Submitted to the Annals of Statistics

LETTER TO THE EDITOR

By David J. Olive

Southern Illinois University

Bali, Boente, Tyler and Wang (2011) gave possibly impressive theory for infinite complexity impractical robust projection estimators, but should have given theory for the practical Fake-projection estimator actually used. This "bait and switch error" occurs far too often in multivariate "robust statistics" papers.

To estimate the first principal direction, the Fake-projection (CR) estimator uses n projections $\boldsymbol{z}_i = \boldsymbol{w}_i/\|\boldsymbol{w}_i\|$ where $\boldsymbol{w}_i = \boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_n$. Note that for $p = 2$ one can select 360 projections (through the origin and a point) on the unit circle that are one degree apart. Then there is a projection that is highly correlated with any projection on the unit circle. If $p = 3$, then 360 projections are not nearly enough to adequately approximate all projections through the unit sphere. Since the surface area of a unit hypersphere is proportional to n^{p-1}, approximations rapidly get worse as p increases.

Theory for the Fake-projection (CR) estimator may be simple. Suppose the data is multivariate normal $N_p(\boldsymbol{0}, diag(p, 1, ..., 1))$. Then $\boldsymbol{\beta} = (1, 0, ..., 0)^T$ (or $-\boldsymbol{\beta}$) is the population first direction. Heuristically, assume $\hat{\boldsymbol{\mu}}_n = \boldsymbol{0}$, although in general $\hat{\boldsymbol{\mu}}_n$ should be a good $\sqrt{n}$ consistent estimator of $\boldsymbol{\mu}$ such as the coordinatewise median. Let $\boldsymbol{b}_o$ be the "best" estimated projection $\boldsymbol{z}_j$ that minimizes $\|\boldsymbol{z}_i - \boldsymbol{\beta}\|$ for $i = 1, ..., n$. "Good" projections will have a $\boldsymbol{y}_i$ that lies in one of two "hypercones" with a vertex at the origin and centered about a line through the origin and $\pm\boldsymbol{\beta}$ with radius r at $\pm\boldsymbol{\beta}$. So for $p = 2$ the two "cones" are determined by the two lines through the origin with slopes $\pm\, r$. The probability that a randomly selected $\boldsymbol{y}_i$ falls in one of the two "hypercones" is proportional to r^{p-1}, and for $\boldsymbol{b}_o$ to be consistent for $\boldsymbol{\beta}$ need $r \to 0$, P(at least one $\boldsymbol{y}_i$ falls in "hypercone") $\to 1$ and $n \to \infty$. If these heuristics are correct, need $r \propto n^{\frac{-1}{p-1}}$ for $\|\boldsymbol{b}_o - \boldsymbol{\beta}\| = O_P(n^{\frac{1}{p-1}})$. Note that $\boldsymbol{b}_o$ is not an estimator since $\boldsymbol{\beta}$ is not known, but the rate of the "best" projection $\boldsymbol{b}_o$ gives an upper bound on the rate of the Fake-projection estimator $\boldsymbol{v}_1$ since $\|\boldsymbol{v}_1 - \boldsymbol{\beta}\| \geq \|\boldsymbol{b}_o - \boldsymbol{\beta}\|$. If the scale estimator is $\sqrt{n}$ consistent, then for a large class of elliptically contoured distributions, a conjecture is that $\|\boldsymbol{v}_1 - \boldsymbol{\beta}\| = O_P(n^{\frac{1}{2(p-1)}})$ for $p > 1$.

There has been a breakdown in the research and refereeing of multivariate "robust statistics." The Rousseeuw Yohai paradigm is to replace the impractical brand name estimator by a practical Fake-estimator that computes no more than a few thousand easily computed trial fits, but no breakdown or large sample theory is given for the Fake-estimator (the "bait and switch error"). 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, 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. Most of the brand name estimators were invented in papers by Rousseeuw, Yohai, Maronna or Tyler.

Problems with these estimators have been pointed out many times. See, for example, Huber and Ronchetti (2009, p. xiii, 8-9, 152-154, 196-197). Also Hawkins and Olive (2002) proved elemental concentration estimators are zero breakdown. Maronna and Yohai (2002) correctly note that elemental concentration estimators are inconsistent if the number of concentration steps is finite, but consistency is not known if the concentration is iterated to convergence. So it is not known whether Fake-MCD and Fake-LTS are consistent.

Since *The Annals of Statistics* frequently publishes papers on "robust statistics," some theory for the Fake-estimators actually used will be given after some notation. Let $p =$ the number of predictors. The Fake-MCD and Fake-S estimators are zero breakdown variants of the elemental concentration and elemental resampling algorithms that use K elemental fits where K

is a fixed number that does not depend on the sample size n. To produce an elemental fit, randomly select h cases and compute the classical estimator $(T_i, \boldsymbol{C}_i)$ (or $T_i = \hat{\boldsymbol{\beta}}_i$ for regression) for these cases, where $h = p$ for multiple linear regression and $h = p+1$ for multivariate location and dispersion. The elemental resampling algorithm uses one of the K elemental fits as the estimator, while the elemental concentration algorithm refines the K elemental fits using all n cases. See Olive and Hawkins (2010, 2011) for more details.

Breakdown is computed by determining the smallest number of cases d_n that can be replaced by arbitrarily bad contaminated cases in order to make $\|T\|$ or $\|\hat{\boldsymbol{\beta}}\|$ arbitrarily large or to drive the smallest or largest eigenvalues of the dispersion estimator $\boldsymbol{C}$ to 0 or ∞. High breakdown estimators have $\gamma_n = d_n/n \to 0.5$ and zero breakdown estimators have $\gamma_n \to 0$ as $n \to \infty$. Note that an estimator can not be consistent for θ unless the number of randomly selected cases goes to ∞, except in degenerate situations. The following theorem shows Fake-MCD and Fake-S are zero breakdown estimators. (If $K_n \to \infty$, 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

Theorem 1 shows that the elemental basic resampling PROGRESS estimators of Rousseeuw and Leroy (1987) and Rousseeuw and van Zomeren (1990) are zero breakdown and inconsistent. Yohai's two stage estimators, such as MM, need initial consistent high breakdown estimators such as LMS, MCD or MVE, but were implemented with the inconsistent zero breakdown elemental estimators such as Fake-LMS, Fake-MCD or Fake-MVE. See Hawkins and Olive (2002, p. 157).

Some workers claim that their elemental estimators search for sets for which the classical estimator can be computed, hence the above trivial results do not hold. (This claim does not excuse the fact that the workers fail to provide any large sample or breakdown theory for their "practical estimators.") For practical estimators, this claim is false since the estimator will not be practical if the program goes into an endless loop or searches all $O(n^p)$ elemental sets when supplied with messy data. For example, the Rousseeuw and Leroy (1987) PROGRESS algorithm starts with a default of $K_d = 3000$ and ends with no more than $K \leq 30000$ elemental sets. Fake-MCD starts with $K_d = 500$.

There is an alternative to the Rousseeuw Yohai paradigm. Use the estimators of Olive and Hawkins (2010, 2011) who avoid the "bait and switch error" by giving theory for the practical HBREG, FCH, RFCH and RMVN estimators actually used in the software.

REFERENCES

Bali, J.L., Boente, G., Tyler, D.E. and Wang, J.L. (2011). Robust functional principal components: a projection-pursuit approach. *Ann. Statist.* **39** 2852-2882.

Hawkins, D.M. and Olive, D.J. (2002). Inconsistency of resampling algorithms for high breakdown regression estimators and a new algorithm (with discussion). *J. Amer. Statist. Assoc.* **97** 136-159.

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

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

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

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. *J. Amer. Statist. Assoc.* **97** 154-155.

OLIVE, D.J., and HAWKINS, D.M. (2010). Robust multivariate location and dispersion. Preprint, see (www.math.siu.edu/olive/preprints.htm).

OLIVE, D.J., and HAWKINS, D.M. (2011). Practical high breakdown regression. Preprint, see (www.math.siu.edu/olive/preprints.htm).

ROUSSEEUW, P.J., and LEROY, A.M. (1987). *Robust Regression and Outlier Detection,* Wiley, New York, NY.

ROUSSEEUW, P.J., and VAN ZOMEREN, B.C. (1990). Unmasking multivariate outliers and leverage points. *J. Amer. Statist. Assoc.* **85** 633-651.

DEPARTMENT OF MATHEMATICS
SOUTHERN ILLINOIS UNIVERSITY
MAILCODE 4408
CARBONDALE, IL 62901-4408, USA.
E-MAIL: dolive@math.siu.edu