An Extension of a Method of Hardin and Rocke, with an Application to Multivariate Outlier Detection via the IRMCD Method of Cerioli

Christopher G. Green[1] and Doug Martin[2]

University of Washington, Seattle, WA, USA

Working Paper
Department of Statistics
University of Washington
Seattle, WA

Revision date: **July 28, 2014**

Abstract

Cerioli (2010) introduced a Mahalanobis distance-based multivariate outlier detection methodology, the Iterated Reweighted Minimum Covariance Determinant (IRMCD) method, that is more accurate, both in the presence and absence of outliers, than previous approaches using MCD-based Mahalanobis distances. The IRMCD method uses an approximation, developed by Hardin and Rocke (2005), for the finite sample distribution of MCD-based squared Mahalanobis distances. This approximation, however, was developed for the maximum-breakdown point case of the MCD estimate; its performance for MCD with other subsample sizes has not been studied to our knowledge. Thus it is not clear that one can rely on the IRMCD approach when using MCD with subsample sizes other than the maximum-breakdown point case. These MCD variants are of interest in applications (e.g., with financial data) where one is hesitant to discard too much of the data in building location and dispersion estimates. Hence, in this technical note, we investigate how well Hardin and Rocke's approximation works for the MCD in general. We use a simulation experiment more extensive that the experiment performed by Hardin and Rocke to construct a modified approximation for use with MCD-based squared Mahalanobis distances. We conclude by validating the IRMCD method for some MCD variants with larger subsample sizes that are of interest to financial practitioners.

Keywords: outlier detection, Mahalanobis distances, minimum covariance determinant, robust estimation

[1]Address for correspondence: Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322, USA. E-mail: `cggreen@uw.edu`

[2]Department of Applied Mathematics, Box 353925, Univerity of Washington, Seattle, WA 98195-3925, USA.

1 Introduction

One traditional method of detecting outliers in multivariate data is through the use of Mahalanobis distances. Mahalanobis distances, introduced by Mahalanobis (1936), measure the distance of an observation from the mean of a distribution, weighted by the correlation information contained in the covariance matrix (Seber, 1984). Formally, let $\mathbf{x}$ be an observation from a multivariate distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$; then the Mahalanobis distance of $\mathbf{x}$ from $\boldsymbol{\mu}$ is defined as

$$(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}). \tag{1}$$

Compare this definition to the usual Euclidean distance of $\mathbf{x}$ from $\boldsymbol{\mu}$:

$$(\mathbf{x}-\boldsymbol{\mu})^T (\mathbf{x}-\boldsymbol{\mu}).$$

The extra $\boldsymbol{\Sigma}^{-1}$ factor captures the fact that the distribution may not look the same in each direction; for example, it may be more dispersed in one direction, so an observation that is far from the mean in a Euclidean sense may not be "unusually" far away once covariance information is taken into account.

When $\mathbf{x}$ is ν-dimensional multivariate normal with known mean and covariance, the population Mahalanobis distance is distributed as a chi-squared χ^2_ν random variable with ν degrees of freedom (Mardia et al., 1979; Seber, 1984). This suggests a test of deviation from the multivariate normal assumption: compare an observation's Mahalanobis distance to an appropriate quantile of the chi-squared distribution. In common practice the unknown mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ are replaced by their classical estimates $\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$, the coordinate-wise sample mean, and

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T, \tag{2}$$

the sample covariance matrix. When the $\mathbf{x}_i$ are multivariate normal, the resulting sample Mahalanobis distances are approximately chi-squared for "moderate" values of n, but "moderate" increases with the dimension p (Andrews et al., 1973; Small, 1978). The exact marginal distribution of the sample Mahalanobis distances was shown to be a scaled Beta distribution by Gnanadesikan and Kettenring (1972) (using an earlier result of Wilks (1962)), but in practice the chi-squared approximation is used for simplicity[3].

Since the classical covariance estimator (2) is not robust to outliers (see, for instance, Maronna et al. (2006)), using it in the Mahalanobis distance metric could lead to some good observations being flagged as outliers (Rousseeuw and van Zomeren, 1990, 1991; Becker and Gather, 1999; Peña and Prieto, 2001; Cuesta-Albertos et al., 2008). Moreover, when there are multiple outliers, the classical Mahalanobis distance metric may lead to masking of moderate outliers by one extreme outlier (Rocke and Woodruff, 1996; Pearson and Chandra Sekar, 1936). So-called "robust" Mahalanobis distances are the result of replacing the classical mean and covariance estimates in Equation (1) with robust estimates of location and dispersion, e.g., the location and dispersion estimate that results from the minimum covariance determinant (MCD) estimator (Rousseeuw, 1985; Rousseeuw and van Driessen, 1999; Hawkins and Olive, 1999). The robust estimate downweights or ignores the outliers, and thus provides a better representation of the location and dispersion of the majority of the data. Non-outlying points should hence be closer to the location estimate than outlying

[3]It is possible as well that some users are not aware that the sample Mahalanobis distances are not exactly chi-squared distributed, or that the accuracy of the approximation depends on the dimension of the data.

points, and outlying points should have larger distances than expected under the multivariate normal model.

The distributional assumption used to test the distances in the classical case, namely that the distances are iid chi-squared random variables, only holds asymptotically in the robust case when the dispersion estimate is consistent for $\mathbf{\Sigma}$ (Mardia et al., 1979; Serfling, 1980; Seber, 1984). The exact finite-sample distribution is not known for any of the common robust dispersion estimates. Thus, determining appropriate cutoff values for the Mahalanobis distance test is more challenging in the robust case than in the classical case. For the commonly used minimum covariance determinant (MCD) estimate of Rousseeuw (1985), the chi-squared approximation is easy to implement and readily available in most software, but it is known to be inaccurate in small to moderate samples, especially when the data contain no outliers (Rousseeuw and van Zomeren, 1991; Becker and Gather, 2001; Hardin and Rocke, 2005; Cerioli et al., 2009; Riani et al., 2009).

Hardin and Rocke (2005) studied how well tests of (non-reweighted) MCD-based Mahalanobis distances detected multivariate outliers. For a given set of n data points of dimension ν and a subset size $h \leq n$, we calculate the MCD estimate of the dispersion matrix of the data points by finding the subset of size h whose (classical) covariance matrix has the smallest determinant (Rousseeuw, 1985). For convenience of notation we will work with the fraction $\alpha_{\mathrm{MCD}} \equiv h/n$, and denote the MCD estimate based on subsets of size h by $\mathrm{MCD}(\alpha_{\mathrm{MCD}})$. Hardin and Rocke focused on the most commonly used version of the MCD estimate, in which the subsample size h is chosen to achieve the maximum possible breakdown point of the estimator[4]; we will denote the resulting ratio h/n as $\alpha_{\mathrm{MCD}}^{\mathrm{MBP}}$, but will denote the corresponding estimate as "MCD" where the context permits us to do so without confusing the reader.

Hardin and Rocke (2005) established that, in the maximum breakdown-point MCD case, the χ^2_ν cutoff values can be too small in small to moderate samples, depending on the dimension ν. This can result in many observations being incorrectly flagged as outliers. They proposed a scaled F distribution for determining appropriate cutoff values for testing robust Mahalanobis distances. Their method gave better results (in terms of detecting an appropriate number of outliers) for the (non-reweighted) MCD-based Mahalanobis distance tests than the standard χ^2_ν-based tests. The simulation study of Cerioli, Riani, and Atkinson (2009) affirmed that the Hardin-Rocke quantiles were more accurate (for testing individual observations for outlyingness) in moderate to large sample sizes than the chi-squared quantiles for the MCD case.

Cerioli et al. (2009) looked at how well MCD-based Mahalanobis distances performed both in an individual testing framework ("is this observation an outlier?") and under a simultaneous testing framework ("are there any outliers in the data?"). First they conducted a simulation experiment in which each observation was tested for outlyingness at some nominal test size (say, $\alpha = 0.01$). Ideally we should see about $\lfloor \alpha n \rfloor$ incorrectly flagged observations; their simulations show this is not the case for the non-reweighted $\mathrm{MCD}(\alpha_{\mathrm{MCD}}^{\mathrm{MBP}})$ with χ^2_ν cutoff values, as well as for the one-step reweighted $\mathrm{MCD}(\alpha_{\mathrm{MCD}}^{\mathrm{MBP}})$ with χ^2_ν cutoff values. The χ^2_ν methodology requires large sample sizes to be reliable for the MCD and the reweighted MCD, with the needed sample size increasing with dimension ν. For small to moderate sample sizes the χ^2_ν cutoff values can give significantly more false positives than expected based on the nominal test size. The Hardin-Rocke F distribution is more accurate for moderate smaple sizes, independent of dimension, though it can be too large for very small samples.

Cerioli et al. (2009) then looked at the accuracy of tests of the intersection null hypothesis

$$H_0 : \{\mathbf{x}_1 \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})\} \cap \cdots \cap \{\mathbf{x}_n \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})\} \tag{3}$$

[4]The formula for h in this case is given by $h = \lfloor (n + \nu + 1)/2 \rfloor$.

that examines whether there are any outliers in the data (as opposed to whether a given observation is outlying). The obvious way to perform this test is via comparison of the largest distance to a χ^2_ν quantile (Becker and Gather, 1999, 2001) or a Hardin-Rocke F quantile, with a Bonferroni-corrected size α/n. Again via a simulation study Cerioli, Riani, and Atkinson showed that the χ^2_ν quantile works poorly for testing the intersection hypothesis with the MCD($\alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$). The reweighted MCD($\alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$) is no better, even if a Bonferroni correction is used in the weighting step in addition to the cutoff statistic. Finally, the Hardin-Rocke F distribution, which worked rather well in the individual testing framework, fails to be of much help in the simultaneous framework. Cerioli et al. suggest that the Hardin-Rocke distributional approximation does not work well for the small quantiles required by the Bonferroni correction.

Cerioli et al. (2009) also developed calibration factors via simulation to correct the size of the MCD and reweighted MCD outlier tests. The simulation is computationally expensive and thus difficult to use for corrections for MCD(α_{MCD}) for $\alpha_{\mathrm{MCD}} \neq \alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$.[5] Fortunately, Cerioli (2010) introduced a simpler correction methodology, the Iterated Reweighted MCD (IRMCD), that does not require simulation for calibration, and is easier to extend to MCD(α_{MCD}) in general.

The follow-up paper (Cerioli, 2010) focuses on the $\alpha_{\mathrm{MCD}} = \alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$ case but also looks at $\alpha_{\mathrm{MCD}} = 0.75$. However, the Hardin-Rocke methodology, which is used as part of the IRMCD methodology developed in that paper, was developed with $\alpha_{\mathrm{MCD}} = \alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$ in mind. We were not aware of any studies examining how well Hardin-Rocke works for other values of α_{MCD}, so we examined the appropriateness of those cutoff values via simulation. Figures 1 and 3 show how the Wishart degrees of freedom m estimates obtained (i) from simulation and (ii) from the Hardin-Rocke adjustment to the asymptotic degrees of freedom estimate (from Croux and Haesbroeck (1999)) compare, while Figures 2 and 4 show how the resulting 0.01 critical values computed using Hardin and Rocke's F distribution compare. For $\alpha_{\mathrm{MCD}} = 0.75$ (Figures 1 and 2), the two degrees of freedom estimates diverge for small to moderate sample size, with the Hardin-Rocke approach leading to values of m that are a bit larger than those obtained from simulation. The quantiles computed using the Hardin-Rocke m are a bit smaller than they should be for moderate sample sizes and much smaller for small sample sizes. The divergence in the Wishart parameter is similar for $\alpha_{\mathrm{MCD}} = 0.95$ (Figure 3), and the critical values computed using the Hardin-Rocke adjusted m are again too small in small to moderate sample sizes (Figure 4).

In general we observe that the Hardin-Rocke adjusted degrees of freedom is still reasonably accurate for producing 0.01 critical values for values of $\alpha_{\mathrm{MCD}} \neq \alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$, but can result in quantiles that are too small (compared to those obtained using simulated degrees of freedom) for small to moderate samples and as α_{MCD} gets closer to 1.[6] This behavior is not surprising: Figure 5 shows how the 0.99 quantile from the Hardin-Rocke scaled F distribution varies with the input parameters m and ν. (Since the quantile increases rapidly as m goes to 0, the plot shows the logarithm of the quantile.) For fixed values of ν, larger values of m lead to smaller quantiles. Thus using the Hardin-Rocke adjustment for α_{MCD} near 1 and/or small samples will result in flagging too many observations as outliers.[7] Therefore prior to using Cerioli's methodology in other studies where $\alpha_{\mathrm{MCD}} \neq \alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$ we had to extend the Hardin-Rocke methodology.

In this technical note, we extend the results of Hardin and Rocke (2005) to arbitrary values of $0.5 \leq \alpha_{\mathrm{MCD}} < 1$. We then verify the calibration procedure presented in Cerioli (2010) for the MCD with $\alpha_{\mathrm{MCD}} =$

[5] The authors provide an easier to use formula based on interpolation from their simulation work, but that only addresses MCD($\alpha^{\mathrm{MBP}}_{\mathrm{MCD}}$).

[6] We observed similar behavior for the 0.025 and 0.05 critical values.

[7] This assumes that Hardin-Rocke's approximation to the distribution of the non-MCD subset Mahalanobis distances is sufficiently accurate. For tests at typical sizes, e.g., 0.01, we know this to be true, but (Cerioli et al., 2009) hinted that this might not be true for smaller (e.g., Bonferroni-corrected) test sizes.

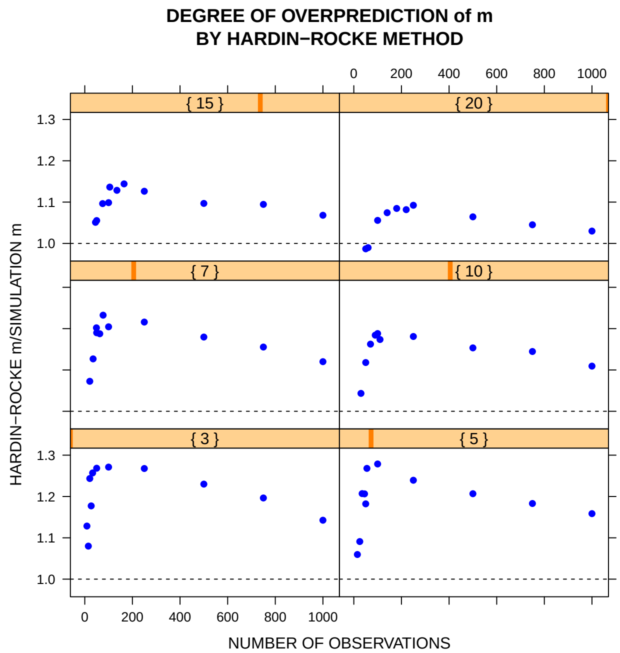

Figure 1: Comparison of Wishart degrees of freedom parameter m estimated via simulation and Hardin-Rocke approach with $\alpha_{\text{MCD}} = 0.75$. The ratio of the degrees of freedom parameters coming from the Hardin-Rocke approach to those resulting from the simulation is shown (stratified by dimension ν). Sample size is plotted on the horizontal axis. The dimension ν for each subgroup is shown in the yellow bars at the top of each subplot.

0.75 and $\alpha_{\text{MCD}} = 0.95$ cases using a corrected Hardin-Rocke methodology for those values of α_{MCD}.

2 Experimental Setup

2.1 Hardin-Rocke Extension

The methodology in (Cerioli, 2010) depends on the Hardin-Rocke methodology, which was only defined in (Hardin and Rocke, 2005) for the $\alpha_{\text{MCD}} = \alpha_{\text{MCD}}^{\text{MBP}}$ case. As discussed above it is remarkably good for $\alpha_{\text{MCD}} \neq \alpha_{\text{MCD}}^{\text{MBP}}$ and moderate sample sizes ($n > 100$); only for $\alpha > 0.95$ or small sample sizes do we see marked divergence from the simulated distribution. Thus, to eliminate any potential effects of using their approximation outside of its design parameters, we need to extend their work to $\alpha_{\text{MCD}} \neq \alpha_{\text{MCD}}^{\text{MBP}}$. We simulated 5000 draws of size n from a multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_v)$ with dimensions $v = 3, 5, 7, 10, 15, 20$ and sample sizes $n = 50, 100, 250, 500, 750, 1000$. We calculated the MCD($\alpha_{\text{MCD}}$) subset of each simulated data set for $0.55 \leq \alpha_{\text{MCD}} \leq 0.95$ in increments of 0.05, as well as maximum breakdown point case $\alpha_{\text{MCD}}^{\text{MBP}}$ and the extreme case of $\alpha_{\text{MCD}} = 0.99$. Preliminary work indicated the need for more information in small sample sizes and α_{MCD} close to 1, so we also included simulations with sample sizes $n = 3v, 5v, 7v, 9v, 11v$ for the above dimensions and values of α_{MCD}, as well as the extreme case of $\alpha_{\text{MCD}} = 0.995$. For each simulated data set and each value of α_{MCD} we computed the covariance of the MCD(α_{MCD}) subset, and then estimated the Wishart degrees of freedom m and the consistency constant c as in (Hardin and Rocke, 2005).

Hardin and Rocke estimated an adjustment to the asymptotic Wishart degrees of freedom m_{asy} (derived

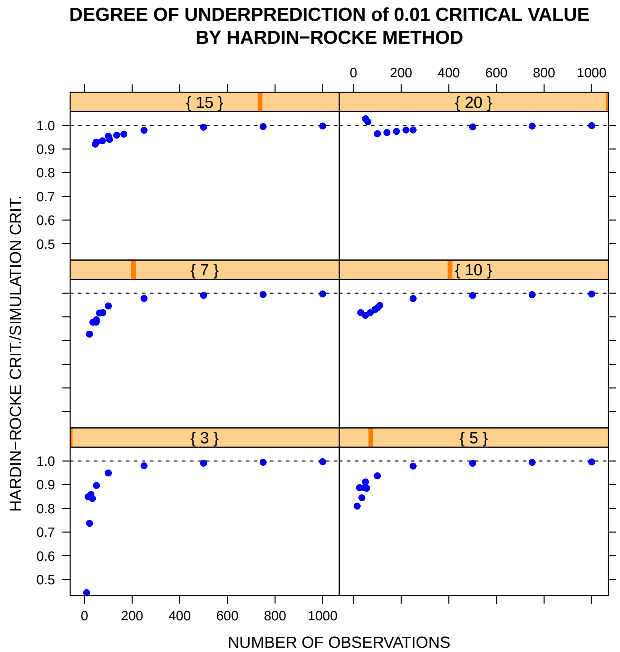

Figure 2: Comparison of 0.01 critical values produced using Wishart degrees of freedom parameter m estimated via simulation and Hardin-Rocke approach with $\alpha_{\text{MCD}} = 0.75$. Critical values are calculated using the scaled F distributional approximation of Hardin and Rocke with each degrees of freedom parameter estimate. The ratio of the Hardin-Rocke critical values to those resulting from the simulation is shown (stratified by dimension ν). Sample size is plotted on the horizontal axis. The dimension ν is shown in the yellow bars at the top of each subplot.

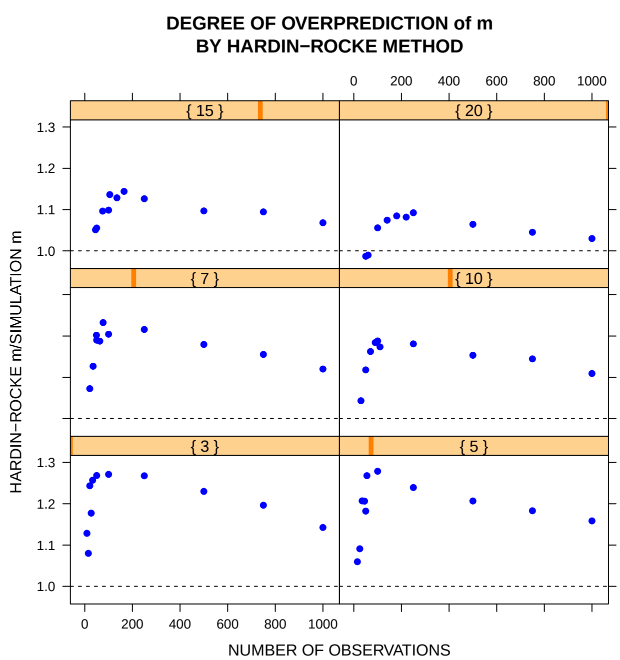

Figure 3: Comparison of Wishart degrees of freedom parameter m estimated via simulation and Hardin-Rocke approach with $\alpha_{\text{MCD}} = 0.95$. The plot setup is identical to that of Figure 1.

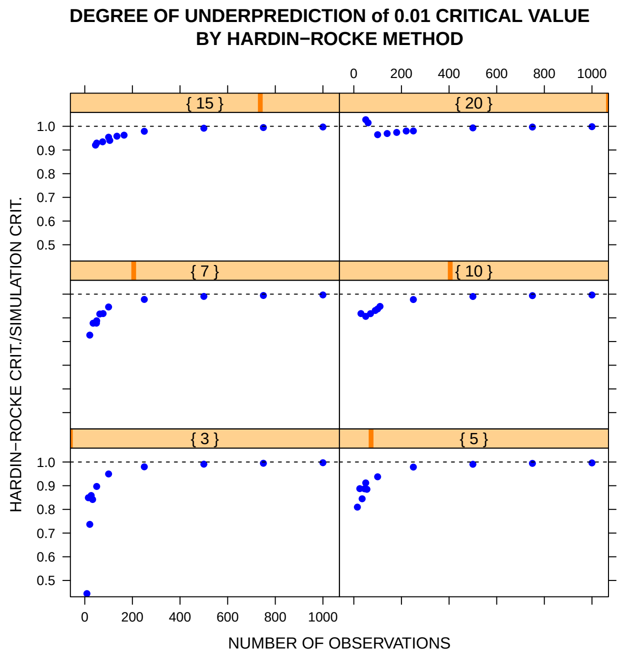

Figure 4: Comparison of 0.01 critical values produced from Wishart degrees of freedom parameter estimated via simulation and Hardin-Rocke approach with $\alpha_{\text{MCD}} = 0.95$. The plot setup is identical to that of Figure 2.

from Croux and Haesbroeck (1999)) using linear regression on a logarithmic scale; their regression equation for the $\alpha_{\text{MCD}} = \alpha_{\text{MCD}}^{\text{MBP}}$ case takes the form

$$\log\left(\frac{m_{\text{sim}}}{m_{\text{asy}}}\right) = \beta_0 + \beta_1\nu + \beta_2 \log n,$$

where m_{sim} is the simulated degrees of freedom and m_{asy} is the asymptotic degrees of freedom. Preliminary investigation (as depicted in Figure 6) of the data suggested a better fit to the general α_{MCD} case could be obtained with a power model of the form

$$\log\left(\frac{m_{\text{sim}}}{m_{\text{asy}}}\right) = \frac{\beta_0 + \beta_1\alpha_{\text{MCD}} + \beta_2\nu}{n^{\beta_3+\beta_4\alpha_{\text{MCD}}}}. \tag{4}$$

This equation can be fit in `R` using nonlinear least squares (available via the `nls` function).

Figure 6 suggests the log ratio decays inversely with a power of sample size that depends on α_{MCD}. (Also, since the asymptotic formula should approach the correct value as $n \to \infty$, we need the log ratio to go to zero as $n \to \infty$.) Dependence on the dimension ν is weak, as is evidenced by the stacking of the points in each plot.[8] Finally the sign of the dependence relation changes for small to moderate sample sizes when α_{MCD} is near 1; here the MCD(α_{MCD}) estimator discards very few observations and becomes more like the sample covariance estimator.[9]

To validate the fitted model, we reran the above simulations with a slightly different parameter set:

[8]This was also observed by Hardin and Rocke.

[9]The change in the shape of the log ratio curves for α_{MCD} near 1 does not appear to be an artifact of the simulation: we ran the experiment for small samples and α_{MCD} near 1 multiple times, and observed very consistent behavior across the experimental runs.

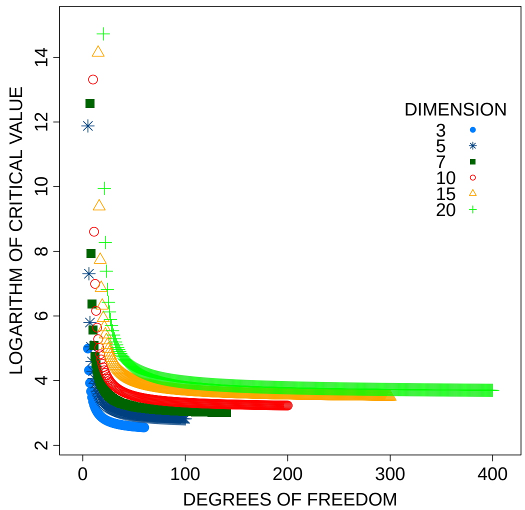

Figure 5: Logarithms of 0.99 quantiles produced from Hardin-Rocke scaled F distribution (vertical axis) as a function of the Wishart degrees of freedom parameter m (horizontal axis). The quantiles are shown for several values of dimension ν (plot symbols and colors).

we simulated 5000 draws of size n from a multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_\nu)$ with dimensions $\nu = 2, 3, 5, 7, 10, 15, 20$ and sample sizes $n = 50, 100, 250, 500, 750, 1000$, as well as $n = 4\nu, 6\nu, 8\nu, 10\nu, 12\nu$. For each sample we computed the MCD(α_{MCD}) subset for $0.55 \leq \alpha_{\text{MCD}} \leq 0.95$ in increments of 0.05, as well as the extreme cases of $\alpha_{\text{MCD}} \in \{0.99, 0.995\}$. Using the output of this experiment we can examine how well the new model predicts the Wishart degrees of freedom parameter m and compare the new model's performance to that of the Hardin-Rocke model for $\alpha_{\text{MCD}} = \alpha_{\text{MCD}}^{\text{MBP}}$.

As further validation of the fitted model, we ran a simulation experiment similar to that used by Hardin and Rocke (2005) to create Tables 1 and 2 in their paper. We generated 5000 draws of size n from a multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_\nu)$ with dimension ν for sample sizes $n = 50, 100, 500, 1000$ and $\nu = 5, 10, 20$. We used the MCD(α_{MCD})-based Mahalanobis distances to detect outliers, for $\alpha_{\text{MCD}} = \alpha_{\text{MCD}}^{\text{MBP}}, 0.65, 0.75, 0.85, 0.95, 0.99$. Points were flagged as outliers using the Hardin-Rocke F distribution with degrees of freedom m calculated using the Hardin-Rocke method and the method developed in this paper. While we know the limitations of this exercise from the work of Cerioli et al., this test does provide another comparison of our method to that of Hardin and Rocke.

The simulations for the Hardin-Rocke extension were performed on a computer running Windows 7 Professional SP 1 with 2 Intel® Xeon™ E5645 6-core processors running at 2.4GhZ and 12GB of RAM. We used R 3.0.2 (64-bit) to conduct the simulations; we used the `parallel` package to distribute simulations amongst the processors. We implemented the simulation and verification steps in two packages, `CerioliOutlierDetection` and `HardinRockeExtension`; these are described in the Appendix. Analysis was performed on a laptop running Windows 7 Ultimate SP 1 with an Intel® Core™ i7-3740QM processor running at 2.7GHz and 32GB of RAM. A full listing of packages used (and their versions) is provided in the Appendix.

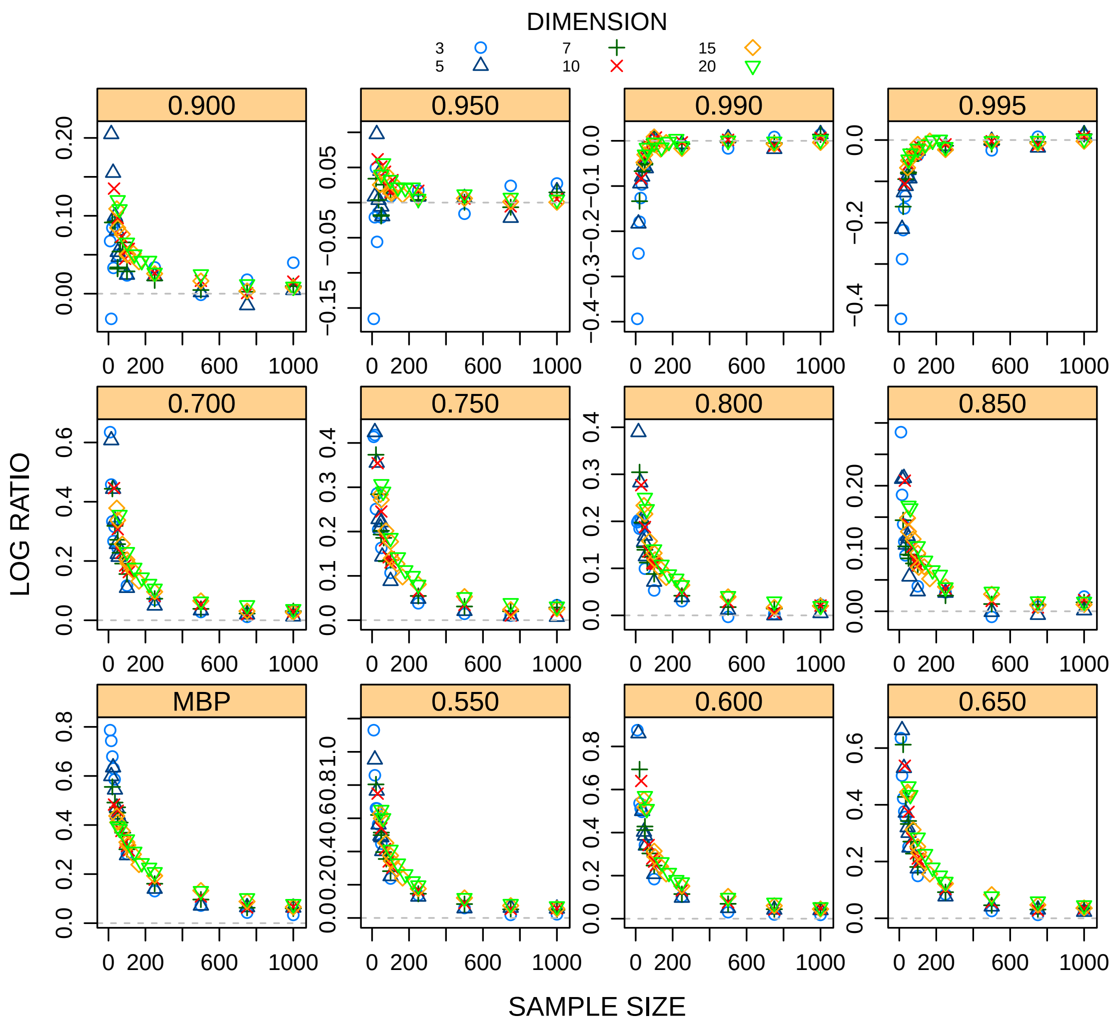

Figure 6: Logarithm of the ratio of the Wishart degrees of freedom calculated via simulation to the degrees of freedom calculated from the asymptotic formula, plotted against sample size and stratified by α_{MCD} (printed in the yellow headers) and dimension (given by the plot symbols in each plot).

2.2 Extension of Cerioli (2010) to Arbitrary $\alpha_{\mathbf{MCD}}$

For use in financial applications (see, for instance, Green and Martin (2014b)) we sought to extend the MCD($\alpha_{\mathrm{MCD}}^{\mathrm{MBP}}$)-based outlier detection procedures (FSRMCD and IRMCD) described in Cerioli (2010) to MCD(α_{MCD}) for any $0.5 \leq \alpha_{\mathrm{MCD}} < 1$. The only hurdle to the extension of these procedures to arbitrary α_{MCD} is the lack of a Hardin-Rocke distribution for $\alpha_{\mathrm{MCD}} \neq \alpha_{\mathrm{MCD}}^{\mathrm{MBP}}$; the needed extension was developed and tested as part of this process and is documented in this paper.

Simulations similar to those in Cerioli (2010) were run to verify the accuracy of the implementation: we drew 5,000 independent samples from an $N(\mathbf{0}, \mathbf{I}_v)$ distribution, and estimated the size of the intersection test (3) as the fraction of samples for which the null hypothesis is incorrectly rejected at the 0.01 level. We focused on the cases $\alpha_{\mathrm{MCD}} \in \{\alpha_{\mathrm{MCD}}^{\mathrm{MBP}}, 0.75, 0.95\}$: the former two for comparison with Cerioli's results, and $\alpha_{\mathrm{MCD}} = 0.95$ for use in other research. The simulations and the analysis were performed on a laptop running Windows 7 Ultimate SP 1 with an Intel® Core™ i7-3740QM processor running at 2.7GHz and 32GB of RAM.

3 Results

3.1 Hardin-Rocke Extension

Figure 7 shows the out-of-sample performance of the proposed improvement to the Hardin-Rocke methodology for the values of α_{MCD} tested.[10] Generally the proposed method is very good when the sample size is between 5 and 20 times the dimension: there is not much bias (the median ratios are close to 0) and not much dispersion in the correction factors (as evidenced by the tight boxplot widths). For small samples ($n < 5\nu$) the new method is generally good for $0.5 < \alpha_{\mathrm{MCD}} < 0.9$; it shows some slight bias downward (meaning the corrected m is smaller than the simulation suggests it should be). The plots show a few outliers (the circles); these correspond to very small samples in very small dimensions (e.g., a sample of size 8 from a bivariate normal). For $\alpha_{\mathrm{MCD}} > 0.95$ we see some overprediction of m in small samples. Finally for the classical covariance case, the method tends to underpredict m unless the sample size is large.

Figure 8 shows how the proposed methodology performs relative to the Hardin-Rocke methodology for $\alpha_{\mathrm{MCD}} = \alpha_{\mathrm{MCD}}^{\mathrm{MBP}}$. The proposed correction is much more accurate (as evidenced by medians closer to 0) and much less variable (as evidenced by smaller boxplot heights). A Mann-Whitney test of the hypothesis that the median difference in the log-ratio of the predicted m to the simulated m between the Hardin-Rocke method and the proposed method is 0 has a p-value of 0.028. If we conduct the same test within each n/ν group, the p-values are as follows: $(0, 5] : 0.002$; $(5, 10] : 1.2 \times 10^{-7}$; $(10, 20] : 0.021$; and $(20, \infty) : 5 \times 10^{-5}$. Thus the new method is generally a modest improvement over Hardin and Rocke (2005) in the $\alpha_{\mathrm{MCD}} = \alpha_{\mathrm{MCD}}^{\mathrm{MBP}}$ case, and a strong improvement for moderate values of n/ν and large values of n/ν.

Tables 1 and 2 show the results of testing how well each method of predicting m translates to outlier detection using the same approach used in Tables 1 and 2 of Hardin and Rocke (2005), but for more choices of α_{MCD}. For larger samples, the Hardin-Rocke method leads to false-detection rates that are a bit smaller than expected as α_{MCD} gets closer to 1; our proposed method gives false-detection rates that are closer to the ideal values for most α_{MCD} values. For moderate samples ($n = 100$), our method gives false-detection rates that are close to ideal, while the Hardin-Rocke method yields false-detection rates that are too small. In small samples ($n = 50$) neither method is particularly accurate: the Hardin-Rocke method tends to yield

[10] Full results are available in Table 7 in the appendix.

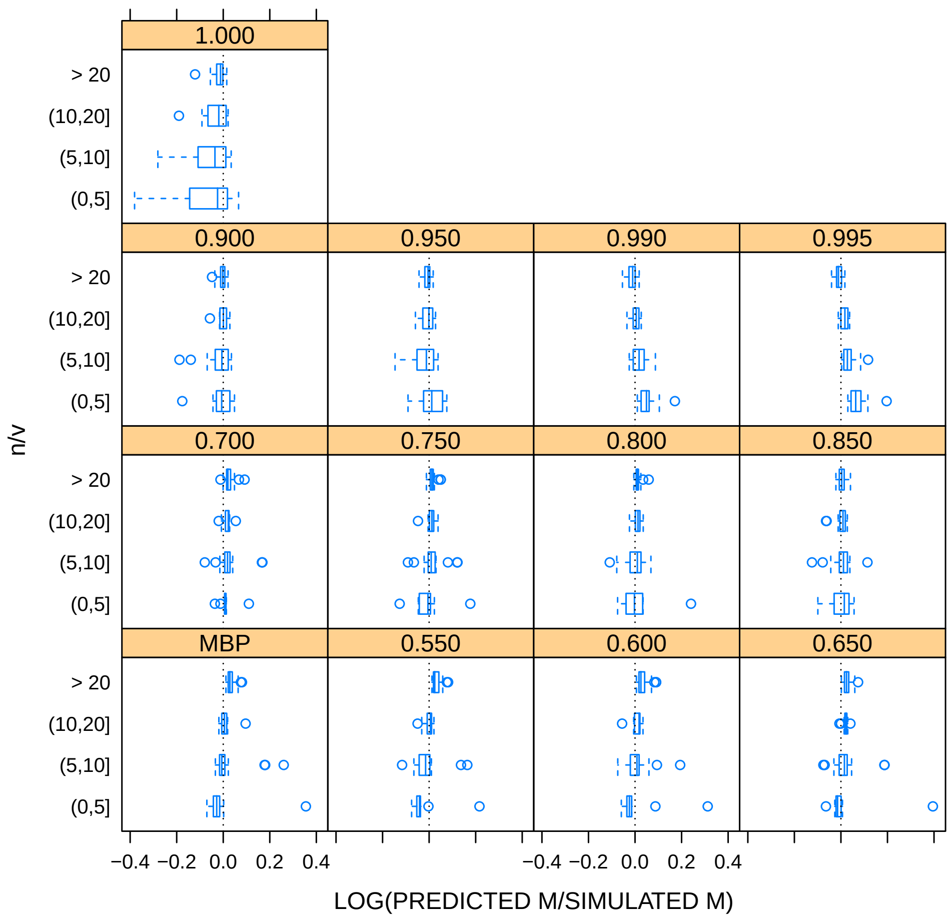

Figure 7: Boxplots showing the range of out of sample performance of the proposed correction methodology, stratified by α_{MCD} (yellow box) and the ratio n/ν of observations to variables (vertical axis). Performance is measured by the logarithm of the ratio of the predicted Wishart degrees of freedom value to the value computed via the simulation methodology used in (Hardin and Rocke, 2005). The dashed lines at 0 correspond to perfect agreement between prediction and simulation.

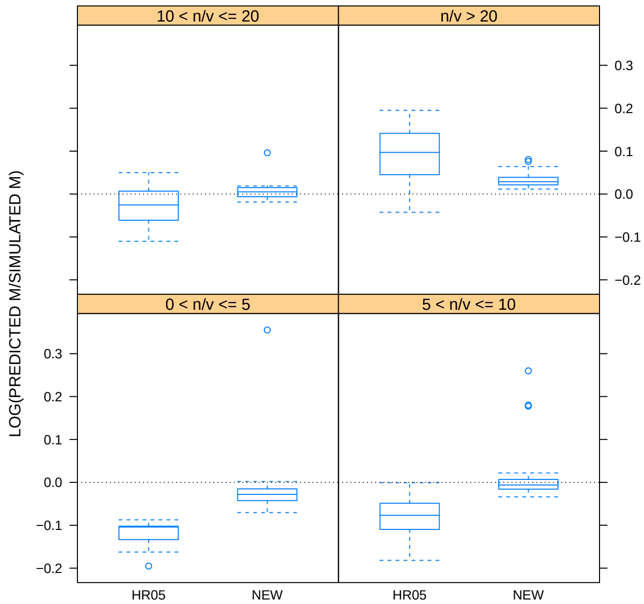

Figure 8: Boxplot showing performance of the proposed correction methodology (NEW) against that of the Hardin-Rocke methodology (HR05) for $\alpha_{\text{MCD}} = \alpha_{\text{MCD}}^{\text{MBP}}$, stratified by the ratio n/ν of observations to variables. Performance is measured by the logarithm of the ratio of the predicted Wishart degrees of freedom value to the value computed via the simulation methodology used in (Hardin and Rocke, 2005).

false-detection rates that are too low, while our method yields rates that are too high. The extreme case of $\alpha_{\text{MCD}} = 0.99$ is particularly challenging for both methods, especially in small-to-moderate samples and higher dimensions.

When the number of observations n is small compared to the dimension ν, the new method still underpredicts the degrees of freedom parameter m a bit. For large samples the new method still overpredicts m, but is more accurate on average than the Hardin-Rocke approach. The final model fit is

$$\log\left(\frac{m_{\text{sim}}}{m_{\text{asy}}}\right) = \frac{12.746 - 14.546\alpha_{\text{MCD}} + 0.127\nu}{n^{0.559+0.149\alpha_{\text{MCD}}}}. \tag{5}$$

Table 3 provides the regression coefficients along with their standard errors.

Table 1: Mean percentage of simulated data, for selected sample sizes n and dimensions ν, with MCD(α_{MCD})-based Mahalanobis distances exceeding the 0.05 quantile produced using the Hardin-Rocke method (HR05) and the proposed correction method (NEW). Ideally, each percentage should be close to 5%. Standard errors are given in parentheses and are also expressed in percentages. Compare to Table 1 of Hardin and Rocke (2005), which showed the results results of using their method in the maximum breakdown point case (MBP) of MCD.

Dimension (ν)	**n = 50**		**n = 100**		**n = 500**		**n = 1000**	
	HR05	NEW	HR05	NEW	HR05	NEW	HR05	NEW
$\alpha_{\mathbf{MCD}}=$ **MBP**								
5	6.10 (0.07)	7.17 (0.07)	6.45 (0.05)	6.29 (0.05)	5.26 (0.02)	5.09 (0.02)	5.06 (0.01)	5.01 (0.01)
10	6.74 (0.07)	8.78 (0.07)	6.70 (0.05)	7.25 (0.05)	5.19 (0.02)	5.11 (0.02)	5.04 (0.01)	4.98 (0.01)
20	5.04 (0.05)	8.34 (0.05)	4.45 (0.04)	7.03 (0.04)	4.69 (0.02)	4.96 (0.02)	4.78 (0.01)	4.94 (0.01)
$\alpha_{\mathbf{MCD}}=$ **0.650**								
5	2.89 (0.04)	6.18 (0.06)	3.22 (0.03)	5.42 (0.04)	4.28 (0.01)	4.98 (0.02)	4.55 (0.01)	5.00 (0.01)
10	4.32 (0.05)	7.85 (0.06)	3.51 (0.03)	6.22 (0.04)	4.04 (0.01)	4.99 (0.02)	4.43 (0.01)	4.97 (0.01)
20	8.51 (0.06)	10.31 (0.06)	2.86 (0.03)	6.63 (0.04)	3.32 (0.01)	4.92 (0.01)	3.94 (0.01)	4.97 (0.01)
$\alpha_{\mathbf{MCD}}=$ **0.750**								
5	1.13 (0.02)	5.00 (0.05)	2.11 (0.02)	5.11 (0.03)	3.93 (0.01)	4.97 (0.01)	4.40 (0.01)	5.01 (0.01)
10	1.50 (0.03)	6.07 (0.05)	1.79 (0.02)	5.41 (0.03)	3.62 (0.01)	5.00 (0.01)	4.18 (0.01)	4.98 (0.01)
20	3.60 (0.04)	8.40 (0.05)	1.12 (0.02)	5.93 (0.03)	2.72 (0.01)	4.93 (0.01)	3.60 (0.01)	4.97 (0.01)
$\alpha_{\mathbf{MCD}}=$ **0.850**								
5	0.52 (0.02)	4.63 (0.04)	1.40 (0.02)	4.95 (0.03)	3.71 (0.01)	4.98 (0.01)	4.26 (0.01)	4.98 (0.01)
10	0.47 (0.01)	4.82 (0.04)	1.06 (0.02)	5.00 (0.03)	3.30 (0.01)	4.99 (0.01)	4.00 (0.01)	4.99 (0.01)
20	1.01 (0.02)	6.47 (0.03)	0.45 (0.01)	5.37 (0.03)	2.33 (0.01)	4.95 (0.01)	3.34 (0.01)	4.99 (0.01)
$\alpha_{\mathbf{MCD}}=$ **0.950**								
5	0.24 (0.01)	3.98 (0.03)	0.99 (0.01)	4.13 (0.02)	3.53 (0.01)	4.75 (0.01)	4.15 (0.01)	4.87 (0.01)
10	0.12 (0.01)	3.10 (0.02)	0.56 (0.01)	4.16 (0.02)	3.07 (0.01)	4.67 (0.01)	3.86 (0.01)	4.80 (0.01)
20	0.14 (0.01)	3.63 (0.01)	0.16 (0.01)	3.68 (0.01)	2.03 (0.01)	4.51 (0.01)	3.12 (0.01)	4.67 (0.00)
$\alpha_{\mathbf{MCD}}=$ **0.990**								
5	0.14 (0.01)	2.01 (0.01)	0.61 (0.01)	2.66 (0.02)	3.07 (0.01)	4.49 (0.01)	3.91 (0.01)	4.74 (0.01)
10	0.07 (0.01)	1.86 (0.01)	0.36 (0.01)	1.37 (0.01)	2.22 (0.01)	3.95 (0.01)	3.35 (0.01)	4.45 (0.01)
20	0.07 (0.01)	1.96 (0.00)	0.09 (0.00)	1.00 (0.00)	1.11 (0.00)	2.85 (0.01)	2.22 (0.01)	3.84 (0.01)

Table 2: Mean percentage of simulated data, for selected sample sizes n and dimensions ν, with MCD(α_{MCD})-based Mahalanobis distances exceeding the 0.01 quantile produced using the Hardin-Rocke method (HR05) and the proposed correction method (NEW). Ideally, each percentage should be close to 1%. Standard errors are given in parentheses and are also expressed in percentages. Compare to Table 2 of Hardin and Rocke (2005), which showed the results results of using their method in the maximum breakdown point case (MBP) of MCD.

Dimension (ν)	**n = 50**		**n = 100**		**n = 500**		**n = 1000**	
	HR05	NEW	HR05	NEW	HR05	NEW	HR05	NEW
α_{MCD}= **MBP**								
5	1.40 (0.03)	1.79 (0.04)	1.69 (0.03)	1.60 (0.03)	1.11 (0.01)	1.06 (0.01)	1.04 (0.01)	1.01 (0.01)
10	1.86 (0.04)	2.87 (0.05)	1.89 (0.03)	2.05 (0.03)	1.07 (0.01)	1.03 (0.01)	1.02 (0.00)	1.00 (0.00)
20	1.42 (0.03)	3.28 (0.04)	1.02 (0.02)	2.00 (0.03)	0.94 (0.01)	1.02 (0.01)	0.95 (0.00)	1.00 (0.00)
α_{MCD}= **0.650**								
5	0.38 (0.01)	1.45 (0.03)	0.46 (0.01)	1.17 (0.02)	0.75 (0.01)	1.02 (0.01)	0.85 (0.00)	0.99 (0.00)
10	0.93 (0.02)	2.34 (0.04)	0.64 (0.01)	1.54 (0.02)	0.72 (0.01)	1.01 (0.01)	0.82 (0.00)	1.00 (0.00)
20	3.32 (0.05)	4.53 (0.05)	0.51 (0.01)	1.79 (0.02)	0.55 (0.00)	1.00 (0.01)	0.69 (0.00)	1.00 (0.00)
α_{MCD}= **0.750**								
5	0.08 (0.01)	1.01 (0.02)	0.21 (0.01)	1.08 (0.02)	0.65 (0.01)	1.00 (0.01)	0.79 (0.00)	1.00 (0.00)
10	0.17 (0.01)	1.50 (0.03)	0.22 (0.01)	1.22 (0.02)	0.59 (0.01)	1.00 (0.01)	0.74 (0.00)	1.00 (0.00)
20	0.82 (0.02)	3.31 (0.04)	0.14 (0.01)	1.40 (0.02)	0.40 (0.00)	1.00 (0.01)	0.60 (0.00)	0.99 (0.00)
α_{MCD}= **0.850**								
5	0.02 (0.00)	0.89 (0.02)	0.10 (0.00)	0.99 (0.01)	0.59 (0.00)	1.00 (0.01)	0.75 (0.00)	1.00 (0.00)
10	0.02 (0.00)	1.04 (0.02)	0.07 (0.00)	1.06 (0.02)	0.51 (0.00)	0.99 (0.01)	0.69 (0.00)	1.00 (0.00)
20	0.12 (0.01)	2.24 (0.03)	0.03 (0.00)	1.18 (0.02)	0.31 (0.00)	0.99 (0.01)	0.53 (0.00)	0.99 (0.00)
α_{MCD}= **0.950**								
5	0.00 (0.00)	0.86 (0.02)	0.05 (0.00)	0.93 (0.01)	0.53 (0.00)	0.99 (0.01)	0.72 (0.00)	0.98 (0.00)
10	0.00 (0.00)	0.79 (0.02)	0.03 (0.00)	0.97 (0.01)	0.44 (0.00)	0.99 (0.01)	0.65 (0.00)	1.01 (0.00)
20	0.01 (0.00)	1.36 (0.02)	0.00 (0.00)	1.06 (0.01)	0.25 (0.00)	1.00 (0.01)	0.48 (0.00)	1.00 (0.00)
α_{MCD}= **0.990**								
5	0.00 (0.00)	0.67 (0.01)	0.03 (0.00)	0.63 (0.01)	0.50 (0.00)	0.90 (0.00)	0.70 (0.00)	0.95 (0.00)
10	0.00 (0.00)	0.72 (0.01)	0.02 (0.00)	0.61 (0.01)	0.42 (0.00)	0.86 (0.00)	0.63 (0.00)	0.92 (0.00)
20	0.00 (0.00)	1.09 (0.01)	0.00 (0.00)	0.66 (0.01)	0.24 (0.00)	0.83 (0.00)	0.47 (0.00)	0.89 (0.00)

Table 3: Estimated coefficients, and their standard errors, for the model described by Equation (4).

Coefficient	Estimate	Std. Error
β_0	12.746	0.305
β_1	-14.546	0.368
β_2	0.127	0.007
β_3	0.559	0.011
β_4	0.149	0.018

3.2 Testing the Implementation of FSRMCD and IRMCD

Table 4 shows the results of testing our implementation of the finite-sample and iteratively reweighted MCD estimators (FSRMCD and IRMCD, respectively) defined in Cerioli (2010). Overall our implementation gives the right sizes empirically, and it produces results consistent with those presented in Table 1 and 2 of that paper.

4 Discussion

Our modified version of the Hardin-Rocke adjustment to the asymptotic degrees of freedom parameter estimate performs very well in general: in the out-of-sample tests portrayed in Figure 7 the median ratio of

Table 4: Results of simulation tests of FSRMCD and IRMCD implementations. The table shows the estimated size for testing the hypothesis of no outliers in the data at the nominal size of 0.01. The size is estimated using $5,000$ simulations for each combination of sample size n and dimension ν. Compare to Table 1 of Cerioli (2010).

Dimension	Method	$n = 40$	$n = 60$	$n = 90$	$n = 125$	$n = 200$	$n = 400$
$\alpha_{\mathbf{MCD}} = \alpha_{\mathbf{MCD}}^{\mathbf{MBP}}$							
$\nu = 5$	FSRMCD	0.017	0.016	0.011	0.014	0.010	0.009
	IRMCD	0.014	0.014	0.011	0.013	0.011	0.012
$\nu = 10$	FSRMCD	0.043	0.019	0.009	0.010	0.010	0.007
	IRMCD	0.041	0.020	0.012	0.012	0.009	0.010
$\nu = 15$	FSRMCD	0.045	0.016	0.008	0.010	0.007	0.009
	IRMCD	0.044	0.017	0.010	0.011	0.010	0.008
$\alpha_{\mathbf{MCD}} = \mathbf{0.75}$							
$\nu = 5$	FSRMCD	0.015	0.016	0.014	0.011	0.012	0.009
	IRMCD	0.015	0.014	0.009	0.009	0.010	0.011
$\nu = 10$	FSRMCD	0.044	0.021	0.009	0.008	0.009	0.010
	IRMCD	0.048	0.021	0.014	0.010	0.011	0.009
$\nu = 15$	FSRMCD	0.047	0.014	0.010	0.010	0.013	0.010
	IRMCD	0.047	0.015	0.009	0.008	0.009	0.007
$\alpha_{\mathbf{MCD}} = \mathbf{0.95}$							
$\nu = 5$	FSRMCD	0.016	0.015	0.013	0.010	0.008	0.007
	IRMCD	0.013	0.018	0.012	0.011	0.014	0.007
$\nu = 10$	FSRMCD	0.040	0.017	0.012	0.010	0.009	0.010
	IRMCD	0.041	0.019	0.011	0.008	0.008	0.008
$\nu = 15$	FSRMCD	0.046	0.012	0.013	0.008	0.008	0.009
	IRMCD	0.049	0.017	0.009	0.011	0.011	0.007

Table 5: Monte Carlo standard deviations of simulation tests of FSRMCD and IRMCD implementations. Standard errors for the quantities in Table 4 can be obtained by dividing the corresponding entries in this table by $\sqrt{5000}$.

Dimension	Method	$n = 40$	$n = 60$	$n = 90$	$n = 125$	$n = 200$	$n = 400$
$\alpha_{\mathbf{MCD}} = \alpha_{\mathbf{MCD}}^{\mathbf{MBP}}$							
$\nu = 5$	FSRMCD	0.128	0.125	0.106	0.118	0.100	0.094
	IRMCD	0.118	0.117	0.103	0.115	0.104	0.107
$\nu = 10$	FSRMCD	0.204	0.138	0.093	0.101	0.100	0.086
	IRMCD	0.197	0.139	0.111	0.111	0.094	0.101
$\nu = 15$	FSRMCD	0.207	0.125	0.090	0.098	0.081	0.093
	IRMCD	0.205	0.129	0.098	0.104	0.101	0.090
$\alpha_{\mathbf{MCD}} = \mathbf{0.75}$							
$\nu = 5$	FSRMCD	0.122	0.125	0.118	0.106	0.111	0.094
	IRMCD	0.120	0.118	0.095	0.092	0.100	0.105
$\nu = 10$	FSRMCD	0.204	0.142	0.094	0.091	0.093	0.098
	IRMCD	0.214	0.143	0.117	0.101	0.106	0.095
$\nu = 15$	FSRMCD	0.212	0.118	0.100	0.098	0.113	0.099
	IRMCD	0.213	0.122	0.093	0.090	0.093	0.083
$\alpha_{\mathbf{MCD}} = \mathbf{0.95}$							
$\nu = 5$	FSRMCD	0.124	0.121	0.112	0.101	0.090	0.086
	IRMCD	0.115	0.134	0.108	0.105	0.118	0.083
$\nu = 10$	FSRMCD	0.196	0.129	0.110	0.100	0.092	0.100
	IRMCD	0.198	0.135	0.105	0.089	0.091	0.088
$\nu = 15$	FSRMCD	0.209	0.111	0.112	0.090	0.091	0.094
	IRMCD	0.216	0.128	0.093	0.103	0.103	0.081

the predicted m to the simulated m is typically about 1. The new method is more accurate, on average, than the Hardin and Rocke (2005) method, and tends to perform more consistently across a variety of sample sizes and dimensions.

For small samples $n < 5\nu$ there is still some bias, i.e., the predicted m tends to be too small for α_{MCD} near 0.5, and too large for α_{MCD} near 1 (but not at $\alpha_{\text{MCD}} = 1$, where it is again too small). Likewise for large samples $n > 20\nu$ the predicted m tends to be too large for α_{MCD} near 0.5 and a little too small for α_{MCD} near 1. The deviations are not terribly large, though; for instance, for small samples and $\alpha_{\text{MCD}} = 0.995$ the predicted value is 1.06 times the simulated value on average, which means a true m of 50 is predicted to be 53; this translates into critical values that are 1-2% too small in dimensions less than 10. In higher dimensions, say larger than 20, the difference in the critical values will be larger and might have a more noticable impact on outlier detection.

Due to the computational requirements of the simulations done here, we were only able to run the full experiment once. Thus, we do not know how variable the simulated c and m can be in general.[11] However, in the process of investigating the behavior of the simulated m for α_{MCD} near 1, we did run the $\alpha_{\text{MCD}} \geq 0.9$ cases several times. As the sample size n gets larger, we observed more variation in the simulated value of m; however this does not seem to translate into much variation in the resulting 0.01 critical values. For small sample sizes ($n < 100$) or when n is a small multiple of p, there can be a wider range of critical values resulting from the simulated m values. Here the MCD estimate (with $\alpha_{\text{MCD}} \geq 0.9$) is discarding very few observations, so a potential improvement to our methodology might consider an alternative approach to

[11] Recall that the commonly used fastMCD procedure of Rousseeuw and van Driessen (1999) involves random sampling as well.

calculating the distribution of the MCD estimate in such cases.

5 Conclusions and Further Research

We have extended the Hardin-Rocke methodology to the general MCD(α_{MCD}) estimator, thereby ensuring that the FSRMCD and IRMCD outlier detection methodologies introduced by Cerioli (2010) give the right test sizes for arbitrary α_{MCD} (as long as the sample size is not very small compared to the dimension).

In another study Green and Martin (2014a) we showed that several other robust dispersion estimates exhibit, to varying degrees, the problems with the size of the Mahalanobis distance test for outliers that Cerioli et al. (2009) found for the MCD estimate. Thus, correction methodologies are also needed for other robust dispersion estimators such as S-estimators and the OGK estimate. A calibration methodology for S-estimators would be valuable, as the bisquare and Rocke-type S-estimators were the recommended approaches from the study performed in Chapter 6.8 of Maronna et al. (2006). Hardin and Rocke (2005) suggest their methodology could be extended to S-estimators, so perhaps a correction methodology analogous to IRMCD could be developed for S-estimators. Likewise, a calibrated methodology for the OGK estimator would also be valuable due to the comparative computational simplicity of the OGK in higher dimensions and its appeal in dealing with componentwise contamination scenarios (as described in Alqallaf et al. (2009)). It is not obvious that Cerioli's approach could be extended to the OGK, however, since that estimator's mathematical underpinnings are quite different from those of the MCD and the S-estimators; a wholly different approach may be needed for the OGK.

We have only considered outlier detection in a multivariate normal framework in this paper. Real data, especially financial data, often exhibit skewness and heavy tails, not all of which can be explained by outliers in the data. An important research direction for the future is outlier detection in more general univariate and multivariate distributions where the data arise from a contaminated skewed and/or heavy-tailed distribution. Some initial work in this direction has been done by Dell'Aquila and Rochetti (2006), who examined the compatibility of robust methods and extreme value theory, and Goegebeur et al. (2014), who proposed a robust estimator for extreme quantiles of heavy-tailed distributions.

6 Appendix

6.1 Simulated Degrees of Freedom and Consistency Factor

Table 6 provides the Wishart degrees of freedom parameter m and consistency factor c calculated via simulation. These values were used to fit the model shown in Equation (5).

Table 6: Wishart degrees of freedom parameter m_{sim} and consistency factor c_{sim} determined via simulation, for various dimensions ν, sample sizes n, and MCD fractions α_{MCD}. The abbreviation MBP indicates the maximum breakdown point fraction $\alpha_{\text{MCD}}^{\text{MBP}}$.

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
3	9	0.550	3.250	0.742	3	27	0.850	13.915	0.770
3	9	0.600	3.250	0.742	3	27	0.900	16.005	0.811
3	9	0.650	3.250	0.742	3	27	0.950	18.484	0.859
3	9	MBP	4.086	0.784	3	27	0.990	21.643	0.917
3	9	0.700	4.086	0.784	3	27	0.995	21.643	0.917
3	9	0.750	4.086	0.784	3	33	MBP	6.748	0.564
3	9	0.800	4.086	0.784	3	33	0.550	6.763	0.564
3	9	0.850	5.522	0.858	3	33	0.600	8.156	0.595
3	9	0.900	5.522	0.858	3	33	0.650	9.025	0.614
3	9	0.950	5.522	0.858	3	33	0.700	10.883	0.656
3	9	0.990	5.522	0.858	3	33	0.750	12.154	0.679
3	9	0.995	5.522	0.858	3	33	0.800	14.803	0.732
3	15	0.550	4.128	0.651	3	33	0.850	16.640	0.762
3	15	MBP	4.744	0.678	3	33	0.900	20.741	0.834
3	15	0.650	4.744	0.678	3	33	0.950	23.545	0.878
3	15	0.700	5.706	0.708	3	33	0.990	27.252	0.930
3	15	0.750	6.838	0.755	3	33	0.995	27.252	0.930
3	15	0.800	6.838	0.755	3	50	MBP	8.512	0.531
3	15	0.850	8.329	0.813	3	50	0.550	9.116	0.542
3	15	0.900	8.329	0.813	3	50	0.600	10.590	0.565
3	15	0.950	10.635	0.887	3	50	0.650	12.304	0.589
3	15	0.990	10.635	0.887	3	50	0.700	15.334	0.633
3	15	0.995	10.635	0.887	3	50	0.750	17.667	0.666
3	21	0.550	4.737	0.607	3	50	0.800	20.565	0.703
3	21	MBP	5.398	0.625	3	50	0.850	25.865	0.766
3	21	0.600	5.398	0.625	3	50	0.900	30.142	0.816
3	21	0.650	6.131	0.648	3	50	0.950	35.444	0.875
3	21	0.700	7.054	0.675	3	50	0.990	43.175	0.949
3	21	0.750	8.099	0.707	3	50	0.995	43.175	0.949
3	21	0.800	9.405	0.745	3	100	MBP	13.157	0.480
3	21	0.850	11.127	0.788	3	100	0.550	14.783	0.499
3	21	0.900	13.104	0.843	3	100	0.600	18.078	0.531
3	21	0.950	15.970	0.910	3	100	0.650	22.212	0.567
3	21	0.990	15.970	0.910	3	100	0.700	27.049	0.607
3	21	0.995	15.970	0.910	3	100	0.750	33.402	0.651
3	27	0.550	6.086	0.585	3	100	0.800	39.272	0.690
3	27	MBP	6.088	0.585	3	100	0.850	48.003	0.744
3	27	0.600	6.777	0.604	3	100	0.900	58.711	0.807
3	27	0.650	7.527	0.623	3	100	0.950	73.020	0.884
3	27	0.700	8.487	0.647	3	100	0.990	91.137	0.969
3	27	0.750	10.785	0.702	3	100	0.995	91.137	0.969
3	27	0.800	12.326	0.733	3	250	MBP	26.447	0.440

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
3	250	0.550	33.438	0.469	3	1000	0.850	472.250	0.733
3	250	0.600	41.651	0.505	3	1000	0.900	596.788	0.798
3	250	0.650	52.025	0.543	3	1000	0.950	743.944	0.878
3	250	0.700	64.455	0.587	3	1000	0.990	922.623	0.965
3	250	0.750	77.957	0.632	3	1000	0.995	956.862	0.980
3	250	0.800	95.964	0.680	5	15	0.550	6.533	0.814
3	250	0.850	118.870	0.739	5	15	0.600	7.340	0.829
3	250	0.900	148.296	0.801	5	15	0.650	7.340	0.829
3	250	0.950	184.108	0.876	5	15	MBP	7.340	0.829
3	250	0.990	225.550	0.961	5	15	0.700	8.390	0.853
3	250	0.995	232.544	0.972	5	15	0.750	8.390	0.853
3	500	MBP	48.759	0.424	5	15	0.800	9.695	0.884
3	500	0.550	61.889	0.457	5	15	0.850	9.695	0.884
3	500	0.600	77.512	0.496	5	15	0.900	11.568	0.931
3	500	0.650	98.140	0.537	5	15	0.950	11.568	0.931
3	500	0.700	123.712	0.580	5	15	0.990	11.568	0.931
3	500	0.750	152.245	0.627	5	15	0.995	11.568	0.931
3	500	0.800	185.621	0.677	5	25	0.550	9.027	0.742
3	500	0.850	228.608	0.734	5	25	MBP	9.758	0.757
3	500	0.900	286.330	0.799	5	25	0.650	10.704	0.773
3	500	0.950	356.292	0.878	5	25	0.700	11.868	0.791
3	500	0.990	447.261	0.966	5	25	0.750	13.050	0.812
3	500	0.995	461.251	0.977	5	25	0.800	14.521	0.837
3	750	MBP	70.533	0.418	5	25	0.850	16.180	0.867
3	750	0.550	89.074	0.454	5	25	0.900	18.346	0.901
3	750	0.600	115.036	0.492	5	25	0.950	21.069	0.944
3	750	0.650	145.219	0.533	5	25	0.990	21.069	0.944
3	750	0.700	182.560	0.578	5	25	0.995	21.069	0.944
3	750	0.750	227.455	0.625	5	35	0.550	10.312	0.704
3	750	0.800	280.237	0.676	5	35	MBP	11.092	0.715
3	750	0.850	348.459	0.734	5	35	0.600	11.939	0.725
3	750	0.900	437.920	0.798	5	35	0.650	12.805	0.737
3	750	0.950	556.105	0.877	5	35	0.700	14.918	0.766
3	750	0.990	688.110	0.963	5	35	0.750	16.087	0.781
3	750	0.995	715.110	0.979	5	35	0.800	18.796	0.819
3	1000	MBP	92.842	0.416	5	35	0.850	20.371	0.840
3	1000	0.550	119.171	0.452	5	35	0.900	24.290	0.890
3	1000	0.600	153.113	0.491	5	35	0.950	26.860	0.919
3	1000	0.650	196.623	0.533	5	35	0.990	29.947	0.954
3	1000	0.700	248.282	0.577	5	35	0.995	29.947	0.954
3	1000	0.750	310.520	0.625	5	45	0.550	12.385	0.684
3	1000	0.800	381.022	0.675	5	45	MBP	12.387	0.684

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
5	45	0.600	13.943	0.703	5	100	0.900	64.378	0.859
5	45	0.650	15.640	0.724	5	100	0.950	77.886	0.917
5	45	0.700	17.732	0.747	5	100	0.990	93.187	0.978
5	45	0.750	20.291	0.774	5	100	0.995	93.187	0.978
5	45	0.800	23.032	0.805	5	250	MBP	40.804	0.562
5	45	0.850	26.558	0.840	5	250	0.550	47.699	0.586
5	45	0.900	30.658	0.881	5	250	0.600	57.000	0.616
5	45	0.950	35.992	0.930	5	250	0.650	68.075	0.648
5	45	0.990	39.416	0.961	5	250	0.700	80.030	0.682
5	45	0.995	39.416	0.961	5	250	0.750	95.994	0.722
5	50	0.550	13.640	0.677	5	250	0.800	113.769	0.761
5	50	MBP	13.633	0.677	5	250	0.850	134.784	0.804
5	50	0.600	15.213	0.694	5	250	0.900	160.658	0.853
5	50	0.650	17.034	0.714	5	250	0.950	192.730	0.910
5	50	0.700	19.053	0.735	5	250	0.990	229.855	0.972
5	50	0.750	22.818	0.772	5	250	0.995	235.649	0.980
5	50	0.800	25.928	0.801	5	500	MBP	74.129	0.542
5	50	0.850	29.398	0.833	5	500	0.550	88.920	0.571
5	50	0.900	33.144	0.869	5	500	0.600	108.685	0.605
5	50	0.950	38.011	0.911	5	500	0.650	131.407	0.640
5	50	0.990	44.466	0.965	5	500	0.700	157.495	0.675
5	50	0.995	44.466	0.965	5	500	0.750	186.799	0.714
5	55	MBP	13.934	0.664	5	500	0.800	221.498	0.756
5	55	0.550	13.807	0.664	5	500	0.850	261.446	0.800
5	55	0.600	15.977	0.688	5	500	0.900	314.777	0.851
5	55	0.650	17.792	0.707	5	500	0.950	384.730	0.911
5	55	0.700	20.756	0.738	5	500	0.990	465.562	0.975
5	55	0.750	23.228	0.761	5	500	0.995	477.168	0.984
5	55	0.800	27.301	0.799	5	750	MBP	109.485	0.536
5	55	0.850	30.427	0.828	5	750	0.550	133.816	0.567
5	55	0.900	36.230	0.879	5	750	0.600	161.796	0.601
5	55	0.950	41.238	0.919	5	750	0.650	194.979	0.636
5	55	0.990	47.936	0.969	5	750	0.700	233.004	0.673
5	55	0.995	47.936	0.969	5	750	0.750	276.904	0.713
5	100	MBP	20.290	0.611	5	750	0.800	328.362	0.754
5	100	0.550	21.592	0.622	5	750	0.850	390.358	0.799
5	100	0.600	25.461	0.648	5	750	0.900	464.325	0.850
5	100	0.650	30.049	0.677	5	750	0.950	561.083	0.910
5	100	0.700	33.982	0.703	5	750	0.990	682.148	0.974
5	100	0.750	39.972	0.738	5	750	0.995	704.263	0.985
5	100	0.800	47.023	0.777	5	1000	MBP	145.046	0.533
5	100	0.850	54.036	0.811	5	1000	0.550	177.539	0.565

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
5	1000	0.600	215.218	0.599	7	49	0.950	38.109	0.930
5	1000	0.650	257.948	0.635	7	49	0.990	43.566	0.972
5	1000	0.700	308.608	0.672	7	49	0.995	43.566	0.972
5	1000	0.750	368.659	0.711	7	50	0.550	17.163	0.753
5	1000	0.800	439.867	0.753	7	50	MBP	18.038	0.759
5	1000	0.850	524.248	0.798	7	50	0.600	18.987	0.766
5	1000	0.900	631.272	0.850	7	50	0.650	20.802	0.782
5	1000	0.950	776.916	0.910	7	50	0.700	22.935	0.800
5	1000	0.990	939.898	0.975	7	50	0.750	25.184	0.820
5	1000	0.995	970.004	0.986	7	50	0.800	27.768	0.842
7	21	0.550	9.634	0.847	7	50	0.850	30.812	0.867
7	21	0.600	10.396	0.853	7	50	0.900	34.459	0.897
7	21	0.650	11.434	0.866	7	50	0.950	38.933	0.930
7	21	MBP	11.439	0.866	7	50	0.990	44.553	0.972
7	21	0.700	11.442	0.866	7	50	0.995	44.553	0.972
7	21	0.750	12.557	0.880	7	63	0.550	19.228	0.724
7	21	0.800	13.753	0.901	7	63	MBP	19.910	0.731
7	21	0.850	13.753	0.901	7	63	0.600	21.522	0.742
7	21	0.900	15.333	0.925	7	63	0.650	24.373	0.762
7	21	0.950	17.222	0.956	7	63	0.700	27.660	0.784
7	21	0.990	17.222	0.956	7	63	0.750	31.228	0.810
7	21	0.995	17.222	0.956	7	63	0.800	34.049	0.829
7	35	0.550	13.357	0.787	7	63	0.850	39.019	0.859
7	35	MBP	14.159	0.793	7	63	0.900	44.313	0.896
7	35	0.650	15.852	0.811	7	63	0.950	51.229	0.940
7	35	0.700	16.825	0.821	7	63	0.990	57.683	0.976
7	35	0.750	19.143	0.846	7	63	0.995	57.683	0.976
7	35	0.800	20.598	0.860	7	77	MBP	22.461	0.708
7	35	0.850	21.985	0.877	7	77	0.550	22.546	0.709
7	35	0.900	25.573	0.915	7	77	0.600	25.799	0.730
7	35	0.950	27.846	0.938	7	77	0.650	28.558	0.748
7	35	0.990	30.677	0.965	7	77	0.700	32.595	0.774
7	35	0.995	30.677	0.965	7	77	0.750	36.206	0.796
7	49	0.550	16.568	0.752	7	77	0.800	42.136	0.829
7	49	MBP	17.500	0.759	7	77	0.850	47.070	0.855
7	49	0.600	18.300	0.766	7	77	0.900	54.801	0.897
7	49	0.650	20.172	0.781	7	77	0.950	62.076	0.933
7	49	0.700	22.149	0.798	7	77	0.990	72.005	0.979
7	49	0.750	24.473	0.818	7	77	0.995	72.005	0.979
7	49	0.800	27.362	0.840	7	100	MBP	26.691	0.687
7	49	0.850	30.380	0.866	7	100	0.550	27.210	0.692
7	49	0.900	33.713	0.895	7	100	0.600	31.673	0.715

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
7	100	0.650	35.379	0.735	7	750	0.950	590.586	0.927
7	100	0.700	40.863	0.762	7	750	0.990	699.522	0.979
7	100	0.750	47.147	0.792	7	750	0.995	719.275	0.988
7	100	0.800	52.753	0.818	7	1000	MBP	181.563	0.601
7	100	0.850	61.008	0.853	7	1000	0.550	215.480	0.629
7	100	0.900	68.568	0.885	7	1000	0.600	258.791	0.660
7	100	0.950	80.515	0.933	7	1000	0.650	305.740	0.692
7	100	0.990	94.003	0.983	7	1000	0.700	360.540	0.725
7	100	0.995	94.003	0.983	7	1000	0.750	423.641	0.759
7	250	MBP	52.723	0.632	7	1000	0.800	493.120	0.795
7	250	0.550	59.386	0.652	7	1000	0.850	574.623	0.834
7	250	0.600	69.362	0.679	7	1000	0.900	672.817	0.877
7	250	0.650	80.933	0.708	7	1000	0.950	804.219	0.927
7	250	0.700	93.993	0.737	7	1000	0.990	950.493	0.980
7	250	0.750	108.703	0.769	7	1000	0.995	977.398	0.989
7	250	0.800	125.927	0.803	10	30	0.550	15.574	0.879
7	250	0.850	145.194	0.839	10	30	0.600	16.520	0.882
7	250	0.900	169.447	0.880	10	30	0.650	17.484	0.889
7	250	0.950	198.936	0.927	10	30	MBP	17.430	0.889
7	250	0.990	232.698	0.978	10	30	0.700	18.564	0.897
7	250	0.995	237.563	0.984	10	30	0.750	19.605	0.907
7	500	MBP	95.703	0.612	10	30	0.800	20.931	0.919
7	500	0.550	111.610	0.637	10	30	0.850	22.516	0.934
7	500	0.600	132.566	0.667	10	30	0.900	24.169	0.952
7	500	0.650	154.554	0.696	10	30	0.950	26.168	0.973
7	500	0.700	181.658	0.729	10	30	0.990	26.168	0.973
7	500	0.750	212.293	0.763	10	30	0.995	26.168	0.973
7	500	0.800	245.824	0.797	10	50	0.550	21.066	0.822
7	500	0.850	286.523	0.836	10	50	MBP	22.968	0.832
7	500	0.900	334.728	0.878	10	50	0.650	24.771	0.844
7	500	0.950	396.522	0.928	10	50	0.700	26.932	0.857
7	500	0.990	467.708	0.981	10	50	0.750	29.302	0.871
7	500	0.995	477.011	0.988	10	50	0.800	31.913	0.888
7	750	MBP	137.325	0.605	10	50	0.850	35.007	0.909
7	750	0.550	162.156	0.632	10	50	0.900	38.786	0.933
7	750	0.600	193.938	0.662	10	50	0.950	43.162	0.961
7	750	0.650	228.363	0.694	10	50	0.990	45.755	0.979
7	750	0.700	268.207	0.726	10	50	0.995	45.755	0.979
7	750	0.750	314.989	0.760	10	70	0.550	26.037	0.789
7	750	0.800	367.534	0.795	10	70	MBP	26.938	0.793
7	750	0.850	429.494	0.834	10	70	0.600	28.625	0.802
7	750	0.900	500.996	0.877	10	70	0.650	31.570	0.817

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
10	70	0.700	34.808	0.833	10	110	0.995	105.035	0.987
10	70	0.750	38.430	0.852	10	250	MBP	65.936	0.698
10	70	0.800	42.415	0.872	10	250	0.550	72.594	0.713
10	70	0.850	47.281	0.895	10	250	0.600	82.798	0.736
10	70	0.900	53.017	0.923	10	250	0.650	94.349	0.761
10	70	0.950	59.984	0.955	10	250	0.700	107.497	0.786
10	70	0.990	65.613	0.982	10	250	0.750	122.330	0.813
10	70	0.995	65.613	0.982	10	250	0.800	139.158	0.841
10	90	0.550	31.061	0.768	10	250	0.850	158.032	0.872
10	90	MBP	31.333	0.768	10	250	0.900	180.708	0.906
10	90	0.600	34.579	0.783	10	250	0.950	208.497	0.945
10	90	0.650	38.496	0.801	10	250	0.990	236.163	0.983
10	90	0.700	42.880	0.819	10	250	0.995	240.007	0.988
10	90	0.750	47.562	0.841	10	500	MBP	118.714	0.676
10	90	0.800	53.210	0.864	10	500	0.550	134.360	0.696
10	90	0.850	59.416	0.889	10	500	0.600	156.698	0.722
10	90	0.900	66.253	0.919	10	500	0.650	180.141	0.748
10	90	0.950	75.344	0.953	10	500	0.700	208.984	0.776
10	90	0.990	84.359	0.986	10	500	0.750	237.320	0.803
10	90	0.995	84.359	0.986	10	500	0.800	272.104	0.834
10	100	MBP	33.638	0.757	10	500	0.850	309.846	0.865
10	100	0.600	37.865	0.776	10	500	0.900	357.849	0.901
10	100	0.650	42.015	0.792	10	500	0.950	412.392	0.941
10	100	0.700	47.235	0.815	10	500	0.990	474.967	0.984
10	100	0.750	52.243	0.834	10	500	0.995	483.598	0.990
10	100	0.800	59.059	0.860	10	750	MBP	168.186	0.668
10	100	0.850	65.595	0.883	10	750	0.550	194.461	0.692
10	100	0.900	74.595	0.916	10	750	0.600	227.278	0.718
10	100	0.950	83.197	0.947	10	750	0.650	262.450	0.745
10	100	0.990	94.726	0.986	10	750	0.700	303.112	0.773
10	100	0.995	94.726	0.986	10	750	0.750	347.535	0.802
10	110	MBP	35.781	0.749	10	750	0.800	398.748	0.832
10	110	0.550	36.566	0.753	10	750	0.850	460.035	0.865
10	110	0.600	40.955	0.770	10	750	0.900	528.355	0.901
10	110	0.650	45.681	0.789	10	750	0.950	611.768	0.942
10	110	0.700	51.269	0.810	10	750	0.990	703.240	0.983
10	110	0.750	57.741	0.832	10	750	0.995	719.156	0.991
10	110	0.800	64.897	0.856	10	1000	MBP	223.575	0.665
10	110	0.850	72.787	0.883	10	1000	0.550	260.175	0.689
10	110	0.900	82.023	0.914	10	1000	0.600	304.915	0.716
10	110	0.950	93.146	0.950	10	1000	0.650	352.404	0.743
10	110	0.990	105.035	0.987	10	1000	0.700	407.844	0.772

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
10	1000	0.750	467.546	0.800	15	100	0.550	42.193	0.830
10	1000	0.800	540.143	0.831	15	100	MBP	44.171	0.835
10	1000	0.850	620.636	0.864	15	100	0.600	46.133	0.840
10	1000	0.900	714.959	0.900	15	100	0.650	50.037	0.852
10	1000	0.950	824.232	0.941	15	100	0.700	54.600	0.864
10	1000	0.990	954.053	0.984	15	100	0.750	60.572	0.882
10	1000	0.995	977.921	0.992	15	100	0.800	65.771	0.898
15	45	0.550	24.411	0.903	15	100	0.850	71.664	0.916
15	45	0.600	26.368	0.909	15	100	0.900	77.931	0.936
15	45	0.650	27.282	0.912	15	100	0.950	86.022	0.960
15	45	MBP	28.351	0.916	15	100	0.990	96.592	0.990
15	45	0.700	29.256	0.920	15	100	0.995	96.592	0.990
15	45	0.750	30.323	0.925	15	105	0.550	42.872	0.825
15	45	0.800	32.753	0.936	15	105	MBP	44.666	0.830
15	45	0.850	34.100	0.943	15	105	0.600	47.191	0.837
15	45	0.900	37.209	0.961	15	105	0.650	51.460	0.849
15	45	0.950	39.079	0.972	15	105	0.700	56.677	0.864
15	45	0.990	41.075	0.985	15	105	0.750	61.260	0.878
15	45	0.995	41.075	0.985	15	105	0.800	68.105	0.897
15	50	0.550	26.400	0.893	15	105	0.850	73.567	0.914
15	50	0.600	28.254	0.899	15	105	0.900	81.870	0.938
15	50	0.650	30.118	0.906	15	105	0.950	90.081	0.961
15	50	MBP	30.130	0.905	15	105	0.990	99.673	0.990
15	50	0.700	31.204	0.909	15	105	0.995	99.673	0.990
15	50	0.750	33.281	0.918	15	135	0.550	52.522	0.807
15	50	0.800	35.738	0.929	15	135	MBP	52.706	0.807
15	50	0.850	37.085	0.935	15	135	0.600	58.014	0.821
15	50	0.900	40.344	0.951	15	135	0.650	64.144	0.836
15	50	0.950	44.138	0.972	15	135	0.700	70.995	0.852
15	50	0.990	46.441	0.985	15	135	0.750	77.767	0.869
15	50	0.995	46.441	0.985	15	135	0.800	85.850	0.889
15	75	0.550	34.122	0.856	15	135	0.850	94.683	0.910
15	75	MBP	36.797	0.864	15	135	0.900	105.400	0.934
15	75	0.650	39.830	0.873	15	135	0.950	116.816	0.962
15	75	0.700	43.179	0.883	15	135	0.990	126.915	0.985
15	75	0.750	46.556	0.896	15	135	0.995	129.944	0.992
15	75	0.800	50.603	0.910	15	165	MBP	59.840	0.792
15	75	0.850	54.970	0.926	15	165	0.550	60.794	0.794
15	75	0.900	60.018	0.945	15	165	0.600	68.455	0.810
15	75	0.950	66.137	0.968	15	165	0.650	74.943	0.825
15	75	0.990	70.934	0.988	15	165	0.700	83.802	0.844
15	75	0.995	70.934	0.988	15	165	0.750	92.298	0.862

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
15	165	0.800	103.036	0.884	15	1000	MBP	270.276	0.729
15	165	0.850	113.508	0.904	15	1000	0.550	306.386	0.748
15	165	0.900	127.447	0.931	15	1000	0.600	352.423	0.770
15	165	0.950	141.267	0.958	15	1000	0.650	403.140	0.793
15	165	0.990	157.417	0.987	15	1000	0.700	457.282	0.816
15	165	0.995	160.681	0.993	15	1000	0.750	520.607	0.840
15	250	MBP	83.017	0.768	15	1000	0.800	587.765	0.865
15	250	0.550	86.392	0.774	15	1000	0.850	660.385	0.892
15	250	0.600	98.002	0.793	15	1000	0.900	747.456	0.921
15	250	0.650	109.802	0.813	15	1000	0.950	847.207	0.954
15	250	0.700	122.544	0.831	15	1000	0.990	954.144	0.988
15	250	0.750	137.531	0.853	15	1000	0.995	973.212	0.993
15	250	0.800	153.622	0.876	20	50	0.550	30.573	0.932
15	250	0.850	169.822	0.898	20	50	0.600	32.486	0.936
15	250	0.900	190.192	0.925	20	50	0.650	33.428	0.939
15	250	0.950	213.333	0.957	20	50	MBP	35.362	0.944
15	250	0.990	235.368	0.986	20	50	0.750	36.589	0.948
15	250	0.995	238.431	0.990	20	50	0.800	38.892	0.955
15	500	MBP	148.482	0.742	20	50	0.850	40.245	0.960
15	500	0.550	163.510	0.757	20	50	0.900	43.091	0.972
15	500	0.600	186.955	0.778	20	50	0.950	44.876	0.980
15	500	0.650	211.263	0.799	20	50	0.990	46.719	0.989
15	500	0.700	237.674	0.821	20	50	0.995	46.719	0.989
15	500	0.750	267.581	0.845	20	60	0.550	34.896	0.919
15	500	0.800	299.694	0.869	20	60	0.600	36.711	0.923
15	500	0.850	335.649	0.894	20	60	0.650	38.906	0.927
15	500	0.900	376.831	0.922	20	60	MBP	38.632	0.928
15	500	0.950	425.942	0.954	20	60	0.700	40.813	0.932
15	500	0.990	478.890	0.988	20	60	0.750	43.164	0.939
15	500	0.995	486.294	0.992	20	60	0.800	45.556	0.946
15	750	MBP	209.241	0.733	20	60	0.850	48.093	0.954
15	750	0.550	234.129	0.750	20	60	0.900	51.111	0.965
15	750	0.600	268.517	0.773	20	60	0.950	54.736	0.980
15	750	0.650	305.563	0.795	20	60	0.990	56.891	0.989
15	750	0.700	344.375	0.818	20	60	0.995	56.891	0.989
15	750	0.750	389.708	0.841	20	100	0.550	48.092	0.877
15	750	0.800	439.510	0.866	20	100	MBP	51.717	0.883
15	750	0.850	494.635	0.892	20	100	0.650	55.813	0.891
15	750	0.900	558.059	0.921	20	100	0.700	60.001	0.900
15	750	0.950	635.867	0.954	20	100	0.750	64.809	0.910
15	750	0.990	713.024	0.987	20	100	0.800	69.697	0.922
15	750	0.995	726.422	0.993	20	100	0.850	75.460	0.936

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
20	100	0.900	81.608	0.953	20	250	0.600	108.803	0.829
20	100	0.950	89.092	0.973	20	250	0.650	119.393	0.844
20	100	0.990	95.923	0.992	20	250	0.700	132.581	0.861
20	100	0.995	95.923	0.992	20	250	0.750	145.793	0.878
20	140	0.550	61.983	0.850	20	250	0.800	161.484	0.897
20	140	MBP	63.573	0.853	20	250	0.850	176.678	0.916
20	140	0.600	66.888	0.860	20	250	0.900	196.281	0.939
20	140	0.650	73.748	0.870	20	250	0.950	216.826	0.964
20	140	0.700	79.668	0.883	20	250	0.990	237.923	0.988
20	140	0.750	86.855	0.896	20	250	0.995	240.751	0.992
20	140	0.800	94.453	0.911	20	500	MBP	166.241	0.782
20	140	0.850	103.241	0.928	20	500	0.550	179.836	0.793
20	140	0.900	112.535	0.947	20	500	0.600	202.427	0.811
20	140	0.950	123.407	0.969	20	500	0.650	226.804	0.829
20	140	0.990	132.881	0.988	20	500	0.700	253.868	0.848
20	140	0.995	135.685	0.993	20	500	0.750	283.506	0.868
20	180	0.550	74.714	0.832	20	500	0.800	315.569	0.889
20	180	MBP	74.541	0.833	20	500	0.850	350.932	0.911
20	180	0.600	81.849	0.844	20	500	0.900	391.789	0.935
20	180	0.650	90.383	0.858	20	500	0.950	436.398	0.962
20	180	0.700	99.133	0.872	20	500	0.990	481.797	0.990
20	180	0.750	108.456	0.887	20	500	0.995	488.413	0.993
20	180	0.800	118.973	0.904	20	750	MBP	237.261	0.772
20	180	0.850	130.720	0.923	20	750	0.550	260.317	0.786
20	180	0.900	143.466	0.943	20	750	0.600	296.913	0.806
20	180	0.950	158.484	0.967	20	750	0.650	334.219	0.825
20	180	0.990	172.969	0.990	20	750	0.700	376.120	0.845
20	180	0.995	176.059	0.994	20	750	0.750	419.596	0.865
20	220	MBP	86.822	0.820	20	750	0.800	467.931	0.887
20	220	0.550	87.525	0.821	20	750	0.850	518.393	0.909
20	220	0.600	96.790	0.835	20	750	0.900	580.222	0.934
20	220	0.650	107.407	0.850	20	750	0.950	651.418	0.962
20	220	0.700	118.567	0.865	20	750	0.990	721.099	0.989
20	220	0.750	130.855	0.882	20	750	0.995	733.108	0.994
20	220	0.800	144.089	0.900	20	1000	MBP	305.754	0.767
20	220	0.850	158.633	0.919	20	1000	0.550	342.274	0.783
20	220	0.900	175.399	0.941	20	1000	0.600	387.670	0.803
20	220	0.950	193.895	0.966	20	1000	0.650	439.643	0.822
20	220	0.990	212.771	0.991	20	1000	0.700	494.906	0.843
20	220	0.995	215.685	0.995	20	1000	0.750	555.149	0.863
20	250	MBP	95.353	0.812	20	1000	0.800	617.888	0.885
20	250	0.550	97.039	0.814	20	1000	0.850	690.712	0.908

Table 6: *(continued)*

ν	n	α_{MCD}	m_{sim}	c_{sim}	ν	n	α_{MCD}	m_{sim}	c_{sim}
20	1000	0.900	771.181	0.933					
20	1000	0.950	866.274	0.961					
20	1000	0.990	965.106	0.990					
20	1000	0.995	981.677	0.994					

6.2 Full Results of Out of Sample Tests of Proposed Modification to Hardin and Rocke (2005) Methodology

Table 7 provides the out of sample results from testing the model shown in Equation (5). The table shows the logarithm of the ratio of the predicted degrees of freedom to the simulated degrees of freedom.

Table 7: Out of sample performance of the proposed improvement to the Hardin-Rocke methodology, as measured by the logarithm of the ratio of the predicted degrees of freedom to the simulated degrees of freedom. Blank cells correspond to combinations of n, ν, and α_{MCD} that were not part of the out of sample tests. The data in this table is depicted in Figure 7.

		$\alpha_{\mathbf{MCD}} =$												
ν	n	MBP	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	0.990	0.995	1.000
Cases where $n \leq 5\nu$														
2	8	0.355	0.216	0.312	0.395	0.110	0.177	0.240	−0.099	−0.029	0.058	0.171	0.196	−0.381
3	12	−0.039	−0.003	0.088	−0.064	0.012	−0.127	−0.053	0.023	−0.176	−0.071	0.059	0.086	−0.263
5	20	−0.070	−0.037	−0.047		0.007	−0.027	−0.075	0.006	−0.044	−0.091	0.032	0.056	−0.144
8	32	−0.042	−0.066	−0.032	0.001	0.014	−0.042	−0.039	−0.030	−0.017	−0.002	0.010	0.030	−0.081
11	44	−0.048	−0.075	−0.059	−0.021	−0.036	−0.047	−0.012	−0.029	0.003	−0.011	0.026	0.043	−0.027
11	50	−0.033	−0.038	−0.035	−0.014	−0.012	−0.008	−0.008	−0.013	−0.016	−0.024	0.019	0.037	−0.022
16	50	−0.021	−0.045	−0.030	−0.026	0.006	0.003	0.003	0.035	0.028	0.025	0.057	0.071	0.018
16	64	−0.023	−0.042	−0.014	−0.012	0.006	−0.003	0.016	0.035	0.018	0.035	0.041	0.056	0.014
22	50	0.002	−0.053	−0.018	−0.019	0.008	0.006	0.033	0.056	0.048	0.076	0.104	0.116	0.066
22	88	−0.015	−0.040	−0.015	0.007	0.007	0.022	0.035	0.028	0.045	0.058	0.061	0.074	0.048

Table 7: *(continued)*

		$\alpha_{\mathbf{MCD}} =$												
ν	n	MBP	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	0.990	0.995	1.000
Cases where $5\nu < n \leq 10\nu$														
2	12	0.260	0.164	0.060	0.187	0.011	0.122	−0.079	0.028	−0.188	−0.059	0.088	0.117	−0.281
2	16	0.180	0.137	0.094	0.025	0.168	0.081	−0.017	0.114	−0.002	−0.123	0.042	0.074	−0.222
2	20	0.178		0.194	0.187	0.166	0.121	0.068	−0.002	−0.069	−0.125	0.052	0.085	−0.172
3	18	−0.014	−0.017	−0.041	−0.070	0.041	−0.015	−0.073	0.037	−0.032	−0.105	0.049	0.079	−0.184
3	24	−0.018	−0.116	−0.074	−0.075	−0.079	−0.091	−0.109	−0.124	−0.140	−0.146	0.022	0.053	−0.154
3	30	−0.001	−0.056	−0.009	0.046	−0.033	−0.003	0.022	−0.078	−0.055	−0.024	0.008	0.041	−0.121
5	30	−0.027	−0.066		−0.004	0.012	−0.066	−0.053	−0.044	−0.037	−0.008	0.013	0.039	−0.095
5	40	−0.034	−0.037	−0.032	−0.030	0.026	−0.001	−0.041	0.000	−0.026	−0.058	−0.009	0.019	−0.078
5	50	−0.012	−0.038	−0.013	0.012	0.027	−0.022	−0.020	−0.019	−0.021	−0.019	−0.025	0.003	−0.077
8	48	−0.029	−0.048	−0.034	−0.017	−0.015	−0.017	−0.023	−0.034	−0.043	−0.046	0.005	0.026	−0.048
8	50	−0.017	−0.048	−0.028	−0.011	0.001	0.001	0.001	−0.002	−0.010	0.001	−0.007	0.015	−0.049
8	64	0.007	0.005	0.024	0.020	0.010	−0.002	0.021	0.008	−0.009	−0.028	−0.012	0.010	−0.042
8	80	0.016		0.009	0.035	0.026	0.006	0.016	−0.011	0.007	−0.014	−0.013	0.010	−0.029
11	66	−0.013	−0.036	−0.028	−0.024	0.009	−0.001	−0.012	0.014	0.001	−0.010	0.003	0.021	−0.025
11	88	−0.018	−0.009	−0.002	0.004	0.003	−0.005	0.008	−0.001	−0.006	−0.013	−0.009	0.010	−0.027
11	110	−0.006	0.003	0.008	0.018	0.022	0.016	0.014	0.001	−0.005	−0.011	−0.009	0.010	−0.010
16	96	−0.004	−0.029	−0.012	−0.005	0.004	0.012	0.011	0.014	0.017	0.018	0.028	0.044	0.019
16	128	−0.008	0.000	0.012	0.022	0.018	0.014	0.027	0.015	0.028	0.020	0.036	0.027	0.012
16	150	−0.006	0.003	0.012	0.008	0.019	0.018	0.009	0.015	0.008	0.001	0.008	0.005	−0.004
16	160	0.006		0.017	0.028	0.033	0.021	0.025	0.024	0.026	0.028	0.022	0.018	0.011
22	132	−0.008	−0.013	0.001	0.012	0.014	0.027	0.026	0.033	0.026	0.035	0.050	0.044	0.030
22	150	0.007	−0.005	0.018	0.018	0.029	0.025	0.031	0.039	0.029	0.039	0.049	0.044	0.034
22	176	0.005	0.010	0.011	0.018	0.029	0.029	0.030	0.038	0.035	0.028	0.033	0.031	0.024
22	220	0.022		0.030	0.033	0.033	0.030	0.029	0.028	0.027	0.030	0.029	0.028	0.026

Table 7: *(continued)*

		$\alpha_{\mathbf{MCD}} =$												
ν	n	MBP	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	0.990	0.995	1.000
Cases where $10\nu < n \leq 20\nu$														
2	24	0.096	−0.006	0.032	0.041	0.054	0.038	0.035	0.028	0.013	0.007	0.003	0.037	−0.190
3	36	−0.014	−0.032	−0.056	0.014	−0.008	−0.048	−0.002	−0.064	−0.012	−0.059	0.004	0.037	−0.091
3	50	−0.019	−0.050	−0.006	0.029	−0.020	−0.004	0.012	−0.061	−0.057	−0.035	−0.035	0.000	−0.087
5	60	−0.005		0.001	−0.006	0.025	0.002	−0.024	0.003	−0.013	−0.037	−0.009	0.020	−0.044
8	96	0.012	0.004	0.034	0.023	0.028	0.004	0.005	0.009	−0.015	−0.006	−0.023	0.000	−0.031
8	150	0.019	0.006	0.017	0.020	0.019	0.010	0.005	0.007	−0.001	0.000	−0.007	−0.011	−0.022
11	132	−0.008	−0.010	−0.005	0.000	0.001	0.000	−0.011	−0.012	−0.016	−0.019	0.006	−0.002	−0.017
11	150	0.003	0.007	0.015	0.022	0.021	0.019	0.014	0.009	0.003	−0.002	0.010	0.003	−0.011
16	192	0.001	0.010	0.019	0.022	0.018	0.014	0.019	0.016	0.015	0.012	0.014	0.015	0.009
16	300	0.018	0.021	0.023	0.027	0.022	0.014	0.014	0.017	0.013	0.019	0.019	0.024	0.016
22	264	0.007	0.011	0.019	0.023	0.024	0.027	0.027	0.027	0.029	0.028	0.027	0.030	0.020
22	300	0.010	0.017	0.018	0.024	0.023	0.021	0.023	0.021	0.020	0.020	0.025	0.030	0.021

Table 7: *(continued)*

ν	n	$\alpha_{\mathbf{MCD}} =$ MBP	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	0.990	0.995	1.000
Cases where $n > 20\nu$														
2	50	0.029	0.021	0.091	0.046	0.068	0.012	0.021	0.022	−0.048	−0.043	−0.054	−0.017	−0.121
2	150	0.076	0.076	0.090	0.074	0.092	0.050	0.059	0.041	0.010	0.012	0.009	−0.002	−0.014
2	300	0.081	0.082	0.083	0.060	0.048	0.049	0.025	0.005	−0.017	−0.024	−0.040	−0.026	−0.033
2	500	0.062	0.057	0.058	0.029	0.012	0.002	0.003	−0.022	−0.036	−0.041	−0.041	−0.035	−0.046
2	750	0.064	0.058	0.071	0.060	0.033	0.002	−0.005	0.001	−0.008	−0.009	−0.013	−0.015	−0.008
2	1000	0.039	0.023	0.006	0.002	−0.012	−0.012	−0.004	−0.008	−0.021	−0.026	−0.027	−0.031	−0.029
3	150	0.037	0.028	0.042	0.050	0.022	0.015	0.013	−0.013	−0.025	−0.029	−0.027	−0.040	−0.055
3	300	0.037	0.042	0.035	0.024	0.013	0.006	0.015	−0.010	−0.030	−0.029	−0.031	−0.016	−0.024
3	500	0.036	0.032	0.027	0.025	0.026	0.023	0.035	0.014	0.004	−0.003	−0.021	−0.019	−0.021
3	750	0.023	0.028	0.024	0.017	0.018	−0.001	0.007	−0.007	−0.008	−0.015	−0.026	−0.025	−0.030
3	1000	0.063	0.044	0.045	0.039	0.029	0.012	0.013	0.016	0.007	−0.001	0.001	0.007	0.004
5	150	0.032	0.023	0.031	0.029	0.038	0.009	0.007	0.000	−0.008	−0.011	−0.019	−0.029	−0.034
5	300	0.054	0.054	0.041	0.034	0.034	0.019	0.008	0.008	−0.009	−0.018	−0.023	−0.013	−0.027
5	500	0.037	0.049	0.037	0.027	0.032	0.040	0.023	0.028	0.019	0.007	0.002	0.006	−0.001
5	750	0.026	0.022	0.011	0.014	0.019	0.006	0.002	0.001	0.000	−0.003	−0.005	−0.006	−0.008
5	1000	0.031	0.035	0.031	0.028	0.016	0.012	−0.003	−0.009	−0.016	−0.027	−0.031	−0.028	−0.028
8	300	0.041	0.039	0.038	0.042	0.032	0.012	0.010	−0.007	−0.010	−0.011	−0.020	−0.012	−0.022
8	500	0.022	0.022	0.018	0.019	0.013	0.004	0.000	−0.013	−0.011	−0.013	−0.019	−0.017	−0.024
8	750	0.021	0.019	0.018	0.017	0.015	0.011	0.011	0.004	0.003	−0.004	−0.011	−0.012	−0.015
8	1000	0.024	0.016	0.010	0.011	0.007	0.005	0.005	−0.001	0.004	−0.005	−0.004	−0.003	−0.004
11	300	0.012	0.016	0.022	0.018	0.020	0.017	0.012	0.010	0.004	0.004	−0.001	0.005	−0.006
11	500	0.012	0.014	0.017	0.015	0.011	0.005	0.007	0.003	−0.002	−0.001	−0.007	−0.003	−0.008
11	750	0.022	0.016	0.015	0.013	0.017	0.010	0.005	0.000	−0.002	−0.007	−0.003	−0.004	−0.007
11	1000	0.018	0.019	0.017	0.008	0.000	−0.005	−0.002	−0.002	−0.002	−0.005	−0.009	−0.010	−0.009
16	500	0.021	0.017	0.016	0.022	0.021	0.014	0.011	0.010	0.011	0.011	0.011	0.014	0.011
16	750	0.018	0.012	0.010	0.010	0.017	0.016	0.015	0.016	0.009	0.001	0.003	0.003	0.001
16	1000	0.022	0.018	0.018	0.015	0.015	0.008	0.009	0.006	0.006	0.003	0.000	0.000	0.000
22	500	0.015	0.020	0.023	0.023	0.023	0.022	0.018	0.016	0.014	0.012	0.013	0.016	0.010
22	750	0.021	0.025	0.026	0.026	0.024	0.022	0.022	0.025	0.021	0.018	0.018	0.017	0.015
22	1000	0.028	0.028	0.024	0.024	0.020	0.018	0.015	0.013	0.009	0.006	0.006	0.008	0.007

6.3 Replicating the Hardin-Rocke Extension Simulations

6.3.1 The `CerioliOutlierDetection` R Package

This pacakge implements the outlier detection methodology of Cerioli (2010) based on Mahalanobis distances and the minimum covariance determinant (MCD) estimate of dispersion. It also implements the extension to Hardin and Rocke (2005) developed in this paper. The development version can be downloaded via `git` or a web browser from Christopher Green's GitHub repository:

`http://christopherggreen.github.io/CerioliOutlierDetection/`

6.3.2 The `HardinRockeExtension` R Package

This package contains scripts to perform the simulations described in this paper. It can be downloaded via `git` or a web browser from Christopher Green's GitHub repository:

(coming soon)

6.3.3 R Session Details

```
> sessionInfo()
R version 3.0.2 (2013-09-25)
Platform: x86_64-w64-mingw32/x64 (64-bit)

locale:
[1] LC_COLLATE=English_United States.1252
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252

attached base packages:
[1] parallel  stats     graphics  grDevices utils     datasets
[7] methods   base

other attached packages:
[1] HardinRockeExtension_1.0      rrcov_1.3-4
[3] pcaPP_1.9-49                  mvtnorm_0.9-9997
[5] abind_1.4-0                   CerioliOutlierDetection_1.0.0
[7] robustbase_0.90-2

loaded via a namespace (and not attached):
[1] DEoptimR_1.0-1 stats4_3.0.2
```

References

Fatemah Alqallaf, Stefan van Aelst, Victor J. Yohai, and Ruben H. Zamar. Propagation of outliers in multivariate data. *Annals of Statistics*, 37(1):311–331, 2009.

D. F. Andrews, R. Gnanadesikan, and J. L. Warner. Methods for assessing multivariate normality. In P. R. Krishnaiah, editor, *Multivariate Analysis*, volume III, pages 95–116. Academic Press, New York, 1973.

C. Becker and U. Gather. The masking breakdown point of multivariate outlier identification rules. *Journal of the American Statistical Association*, 94(447):947–955, 1999.

C. Becker and U. Gather. The largest nonidentifier outlier: A comparison of multivariate simultaneous outlier identification rules. *Computational Statistical & Data Analysis*, 36:119–127, 2001.

Andrea Cerioli. Multivariate outlier detection with high-breakdown estimators. *Journal of the American Statistical Association*, 105(489):147–156, 2010.

Andrea Cerioli, Marco Riani, and Anthony C. Atkinson. Controlling the size of multivariate outlier tests with the mcd estimator of scatter. *Statistical Computing*, 19:341–353, 2009. doi: 10.1007/s11222-008-9096-5.

Christopher Croux and Gentiane Haesbroeck. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*, 71:161–190, 1999.

J. A. Cuesta-Albertos, C. Matran, and A. Mayo-Iscar. Trimming and likelihood: robust location and dispersion estimation in the ellipitcal model. *The Annals of Statistics*, 36:2284–2318, 2008.

Rosario Dell'Aquila and E. Rochetti. *Robust statisics and econometrics with economic and financial applications*. John Wiley & Sons, New York, 2006. Not yet published as of 8/6/2008.

R. Gnanadesikan and J. R. Kettenring. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28(1):81–124, 1972. ISSN 0006341X. URL `http://www.jstor.org/stable/2528963`.

Yuri Goegebeur, Armelle Guillou, and Andréhette Verster. Robust and asymptotically unbiased estimation of extreme quantiles for heavy tailed distributions. *Statistics and Probability Letters*, 87:108–114, 2014.

Christopher G. Green and R. Douglas Martin. Untitled paper on cerioli et al. (2009). Working paper, 2014a.

Christopher G. Green and R. Douglas Martin. Diagnosing the presence of multivariate outliers in financial data using calibrated robust mahalanobis distances. Working paper, 2014b.

J. Hardin and D. M. Rocke. The distribution of robust distances. *Journal of Computational and Graphical Statistics*, 14:928–946, 2005.

D. M. Hawkins and D. J. Olive. Improved feasible solution algorithms for high breakdown estimation. *Computational Statistics and Data Analysis*, 30:1–11, 1999.

P. C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2:49–55, 1936.

K. Mardia, J. Kent, and J. Bibby. *Multivariate Analysis*. Academic Press, New York, 1979.

Ricardo Maronna, R. Douglas Martin, and Victor Yohai. *Robust Statistics: Theory and Practice*. John Wiley & Sons, New York, 2006.

Daniel Peña and Francisco J. Prieto. Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, 43(3):286–300, 2001. ISSN 00401706. URL `http://www.jstor.org/stable/1271215`.

E. Pearson and C. Chandra Sekar. The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, 28:308–320, 1936.

Marco Riani, Anthony C. Atkinson, and Andrea Cerioli. Finding an unknown number of multivariate outliers. *Journal of the Royal Statistical Society*, 71, 2009.

D. Rocke and D. Woodruff. Identification of outliers in multivariate data. *Journal of the American Statistical Association*, 91:1047–1061, 1996.

Peter J. Rousseeuw. Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, editors, *Proc. of the 4th Pannonian Symp. on Math. Stat., Bad Tatzmannsdorf, Austria, 4-10 September, 1983*, volume B, pages 283–297. Kluwer Academic Publishers, 1985.

Peter J. Rousseeuw and Katrien van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999. ISSN 00401706. URL `http://www.jstor.org/stable/1270566`.

Peter J. Rousseeuw and B. C. van Zomeren. Unmasking multivariate outlier and leverage points. *Journal of the American Statistical Association*, 85:633–639, 1990.

Peter J. Rousseeuw and B. C. van Zomeren. Robust distances: Simulation and cutoff values. In W. Stahel and S. Weisberg, editors, *Directions in Robust Statistics and Diagnostics*, volume 2. Springer-Verlag, New York, 1991.

G. A. F. Seber. *Multivariate Observations*. John Wiley and Sons, New York, 1984.

R. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1980.

N. J. H. Small. Plotting squared radii. *Biometrika*, 65:657–658, 1978.

S. Wilks. *Mathematical Statistics*. Wiley, New York, 1962.