WP. 3
ENGLISH ONLY

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)

CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Tarragona, Spain, 26-28 October 2011)

Topic (i): Disclosure risk assessment

Disclosure risk for high dimensional business microdata

Prepared by Flavio Foschi, ISTAT, Italy

Disclosure risk for high dimensional business microdata

Flavio Foschi

Istat, P. Indipendenza 4, 00185 Rome, Italy, foschi@istat.it

Abstract: A strategy for assessing continuous microdata disclosure risk can be based on the robust fitting of finite multivariate gaussian mixture, by using an unsupervised learning approach so that no prior information about the "true" number of mixture components is requested. Computational feasibility can be pursued by means of approximate methods. After a description of the proposal, some simulation results are ilustrated together with an application to business census-like microdata stemming from the Enterprises' System of Accounts survey by the Italian National Institute of Statistics.

1 Introduction

Since for continuous microdata "all units are unique (rare) with respect to a small set of quantitative variables" (AA.VV., 2010), the detection of statistical units which show a low degree of consistency w.r.t. the generating process of the majority of observations can constitute a relevant target. Several studies in the literature on multivariate outlier detection cope with the problem of distinguishing abnormal behaviours of observations under the condition data follow at least approximately, eventually by means of suitable transformations, a Gaussian distribution. By analogy with Neykov, Filzmoser, Dimova and Neytchev (2007), the strategy we propose for detecting observations featured by some kind of heterogeneity, without restrictive assumptions about the distribution of "clean" data, is based on the robust fitting of finite multivariate Gaussian mixture according to the FAST–TLE algorithm described by Neykov and Mueller (2003), whereas survey weight and missing value issues (see Todorov, Templ and Filzmoser, 2011) are not addressed at present, being a subject of future developments. Section two gives a short review of some basic concepts, handling a description of the proposal together with some technical details about the extension of robustness tools to finite Gaussian mixtures; the rest of the paper discusses two applications. The first is a simulation exercise directed to assess the performance of the proposed method against several patterns of data contamination, tailored to give some insight about the second application. The latter (at experimental stage) constitutes the primary aim of this paper and concerns business census-like microdata stemming from the Enterprises' System of Accounts survey by the Italian National Institute of Statistics.

2 Robustness and Finite Gaussian mixtures

The purpose of the robustness is to safeguard against deviations from the assumptions, so that robust estimates are consistent even if the underling model is only approximately true. As pointed out by Huber and Ronchetti (2009), since the purpose of diagnostics is to identify deviations from the assumptions, good diagnostic tools are usually robust. A key quantity is the breakdown point, the smallest number of observations necessary to make an estimate aberrant. Considering an estimator T, a sample $\mathbf{X}$ of n observations on p variables and all possible corrupted samples $\mathbf{X}'$ obtained substituting m original data points, denote

$$b(m) = \sup_{\mathbf{X}'} \left\| T\left(\mathbf{X}'\right) - T\left(\mathbf{X}\right) \right\|$$

Then, the finite sample breakdown fraction of T at the sample $\mathbf{X}$ is defined as

$$\varepsilon_n\left(T, \mathbf{X}\right) = \min\left\{ \frac{m}{n} : b\left(m\right) = \infty \right\}$$

Hence the higher $\varepsilon_n(T, \mathbf{X})$, the higher the robustness of T at the sample $\mathbf{X}$. A breakdown fraction greater than 0.5 seems meaningless: the concept of data contamination should not be referred to the majority of sample units. When data are distributed according to a multivariate Gaussian law, the robust Mahalanobis distance constitutes a popular tool for detecting contaminated units. To that aim, the Minimum Covariance Determinant (MCD) method selects iteratively a subset of h observations whose the covariance matrix determinant is minimum. The location estimate is computed averaging the best h points, whereas the scatter matrix is a multiple of their covariance matrix (see section 2.3). The breakdown point for the scatter matrix is the smallest number of observations which makes either the largest eigenvalue not finite or the minimum eigenvalue equal to zero. Lopuhaä and Rousseeuw (1991) prove that MCD estimates attain the highest breakdown fraction 0.5 when h=0.5(n+p+1). Several improvements of the MCD method have been proposed and some of them are discussed by Cerioli (2010). The FAST-TLE algorithm (Neykov and Müller, 2003) encompasses a wide class of robust estimators and can be expressed as a sequence of steps which yields an iterative process. Let $\ell(\ldots)$ a log-likelihood function and $\nu(i)$, i= 1,…,n, a permutation of indices such that:

$$\ell\left(\mathbf{x}_{\nu(i)}; \theta\right) \geq \ell\left(\mathbf{x}_{\nu(i+1)}; \theta\right), \quad \mathbf{x}_{\nu(i)} \in \mathbb{R}^p, \quad i=1,\cdots,n$$

Hence, given h<n, the maximum trimmed likelihood estimator of θ is defined by:

$$TL_h = \arg\max_{\theta} \sum_{i=1}^{h} \ell\left(\mathbf{x}_{\nu(i)}; \theta\right) \quad (1)$$

An iterative process gives an approximate solution to (1). Given a subset of h observations and the estimate $\theta^{(r)}$ associated to the r^{th} iteration, (r=1,…, R):

1. define $Q^{(r)} = \sum_{i=1}^{h} \ell\left(\mathbf{x}_{\nu(i)}; \theta^{(r)}\right)$,
2. for $i=1,\ldots,n$, sort $\ell\left(\mathbf{x}_{\nu(i)}; \theta^{(r)}\right)$ in descending order and select the indices $\nu(1),\ldots,\nu(h)$ related to the first h values,
3. compute $\theta^{(r+1)} = \arg\max_{\theta} \sum_{i=1}^{h} \ell(\mathbf{x}_{\nu(i)}; \theta)$,
4. by using $\theta^{(r+1)}$, return to the step 1.

Each iteration constitutes a concentration step (C-step) since it selects the subset of h observations featured by the larger h log-likelihood contributions. Apart the difference on objective functions to optimize, the similarity w.r.t. MCD is evident: when a multivariate Gaussian distribution is assumed as statistical model, that is $\theta=\{\mu, \Sigma\}$, MCD C-steps select the h observations with the smallest distances and the scatter matrix with the lower determinant (Rousseeuw and Van Driessen, 1999). The FAST-TLE framework represents a natural way to mimic robustness features of MCD estimators when finite Gaussian mixtures are used:

$$\mathcal{L}\left(\mathbf{x}_{\nu(i)}; \theta\right) = \sum_{g=1}^{G} \varphi_g N\left(\mathbf{x}_{\nu(i)}; \mu_g, \Sigma_g\right), \quad \forall g \ \varphi_g \geq 0, \quad \sum_{g=1}^{G} \varphi_g = 1 \qquad (2)$$

The immediate consequence is the possibility to extend findings about robustness against data contamination to non normal data.

2.1 The breakpoint of TLE

A fundamental result due to D.L. Vandev, establishes a link between the *d-fullness* of $\mathcal{L}(\mathbf{x}_{\nu(i)}; \theta)$ and the breakdown point. A function $f(\theta)$ on the topological space Θ is subcompact if the sets $\{\theta \in \Theta: f(\theta) \leq C\}$ are compact or empty for every constant C; a finite set $F=\{f_1(\theta),\ldots,f_n(\theta)\}$ of n functions is *d-full* if for each subset of F having cardinality d, the supremum is a subcompact function. Roughly speaking, the value d represents the number of observations necessary to make a unique guess for the parameter θ. Then, it is possible to prove that if the set $F=\{-\ell(\mathbf{x}_i; \theta), i=1,\ldots,n\}$ is *d-full* and $0.5(n+d)<h<(n-d)$, the breakdown point is not less than $(n-h)/n$. Vandev and Neykov (1993) show that if $\mathcal{L}(\mathbf{x}_i; \theta)$ follows a multivariate Gaussian distribution and $\theta=\{\mu, \Sigma\}$, then $d=(p+1)$, whereas Neykov, Filzmoser, Dimova and Neytchev (2007), for finite Gaussian mixtures having G components and distinct parameters achieve $d=G(p+1)$; the breakpoint obtained from $d=G(p+1)$ is referred to parameters $\theta_g=\{\mu_g, \Sigma_g\}$, $g=1,\ldots,G$.

2.2 Fast approximations on large datasets

Since each C-step involves many estimations conditionally on different numbers G of mixture components, several parameterizations of the covariance matrix (see Fraley

and Raftery, 2002, 2006) and the model selection through the Bayesian Information Criterion, using fewer C-steps improves the speed of the algorithm. Rousseeuw and Van Driessen (1999) suggest a strategy depending on data size which can be easily applied to our framework. Defining G_{max} the highest candidate value for the number of mixture components:

1. for small n:
   a. select J_1 initial subsets having cardinality G_{max} (p+1): note that the probability to select clean data is inversely related to the number of observations,
   b. for each subset, number G and parameterization of the covariance matrix, compute $\theta_g=\{\mu_g, \Sigma_g\}$, $g=1,\ldots,G$, select the best model, sort $\ell(\mathbf{x}_i;\theta)$ ($i=1,\ldots,n$) in descending order and keep the indices $\nu(1),\ldots,\nu(h)$ related to the first h values: the cardinality becomes h,
   c. for each subset carry out a very small number of C-steps,
   d. for the subsets featured by the J_2 highest log-likelihoods carry out C-steps until convergence and select the best model.
2. for larger n:
   a. partition data in J_0 non-overlapping subsets having roughly the same size and:
      - select J_1/J_0 subsets having cardinality G_{max} (p+1),
      - carry out the step b. previously described, obtaining subsets with cardinality h/J_0,
      - perform the step c. as above and select best J_3 results from each of the J_1/J_0 subsets.
   b. Pool the J_0 non-overlapping subsets and, for each of the J_3 J_0 outcomes, carry out a very small number of C-steps by using subsets having cardinality h; then retain the best J_3 results.
   c. In the full dataset, starting from the J_3 outcomes achieved from b., take several C-steps and keep the best result.

2.3 Corrections for consistency and small sample bias

If $\mathbf{x}$ is constrained to lie in the subspace $0\leq(\mathbf{x}-\mu)^t\Sigma^{-1}(\mathbf{x}-\mu)\leq\eta(h)$, where $\eta(h)$ can be considered as the h^{th} order statistic of a sample of n squared Mahalanobis distances, then (Tallis, 1963):

$$E(\mathbf{x})=\mu \quad \mathrm{Cov}(\mathbf{x})=c_1(h)^{-1}\Sigma$$
$$c_1(h)^{-1}=P\left\{\chi^2_{p+2}<\eta(h)\right\}\Big/P\left\{\chi^2_p<\eta(h)\right\}$$

Hence, to inferential aims about the whole sample, the inflation factor $c_1(h)$ for Σ has to be computed. An approximation that works well even for small samples is:

$$c_1(h) = (h/n) \Big/ P\left\{\chi^2_{p+2} < \chi^2_{p,\alpha=h/n}\right\}$$

That correction is not sufficient to make the scatter matrix unbiased in small samples. The problem is addressed by Pison, Van Aelst and Willems (2002), through simulation results based on affine equivariance of multivariate estimators of location and shape. Affine equivariance means that affine transformations (linear transformations followed by translations) of data make estimators changed accordingly. If **A** is a non singular $p \cdot p$ matrix, **c** a $p \cdot 1$ vector and $\mathbf{X}=(\mathbf{x}_1,\ldots,\mathbf{x}_n)$:

$$\hat{\mu}\left(\mathbf{XA} + \mathbf{1}_n \mathbf{c}^t\right) = \hat{\mu}\left(\mathbf{X}\right)\mathbf{A} + \mathbf{1}_n \mathbf{c}^t$$
$$\hat{\Sigma}\left(\mathbf{XA} + \mathbf{1}_n \mathbf{c}^t\right) = \mathbf{A}^t \hat{\Sigma} \mathbf{A}$$

That feature is invoked to make analyses independent from coordinate transformations: since squared Mahalanobis distances are no affected by affine data transformation, they were able to obtain empirical estimates for the bias of the MCD scatter matrix and fit formulas to approximate the correction factor at any n and p. The rationale behind their results can be explained as it follows: by using standard Gaussian distribution, for fixed n and p, unbiased estimates of Σ should give $E\left(\hat{\Sigma}\right)=\mathrm{I}_p$, which in turn implies the p^{th} root of the determinant $\left|\hat{\Sigma}\right|$ is equal to one. Hence, performing k Monte Carlo replications, the correction factor estimate for the small sample bias becomes

$$c_2(h) = \left(k^{-1} \sum_{J=1}^{k} \left|\hat{\Sigma}^{(j)}\right|^{1/p} \right)^{-1}$$

In our setting, h and n are related to the TLE on the whole, while corrections can be computed for each mixture component. By assigning the trimmed $n-h$ observation to components g=1,..,G according to a suitable criterion (i.e. the minimization of the Mahalanobis distance), the number h_g of statistical units which maximize posterior probabilities $p(g \mid \mathbf{x}_{v(i)})$ and the number n_g of statistical units which best fit the assignation criterion are used in place of h and n.

2.4 Inferences

Cerioli and Farcomeni (2011) give an in-depth analysis of error rates in multiple hypotheses testing for multivariate outlier detection. In our framework, since each observation, conditionally on its membership to the g^{th} mixture component, is assumed to come from a Gaussian distribution, if investigated hyphoteses are

$$H_{0,i} : \mathbf{x}_{g,i} \sim N\left(\mu_g, \Sigma_g\right) \qquad (3)$$

a multiple test problem arises. Given robust estimators for μ_g and Σ_g, tests (3) are based on estimated robust squared Mahalanobis distances

$$D_{g,i}^{\ 2} \equiv \left(\mathbf{x}_{g,i} - \hat{\mu}_g\right)^t \hat{\Sigma}_g^{-1} \left(\mathbf{x}_{g,i} - \hat{\mu}_g\right)$$

The outcomes related to the g^{th} mixture component can be illustrated as follows:

	H_0 not rejected	H_0 rejected	Total
H_0 True	$n_{g,0/0}$	$n_{g,1/0}$	$n_{g,0}$
H_0 False	$n_{g,0/1}$	$n_{g,1/1}$	$n_{g,1}$
Total	$n_g - R$	R	n_g

Quantities $n_{g,0/1}$ and $n_{g,1/0}$ are indicated as the amount of "masking" (M) and "swamping" (S). The latter is controlled by the level α of the test whereas the former depends on the power. Becker and Gather (1999) proposed to control false rejections changing the null hyphotesis and adjusting the size of the critical region:

$$H_0 : \bigcap_{i=1}^{n_g} \left\{ \mathbf{x}_{g,i} \sim N\left(\mu_g, \Sigma_g\right) \right\} \qquad \alpha_{n_g} = 1 - (1-\alpha)^{1/n_g} \tag{4}$$

Thus, under H_0, with probability $1-\alpha$, no observation lies in the critical region. By analogy to Hardin and Rocke (2005) findings about MCD estimates, since extreme observations are approximately independent from location and scale estimates obtained from a clean subset of data, the intersection between multiple tests sharing same estimates should be negligible. A simple way to correct the low power problem is applied by Cerioli (2010), considering that an increase of swamping can be accepted to alleviate the amount of masking if the absence of contamination is confuted: when H_0 in (4) is rejected, each observation is tested at the α level according to (3). Distributional results about squared Mahalanobis distance estimates are well known. Those related to observations used in parameters estimation are distributed as scaled Beta (Wilks, 1963):

$$D_{g,i}{}^2 \sim \frac{\left(n_g - 1\right)^2}{n_g} Beta\left(\frac{p}{2}, \frac{n_g - p - 1}{2}\right), \quad i \in g \bigcap \{1, \cdots, h\} \tag{5}$$

Remaining data points are distributed according to prediction regions for multivariate normal having a given probability to contain the next $\mathbf{x}_i$ observation (Chew, 1966):

$$D_{g,i}{}^2 \sim \frac{\left(n_g^2 - 1\right) p}{\left(n_g - p\right) n_g} F\left(p, n_g - p\right), \quad i \in g \bigcap \{h+1, \cdots, n\} \tag{6}$$

Observations whose maximum p-value over g=1,…,G, falls in the critical region of tests (4) or (3) are labelled as atypical.

3 Some empirical results.

A simulation study allows to get some insight about the effectiveness of the proposed method. Let clean data be distributed as a mixture of two spherical normal random

variables, using n=90 (or n=180), p=6, μ_g={0·$\mathbf{1}_p$, 2c·$\mathbf{1}_p$} with c=$(\chi^2_{p,\,0.99}/p)^{1/2}$, φ_g={5/9, 4/9}. Added contaminated data points (henceforth *cdp*) are 20% of the observations in the smallest clean component: approximately 8 (or 16). By analogy with the simulation settings of Hardin and Rocke (2004), three patterns of contamination are considered: separate, radial, diffuse. The former consists in a cluster of observations generated by a $N(4c\cdot\mathbf{1}_p, \mathbf{I}_p)$, while the radial one forms an anulus around the clean data and is constructed by points from Gaussian laws with 2 times the original covariances; the diffuse contamination is drawn by a normal law having location and shape of the entire dataset. An acceptance-rejection rule ensures disperse points fall outside both clean data $\chi^2_{p,\,0.99}$ ellipses.

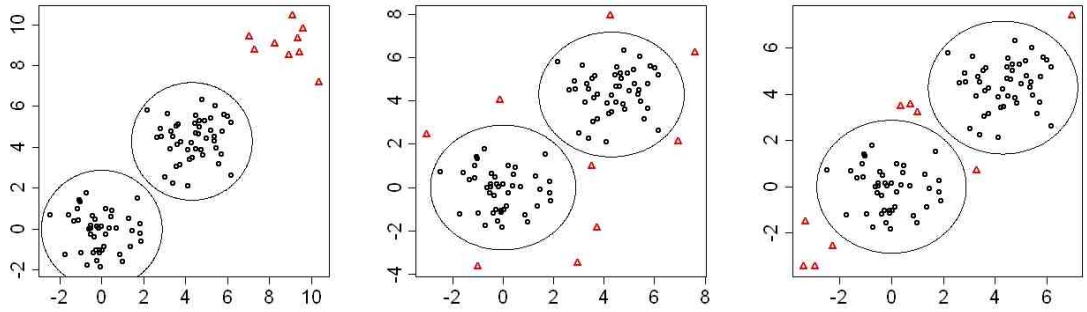

Fig 1 Two dimensional examples: separate, radial, diffuse contaminations (symbol △).

For each pattern, 500 Monte Carlo replicates were performed by means of the software *R* (R Development Core Team, 2010), by using the package *Mclust* (Fraley and Raftery, 2006) for parameters estimation and model selection within C-steps. The breakpoint fraction used is ε_n=0.25 whereas α=0.05. Table 1 shows a negligible swamping while, for n=90, masking is evident for diffuse and radial contaminations because of the small sample size that affects the power and the small number of contaminated observations which inflates the respective proportions. On the whole, considering the severity of those patterns, the performances seem acceptable.

	N=90+contaminated obs.		N=180+contaminated obs.	
	Type I error	Type II error	Type I error	Type II error
Separate cont.	0.0099	0.0359	0.0142	0.0096
Radial cont.	0.0058	0.1538	0.0098	0.0172
Diffuse cont.	0.0062	0.1425	0.0095	0.0234

Table 1 Fractions of I and II type errors ($|S|/n$ and, respectively, $|M|/|cdp|$).

4 Application to business microdata.

The Enterprises System of Account survey (ESA) is a structural-type survey carried out by Istat on a census basis. All economic activities are surveyed, excluding agriculture, zootechnics, hunting and fishing, financial activities (except for the financial intermediation and insurance auxiliary ones), public administration and

associative organization activities as well activities carried out by families and cohabitations. Data of 2004 ESA survey are stratified by economic activity (NACE Rev. 1.1) and size class. To our aims, enterprises classified according to their main activity are centre of interest while functional units (featured by different lines of production) are ignored for simplicity. Hence, a set of 10,313 records is analysed. Hundreds of variables are gathered in ESA survey and 68 of them are related to the profit and loss account. A realistic disclosure scenario could involve a more restrict set of balance items, i.e. focusing on subtotals more easy to learn or gauge: a convenient selection can maintain a large extent of relevant information. Since *turnover*, *cost of materials together with power consumptions and goods to resale*, *cost for services*, *staff costs*, *number of workers*, *earnings*, are often sufficient to achieve a "fingerprint" of an examined enterprise, they are assumed as (perhaps too pessimistic) scenario variables. Considering first four NACE digits and two size classes (henceforth *W*) in term of workers, $[100, 499] \cup [500, +\infty)$, 660 strata follow.

10%	20%	30%	40%	50%	60%	70%	80%	90%
1	2	2	4	5	8	13	20	38

Table 2 Quantiles of stratum sizes (four NACE digits and two size classes).

For reducing data sparsity it is possible to collapse strata having few units into a pooled stratum. Given the minimum number of units requested, a collapsing sequence is applied. Firstly, when necessary, NACE digits are step by step decreased from 4 to 1; finally, if a stratum is still unsatisfactory, classes *W* are ignored; thus, the 1st digit of NACE represents the lowest admissible data resolution. In a given step, a statistical unit belonging to a stratum which unfits the threshold is assigned to the next larger stratification level; the latter does not include units previously allocated into strata above the minimum number of observations. A threshold of 80 observations (selected by trial and error to attenuate size differences), gives 76 strata.

	NACE 4+*W*	NACE 3+*W*	NACE 2+*W*	NACE 1+*W*	NACE 1
N. of strata	21	18	22	10	5
N. of obs.	3346	2161	3032	1302	472

Table 3 Summary about collapsed stratification.

1%	5%	10%	25%	50%	75%	90%	95%	99%
77	82	84	94	116	156	200	226	357

Table 4 Quantiles of collapsed stratum sizes.

Without resorting to the collapse of lighter strata, estimations would be not feasible or misleading. The price to pay is the voluntary aggregation of heterogeneous observations. Mixture models are suitable to deal with a wide range of distributional shapes and this circumstance contributes to make less problematic the interpretation of results even though the genuine nature of the detected heterogeneity could always be questioned because of the subjectiveness implicit in the collapsing strategy. The

total number of enterprises declared not compliant w.r.t. the distribution of the remaining ones, by using ε_n=0.25 and α=0.05, is 2143. Conditionally on respective strata, the fraction of candidate atypical units is showed in table 5. The results match the conjecture large size companies are featured by weak exchangeability.

1%	5%	10%	25%	50%	75%	90%	95%	99%
0.117	0.142	0.155	0.184	0.204	0.228	0.262	0.273	0.288

Table 5 Quantiles of the number of "atypical" units divided by the stratum size.

By using the original strata definition according to four NACE digits and two size classes in term of workers, table 6 shows some order statistics about the size of strata to whom suspect units belong. The comparison w.r.t. table 2 makes evident no more than 25% of those units fall in the first 70% of original strata.

5%	10%	25%	50%	60%	70%	75%	80%	85%	90%	95%
4	6	14	38	54	83	117	126	160	200	564

Table 6 Quantiles of sizes featuring original strata the "atypical" units belong to.

5 Conclusions

The detection of statistical units lacking in consistency w.r.t. the process generating the majority of observations, represents a relevant first step to assess the disclosure risk of business microdata. To that aim, in a simplified framework (ignoring survey weight and missing value issues), this paper proposes robust finite Gaussian mixtures and strata collapsing to perform reliable hypotheses tests. It is to note that task is not sufficient to run out the risk assessment. Since those tests are uninformative about the kind of data heterogeneity, any judgment requests further analyses and can imply subjective choices: i.e. if atypical data form a separate cluster, then a threshold should distinguish the harmless heterogeneity (intuitively, the larger the cluster size, the lower the risk). Matters not investigated here will be dealt in future works as well further studies to achieve a global strategy on disclosure risk evaluation.

Acknowledgments. Istat is not responsible for any view or result presented. The author was supported by the European Project BLUE-ETS funded by the European Commission under the 7th Framework Programme, Theme 8, and ESSnet on common tools and harmonised methodology for SDC in the ESS.

References

AA.VV. (2010). *Handbook on Statistical Disclosure Control.* ESSnet on Statistical Disclosure Control. http://neon.vb.cbs.nl/casc/handbook.htm. 18/08/2011

Becker, C., Gather, U. (1999). The masking breakdown point of multivariate outlier identification rules. *Journal of the American Statistical Association.* 94. 947–955

Cerioli, A. (2010). Multivariate outlier detection with high-breakdown estimators. *Journal of the American Statistical Association.* 105. 147–156

Cerioli, A., Farcomeni, A. (2011). Error rates for multivariate outlier detection. *Computational Statistics and Data Analysis*. 55. 544–553

Chew, V. (1966). Confidence, Prediction, and Tolerance Regions for the Multivariate Normal Distribution. *Journal of the American Statistical Association*. 61. 605–617.

Fraley, C., Raftery, A., E. (2002). Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association.* 97. 611–631.

Fraley, C., Raftery A., E. (2006). MCLUST version 3 for R: normal mixture modelling and model-based clustering. *Technical report no. 504, Department of Statistics, University of Washington* (revised 2009).

Hardin, J., Rocke, D., M. (2004). Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. *Computational Statistics & Data Analysis.* 44. 625–638.

Hardin, J., Rocke, D., M. (2005). The distribution of robust distances. *Journal of Computational and Graphical Statistics*. 14. 910–927.

Huber, P., J., Ronchetti, E. (2009). *Robust statistics*. John Wiley and Sons.

Lopuhaä, H., P., Rousseeuw, P., J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*. 19(1). 229–248.

Neykov, N., M., Müller, C., H. (2003). Breakdown Point and Computation of Trimmed Likelihood Estimators in Generalized Linear Models. In: Dutter, R., Filzmoser, P., Gather, U., Rousseeuw, P., J. (Eds.), *Developments in robust statistics.* Physica-Verlag, Heidelberg. 277–286.

Neykov, N., M., Filzmoser, P., Dimova, R. & Neytchev, P. (2007). Robust fitting of mixtures using the Trimmed Likelihood Estimator. *Computational Statistics & Data Analysis.* 17(3). 299–308.

Pison, G., Van Aelst, S. & Willems G. (2002). Small sample corrections for LTS and MCD. *Metrika*. 55. 111–123.

R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3–900051–07–0. http://www.R-project.org/.

Rousseeuw, P., J., Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics.* 41, 212–223.

Tallis, G., M. (1963). Elliptical and radial truncation in normal samples. *Annals of Mathematical Statistics.* 34. 940–944.

Todorov, V., Templ, M. & Filzmoser, P. (2011). Detection of multivariate outliers in business survey data with incomplete information. *Advances in Data Analysis and Classification.* 5(1). 37–56.

Vandev, D., L., Neykov N., M. (1993). Robust maximum likelihood in the Gaussian case. In: Morgenthaler, S., Ronchetti, E. and Stahel, W. A (Eds.), *New Directions in Statistical Data Analysis and Robustness*. Basel. Birkhauser Verlag. 257–264.

Wilks, S., S. (1963). Multivariate Statistical Outliers. *Sankhyā, A*. 25(4). 407–426.