Chapter 3

High Breakdown Estimation for Multivariate SPC Data

Robust estimation methods for univariate quality control data (such as those based on a median or trimmed mean) are straightforward and have received attention in past research (Rocke, 1989; Rocke, 1992; Tatum, 1997; de Mast and Roes, 2004; Davis and Adams, 2005). Robust methods for multivariate data are not as straightforward, nor as easily implemented. Robust estimation methods have been widely used in a regression context but they have only recently been introduced to multivariate quality control applications. Because of the differences that can result from competing methods, the choice of which robust estimator to use has not been made clear from previous studies (Wisnowski, Simpson, and Montgomery, 2002; Vargas, 2003).

To evaluate the performance of competing methods for Phase I applications the probability of a signal is the preferred measure. When the data come from an in-control process then the probability of a signal should be close to a specified nominal value. When data come from an out-of-control process then the probability of a signal should be large to ensure that the out-of-control points are not included in the calculation of the control limit for Phase II.

In this chapter we give a brief overview of high breakdown estimation methods and various high breakdown estimation methods based on the minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD) for multivariate Phase I applications. A comprehensive simulation study allows us to determine the conditions under which each method is preferred. We also give control limit values for practical use.

3.1 Properties of Estimators

There are four major measures or properties that can be used to determine the usefulness of a multivariate estimator. The first, the breakdown point, has many different definitions, but the definition used here is the finite sample replacement breakdown point as defined by Donoho and Huber (1983). This value, π, is the smallest fraction of arbitrarily large bad data points that can be present before the estimator is impacted. As the sample size increases, π will often converge to an asymptotic breakdown point. The asymptotic breakdown point is often used to compare different estimators.

Classical estimation methods have low breakdown points while the high breakdown estimators considered here have breakdown points that approach 50%, the maximum possible value. The higher the breakdown point, the more resistant the estimator is to bad data. In other words, the less susceptible it is to the "masking effect".

The second property to consider is that of affine equivariance. Changing the measurement scale should not impact the properties of the estimator. Lopuhaä and Rousseeuw (1991) showed that the maximum possible asymptotic breakdown point for an affine equivariant estimator is 50%. The estimators of location and dispersion that are considered here are all affine equivariant (Rousseeuw and Leroy, 1987). For an example of non-affine equivariant

estimators, see Maronna and Zamar (2002) who found that alternative estimators can be found by relaxing the restriction of affine equivariance.

The third property is the statistical efficiency of the estimator. This concerns how well it makes use of all the good data available. For the univariate case it is well known that while the median is very robust, it is also very inefficient when compared to the mean. There often has to be some tradeoff between increasing the breakdown point and the decreasing efficiency.

Finally, it should be possible to calculate the estimators with a reasonable amount of computing power in a reasonable amount of time. It should not always be expected that a reasonable time to compute the estimators be only a few seconds. It is good to spend the necessary time to get good estimators that give accurate information in the spirit of the following statement: "Statistical analysis is generally just a small part of the effort and cost of any data gathering and analysis . . . we consider it clearly far better to use an analysis that takes 10 hours but finds all the outliers than one that takes 10 seconds yet misses most outliers" (Hawkins and Olive, 2002, p. 146).

3.2 High Breakdown Estimation

Robust estimation methods can be used in two different approaches. The first approach is to use the robust estimators in place of classical estimators. This has been the primary focus of a large amount of research dedicated to robust estimation procedures and is most useful in a regression context where the data does not necessarily have a given time order. Here the goal is to identify, for descriptive and predictive purposes, a good model that has not been unduly influenced by outliers. This approach has a higher priority on efficiency.

The second approach is to use the robust estimators to identify and remove outliers and then use classical estimators on the remaining "good" data points. Phase I quality control applications (both univariate and multivariate) have predominantly utilized this second approach. This second approach seems to be a reasonable trade off between the good efficiency of the classical estimates and the high breakdown point of resistant methods. Under this framework robust methods that are efficient are not as useful if they have lower breakdown points. In this second approach, the statistical properties are not as well defined and some authors have disapprove of such "ad hoc" type methods (de Mast and Roes, 2004).

When using the second approach, the computability and breakdown point of the estimator become more important. As a consequence, statistical efficiency is not as crucial because the resistant estimators will eventually be replaced by classical estimators. Therefore estimators based on the minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD) are considered here. Algorithms for computing them are more plentiful, they are affine equivariant, and most importantly, they have high breakdown points. They have lower statistical efficiency because they only use slightly more than half of the available points, but this is of minor concern in Phase I analysis, especially when the Phase I data set is sufficiently large. The main concern in our Phase I setting is to provide protection against outliers.

There is a wide variety of robust estimation methods that are not considered here for multivariate data. For example, methods based on M-estimation have been widely used in a regression context. M-estimation seeks to appropriately down weight outliers in order to minimize their impact. As such, they are more efficient than the high breakdown methods considered here, but they have lower breakdown points that get even worse as the number of dimensions increases. Other methods include S-estimation, the projection methods of Stahel-

Donoho (Rousseeuw and Leroy, 1987, Section 7.1.c), and the sequential point addition type methods of Hadi (1992, 1994) and Atkinson (1993). These other methods are usually applied to regression problems.

3.3 Minimum Volume Ellipsoid Estimator

The minimum volume ellipsoid (MVE) estimator, first proposed by Rousseeuw (1984), has been studied extensively for non-control chart settings and frequently used in the detection of multivariate outliers. One seeks to find the ellipsoid of minimum volume that covers a subset of at least h data points. Subsets of size h are called halfsets because h is often chosen to be just more than half of the m data points. The location estimator is the geometrical center of the ellipsoid and the estimator of the variance-covariance matrix is the matrix defining the ellipsoid itself, multiplied by an appropriate constant to ensure consistency (Rousseeuw and van Zomeren, 1990; Rousseeuw and Van Zomeren, 1991; and Rocke and Woodruff, 1996). Thus the MVE estimator of location and dispersion do not correspond to the sample mean vector and sample variance-covariance matrix of a particular halfset.

To achieve the highest breakdown point possible, Davies (1987) and Lopuhaä and Rousseeuw (1991) showed that the integer value of $h = (m+p+1)/2$ should be used for the MVE. This will achieve a breakdown value of $\frac{[(m-p+1)/2]}{m}$ percent which converges to 50% as $m \to \infty$. The value of h can be increased, to say, $.75m$, if it is believed that the percentage of bad data is low. This will increase the efficiency of the MVE estimator. However caution must be exercised because the consequences of having a value of h higher than the number of good data points is more severe (contaminated estimates) than the consequences of having a value of h lower than the number of good data points (loss of statistical efficiency but

still giving good estimates). For this reason, h is often set to achieve the highest breakdown point possible, as is the case here.

Finding the MVE estimators is essentially a two-part process. The first part is to find the best halfset consisting of h points. The second part requires finding the ellipsoid of minimum volume that covers the halfset. For a given halfset there are many ellipsoids that cover it. Titterington (1975) found that the solution to this second step is equivalent to finding a D-optimal design for a design region where the points in a halfset are the design points. As a consequence, iterative algorithms to find D-optimal designs could be used to find the best covering ellipsoid for the best halfset. The first step is referred to as the "subset" problem and the second step is referred to as the "covering problem" (Agullo, 1996).

While the idea of the MVE is very intuitive, actually finding the MVE estimator can be very difficult in practice. As the sample size (m) and data dimension (p) increase, the required computational effort increases dramatically. For example, if $m = 30$ and $p = 3$, so that $h = \frac{30+3+1}{2} = 17$, then there are a total of $\frac{30!}{13!\,17!} = 119,759,850$ halfsets that could potentially be the basis for the MVE estimator. Even when this halfset is found, it takes additional calculations to find the best covering ellipsoid for the particular halfset. As a consequence of the computational difficulty, Rousseeuw and Leroy (1987) proposed an approximate method to find the MVE estimators by a subsampling algorithm.

3.3.1 Subsampling Algorithm

The subsampling algorithm is very commonly used, is widely accessible, and is the basis of the MVE functions of software packages such as S-Plus and $SAS^{®}$. This subsampling algorithm takes a fixed number of random subsets, known as elemental subsets, each containing $p + 1$ points. For each elemental subset, the sample mean vector and sample variance-covariance

matrix are calculated, which determines the shape of an ellipsoid. This ellipsoid is then increased in size by multiplying by a constant until the inflated ellipsoid covers at least h data points. The inflated ellipsoid with the smallest volume is then used to obtain the MVE estimates.

Rousseeuw and Leroy (1987, p. 199) recommended doing a minimum of 500 subsamples for small datasets with low dimensions. More subsamples should be used as m and p increase. Rousseeuw and Leroy (1987, p. 260) showed that if ϵ is the true proportion of outliers in the dataset then a probabilistic argument can be used to determine the number of random subsamples (j) needed to ensure with a high probability (α) that at least one contains only good points. The probability that at least one sample contains only good points is

$$\alpha = 1 - (1 - (1 - \epsilon)^{p+1})^j, \tag{3.1}$$

and (3.1) can be rewritten to solve for j as

$$j = \frac{\ln(1 - \alpha)}{\ln(1 - (1 - \epsilon)^{p+1})}. \tag{3.2}$$

Use of (3.2) shows that when $p \leq 5$ and $\epsilon \leq .50$ then 500 subsamples will ensure that α will be greater than .999. If $p \leq 10$ and $\epsilon \leq .50$ then $10,000$ subsamples will ensure that α will be greater than .99.

A similar argument along these lines to determine the probability that a particular halfset will contain only good points shows that the number of halfsets that need to be considered is very large indeed. To see this value, replace the value of $p + 1$ (the size of the elemental subset) in (3.2) with the integer value of $h = (m+p+1)/2$ (the size of the halfset). Table 3.1 shows the number of halfsets that need to be considered when $p = 3$ to have a 95% chance of obtaining one containing only good points. This illustrates the difficulty in finding a good MVE estimator, particularly as the sample size and the proportion of bad data increase.

Table 3.1: Number of halfsets that need to be considered to ensure that one only contains good points with probability .95 for sample size (m) and proportion of outliers (ϵ) for p=3.

	ϵ									
m	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
20	4	10	20	43	94	215	526	1375	3909	1.23E+04
30	6	17	46	132	398	1287	4538	1.77E+04	7.77E+04	3.93E+05
40	8	29	106	405	1678	7661	3.91E+04	2.28E+05	1.54E+06	1.26E+07
50	11	51	240	1238	7076	4.56E+04	3.37E+05	2.93E+06	3.07E+07	4.02E+08
60	14	86	542	3780	2.98E+04	2.71E+05	2.91E+06	3.76E+07	6.09E+08	1.29E+10
70	19	147	1224	1.15E+04	1.26E+05	1.61E+06	2.50E+07	4.84E+08	1.21E+10	4.12E+11
80	25	249	2759	3.52E+04	5.30E+05	9.60E+06	2.16E+08	6.23E+09	2.41E+11	1.32E+13
90	32	423	6219	1.07E+05	2.23E+06	5.71E+07	1.86E+09	8.01E+10	4.78E+12	4.22E+14
100	42	717	1.40E+04	3.28E+05	9.40E+06	3.40E+08	1.60E+10	1.03E+12	9.50E+13	1.35E+16

In addition, it is important to recognize that while this subsampling approach is computationally easier, it is only an approximation. Even when elemental subsets with only good points are used, this does not ensure that the resulting halfsets will only have good points. This is because the resulting ellipsoid that covers the halfset is proportional to the ellipsoid for the corresponding elemental subset, which is not necessarily the minimum volume ellipsoid for the halfset. An exhaustive calculation using all possible elemental subsets will still yield an approximate estimator (Cook and Hawkins, 1990). This is because the ellipsoid for the MVE estimator is not necessarily proportional to the ellipsoid for any of its elemental subsets. In fact, it is almost certain that the MVE is not proportional to the ellipsoid of one of its elemental subsets.

While the number of possible elemental subsets is smaller than the number of possible halfsets, there is still the same drastic increase in number of possible elemental subsets as m and p increase. For the example shown earlier with $m = 30$ and $p = 3$, there are a total of $\frac{30!}{4!\,26!} = 27,405$ elemental subsets that need to be considered if an exhaustive calculation were done. This increase in the number of elemental subsets needed to ensure a good approximate estimate limits the types of problems that can be analyzed using the MVE.

Finally, there are potential repeatability issues with this subsampling approach. If an

exhaustive calculation using all possible elemental subsets is not done, then two different analyses on the same data set will likely yield different results. We believe this to be a potential drawback of the method because it removes the objectiveness of the analysis. One could conceive of a situation where if the initial analysis done by a consultant was not to a client's satisfaction, that the analysis could simply be redone with a different set of randomly generated subsets to obtain different results. The lack of repeatability is not a problem for smaller sample sizes because all possible subsamples can be used. For larger samples sizes, the difference in results from multiple analyses of the same data tends to be more severe as m and p get larger, because for a fixed number of random elemental subsets, the proportion of subsets that can be feasibly calculated relative to the total number of available subsets gets smaller. However, our simulation studies later in this chapter show that the MVE is preferred for smaller sample sizes, thus repeatability will not impact our conclusions.

For an example of the repeatability issues of the subsampling method, Vargas (2003) calculated the $T^2_{mve,i}$ statistics based on the MVE estimators using the subsampling algorithm for the data of Quesenberry (2001). Table 3.2 shows the results obtained by Vargas using S-PLUS which we call Method 1, our results using the "call mve" function of $SAS^{®}$ for 500 subsamples which we call Method 2, and our results using all $\frac{30!}{3!\,27!} = 4,060$ possible subsamples, which we call Method 3. Notice here that different values are obtained depending on the number of subsamples used. It is not clear what number of subsamples or what covering methods that Vargas (2003) used.

3.3.2 Alternative Approaches

To avoid some of the difficulties with the subsampling approach, an exact method to calculate the MVE estimators was proposed by Cook, Hawkins, and Weisberg (1993). It considers

Table 3.2: Comparison of $T^2_{mve,i}$ obtained by MVE subsampling algorithm for Quesenberry (2001) data. Method 1 is the results from Vargas (2003), Method 2 is the call MVE function from $SAS^{®}$ for 500 subsamples and Method 3 is all possible subsamples.

Observations	Method 1	Method 2	Method 3
1	0.835	0.860	0.921
2	25.770	30.803	24.960
3	0.432	0.484	0.353
4	2.398	2.700	2.614
5	1.434	1.420	1.506
6	0.227	0.274	0.313
7	1.143	1.282	1.292
8	1.039	1.126	0.928
9	0.064	0.066	0.094
10	0.867	1.064	1.034
11	0.878	0.970	0.768
12	1.175	1.332	1.033
13	0.467	0.560	0.585
14	5.712	6.815	6.101
15	0.183	0.220	0.121
16	5.117	5.212	4.949
17	2.268	2.194	2.303
18	3.060	3.014	3.151
19	1.702	1.865	1.868
20	6.736	6.770	6.569
21	1.885	1.818	1.899
22	6.385	7.896	5.952
23	0.380	0.367	0.390
24	1.012	1.119	1.146
25	1.637	1.580	1.631
26	0.476	0.477	0.440
27	0.576	0.604	0.509
28	4.622	5.648	4.265
29	3.329	3.977	3.044
30	0.181	0.182	0.218

all the possible halfsets and would require an enormous amount of computation even for modest sample sizes in lower dimensions. Once the best halfset is found the "covering" solution is found using the approach of Titterinton (1975). To speed up the algorithm they proposed a modification based on the fact that the ellipsoid cannot decrease in volume with each successive iteration. The volume is measured by the determinant so in the algorithm if a subset of points yields a value for the determinant larger than the current best value, then the halfset is not evaluated any further. This modification allows the calculation of the exact MVE without the explicit calculation of the minimum covering for every halfset. This

speeds up the algorithm considerably, as a great majority of halfsets do not require explicit calculation. Cook, Hawkins, and Weisberg (1993) found that for typical datasets fewer than 1% of the possible halfsets require explicit evaluation. However, even with this speedup of the algorithm, this exact method is still only feasible for small datasets where $m \leq 30$ and $p \leq 5$ (Cook, Hawkins, and Weisberg, 1993).

Agullo (1996) proposed an exact method to calculate the MVE estimators based on a more computationally efficient branch and bound method. Similar to the modification proposed by Cook, Hawkins, and Weisberg (1993) to speed up their algorithm, the branch and bound method utilizes the fact that the volume of a subset of points cannot decrease as additional points are added. In other words, the volume is monotonically non-decreasing as points are added to the subset. For example, consider the situation with $p = 2$, $m = 30$, and $h = 16$. During the search if a subset of 8 points is found to have a higher volume (as measured by a determinant) than the best halfset found to that point, then no further halfsets containing those 8 points need to be considered. This reduces substantially the number of halfsets for which a determinant is calculated. Once the best halfset is found, Agullo (1996) recommended using an algorithm by Atwood (1973) that is faster than the approach of Titterington (1975) to solve the "covering problem". The branch and bound algorithm can be sped up by ordering the data prior to beginning the search. As a result, the branch and bound method is computationally feasible for datasets where $m \leq 100$ and $p \leq 5$.

Other computationally feasible methods to find an approximate MVE have been proposed. For example, Hawkins (1993) proposed a feasible solution algorithm (FSA). This algorithm considers a randomly selected halfset (called a random start) and then makes use of swapping techniques to find a better halfset for which its covering ellipsoid is found. Then the procedure

is repeated for many randomly selected halfsets, each of which converges to a local feasible solution. The MVE estimators are based on the minimum of the local solutions. If enough randomly selected halfsets are used, this algorithm will eventually yield an exact solution, but this will not be guaranteed for a finite number of halfsets. If we denote by θ the proportion of initial halfsets that will yield the best halfset, then the probability of finding the exact result, Pr(exact), is $1 - (1 - \theta)^N$ where N is the number of random starts. This expression can be use to determine the number of random starts that is needed to achieve a certain probability of getting the exact results. Hawkins (1993) showed that for many common datasets previously studied in the literature (with $m \leq 50$ and $p \leq 5$) that the N required to achieve a high probability of success is often less than 100 (Hawkins, 1994) so the computation time is substantially smaller than those of Cook, Hawkins, and Weisberg (1993) and Agullo (1996).

Croux and Haesbroeck (1997, 2002) showed how the efficiency of the subsampling approach can be improved for the MVE. Instead of just picking the optimal elemental subset that gives the minimum volume, they first computed the ordered minimum volumes and then averaged some of the smallest ones. These estimators still retain consistency, affine equivariance, and have the highest possible breakdown point. However, it is still an approximate method and if an exhaustive calculation is not done, this averaged approach still has the repeatability problem.

Methods to find the MVE based on a heuristic search algorithms were proposed by Woodruff and Rocke (1993). These search algorithms reduce the amount of computing time needed to solve the "subset" problem and include genetic algorithms, simulated annealing, and their corresponding enhanced versions.

Many of these alternative algorithms can be shown to be more computationally efficient

than the subsampling method and do not suffer from the repeatability issue. They have been shown to give good results but in our comparison to be discussed later, the alternative algorithms are not considered because the algorithms to compute them are not easily accessible. In addition, because of the prevalence of the subsampling method to obtain the MVE estimator, we wish to show where it does not work very well.

3.4 Minimum Covariance Determinant Estimator

An alternative high breakdown estimation procedure to the MVE is an estimator based on the minimum covariance determinant (MCD), which was first proposed by Rousseeuw (1984). It is obtained by finding the halfset that gives the minimum value of the determinant of the variance-covariance matrix. The resulting estimator of location is the sample mean vector of the points that are in the halfset and the estimator of the dispersion is the sample variance-covariance matrix of the points multiplied by an appropriate constant to ensure consistency, just as was done for the MVE. Thus in contrast to the MVE, the MCD estimators correspond to the sample mean and sample variance-covariance matrix of a specific halfset. Thus once the best halfset is determined, the estimators are calculated using the expressions in (2.2) and (2.3) where m is replaced with h. Because the MCD estimators are simple to calculate once the best halfset is found, they can be easier to compute than for the MVE.

The MCD estimators are intuitively appealing because a small value of the determinant corresponds to a small eigenvalue, which suggests near linear dependencies of the data in the p-dimensional space. A near linear dependency suggests that there is a group of points that are similar to each other, which would be a desirable group of points to use for the estimator, particularly if there are outliers present.

Like the MVE, the MCD estimators have the same maximum breakdown point, which is achieved when h is the integer value of $(m + p + 1)/2$. Furthermore, the MCD estimators can be very computationally difficult to obtain because of the rapid increase in the number of potential halfsets that need to be considered as m and p increase. As a result, the approximate methods and algorithms to obtain MVE estimates can also be used to obtain the MCD estimates. For example, MCD estimates can be computed via the exact method of Cook, Hawkins, and Weisberg (1993). The branch and bound method of Agullo (1996) can also be used, as shown in Agullo (2001). The subsampling approach of Rousseeuw and Leroy (1987) can be used to get an approximate MCD estimates which would have the same repeatability issues as the approximate MVE obtained via subsampling. The feasible solution algorithm of Hawkins (1993) can be implemented for the MCD, as shown by Hawkins (1994). An improved version of the feasible solution algorithm for the MCD was proposed by Hawkins and Olive (1999).

3.5 Hybrid Algorithms

Other high breakdown estimation methods for detecting multivariate outliers are hybrid algorithms that combine various components of earlier methods with modifications. Two notable ones are the hybrid algorithm of Rocke and Woodruff (1996) and the FAST-MCD algorithm of Rousseeuw and Van Driessen (1999).

The hybrid algorithm of Rocke and Woodruff (1996) is a combination of the data partitioning methods of Woodruff and Rocke (1994), the FSA algorithm involving the MCD from Hawkins (1994), a sequential point addition algorithm, and M-estimation. This hybrid algorithm is very effective in detecting a larger percentage of outliers. A more complete

explanation of the algorithm and the justification for its various components can be found in Rocke and Woodruff (1997).

Rousseeuw and Van Driessen (1999) proposed a hybrid algorithm which they called the FAST-MCD that is based on an iterative scheme and the MCD estimators. The algorithm can be described as follows:

1. Start with a fixed number, A, of random elemental subsets and use them to construct corresponding halfsets.

2. Carry out two concentration steps (C-step) on the A halfsets and select a small number of "best" ones.

3. For the "best" halfsets, carry out C-steps until convergence and the FAST-MCD estimators are based on the halfset with the lowest determinant of the covariance matrix.

The C-steps are based on the fact that for any given halfset and its estimates of location and dispersion, a better (or at least equivalent) solution can be found by reordering the observations of the full dataset according to their Mahalanobis distances from the sample mean based on the sample variance-covariance matrix. A new and improved halfset of the reordered points is found by selecting from the full dataset those with the smallest Mahalanobis distances. The new halfset will have a smaller determinant of the variance-covariance matrix than the determinant of the original halfset. So each C-step yields a halfset that is more concentrated than the previous halfset. If enough C-steps are done on enough halfsets, convergence to the exact MCD estimator results. Because not all halfsets are considered, the FAST-MCD will be an approximate method unless a large enough number of initial halfsets are considered.

The FAST-MCD method is able to handle large data sets within a reasonable amount of time. In fact, Rousseeuw and Van Driessen (1999) successfully analyzed a data set with $m = 132,402$ and $p = 27$, which is certainly beyond the capabilities of all the high breakdown estimation algorithms discussed previously. For smaller datasets that they analyzed (all with $m \leq 75$ and $p \leq 5$), the FAST-MCD algorithm resulted in estimates that were equivalent to the exact MCD estimates. This means that the number of halfsets considered was large enough to achieve convergence to the exact MCD estimates. It remains to be seen how large m and p can be and still obtain the exact result with a high probability. The control charts that are considered here generally use smaller values of m and p suggesting that the FAST-MCD for practical purposes is likely to give the exact result.

Because of the concentration steps, the FAST-MCD algorithm does not have the repeatability issues that are present in the subsampling algorithm. Thus the FAST-MCD serves as a better algorithm to obtain the MCD estimator than the subsampling algorithm for the MVE estimator especially as m and p increase.

3.6 Asymptotic Properties

The MCD and MVE estimators have been used historically as a starting point for other robust estimation procedures, such as M-estimation. As such, it has not been as important that the MCD and MVE estimators be exact. However, in Phase I quality control applications, the MCD and MVE are used directly to determine multivariate outliers and thus it becomes more important that they be sufficiently accurate. It is also important to have some understanding of the distributions of the MCD and MVE estimators in order to be able to obtain an appropriate control limit for the $T^2_{mve,i}$ and $T^2_{mcd,i}$ statistics. The distributions

of the exact MCD and MVE estimators of location and dispersion are not known in closed form. So when quantiles are needed from the distributions to calculate a control limit, they have been found via simulation (See Vargas, 2003; or Williams, Woodall, and Birch, 2003).

However, the asymptotic distributions of the MVE and MCD estimators can be derived. Davies (1987, 1992) showed that the exact MVE estimators of location and dispersion are consistent for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given that the $\mathbf{x}_i$ are independently and identically distributed with a common distribution. Butler, Davies, and Jhun (1993) showed the corresponding result for the exact MCD estimators of location and dispersion. However, the MCD estimators converge to its population counterparts at a rate of $n^{-1/2}$ while the MVE estimators converge at a slower rate of $n^{-1/3}$, thus the MCD estimators are more efficient. In addition, the distribution of the MCD estimator of location converges to a normal distribution, which is not necessarily the case for the MVE estimator of location. Thus, the asymptotic properties of the MCD estimators are superior to those of the MVE estimators. An intuitive reason for the superior convergence properties of the MCD can be found by noting that as $\epsilon \to 0$ the location MVE estimator converges to the center of the ellipsoid covering all the data while the location MCD estimator converges to the mean vector of all the points.

The asymptotic distributions of the $T^2_{mve,i}$ and $T^2_{mcd,i}$ statistics follow directly from the consistency of the MVE and MCD estimators, as seen in the following theorems.

Theorem 3.1 As $m \to \infty$, the distribution of $T^2_{mve,i}$ converges in distribution to a χ^2_p distribution for $i = 1, \ldots, m$.

Proof. The assumption of multivariate normality satisfies the conditions of Theorem 3 of Davies (1987), therefore the MVE estimators are consistent, i.e., they converge in probability to their parameter values, so we write $\mathbf{x}_{mve} \xrightarrow{p} \boldsymbol{\mu}$ and $\mathbf{S}_{mve} \xrightarrow{p} \boldsymbol{\Sigma}$ as $m \to \infty$. We note that if

a sequence of random variables converges to a constant then a transformation of the sequence converges to the transformation of the constant. This implies that $\mathbf{S}_{mve}^{-1} \xrightarrow{p} \boldsymbol{\Sigma}^{-1}$ because the matrix inverse is just a transformation. Thus we then have

$$T_{mve,i}^2 = (\mathbf{x}_i - \mathbf{x}_{mve})' \mathbf{S}_{mve}^{-1} (\mathbf{x}_i - \mathbf{x}_{mve}) \xrightarrow{p} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \sim \chi_p^2$$

Theorem 3.2 As $m \to \infty$, the distribution of $T_{mcd,i}^2$ converges in distribution to a χ_p^2 distribution for $i = 1, \ldots, m$.

Proof. Same as the proof of the previous theorem but replacing Theorem 3 of Davies (1987) with Theorem 3 of Butler, Davies, and Jhun (1993) to show the consistency of the MCD estimators.

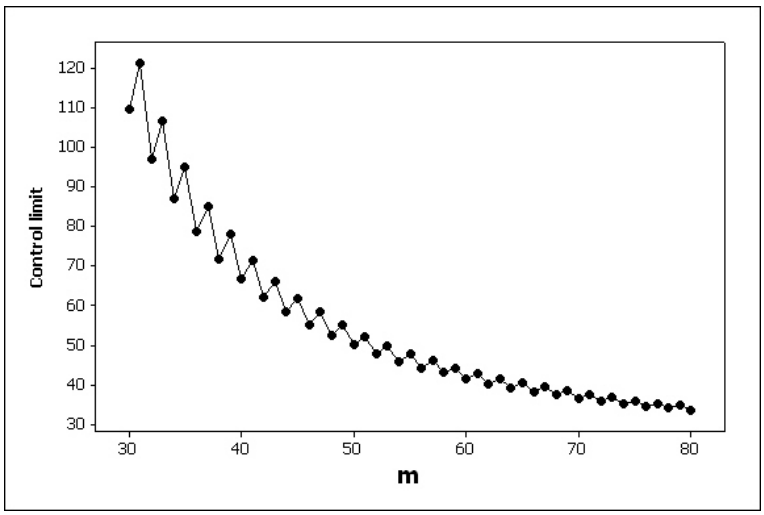
It should be noted that because the subsampling algorithm to obtain the MVE estimators and the FAST-MCD algorithm are approximations, their asymptotic distributions are not necessarily χ_p^2. If it were computationally feasible to compute exactly the MVE and MCD estimators, the control limit could be easily approximated using the corresponding quantile of the χ_p^2 distribution when the Phase I sample size is large. It should also be noted that as the proportion of bad points, ϵ, goes to 0, the $T_{mcd,i}^2$ statistic converges to the $T_{1,i}^2$ statistic which has a χ_p^2 distribution.

3.7 Control Limits

As mentioned in the previous section, the distribution of the T^2 statistic based on the MVE and MCD estimators are only known asymptotically, thus implementation of the Phase I control chart requires control limits to be generated via simulation. Tables 3.3 - 3.6 contain

the control limits for T^2 statistic based on the MVE estimators obtained via the subsampling method and for the MCD estimators obtained via the FAST-MCD method. To obtain the simulated control limits, 200,000 data sets were generated for each combination of m and p with a zero mean vector and the identity covariance matrix. Datasets rather than individual values are simulated because of the dependence of the Phase I T^2 statistics within a dataset (Mason and Young, 2002). Due to the invariance of the $T^2_{mve,i}$ and $T^2_{mcd,i}$ statistics, these limits will be applicable for any values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The T^2 statistics for each observation in the data set were calculated and the maximum value attained for each data set was recorded. The 95^{th} percentile of this generated empirical distribution is the simulated control limit. As will be seen shortly, use of control charts based on $T^2_{mve,i}$ and $T^2_{mcd,i}$ will be preferred for different situations. Thus the control limits are only provided for the situations where the particular estimator is preferred.

Figure 3.1: Plot of simulated control limits for the MCD estimator versus the size of the data set, m, when $p = 3$.



The control limits are dependent on the integer value of h used and are not monotonic functions of m. For example, consider Figure 3.1, which shows the scatterplot of the control

limit of the chart based on $T^2_{mcd,i}$ vs. m for $p = 3$ where the overall probability of a signal is .05. The sawtooth pattern here is also present when using control charts based on $T^2_{mve,i}$ and is due to the fact that the integer value of h is the same for successive values of m.

It should be noted that the control limits are only appropriate for the particular algorithm used. That is, the limits for $T^2_{mcd,i}$ are appropriate when the FAST-MCD algorithm is used and the limits for $T^2_{mve,i}$ are appropriate when the MVE estimator with subsampling is used. Here the number of subsamples for the MVE estimator is the default number based on the $SAS^{®}$ MVE algorithm. A difference in the algorithm changes the variability of the results from that algorithm and thus the generated control limits would vary. Because the resulting estimates can vary depending on which robust estimation algorithm is used, it is helpful to think of the "algorithm as the estimator" as discussed by Woodruff and Rocke (1994, p. 889). These control limits are based on the integer value of $h = (m + p + 1)/2$, which gives the maximum possible breakdown point. Using a different value of h will change the appropriate control limit.

3.8 Simulation Study

We made some performance comparisons of the high breakdown estimators. Wisnowski, Simpson and Montgomery (2002) did a performance study via simulation to compare various types of robust estimation procedures. They compared a sequential point addition algorithm of Hadi (1992, 1994), M-estimation, the approximate MVE calculated by the subsampling method, the FAST-MCD and the hybrid algorithm of Rocke and Woodruff (1996). However, their comparisons of the MVE obtained via subsampling and the FAST-MCD involve combinations of p, m, and the proportion of outliers (ϵ) obtained via factorial designs and thus

only used 2 design points for each "factor". The design points were $m = 40, 60$, $p = 2, 6$, and $\epsilon = 10\%, 20\%$ and were only evaluated for large shifts. Use of only two design points will be insufficient to determine which method is best because there is not a method that will be superior for all combinations. Wisnowski, Simpson, and Montgomery (2002) concluded that the hybrid algorithm performed best and that the FAST-MCD was slightly better than the MVE based on simulation runs involving 1000 datasets. They also considered various other outlier situations not considered here such as outliers scattered in random directions, clusters of outliers in all p variables, clusters of outliers in one of p variables, clusters of outliers in some of the p variables, and multiple clusters in close proximity. Vargas (2003) did a simulation study to compare use of control charts based on $T^2_{1,i}$, $T^2_{2,i}$, $T^2_{mve,i}$ obtained via subsampling, and $T^2_{mcd,i}$ obtained via the FAST-MCD. He concluded that the chart based on $T^2_{mve,i}$ gave the best performance in terms of probability of a signal when outliers are present. However, his comparisons between the MVE estimator obtained by subsampling and the FAST-MCD only covered the case for $p = 2$ and $m = 30$.

We performed a similar study to those of Wisnowski, Simpson, and Montgomery (2002) and Vargas (2003) to compare the MVE subsampling and FAST-MCD algorithms. Our study involves more combinations of p, m, and k. Of the m observations, k of them are random data points generated from the out-of-control distribution, and the other $n - k$ observations were generated from the in-control distribution. For a particular combination of p, m, and k, a number of datasets were generated.

The in-control distribution is a multivariate normal where it can be assumed that $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$ without loss of generality. The out-of-control distribution is a multivariate normal with the same variance-covariance matrix but where the mean vector has been shifted by

some amount. This amount depends on a value of the non-centrality parameter, given by

$$(\boldsymbol{\mu_1} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu}), \tag{3.3}$$

where $\boldsymbol{\mu_1}$ is the vector to which the mean vector has shifted. Thus the conclusions here will hold no matter the direction of the shift as long as all the outliers have shifted in the same direction. The larger the value of the non-centrality parameter, the more extreme the outliers are. The proportion of datasets that had a least one T^2 statistic greater than the control limit was calculated and this proportion becomes the estimated probability of a signal. The control chart based on $T^2_{1,i}$ was included in our study as a reference because of its common usage.

Figures 3.2 - 3.5 show the probability of a signal for different values of the non-centrality parameter and for some of the values of m and k considered in our study. For $p = 2, 3,$ and 5 a total of $100,000$ datasets of size m were generated for each combination of m, k, and value of the non-centrality parameter. For $p = 10$, $50,000$ datasets were generated for each combination.

As expected, when the value of the non-centrality parameter is small, the probability of a signal is close to .05 which is what would be the case for an in-control process. As the value of the non-centrality parameter increases the probability of a signal will increase. If not, then this indicates that the estimator has broken down and is not capable of detecting the outliers. In general, for small values of m, use of the MVE performs best, unless the number of outliers is large. As m increases, use of MCD is more likely to be superior. The actual breakdown point of the MVE method is smaller than that of the MCD although in theory they should have similar breakdown points. It is clear that the method based on $T^2_{1,i}$ possesses little ability to detect multiple outliers. As p increases for a fixed value of m, the breakdown points of the MVE and MCD get smaller. This suggests that the larger p is, the

larger m will need to be in order to detect outliers. In general, there was always one estimator that was found to be superior across all the values of the non-centrality parameter as long as the proportion of outliers was not so large as to cause the estimators to break down. This greatly simplifies the conclusions that can be made about when the MVE method or MCD method is preferred.

Figures 3.6 and 3.7 summarize the results of the probability of signal graphs from the simulation study by showing which of the three estimators (Standard, MVE, MCD) is preferred for the various combinations of m, p, and ϵ. Based on Figures 3.6 and 3.7, some broad recommendations can made made.

The standard estimator should be used if at most one outlier is expected. When $m \leq 50$ the MVE will be the best estimator unless the percentage of outliers is greater than 25%. When $m > 50$, the MCD is preferred as long as the percentage of outliers is less than 40%. As p increases, then the percentage of outliers that can be detected by the MVE estimator will decrease until it is only 10% for $p = 10$. It is true for both the MVE and MCD that the number of outliers that can be detected decreases as p increases.

Thus for Phase I applications where the number of outliers is unknown, the control chart based on the MVE should only be used for smaller sample sizes for which it is also computationally feasible. The control chart based on the MCD should be used for larger sample sizes or when it is believed that there is a large number of outliers. As more variables are monitored (p), the larger the sample size will need to be to ensure that the estimator does not break down and lose its ability to detect the outliers.

Table 3.3: Control limits for $T^2_{mve,i}$ statistic obtained via subsampling to maintain an overall in-control probability of signal = 0.05.

	p								
m	2	3	4	5	6	7	8	9	10
20	30.15	38.48	63.28	62.83	97.05	89.44	149.47	140.90	284.29
21	25.46	43.74	48.11	72.39	68.58	101.70	100.27	154.19	156.86
22	28.70	35.21	54.51	54.34	75.19	75.16	107.58	107.21	167.24
23	24.22	39.88	43.49	60.38	59.66	80.36	82.58	116.23	118.72
24	26.50	33.40	47.82	49.27	64.11	65.27	87.01	89.92	126.12
25	23.61	36.44	39.98	53.04	53.40	69.42	71.86	95.80	100.59
26	25.72	31.76	43.08	45.71	57.23	58.87	74.65	79.58	105.12
27	23.18	34.34	37.64	47.94	49.41	62.19	65.45	83.00	88.05
28	24.88	30.48	40.20	42.68	52.25	54.66	67.87	72.59	91.04
29	22.91	32.66	36.01	44.97	46.83	56.67	60.81	74.57	79.35
30	24.31	29.52	37.97	40.21	48.48	51.65	62.15	67.41	81.74
31	22.50	31.45	34.42	42.00	44.42	53.23	57.06	68.67	73.84
32	23.70	28.74	36.37	38.84	46.18	49.00	58.47	63.16	75.60
33	22.26	30.32	33.39	40.13	42.75	50.21	53.97	64.14	69.49
34	23.48	28.02	34.70	37.41	44.13	46.78	55.51	60.12	70.52
35	22.03	29.35	32.46	38.60	41.24	48.07	52.14	60.81	65.62
36	23.05	27.58	33.54	36.27	42.18	45.46	52.86	57.84	67.13
37	21.79	28.71	31.67	37.43	40.11	46.58	50.49	58.23	63.12
38	22.71	26.98	32.69	35.43	41.01	44.50	51.26	55.38	63.86
39	21.58	28.04	30.95	36.24	39.22	45.29	48.84	56.33	60.83
40	22.48	26.50	32.09	34.84	40.04	43.02	49.61	53.67	61.23
41	21.52	27.51	30.68	35.43	38.44	43.81	47.65	54.35	58.71
42	22.48	26.19	31.28	34.00	39.26	42.23	48.24	52.15	59.35
43	21.27	26.98	30.10	34.88	37.65	42.88	46.58	52.76	57.20
44	22.02	25.90	30.80	33.76	38.33	41.59	47.10	50.84	57.59
45	21.25	26.61	29.68	34.16	36.94	42.10	45.85	51.60	55.84
46	21.92	25.57	30.34	32.99	37.72	40.85	46.27	50.00	56.18
47	21.20	26.32	29.29	33.67	36.54	41.39	44.85	50.70	54.70
48	21.78	25.38	29.87	32.64	37.02	40.22	45.35	49.32	54.89
49	21.08	25.90	28.91	33.21	36.05	40.58	44.23	49.47	53.49
50	21.68	25.24	29.46	32.25	36.49	39.58	44.56	48.20	53.75

Table 3.4: Control limits for $T^2_{mve,i}$ statistic obtained via subsampling to maintain an overall in-control probability of signal = 0.05.

	p					p				
m	2	3	4	5		m	2	3	4	5
51	20.97	25.86	28.64	32.73		76	20.70	23.76	27.09	29.80
52	21.53	24.96	29.12	31.99		77	20.31	24.00	26.70	30.01
53	20.90	25.40	28.39	32.33		78	20.66	23.68	26.91	29.70
54	21.33	24.86	28.90	31.64		79	20.32	23.99	26.66	30.00
55	20.77	25.26	28.22	32.05		80	20.69	23.62	26.89	29.68
56	21.27	24.58	28.54	31.44		81	20.37	23.85	26.65	29.95
57	20.72	25.06	27.95	31.73		82	20.69	23.62	26.82	29.51
58	21.24	24.52	28.35	31.14		83	20.34	23.76	26.52	29.79
59	20.62	25.00	27.79	31.59		84	20.57	23.49	26.73	29.52
60	21.16	24.34	28.16	30.97		85	20.29	23.75	26.54	29.67
61	20.64	24.80	27.67	31.26		86	20.56	23.46	26.70	29.45
62	21.08	24.29	28.01	30.78		87	20.25	23.72	26.41	29.62
63	20.60	24.66	27.40	31.06		88	20.58	23.45	26.67	29.40
64	21.09	24.21	27.84	30.60		89	20.30	23.65	26.38	29.51
65	20.59	24.61	27.29	30.88		90	20.57	23.41	26.63	29.21
66	20.94	24.06	27.60	30.42		91	20.33	23.63	26.35	29.50
67	20.45	24.46	27.24	30.69		92	20.50	23.48	26.49	29.26
68	20.91	23.96	27.46	30.29		93	20.24	23.64	26.31	29.40
69	20.47	24.38	27.10	30.55		94	20.48	23.28	26.51	29.18
70	20.84	23.95	27.43	30.12		95	20.23	23.51	26.23	29.31
71	20.44	24.30	27.03	30.45		96	20.46	23.27	26.41	29.12
72	20.79	23.85	27.28	30.11		97	20.23	23.56	26.22	29.25
73	20.42	24.16	26.90	30.28		98	20.48	23.32	26.38	29.12
74	20.71	23.76	27.14	29.98		99	20.23	23.51	26.21	29.23
75	20.40	24.01	26.82	30.15		100	20.40	23.25	26.35	29.08

Table 3.5: Control limits for $T^2_{mcd,i}$ statistic obtained via the FAST-MCD algorithm to maintain an in-control overall probability of signal = 0.05.

	p								
m	2	3	4	5	6	7	8	9	10
20	116.35	221.31	512.57	573.00	1263.75	1261.29	3216.16	2732.59	8607.95
21	101.69	267.12	356.97	732.92	780.43	1749.17	1635.04	4263.75	3475.45
22	96.48	190.82	426.29	494.98	1001.39	1021.63	2306.46	2099.58	5394.73
23	88.45	225.37	312.60	614.28	658.21	1338.25	1316.79	2925.82	2617.15
24	82.50	165.93	363.30	433.36	824.27	854.21	1717.70	1654.24	3699.36
25	76.88	191.09	276.33	524.46	573.24	1080.57	1081.13	2153.21	2006.16
26	71.17	142.78	310.01	386.96	689.36	729.03	1358.16	1343.84	2640.49
27	67.76	161.40	244.43	451.12	507.49	887.25	919.94	1676.68	1606.62
28	63.58	124.24	266.56	341.02	592.22	638.91	1103.96	1118.05	2010.77
29	60.79	140.21	218.91	397.85	453.10	752.13	784.03	1328.16	1330.95
30	57.87	109.75	229.62	306.61	521.58	565.05	920.97	938.37	1584.42
31	56.06	121.26	192.12	348.47	403.29	649.23	682.26	1100.83	1109.26
32	52.81	96.93	198.93	273.13	461.13	502.58	791.71	811.53	1278.29
33	51.65	106.74	169.80	306.25	368.58	567.87	603.86	927.22	943.58
34	48.73	87.04	174.05	243.99	402.76	446.01	678.92	709.17	1064.44
35	47.82	94.92	151.74	272.18	331.76	502.06	535.50	787.65	798.08
36	45.88	78.60	151.54	219.12	359.95	403.76	592.05	618.53	893.32
37	45.39	85.04	135.32	240.93	299.32	446.20	478.61	685.69	703.16
38	43.28	71.83	134.45	195.70	320.15	368.05	523.59	550.52	768.13
39	42.95	77.90	121.48	214.69	273.21	398.36	435.75	598.72	621.28
40	40.83	66.67	118.92	176.28	283.51	331.61	466.44	493.09	670.10
41	40.89	71.31	110.40	190.07	245.31	358.95	391.38	532.95	552.36
42	39.16	62.05	108.37	158.55	255.07	300.85	417.22	444.66	591.07
43	39.04	66.12	100.11	171.10	223.88	325.15	355.00	476.19	497.57
44	37.40	58.53	98.07	142.79	229.56	273.68	377.42	401.43	527.47
45	37.57	61.83	91.20	153.44	203.37	294.03	325.84	432.77	446.93
46	35.95	55.21	90.13	129.36	205.50	251.08	340.95	367.12	474.81
47	35.91	58.29	85.00	137.97	184.30	266.10	297.59	388.44	408.06
48	34.83	52.52	82.86	118.33	185.14	228.42	311.98	337.64	427.32
49	34.96	55.05	78.76	125.40	170.19	240.78	272.27	351.29	373.57
50	33.63	50.05	76.83	109.48	166.78	208.99	281.82	307.01	388.70
51	34.00	52.21	74.30	114.33	153.54	221.02	251.74	322.26	340.80
52	32.80	47.63	72.24	100.22	152.20	190.70	260.23	284.55	353.93
53	33.00	49.94	69.68	104.77	141.09	199.71	234.03	297.29	316.36
54	32.08	45.89	68.06	93.16	138.70	175.65	236.04	261.73	326.26
55	32.51	47.86	65.91	97.77	129.80	185.02	215.83	273.59	291.71
56	31.44	44.24	64.20	86.75	128.02	160.04	216.92	242.96	299.15
57	31.42	46.12	62.25	90.67	119.58	169.09	197.30	250.11	270.43
58	30.60	43.00	60.88	81.39	117.37	149.04	198.86	224.10	277.98
59	30.85	44.21	59.22	84.91	111.41	154.91	182.85	233.49	254.14
60	29.93	41.51	58.09	76.89	108.83	137.20	182.80	208.72	257.63

Table 3.6: Control limits for the $T^2_{mcd,i}$ statistic obtained via the FAST-MCD algorithm to maintain an overall in-control probability of signal = 0.05.

	p								
m	2	3	4	5	6	7	8	9	10
61	30.37	42.94	56.89	79.90	103.46	142.42	170.75	217.32	236.53
62	29.60	40.05	55.66	72.70	100.76	128.10	169.24	193.12	240.25
63	29.81	41.60	54.41	75.19	96.07	132.49	158.53	200.15	219.66
64	28.97	39.30	53.53	69.09	94.22	118.68	157.60	180.96	222.52
65	29.31	40.50	52.23	71.24	90.69	123.39	147.74	186.37	206.70
66	28.52	38.19	51.52	65.55	88.53	110.76	146.46	169.88	209.14
67	28.78	39.42	50.62	67.75	85.59	114.25	137.55	174.27	193.82
68	28.20	37.55	49.63	62.58	83.68	103.68	136.62	158.56	193.39
69	28.43	38.56	49.04	64.45	81.00	108.00	128.53	162.14	181.61
70	27.83	36.53	48.24	60.42	79.45	97.71	126.77	148.48	181.76
71	28.03	37.58	47.41	61.90	77.35	100.50	120.42	151.79	170.44
72	27.50	35.87	46.58	58.14	75.28	92.39	118.68	139.10	170.02
73	27.69	36.77	46.03	59.70	73.34	94.81	113.96	142.75	161.94
74	27.05	35.22	45.36	56.10	72.34	87.58	112.12	131.52	160.34
75	27.38	35.89	45.00	57.47	70.37	89.91	107.88	133.60	152.33
76	26.85	34.45	44.20	53.92	68.94	83.38	105.06	123.58	151.70
77	27.10	35.30	43.74	55.49	67.63	84.98	101.60	125.70	144.50
78	26.57	34.19	43.17	52.45	66.11	79.10	99.23	116.99	142.80
79	26.76	34.80	42.70	53.80	64.74	80.83	95.81	118.40	136.02
80	26.24	33.46	42.03	50.93	63.46	75.95	94.29	110.74	134.38
81	26.67	34.10	41.72	52.00	62.33	77.57	92.27	112.47	128.79
82	26.02	32.99	41.20	49.55	61.36	72.67	90.02	105.60	127.13
83	26.25	33.71	41.16	50.67	60.19	74.76	87.58	107.04	123.17
84	25.84	32.56	40.45	48.36	59.59	70.01	85.75	99.78	120.54
85	26.05	33.11	40.34	49.19	58.18	71.34	83.71	102.45	116.80
86	25.61	32.07	39.66	47.17	57.70	67.48	82.18	94.92	114.90
87	25.90	32.71	39.58	47.94	56.54	68.76	80.03	97.42	111.43
88	25.50	31.90	38.79	46.19	55.73	65.49	78.83	91.25	109.30
89	25.70	32.36	38.75	46.82	54.98	66.34	77.46	92.95	106.56
90	25.24	31.45	38.26	45.19	54.11	63.28	75.59	87.85	104.28
91	25.53	31.93	38.17	46.07	53.68	64.00	74.57	89.18	101.58
92	25.12	31.07	37.77	44.15	52.94	61.22	73.18	84.42	99.39
93	25.37	31.59	37.50	45.08	52.28	62.35	71.91	85.44	97.31
94	24.93	30.73	37.17	43.48	51.37	59.57	70.76	81.17	95.86
95	25.12	31.17	37.14	44.17	51.16	60.49	69.53	81.71	93.73
96	24.76	30.47	36.56	42.67	50.50	58.00	68.21	78.23	92.06
97	25.03	30.83	36.51	43.29	50.09	58.80	67.09	79.56	89.75
98	24.60	30.15	36.18	42.05	49.29	56.77	66.18	75.89	88.54
99	24.81	30.62	35.99	42.52	48.79	57.35	65.43	76.20	86.71
100	24.47	29.90	35.68	41.30	48.28	55.25	64.40	73.49	85.16

Figure 3.2: Probability of signal for various combinations of m and k for $p = 2$. The circles and solid line correspond to the MVE, the triangles and dashed line correspond to the MCD, and the squares and dotted line correspond to the standard estimator.

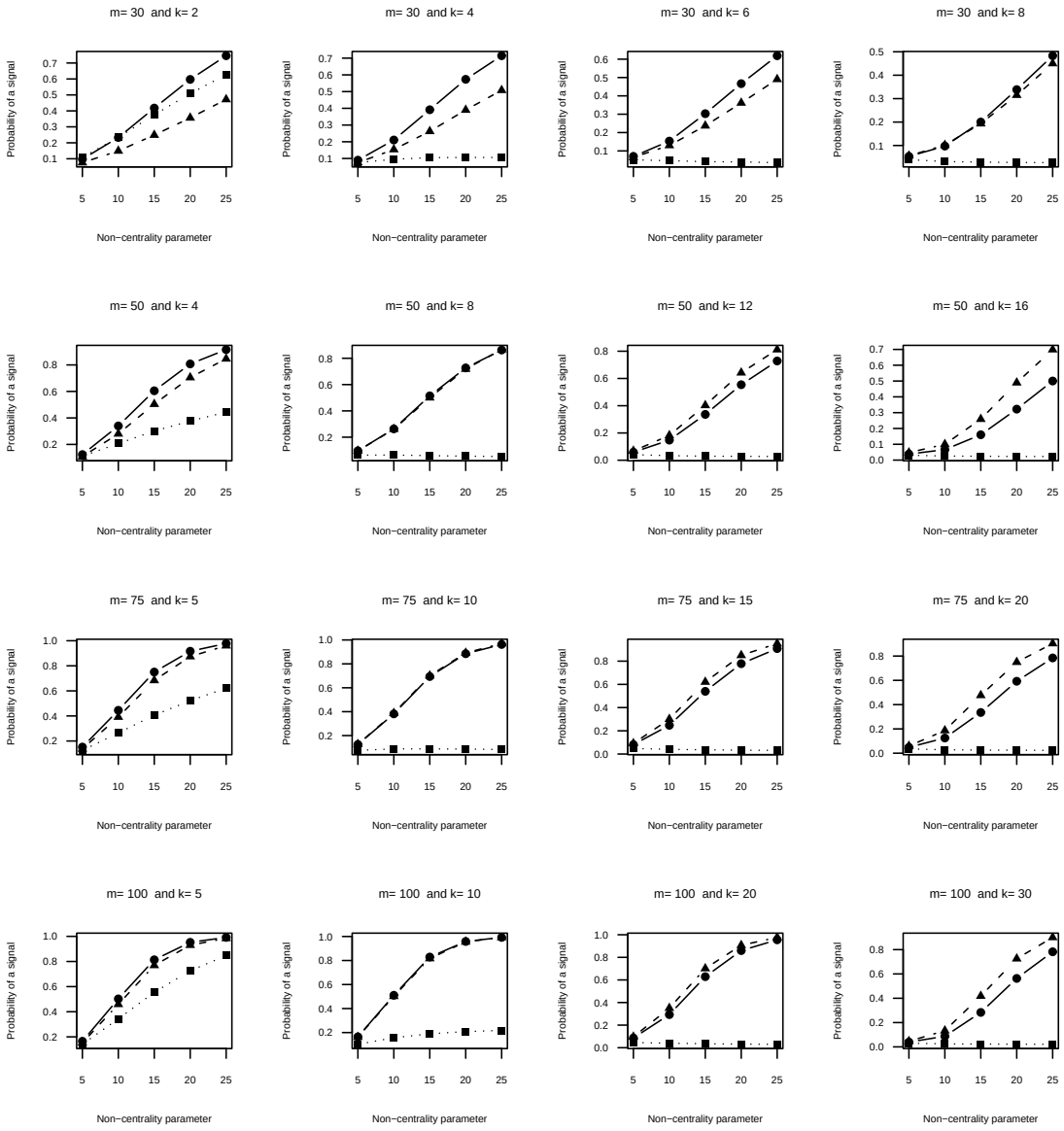


Figure 3.3: Probability of signal for various combinations of m and k for $p = 3$. The circles and solid line correspond to the MVE, the triangles and dashed line correspond to the MCD, and the squares and dotted line correspond to the standard estimator.

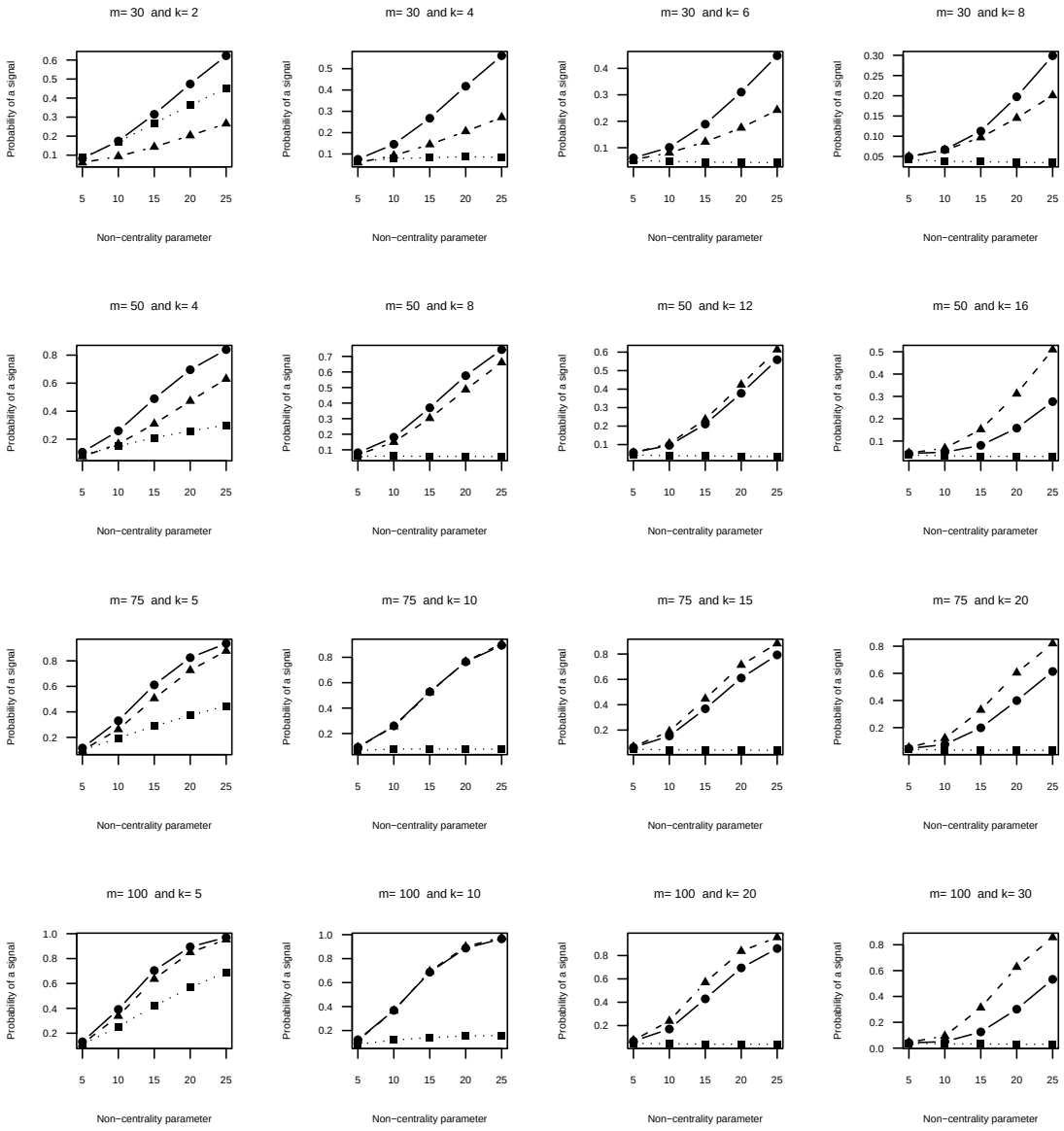


Figure 3.4: Probability of signal for various combinations of m and k for $p = 5$. The circles and solid line correspond to the MVE, the triangles and dashed line correspond to the MCD, and the squares and dotted line correspond to the standard estimator.

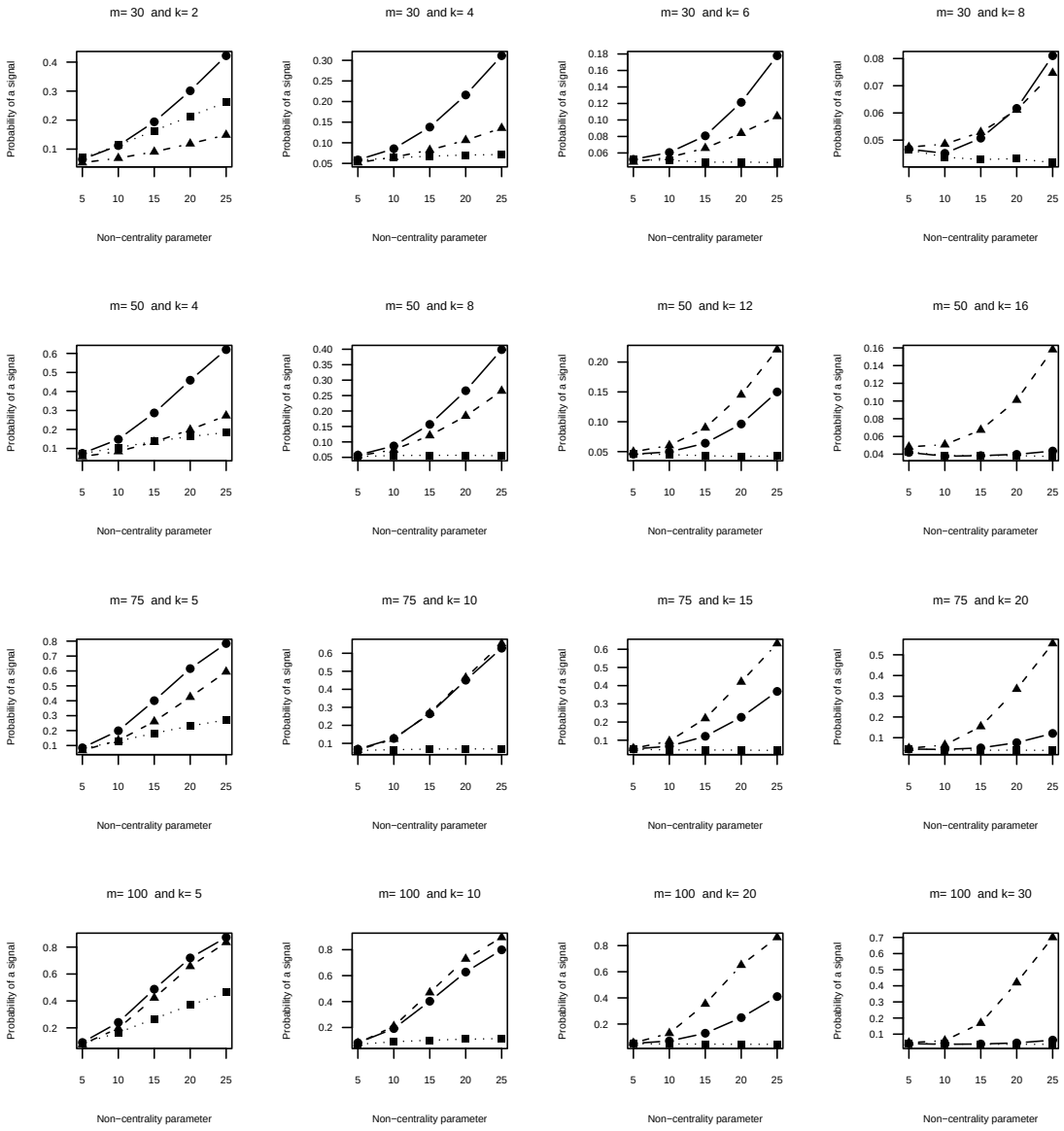


Figure 3.5: Probability of signal for various combinations of m and k for $p = 10$. The circles and solid line correspond to the MVE, the triangles and dashed line correspond to the MCD, and the squares and dotted line correspond to the standard estimator.

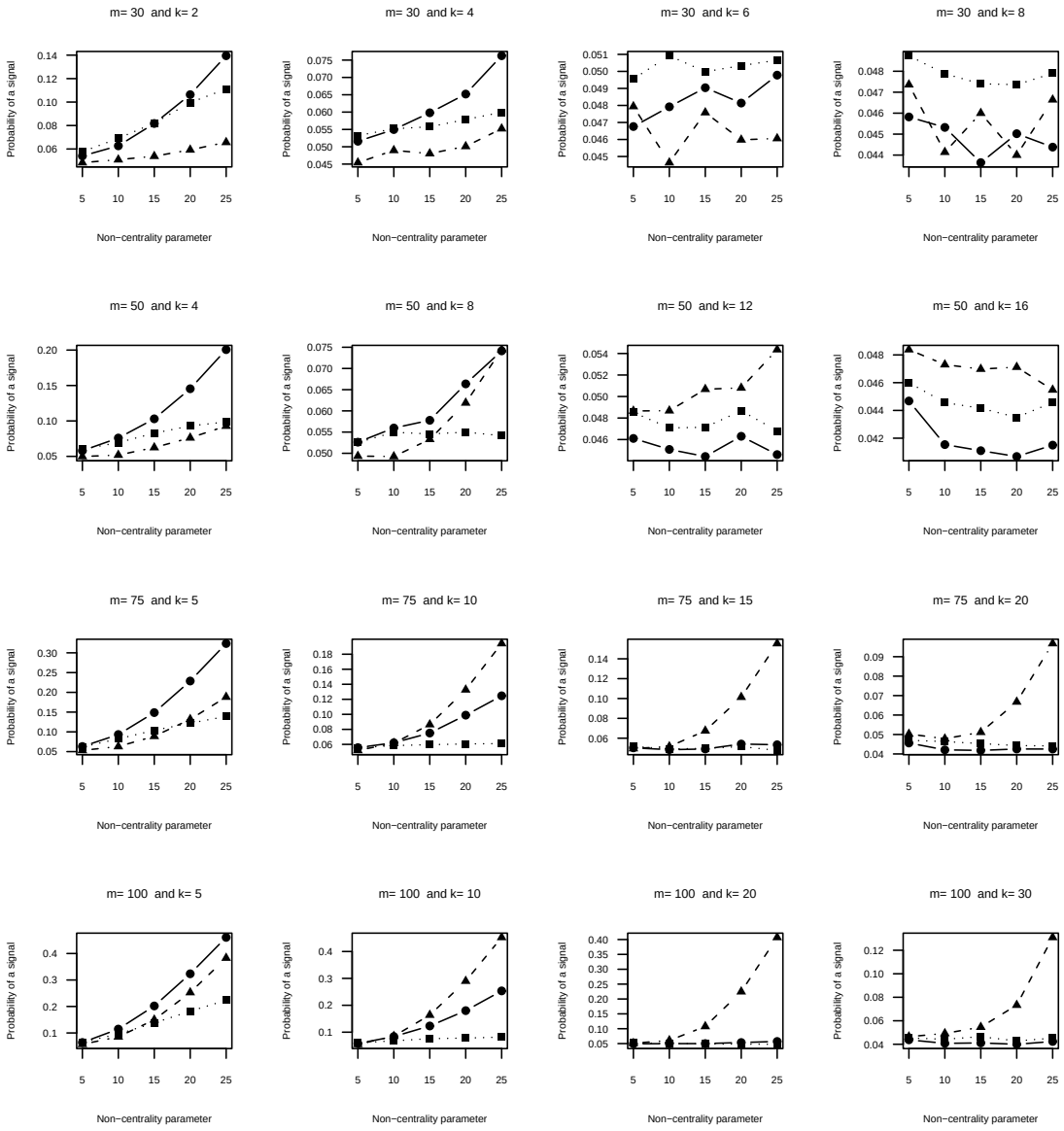


Figure 3.6: Summary of preferred estimator for $p = 2$ and $p = 3$. The unlabelled area is where the MVE and MCD methods perform equally well.

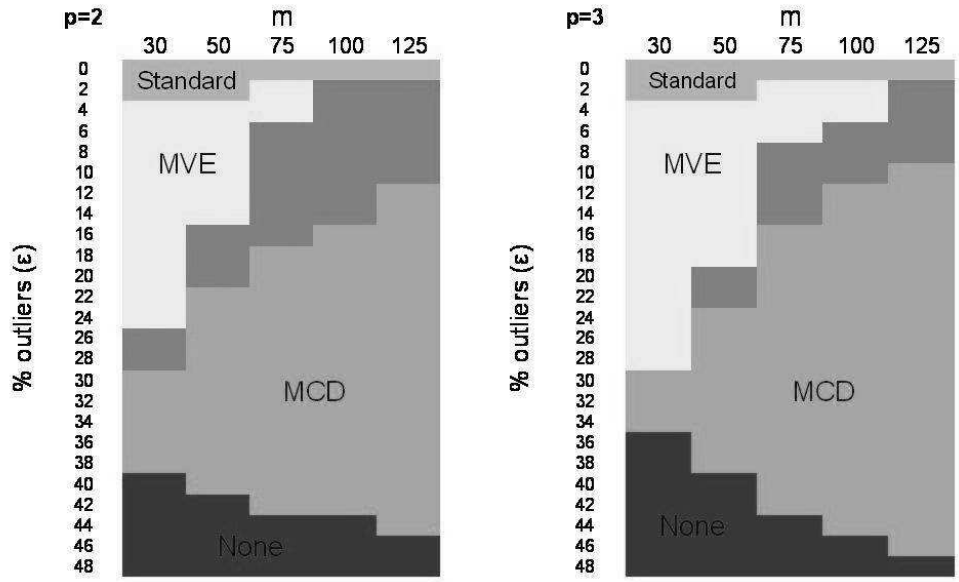


Figure 3.7: Summary of preferred estimator for $p = 5$ and $p = 10$. The unlabelled area is where the MVE and MCD methods perform equally well.