
FROBT	0.068	0.43	0.44	0.23	0.14	0.015	4.09	15.3%
FROBLOG	0.075	0.33	0.33	0.26	0.14	0.015	3.70	15.1%
FROBLOGT	0.071	0.18	0.18	0.18	0.13	0.016	1.70	12.6%

freedom. For a Chi-squared distribution centred about zero with six degrees of freedom and for a Chi-squared distribution centred about zero with ten degrees of freedom FROBLOGT is best above 25% contamination. Due to page limitations, these simulations are not presented here.

4 Conclusion

Using the traditional ML divergence measure to estimate the model parameters under multivariate normality assumptions does not deliver a good fit when there is contamination in the data. This ML measure was tailored for high efficiency under the ideal non-contaminated model so the result is not surprising. It trades off robustness for high efficiency when the endogenous variables have a multivariate normal distribution with zero mean. Our simulated data represents a mixture of two normal distributions, thus violating the normality assumption. This violation does not have a detrimental effect on both types of robust estimators—the ones previously considered by Penev and Prvan [5] as well as the new ones presented here. Replacing \mathbf{S} in the traditional ML divergence measure with the robust estimator \mathbf{U} of the covariance matrix delivered a better performance. The t-centre approach for calculating the recommended robust proximity matrix $\sum_i w_i \mathbf{S}_i$, with the recommended weights, avoids the inconvenience of not knowing the actual percentage of contamination in the data. The averaging of the minimum covariance determinant estimators over different levels of contamination avoids committing to (or having to estimate) a specified level of contamination.

For large contamination, our new FROBLOGT divergence measure performed much better than the other divergence measures considered. This is due to our construction of the divergence measure: (i) the numerator is a squared log-Euclidean metric with superb properties as a squared Riemannian metric on the Riemannian manifold of symmetric positive definite matrices; (ii) the norm in the denominator serves to make ensures that non-robust estimators in the metric are penalized and hence rendered non-competitive.

On the basis of our experimentation, we recommend FROBLOGT as the divergence measure of choice when a significant contamination is present. Further investigations of other divergence measures are possible. Another research question is related to deriving analytically the influence functions of the proposed robust estimators. This is a difficult task in general but for a simple representative class of structural equation models it should be possible and will allow the comparison of robustness properties via comparing the influence functions. Although such a comparison is asymptotic in spirit and may not be quite valid for specific fixed sample sizes, it still represents a valuable complementary view of robustness in structural equation models.

References

- [1] K. A. Bollen. *Structural equations with latent variables*. Wiley, New York, 1989. <http://au.wiley.com/WileyCDA/WileyTitle/productCd-0471011711.html> C341
- [2] R. W. Butler, P. L. Davies and M. Jhun. Asymptotics for the minimum covariance determinant estimator. *Ann. Stat.*, 21(3):1385–1400, 1993. <http://www.jstor.org/stable/2242201> C344
- [3] N. J. Higham. *Functions of matrices: Theory and computation*. SIAM, Philadelphia, 2008. doi:10.1137/1.9780898717778 C342
- [4] S. Jayasumana, R. Hartley, M. Salzmann, H. Li and M. Harandi. Kernel methods on the Riemannian manifold of symmetric positive definite matrices. *Proc. CVPR IEEE 2013*, 73–80, 2013. doi:10.1109/CVPR.2013.17 C343
- [5] S. Penev and T. Prvan. Robust estimation in structural equation models using Bregman divergences. *CTAC2012, ANZIAM J.*, 54:C574–C589, 2013. <http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/6306> C341, C342, C352

- [6] B. C. Vemuri, M. Liu, S.-I. Amari, and F. Nielsen. Total Bregman divergence and its applications to DTI analysis. *IEEE T. Med. Imaging.*, 30(2):475–483, 2011. doi:[10.1109/TMI.2010.2086464](https://doi.org/10.1109/TMI.2010.2086464) C342, C343, C344
- [7] S. Verboven and M. Hubert. LIBRA: a MATLAB library for robust analysis, *Chemometr. Intell. Lab.*, 75(2):127–136, 2005. doi:[10.1016/j.chemolab.2004.06.003](https://doi.org/10.1016/j.chemolab.2004.06.003) C346

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.