Letter to the Editor: DetMCD is Yet Another Fake-MCD Estimator

David J. Olive*
Department of Mathematics, Southern Illinois University

October 25, 2012

The recent JCGS paper Hubert, Rousseeuw and Verdonck (2012) follows the Rousseeuw Yohai paradigm of replacing an impractical brand name robust estimator by a practical Fake-brand name estimator that is not yet backed by breakdown or large sample theory. A Fake-brand name estimator produces some easily computed trial fits and then uses a criterion from the brand name estimator to select one of the trial fits to create the final Fake-brand name estimator. Hence the brand name MCD estimator is replaced with a Fake-MCD estimator, such as Det-MCD, that selects the trial fit that has the dispersion estimator with the smallest determinant. Most of the literature follows the Rousseeuw Yohai paradigm, using estimators like Fake-MCD, Fake-LTS, Fake-MVE, Fake-S, Fake-LMS, Fake-τ, Fake-Stahel-Donoho, Fake-Projection, Fake-MM, Fake-LTA, Fake-Constrained M, ltsreg, lmsreg, cov.mcd, cov.mve or OGK that are not backed by theory. Maronna, Martin and Yohai (2006, ch. 2, 6) and Hubert, Rousseeuw and Van Aelst (2008) provide references for the above estimators.

Problems with these estimators have been pointed out many times. See, for example, Huber and Ronchetti (2009, pp. xiii, 8-9, 152-154, 196-197) and Hawkins and Olive (2002) with discussion by Hubert, Rousseeuw and Van Aelst (2002) and Maronna and Yohai (2002).

MCD was shown to be $\sqrt{n}$ consistent and asymptotically normal by Cator and Lopuhaä (2010), but MCD takes too long to simulate for $p > 2$ and too long to compute for $p > 5$ variables. The three fastest MCD algorithms are the Agulló (1998) and Pesch (1999) branch and bound algorithms that take a few hours to compute for $n = 100$ and $p = 4$, and the Bernholt and Fischer (2004)

*David J. Olive is Associate Professor, Department of Mathematics, Southern Illinois University, Mailcode 4408, Carbondale, IL 62901-4408, USA. E-mail address: dolive@.siu.edu.

MCD algorithm with complexity higher than $O(n^{p^2/2})$.

For many years, Fake-brand name estimators used the classical estimator computed from K randomly selected elemental sets (of size $h = p$ for regression and $h = p + 1$ for multivariate location and dispersion) as the trial fits. Here K is some fixed number such as $K = 500$ that does not depend on the sample size n. Let d_n be the number of n cases that have been replaced by arbitrarily bad contaminated cases, then the contamination fraction is $\gamma_n = d_n/n$. Zero breakdown estimators have $\gamma_n \to 0$ as $n \to \infty$. Elemental concentration estimators use elemental fits as starts and apply concentration steps to produce trial fits called attractors.

It is a massive error to claim without proof, as done by Hubert, Rousseeuw and Van Aelst (2008), that Fake-MCD efficiently computes MCD. Although Hubert, Rousseeuw and Verdonck (2012) has a section called "Deterministic MCD Algorithm" and page 619 claims MCD became efficient with FASTMCD, Table 1 and p. 635 prove that neither DetMCD nor FASTMCD computes MCD, since if the two estimators were computing MCD, their objective functions would be equal. Since the FASTMCD objective function is sometimes the smallest and the DetMCD objective function is sometimes the smallest, neither estimator can be MCD. Note that Hubert, Rousseeuw and Verdonck (2012) and Rousseeuw and Van Driessen (1999) fail to show that FASTMCD or Det-MCD compute the MCD estimator, are asymptotically equivalent to MCD, have a limiting distribution, are consistent or are high breakdown! Hence the p. 619 claim that MCD has been applied in various fields is FALSE. FASTMCD, which HAS NO THEORY, has been applied in various fields. Thus the impractical MCD estimator has been replaced by the practical FASTMCD and DetMCD estimators that are not backed by theory. Since FASTMCD is neither fast nor the MCD estimator, a more descriptive name is Fake-MCD.

Page 619 notes that FASTMCD needs to draw many elemental subsets to obtain one that is outlier free. Actually 500 are drawn and the outlier resistance decreases rapidly as the number of variables p increases. This claim can be seen in table 2 of Rousseeuw and Van Driessen (1999), where the p = 30 entry is almost certainly p = 20. Also p = 20 agrees with the approximate amount of outliers an elemental concentration algorithm can tolerate given in Hawkins and Olive (2002) if p+1 replaces p.

False claims that impractical high breakdown estimators such as MCD can be computed with a practical estimator like FAST-MCD that have no theory have plagued the field of Robust Statistics for nearly 30 years, and it will be difficult to make progress until papers that give theory to the estimator that is actually used, as done in Olive and Hawkins (2010), are published, and papers that give no theory are rejected or retracted.

The following theorem shows that the estimators from Rousseeuw (1984), Rousseeuw and Leroy (1987) and Rousseeuw and van Zomeren (1990) are zero breakdown and inconsistent. These elemental basic resampling estimators use the PROGRESS algorithm where $K \leq 30000$ and the default is $K = 3000$. The theorem also shows that a variant of the Rousseeuw and Van Driessen (1999) Fake-MCD estimator that uses $K = 500$ elemental starts is zero breakdown. (If we let $K \equiv K_n \to \infty$, then the elemental estimator is zero breakdown if $K_n = o(n)$. A necessary condition for the elemental basic resampling estimator to be consistent is $K_n \to \infty$.)

Theorem 1: a) The elemental basic resampling algorithm estimators are inconsistent. b) The elemental concentration and elemental basic resampling algorithm estimators are zero breakdown.

Proof: a) Note that you can not get a consistent estimator by using Kh randomly selected cases since the number of cases Kh needs to go to ∞ for consistency except in degenerate situations.

b) Contaminating all Kh cases in the K elemental sets shows that the breakdown value is bounded by $Kh/n \to 0$, so the estimator is zero breakdown. QED

The Det-MCD estimator chooses six intelligently chosen starts for concentration. This idea is not new, and Det-MCD is also described in Verdonck, Hubert, and Rousseeuw (2010). Olive (2004, p. 100) suggests that a simple modification to the median ball algorithm (MBA) estimator is to use several starts such as $(MED(\boldsymbol{X}), diag([MAD(X_1)]^2, ..., [MAD(X_p)]^2))$ where $MED(\boldsymbol{X})$ is the coordinatewise median and $MAD(X_i)$ is the median absolute deviation of the ith variable; OGK; $(MED(\boldsymbol{X}), \boldsymbol{I})$ where $\boldsymbol{I}$ is the identity matrix resulting in the median ball (MB) estimator; and the classical sample mean and covariance matrix estimator $(\overline{\boldsymbol{x}}, \boldsymbol{S})$ resulting in the Devlin, Gnanadesikan and Kettenring (1981) DGK estimator (called iterative trimming in Rousseeuw and Leroy (1987, p. 254) who suggest that the breakdown value of DGK is at most $1/p$). Similar methods are described in Gnanadesikan and Kettenring (1972) and Gnanadesikan (1977, p. 134).

The Olive Hawkins paradigm is to develop practical robust estimators backed by large sample and breakdown theory. Olive (2012 ∮ 4.4; 2014 ∮ 9.2, 10.7) and Olive and Hawkins (2010, 2011) develop the practical $\sqrt{n}$ consistent robust estimators FCH, RFCH, RMVN, and HBREG, using concentration on a few intelligently chosen starts. Also see Zhang, Olive and Ye (2012). This theory is sketched below where $D_i^2(T, \boldsymbol{C})$ is the ith squared Mahalanobis distance.

The following theorem implies that applying k concentration steps to a high breakdown start $(T_{-1}, \boldsymbol{C}_{-1})$ results in a high breakdown attractor $(T_k, \boldsymbol{C}_k)$. Hence the Olive (2004) MB estimator is high breakdown. Note that the number of steps is fixed, e.g., use $k = 5$ concentration steps. It is not known whether the result holds if concentration is iterated to convergence as done for the Det-MCD estimator. It is also not clear that any of the six starts for Det-MCD are high breakdown. Croux, Dehon, and Yadine (2010) appear to show that the sign covariance matrix estimator is high breakdown, and it may be possible to adapt their proof for one of the Det-MCD starts.

Theorem 2. Suppose $(T, \boldsymbol{C})$ is a high breakdown estimator where $\boldsymbol{C}$ is a symmetric, positive definite $p \times p$ matrix if the contamination proportion d_n/n is less than the breakdown value. Then the concentration attractor $(T_k, \boldsymbol{C}_k)$ is a high breakdown estimator if the coverage $c_n \approx n/2$ and the data are in general position.

The assumption below gives the class of distributions for which FCH, RFCH and RMVN have been shown to be $\sqrt{n}$ consistent. Distributions where the MCD functional is unique are called "unimodal," and rule out, for example, a spherically symmetric uniform distribution.

Assumption (E1): The $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid from a "unimodal" elliptically contoured $EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ distribution with nonsingular covariance matrix $\text{Cov}(\boldsymbol{x}_i)$ where g is continuously differentiable with finite 4th moment: $\int (\boldsymbol{x}^T\boldsymbol{x})^2 g(\boldsymbol{x}^T\boldsymbol{x}) d\boldsymbol{x} < \infty$.

Theorem 3, Lopuhaä (1999). Suppose $(T, \boldsymbol{C})$ is a consistent affine equivariant estimator of $(\boldsymbol{\mu}, s\boldsymbol{\Sigma})$ with rate n^δ where $s > 0$ and $0 < \delta \leq 0.5$. Assume (E1) holds. Then the classical estimator $(T_0, \boldsymbol{C}_0)$ applied to the cases with $D_i^2(T, \boldsymbol{C}) \leq h^2$ is a consistent affine equivariant estimator of $(\boldsymbol{\mu}, a\boldsymbol{\Sigma})$ with the same rate n^δ where $a > 0$. The constant a depends on the positive constants s, h^2, p and the elliptically contoured distribution, but does not otherwise depend on the consistent start $(T, \boldsymbol{C})$.

Let $\delta = 0.5$. Applying the above theorem iteratively for a fixed number k of steps produces a sequence of estimators $(T_0, \boldsymbol{C}_0), ..., (T_k, \boldsymbol{C}_k)$ where $(T_j, \boldsymbol{C}_j)$ is a $\sqrt{n}$ consistent affine equivariant estimator of $(\boldsymbol{\mu}, a_j\boldsymbol{\Sigma})$ where the constants $a_j > 0$ depend on s, p, h and the elliptically contoured distribution, but do not otherwise depend on the consistent start $(T, \boldsymbol{C}) \equiv (T_{-1}, \boldsymbol{C}_{-1})$.

Concentration applies the classical estimator to $c_n \approx n/2$ cases with $D_i^2(T, \boldsymbol{C}) \leq D_{(c_n)}^2(T, \boldsymbol{C})$.

Olive and Hawkins (2010) show that if $(T, \boldsymbol{C})$ is a $\sqrt{n}$ consistent affine equivariant estimator of $(\boldsymbol{\mu}, s\boldsymbol{\Sigma})$ then $(T, \tilde{\boldsymbol{C}}) \equiv (T, D^2_{(c_n)}(T, \boldsymbol{C})\ \boldsymbol{C})$ is a $\sqrt{n}$ consistent affine equivariant estimator of $(\boldsymbol{\mu}, b\boldsymbol{\Sigma})$ where $b = D^2_{0.5}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the population median of the population squared distances, and that $D^2_i(T, \tilde{\boldsymbol{C}}) \leq 1$ is equivalent to $D^2_i(T, \boldsymbol{C}) \leq D^2_{(c_n)}(T, \boldsymbol{C}))$. Hence Lopuhaä (1999) theory applied to $(T, \tilde{\boldsymbol{C}})$ with $h = 1$ is equivalent to theory applied to the concentration estimator using the affine equivariant estimator $(T, \boldsymbol{C}) \equiv (T_{-1}, \boldsymbol{C}_{-1})$ as the start. Since b does not depend on s, concentration produces a sequence of estimators $(T_0, \boldsymbol{C}_0), ..., (T_k, \boldsymbol{C}_k)$ where $(T_j, \boldsymbol{C}_j)$ is a $\sqrt{n}$ consistent affine equivariant estimator of $(\boldsymbol{\mu}, a\boldsymbol{\Sigma})$ where the constant $a > 0$ is the same for each j. Then Olive and Hawkins (2010) show that the DGK and MCD are $\sqrt{n}$ consistent estimators of $(\boldsymbol{\mu}, a_{MCD}\boldsymbol{\Sigma})$. Note that the DGK estimator is practical to compute but has a much lower breakdown value than the impractical MCD estimator. Boente (1987) has some large sample theory for estimators similar to DGK except that a continuous weight function is used instead of zero one weighting.

Hence a fixed number of concentration steps applied to a $\sqrt{n}$ consistent affine equivariant start results in a $\sqrt{n}$ consistent attractor. For Det-MCD, a similar result is needed when the start is not affine equivariant and concentration is iterated to convergence. It is not clear that any of the Det-MCD starts are $\sqrt{n}$ consistent. Proofs for OGK are not given. Boente and Fraiman (1999) claim that the sign covariance matrix is consistent, also see Taskinen, Koch, and Oja (2012).

Theorem 4, Olive and Hawkins (2010). Assume (E1) holds. a) Then the DGK estimator and MCD estimator are $\sqrt{n}$ consistent affine equivariant estimators of $(\boldsymbol{\mu}, a_{MCD}\boldsymbol{\Sigma})$.

b) The FCH, RFCH and RMVN estimators are $\sqrt{n}$ consistent estimators of $(\boldsymbol{\mu}, c_i\boldsymbol{\Sigma})$ for $c_1, c_2, c_3 > 0$ where $c_i = 1$ for multivariate normal data. If the clean data are in general position, then T_{FCH} is a high breakdown estimator and $\boldsymbol{C}_{FCH}$ is nonsingular even if nearly half of the cases are outliers.

It will be a massive undertaking to modify the theory to show whether Det-MCD has any good large sample or breakdown properties. For applications comparing estimators from the Olive Hawkins paradigm with those from the Rousseeuw Yohai paradigm, see Alkenani and Yu (2012), Ng and Wilcox (2010), Özdemira and Wilcox (2012), Park, Kim and Kim (2012), Wilcox (2012), and Reyen, Miller and Wegman (2009) who simulate the OGK and MBA estimators for $p = 100$ and n up to 50000. The OGK complexity is $O[p^3 + np^2 \log(n)]$ while that of MBA, FCH, RFCH and RMVN is $O[p^3 + np^2 + np\log(n)]$. These four estimators are roughly 100 times faster than Fake-MCD and also much faster than Det-MCD.

References

[1] Agulló, J., (1998). Computing the minimum covariance determinant estimator. Unpublished manuscript, Universidad de Alicante.

[2] Alkenani, A., and Yu, K. (2012), A comparative study for robust canonical correlation methods. *Journal of Statistical Computation and Simulation,* to appear.

[5] Bernholt, T., and Fischer, P. (2004). The complexity of computing the MCD-estimator. *Theoretical Computer Science* **326**, 383-398.

[4] Boente, G. (1987). Asymptotic theory for robust principal components. *Journal of Multivariate Analysis* **21**, 67-78.

[5] Boente, G., and Fraiman, R. (1999). Discussion of 'Robust principal component analysis for functional data' by Locantore et al. *Test* **8**, 28-35.

[6] Cator, E.A., and Lopuhaä, H.P. (2010). Asymptotic expansion of the minimum covariance determinant estimators. *Journal of Multivariate Analysis* **101**, 2372-2388.

[7] Croux, C., Dehon, C., and Yadine, A. (2010). The k-step spatial sign covariance matrix. *Advances in Data Analysis and Classification* **4**, 137-150.

[8] Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association* **76**, 354-362.

[9] Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York, NY.

[10] Gnanadesikan, R., and Kettenring, J.R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28**, 81-124.

[11] Hawkins, D.M., and Olive, D.J. (2002). Inconsistency of resampling algorithms for high breakdown regression estimators and a new algorithm (with discussion). *Journal of the American Statistical Association* **97**, 136-159.

[12] Huber, P.J., and Ronchetti, E.M. (2009). *Robust Statistics,* 2nd ed., Wiley, Hoboken, NJ.

[13] Hubert, M., Rousseeuw, P.J., and Van Aelst, S. (2002). Comment on 'Inconsistency of resampling algorithms for high breakdown regression and a new algorithm' by D.M. Hawkins and D.J. Olive. *Journal of the American Statistical Association* **97**, 151-153.

[14] Hubert, M., Rousseeuw, P.J., and Van Aelst, S. (2008). High breakdown multivariate methods. *Statistical Science* **23**, 92-119.

[15] Hubert, M., Rousseeuw, P.J., and Verdonck, T. (2012). A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics* **21**, 618-637.

[16] Lopuhaä, H.P. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *The Annals of Statistics* **27**, 1638-1665.

[17] Maronna, R.A., Martin, R.D., and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*, Wiley, Hoboken, NJ.

[18] Maronna, R.A., and Yohai, V.J. (2002). Comment on 'Inconsistency of resampling algorithms for high breakdown regression and a new algorithm' by D.M. Hawkins and D.J. Olive. *Journal of the American Statistical Association* **97**, 154-155.

[19] Ng, M., and Wilcox, R.R. (2010). The small-sample efficiency of some recently proposed multivariate measures of location. *Journal of Modern Applied Statistical Methods* **9**, 28-42.

[20] Olive, D.J. (2004), A resistant estimator of multivariate location and dispersion. *Computational Statistics and Data Analysis* **46**, 99-102.

[21] Olive, D.J. (2012). *Robust Multivariate Analysis.* Online course notes, see (www.math.siu.edu/olive/multbk.htm).

[22] Olive, D.J. (2014). *Robust Statistics.* To appear, incorporates the two preprints below.

[23] Olive, D.J., and Hawkins, D.M. (2010). Robust multivariate location and dispersion. Preprint, see (www.math.siu.edu/olive/preprints.htm).

[24] Olive, D.J., and Hawkins, D.M. (2011). Practical high breakdown regression. Preprint, see (www.math.siu.edu/olive/preprints.htm).

[25] Özdemira, A.F., and Wilcox, R. (2012). New results on the small-sample properties of some robust univariate estimators of location. *Communications in Statistics: Simulation and Computation* **41**, 1544-1556.

[26] Park, Y., Kim, D., and Kim, S. (2012). Robust regression using data partitioning and M-estimation. *Communications in Statistics: Simulation and Computation* **8**, 1282-1300.

[27] Pesch, C. (1999). Computation of the minimum covariance determinant estimator. In *Classification in the Information Age, Proceedings of the 22nd Annual GfKl Conference, Dresden 1998,* eds. Gaul W., and Locarek-Junge, H., Springer, Berlin, 225-232.

[28] Reyen, S.S., Miller, J.J., and Wegman, E.J. (2009). Separating a mixture of two normals with proportional covariances. *Metrika* **70**, 297-314.

[29] Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association* **79**, 871-880.

[30] Rousseeuw, P.J., and Leroy, A.M. (1987). *Robust Regression and Outlier Detection,* Wiley, New York, NY.

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

[32] Rousseeuw, P.J., and van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association* **85**, 633-651.

[33] Taskinen, S., Koch, I., and Oja, H. (2012). Robustifying principal component analysis with spatial sign vectors. *Statistics & Probability Letters* **82**, 765-774.

[34] Verdonck, T., Hubert, M., and Rousseeuw, P.J. (2010). DetMCD in a calibration framework. In *Proceedings in Computational Statistics, 2010*, eds. Lechevallier, Y., and Saporta, G., Springer-Verlag, Heidelberg, 589-596.

[35] Wilcox, R.R. (2012). *Introduction to Robust Estimation and Testing,* 3rd ed., Academic Press, San Diego, CA.

[36] Zhang, J., Olive, D.J., and Ye, P. (2012). Robust covariance matrix estimation with canonical correlation analysis. *International Journal of Statistics and Probability* **1**, 119-136.