

Robustness of the One-Class Peeling method to the Gaussian Kernel Bandwidth

Lina Lee¹, Maria L. Weese², Waldyn G. Martinez², and L. Allison Jones-Farmer²

¹Department of Statistics, Virginia Tech, Blacksburg, VA

²Department of Information Systems & Analytics, Miami University, Oxford, OH

Abstract

The One Class Peeling (OCP) method is an outlier detection method for Phase I analysis of large multivariate samples. The two stage OCP method is based on support vector data description and employs the Gaussian kernel function to create sequential flexible boundaries to peel away observations to estimate the center of a data set. Subsequently the distance from that estimated center is used as the monitoring statistic. In this work we study the effect of the bandwidth parameter for the Gaussian Kernel on the performance of the OCP method. We find that the OCP method is robust to the value of the bandwidth parameter in terms of in-control and out-of-control simulation performance. We verify these results on example data. The result is a recommendation for a small value of the bandwidth value, providing a computational advantage to the OCP method.

Keywords: Convex Hull Peeling, Outlier Detection, Phase I, Support Vector Data Description

1 Introduction

Statistical Process Control is typically divided into two important stages, Phase I, where a baseline in-control state is established and Phase II where a process is monitored for changes from the in-control state. Phase I consists of retrospectively analyzing process data to determine if abnormal, or out of control, behavior occurred. Examples of abnormal process behavior include process mean shifts, change points and extreme observations or outliers. This work focuses on the problem

of detecting outliers in a retrospective sample. While there are many methods for robust outlier detection¹⁻⁶, outlier detection in high-dimensional, multivariate data is more challenging. Martinez et al.⁷ introduced the One Class Peeling (OCP) method to detect outliers in high-dimensional data for the purpose of establishing a baseline sample in the Phase I stage of process control. The OCP method is particularly appealing since it does not require covariance estimation and works well when larger percentage of outliers are present. There are very few Phase I methods available that will work with large, multivariate, data sets when the number of observations (n) is less than the dimension of the data (p) and are computationally feasible. One of those methods is the Robust Minimum Diagonal Product (RMDP)⁸. The RMDP method uses only the diagonal elements of the sample covariance matrix and therefore can be applied efficiently in settings where p is large. Martinez et al.⁷ compared the OCP method with an improved version of the Ro et al.⁸ method and showed OCP outperformed the RMDP. For a review of Phase I methods, see Jones-Farmer et al.⁹. In this work we investigate properties of the OCP method.

The OCP method is a two stage method using an iterative peeling approach, similar to convex hull peeling¹⁰. In the first stage, iterative boundaries are constructed via Support Vector Data Description (SVDD)¹¹. These boundaries are used to “peel” observations from the data until a specified number of observations remain. The center of the data is then robustly estimated using the remaining observations. A distance measure between each observation and the center is used to evaluate outlyingness. Potential outliers are flagged using an empirical threshold established to control the false positive rate to a user specified level. The boundaries, or peels, constructed via SVDD are found from

$$\begin{aligned} & \underset{R, \xi_i}{\text{minimize}} && F(R, \xi_i) = R^2 + C \sum_i \xi_i \\ & \text{subject to} && \|\mathbf{x}_i - \boldsymbol{\mu}\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0 \quad \forall i, \end{aligned} \tag{1}$$

where C influences the size of the boundary based on the fraction of observations that are outside the boundary. Martinez et al.⁷ suggest setting the parameter C to create boundaries that include all possible points.

Using Lagrange multipliers, Tax and Duin¹² showed that incorporating the constraints from

Equation (1) results in the need to maximize

$$L = \sum_i \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (2)$$

with respect to α_i such that $0 \leq \alpha_i \leq C$. Maximizing Equation (2) results in a set of α_i such that when an observation, \mathbf{x}_i satisfies $\|\mathbf{x}_i - \boldsymbol{\mu}\|^2 < R^2$ in Equation (1), the corresponding $\alpha_i = 0$, and these observations are inside of the boundary. When, $\|\mathbf{x}_i - \boldsymbol{\mu}\|^2 = R^2$ and $\alpha_i > 0$ these observations are the support vectors providing the description of the data and, thus, the one-class boundary. Boundaries constructed with Equation (2) are hyper-spheres¹². In order to construct more flexible boundaries the inner products in Equation (2) are replaced with a kernel function. The OCP method can be used with many kernel functions, but we consider the implementation using the Gaussian kernel function⁷ defined as

$$\text{KSG}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{s^2}\right), \quad (3)$$

where the parameter s must be pre-specified by the user. Weese et al.¹³ studied the effect of s when using the Gaussian kernel function with the k -chart¹⁴ and concluded s should be set equal to the dimension of a data set, p , when the data are scaled to unit variance.

In this work, we investigate the choice of s and the performance of the OCP method in terms of Finite Sample Replacement Breakdown Point (FSRBP) and performance detection rate in situations where outliers are present (out-of-control) and when they are not (in-control). We find that the performance of the OCP method is invariant to the choice of s , making the OCP method more robust than was originally presented in Martinez et al.⁷. The implication of this finding is the shape of the boundary using a Gaussian kernel function is not important to the performance of the OCP method. With this knowledge, the value of s can be set to minimize the computational requirements of the method, thus improving the speed of application of the OCP in practice.

The rest of this paper is organized as follows. Section 2 describes the OCP method in further detail. Section 3 describes the behavior of the estimation of the center when s is varied. Section 4 illustrates the performance of the method in in-control and out-of-control situations. Section 5

illustrates the robustness of the method on an example. Finally, in Section 6, we offer concluding remarks.

2 The OCP Method

The OCP method is a two-stage method that consists of (1) robustly estimating the center of the data; and (2) comparing the distance of observations from the center to a threshold to determine outlyingness.

To estimate the center, boundaries are constructed using Equation (2) where the dot product is replaced by Equation (3). The parameter C , which helps to control the boundaries is set so that the boundaries are inclusive of all points. Once the boundary is constructed, those points identified as support vectors ($\alpha_i > 0$) will be peeled, or removed from the data and another boundary is constructed with the remaining data. This process of generating boundaries and peeling the exterior observations is repeated until a specified number of observations are left. The default is to peel until $n = 2$. The mean estimator $\hat{\boldsymbol{\mu}}_{\text{OCP}}$ is determined by averaging the last remaining observations.

In the second step of the method, the Gaussian kernel is again used to calculate a distance from the estimated center using

$$\text{KD}(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{OCP}}) = 1 - \text{KS}_G(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{OCP}}).$$

The distance value $\text{KD}(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{OCP}})$ ranges from 0 to 1 and is 0 if and only if $\mathbf{x}_i = \hat{\boldsymbol{\mu}}_{\text{OCP}}$. To determine outlyingness the kernel distances are first re-scaled by

$$\text{sRKD}_i = \frac{\text{KD}(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{OCP}}) - \text{med}(\text{KD})}{\text{MAD}(\text{KD})}.$$

Here, $\text{med}(\text{KD})$ is the median of the vector of N kernel distances, where N is the sample size, and $\text{MAD}(\text{KD})$ is the median absolute deviation of the N kernel distances. Outliers are identified as those with sRKD_i greater than a threshold, h . Martinez et al.⁷ provided an algorithm and code to determine the threshold value, h , for a desired false positive rate based on the underlying

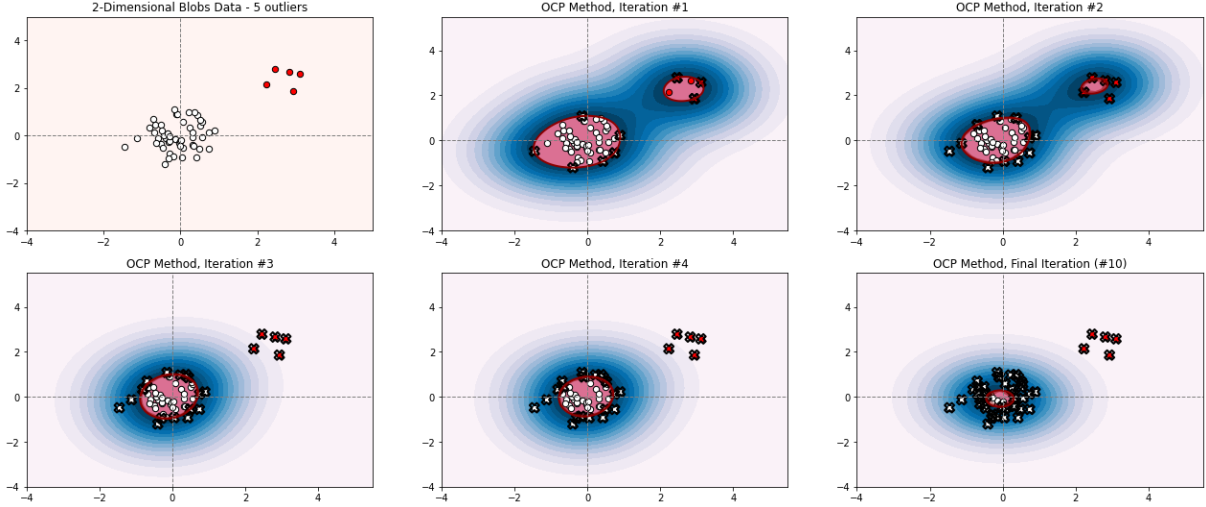


Figure 1: OCP Example on a 2-dimensional blobs data set. 50 inlier observations generated with $\mu_0 = [0, 0]$, and 5 outlier observations generated with $\mu_1 = [2.5, 2.5]$. Contours indicate kernel distance from support vectors.

distribution of the data, an estimate of the amount of correlation present, and the size of the data. If the user is unable to provide these characteristics about the data, Martinez et al.⁷ provided a robust value of h . Figure 1 illustrates how the OCP method works using a synthetic 2-dimensional data set centered at $\mu_0 = [0, 0]$, with 5 outlying points centered at $\mu_1 = [2.5, 2.5]$. Figure 1 shows how the OCP method initially creates boundaries around both the inlier and outlier groups peeling away the most unusual observations. By the third iteration OCP has peeled all outlying observations, and by the final iteration $\hat{\mu}_{\text{OCP}}$ is close to the mean of the inlier data, μ_0 .

The bandwidth parameter s in Equation (3) is an important parameter in the OCP algorithm. The value of s changes the shape of the flexible boundary generated from SVDD and has been shown to dramatically affect performance of the k-chart¹³, which is a control chart based on a single SVDD boundary. Figure 2 shows three one-class boundaries generated on a subset of the Wine Quality data set¹⁵ using different values of s . The example illustrates how smaller values of s produce boundaries that fit more tightly to the perimeter of the data.

Weese et al.¹³ suggested a value of $s = p$ for use with the k-chart when the data are scaled to unit variance. Martinez et al.⁷ used only $s = p$ to evaluate the performance of the OCP method. In this work we investigate how changing the value of s affects the performance of the OCP method

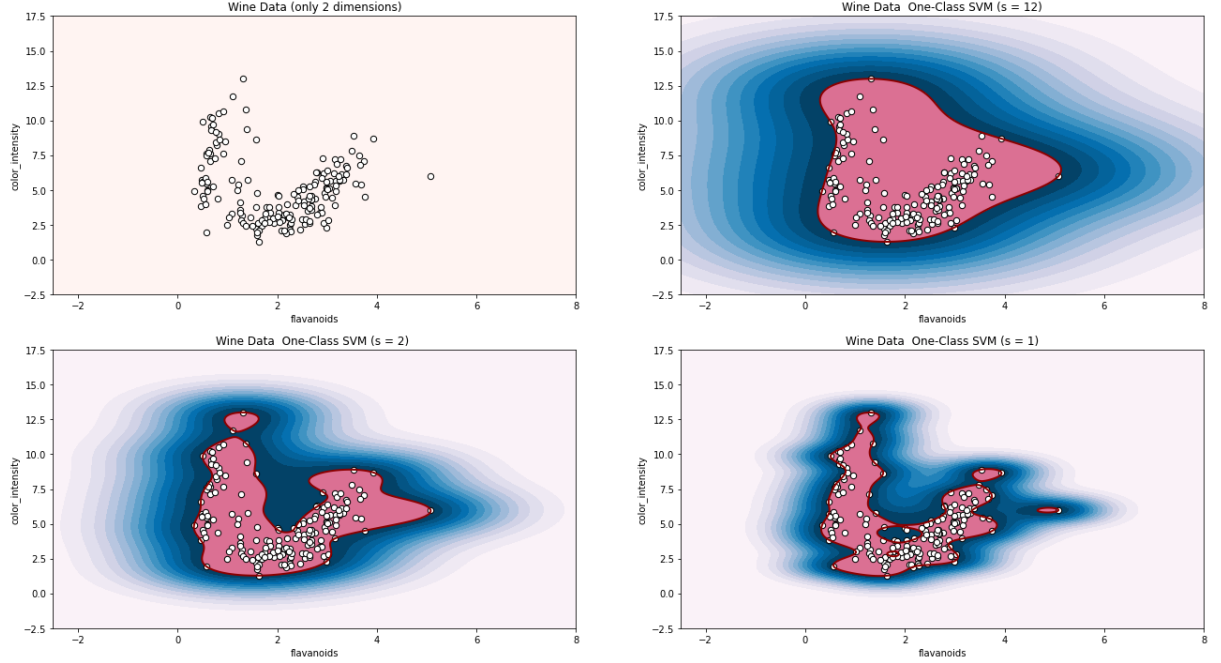


Figure 2: Example with different values of s on two variables from the Wine Quality¹⁵ data set.

in both (1) determining the center of data and (2) correctly identifying outlying observations. We investigate the effect of changing s on the robust mean estimator, $\hat{\mu}_{\text{OCP}}$, by evaluating the empirical Finite Sample Replacement Breakdown. In addition, we evaluate the performance of the OCP method in correctly identifying outlying observations in both out-of-control scenarios (outliers present) and in-control scenarios (no outliers present).

3 Empirical Finite Sample Replacement Breakdown

The robustness of a mean estimate to the presence of outliers can be measured in terms of the Finite Sample Replacement Breakdown Point (FSRBP). Donoho and Huber¹⁶ defined the FSRBP of a location estimator, \mathbf{t}_N , based on a data set, S , as the smallest fraction, m/N , of outliers that can take the estimate “over all bounds” :

$$\epsilon^*(\mathbf{t}_N, S) = \min_{1 \leq m \leq N} \left\{ \frac{m}{N} : \sup_{S_O} \|\mathbf{t}_N(S) - \mathbf{t}_N(S_O)\| = \infty \right\},$$

where the supremum is taken over all possible corrupted samples, $S_O = \{\mathbf{y}_1, \dots, \mathbf{y}_m\}$, for $\mathbf{y}_i \in \mathbb{R}^p$, obtained by replacing m points from S with arbitrary values. It is not possible to analytically evaluate the FSRBP of the OCP estimator⁷; thus empirical methods must be used to evaluate the breakdown properties of $\hat{\boldsymbol{\mu}}_{\text{OCP}}$. Martinez et al.⁷ evaluated the empirical FSRBP of $\hat{\boldsymbol{\mu}}_{\text{OCP}}$ when $s = p$ and showed that it begins to break down at around 30-35% contamination. Notably, Martinez et al.⁷ also computed the empirical FSRBP of $\hat{\boldsymbol{\mu}}_{\text{OCP}}$ for several values of remaining observations, n after the peeling process. Their results suggest that $\hat{\boldsymbol{\mu}}_{\text{OCP}}$ is most robust when $n = 2$.

To empirically evaluate FSRBP of the OCP method when s changes we vary probability distribution; the original sample size of the data, N ; dimension, p ; correlation structure, $\boldsymbol{\Sigma}$; and bandwidth, s . We generate correlated and uncorrelated samples of from multivariate normal, log-normal and t(df=10)-distributions. The uncorrelated samples have a covariance of $\boldsymbol{\Sigma} = \mathbf{I}$ and correlated samples have a correlation matrix generated randomly such that it is positive semi-definite with correlations uniformly ranging from $[-1, 1]$. The bandwidth parameter is adjusted according to the dimension, p , ranging from values $s = 0.1p$ to $s = 3p$. Table 1 provides a summary of the simulation conditions for the FSRBP simulations.

Distribution	Correlation	Bandwidth
Normal	uncorrelated	$0.1p$
Lognormal	correlated	$0.5p$
t(df=10)		$0.75p$
		p
		$1.5p$
		$3p$

Table 1: The summary of simulation conditions for FSRBP simulations.

We generate p dimensional samples of size N containing both in-control and out-of-control points. Out-of-control samples, S_O , contain a p dimensional sample of size m outliers. The in-control samples, S_I , contain $N - m$ inliers. The in-control samples are generated from each distribution with mean of $\boldsymbol{\mu}_0 = \mathbf{0}$ and covariance of $\boldsymbol{\Sigma}$. For a given probability distribution, $F_X(\cdot)$, 500 replications are summarized to obtain the breakdown percentages. We generate the p -dimensional outlier sample, S_O , of size m from $F_Y(\cdot)$, with a shifted mean $\boldsymbol{\mu} + \boldsymbol{\delta}$ where $\delta_i = 20\sigma_{ii}$, for $i = 1, \dots, p$ for the multivariate normal and t-distributions, and $\delta_i = e^{20\sigma_{ii}}$ for the multivariate lognormal dis-

tribution. The value of δ_i is chosen such that the means of the inlier and outlier distributions are as far apart as numerically measurable. The estimator, $\hat{\boldsymbol{\mu}}_{\text{OCP}}$ is considered to breakdown when $.05 > 1 - \int_{-\infty}^{\hat{\mu}_1} \dots \int_{-\infty}^{\hat{\mu}_p} f_X(x_1, \dots, x_p) dx_1 \dots dx_p$, where $\hat{\boldsymbol{\mu}}_{\text{OCP}}^\top = [\hat{\mu}_1, \dots, \hat{\mu}_p]$ is the one-class peeling estimate of the mean of $F_X(\cdot)$. We measure the percentage breakdown as a function of changing values of s .

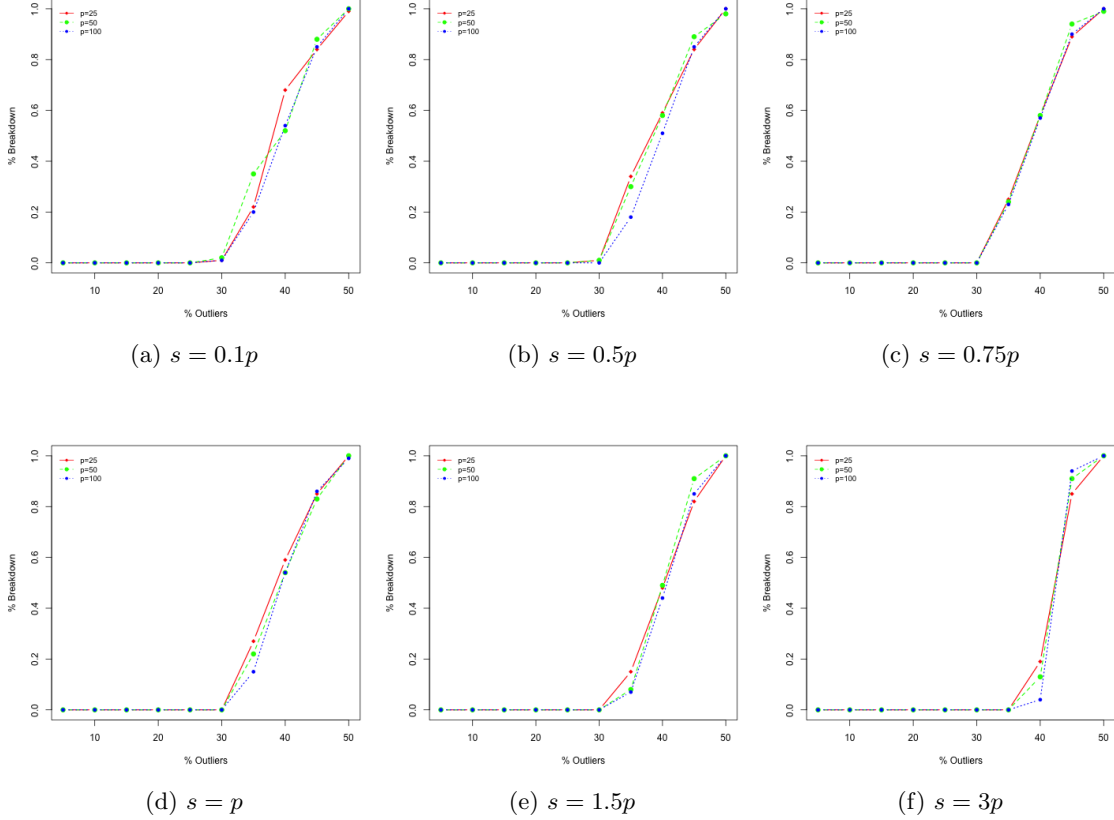


Figure 3: Breakdown points for the uncorrelated normal distribution case

Figures 3 and 4 show the FSRBP for both uncorrelated and correlated normal distributions with samples with $N = 50$ with s increasing from $s = 0.1p$ to $s = 3p$. Upon first glance we notice the FSRBP performance is very stable around 30% as s increases until $s = 3p$ where we notice the breakdown point increases to about 35% but in general the the breakdown performance does not appreciably change with p . When samples were generated from the uncorrelated and correlated lognormal and $t(\text{df}=10)$ - distributions, the results are similar (see Appendix A). From

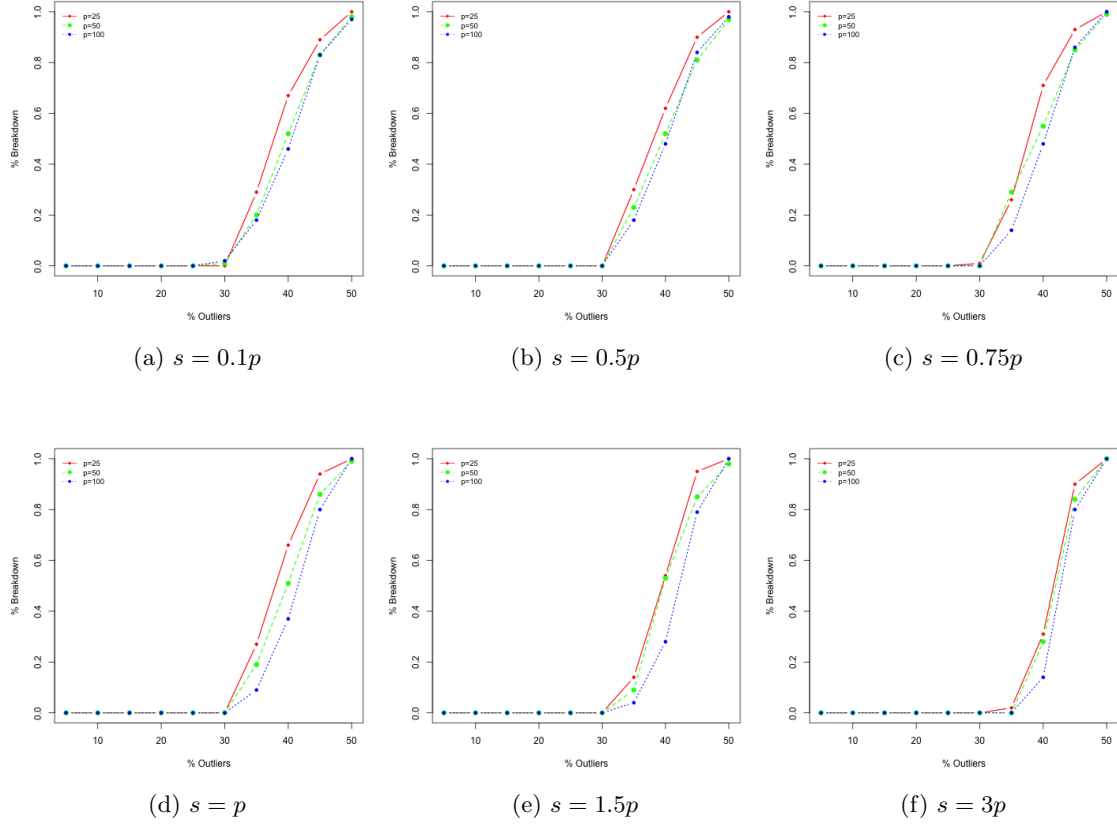


Figure 4: Breakdown points for the correlated normal distribution case

these simulations we conclude that the empirical FSRBP of the $\hat{\mu}_{\text{OCP}}$ does not change with respect to the bandwidth parameter, s ; thus, $\hat{\mu}_{\text{OCP}}$ is robust to s .

4 Outlier Detection Performance Simulations

In this section we examine the effect of s on the ability of the OCP method to correctly identify in- and out-of-control observations. We generate data from normal, lognormal and $t(\text{df}=10)$ -distributions for various values of N and p . We generate Σ with specific values of the off diagonals, either $\rho = 0$ or $\rho = 0.75$ to clearly illustrate the OCP performance under two different correlation structures. Table 2 summarizes the simulation settings for evaluating the in-control and out-of-control performance of the OCP for different values of s .

Sample Size	Distribution	Correlation	Bandwidth
N=50,p=50	Normal	$\rho = 0$	$0.1p$
N=50,p=100	Lognormal	$\rho = 0.75$	$0.5p$
N=100,p=100	t(df=10)		$0.75p$
			p
			$1.5p$
			$3p$

Table 2: The summary of performance simulation conditions

We perform 1000 replications for each of the 18 combinations of sample size, dimension, distribution, and correlation for each of the 6 different bandwidth values. For the in-control cases, we measure the false positive rate $\hat{\alpha}_i = \frac{FP_i}{N} * 100$, where FP_i is the number of false positives, and the statistical distance between the in-control mean, $\boldsymbol{\mu}_0$ and the estimated mean, $\hat{\boldsymbol{\mu}}_{\text{OCP}}$, see Equation (4). To classify observations as outlying, we use the exact empirical threshold values presented in Martinez et al.⁷ in Appendix Table D, which set the empirical false positive rate at 5%.

$$\text{Distance} = \sqrt{(\boldsymbol{\mu}_0 - \hat{\boldsymbol{\mu}}_{\text{OCP}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_0 - \hat{\boldsymbol{\mu}}_{\text{OCP}})} \quad (4)$$

To generate out-of-control samples we shift 10% of observations to a new mean, $\boldsymbol{\mu}_1$, where $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{\delta}$. The value of $\boldsymbol{\delta}$ is assigned such that $(\boldsymbol{\mu}_0 + \boldsymbol{\delta})$ is unlikely to be generated from the same distribution as the in-control data. To ensure equivalent shifts in $\boldsymbol{\mu}_1$ across the different distributions, we set the value of $\boldsymbol{\delta}$ such that $0.023 = 1 - \int_{-\infty}^{\mu_1 + \delta_1} \dots \int_{-\infty}^{\mu_p + \delta_p} f_x(x_1, \dots, x_p) dx_1 \dots dx_p$ where $f_x(\cdot)$ is the in-control density function and $\boldsymbol{\mu}_0 = [\mu_1, \dots, \mu_p]^\top$. Martinez et al.⁷ showed that the performance of the OCP method is stable with regard to the percentage of outliers; thus, we generate all out-of-control samples with 10% outliers. Out-of-control performance is measured using the classification rate, $\text{CR}_i (\%) = \frac{\text{TP}_i + \text{TN}_i}{N} \times 100$ and detection rate, $\text{DR}_i (\%) = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \times 100$. Here TP_i is the number of true positives and TN_i is the number of true negatives. In addition, we compute the distance from estimated mean, $\hat{\boldsymbol{\mu}}_{\text{OCP}}$ to the true mean $\boldsymbol{\mu}_0$ using the statistical distance as defined by (4).

4.1 In-Control Simulation Results

Figure 5 shows the empirical false positive (FP) rate over 1000 simulations for $N = 50$ and $p = 50$ samples of correlated and uncorrelated observations from each of the three distributions. We immediately notice that the FP is very stable across the various bandwidth values both in terms of the mean and the distribution of values over the simulations. The results indicate that the value of s does not affect the in-control performance of the OCP method. This pattern is repeated for the other two sample sizes tested, see Appendix B.1.

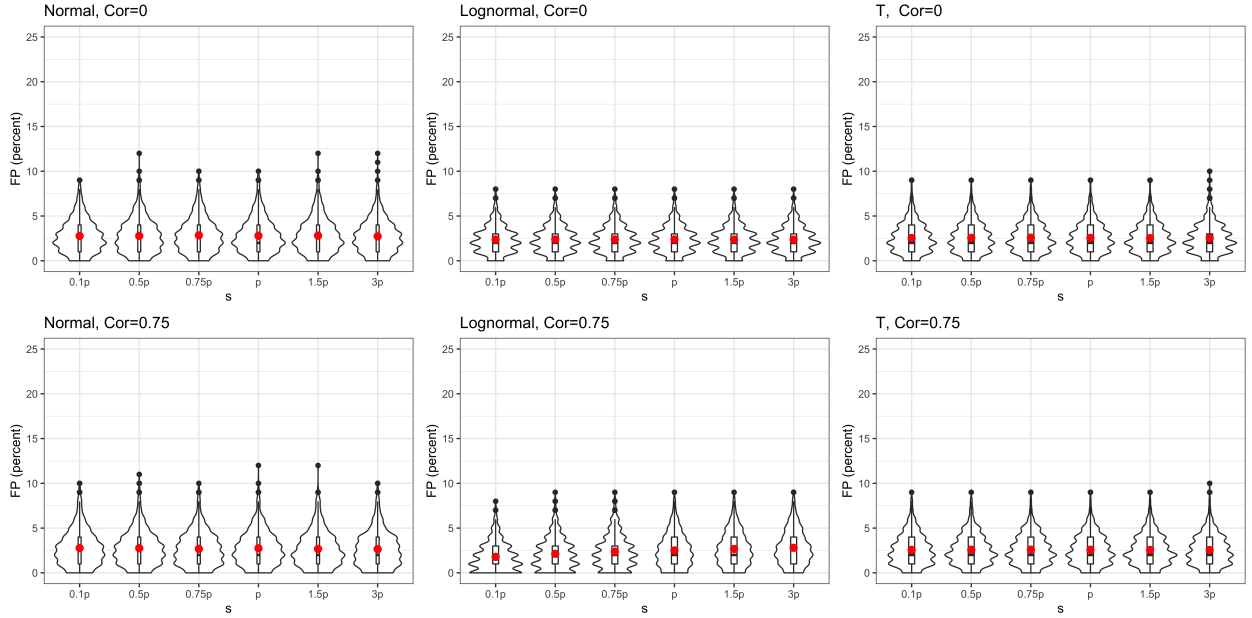


Figure 5: False Positive Rate for in-control samples with $N=50$, $p=50$. The red dot indicates the mean.

Figure 6 shows the distribution of the statistical distances between the true mean, μ_0 and the estimated mean, $\hat{\mu}_{\text{OCP}}$ for different values of s . This plot gives an indication of the bias of the estimator, where an average distance of zero indicates an unbiased estimator. Here, the results suggest a consistent and small bias for normally distributed cases, regardless of s or the correlation structure of the data. Similar results are shown for the t-distribution cases. The lognormal results display a consistent, but larger bias than the symmetric distributions when no correlation is present, indicating a different measure of central tendency should be used with skewed

distributions. Although the presence of high correlation in the lognormal distribution seems to mitigate the bias in the estimator, we observe a similar pattern for larger sample sizes, see Appendix B.2. The presence of some bias in $\hat{\mu}_{\text{OCP}}$ does not affect the empirical false positive rate which is stable at about 5% on average, or the ability of the OCP to correctly identify outliers (see Section 4.2).

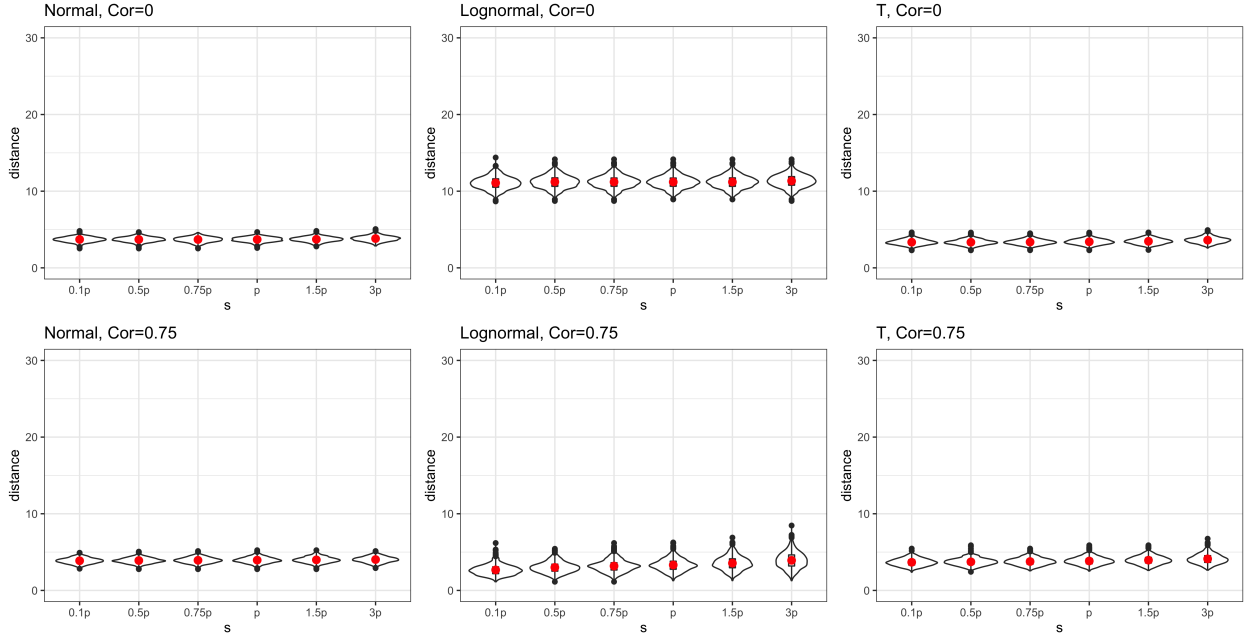


Figure 6: In-control statistical distance between $\hat{\mu}_{\text{OCP}}$ and μ_0 for samples with $N=50$, $p=50$. The red dot indicates the mean.

4.2 Out-of-Control Simulation Results

Figure 7 shows the empirical classification rates for samples of size $N = 50$ and $p = 50$ for the distribution and correlation combinations as the bandwidth, s changes. The distribution of the classification rates are noticeably similar across all values of s , with minor changes in the lower tails of the distribution. The results suggest that, on average, the classification rate is stable with respect to the bandwidth parameter, s . Results from additional sample sizes are available in Appendix C.2 and show the same pattern.

Figure 8 shows the detection rate as the bandwidth changes and again we see that although the tail of the distribution of the detection rate changes with increased correlation, the performance does not vary across changing values of s . It is interesting to note that for the highly correlated cases, although the mean detection rate remains high across the distributions, all distributions showed some cases of poor performance in the simulation. Additional results are given in Appendix C.1.

Lastly we assess the statistical distance between the estimated center and the true center when outliers are present in the sample. Figure 9 shows a similar pattern to Figure 6 indicating the bias of the estimator is not changed with changing values of s or with outliers present in the sample. We observe the same behavior with the uncorrelated lognormal data as we did when the samples contained no outliers. The similarity of the out-of-control statistical distances to those of the in-control cases is not surprising as the outliers are likely to be peeled, not affecting the estimator, $\hat{\mu}_{OCP}$.

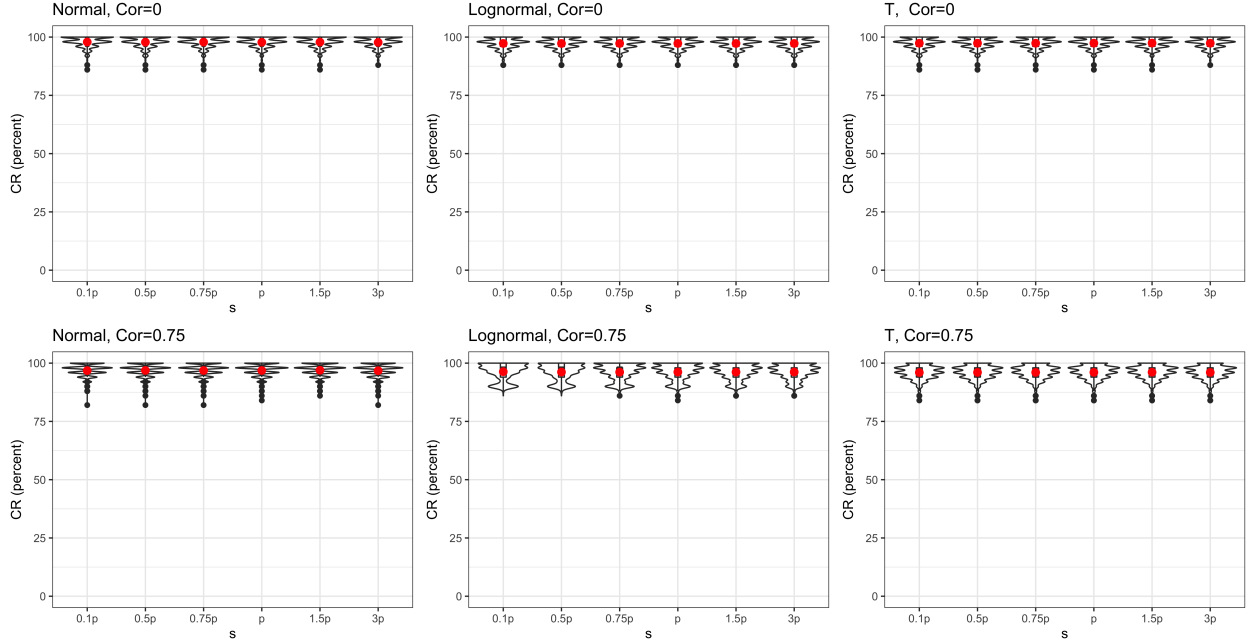


Figure 7: Classification Rate for samples with $N=50$, $p=50$, 10% outliers. The red dot indicates the mean.

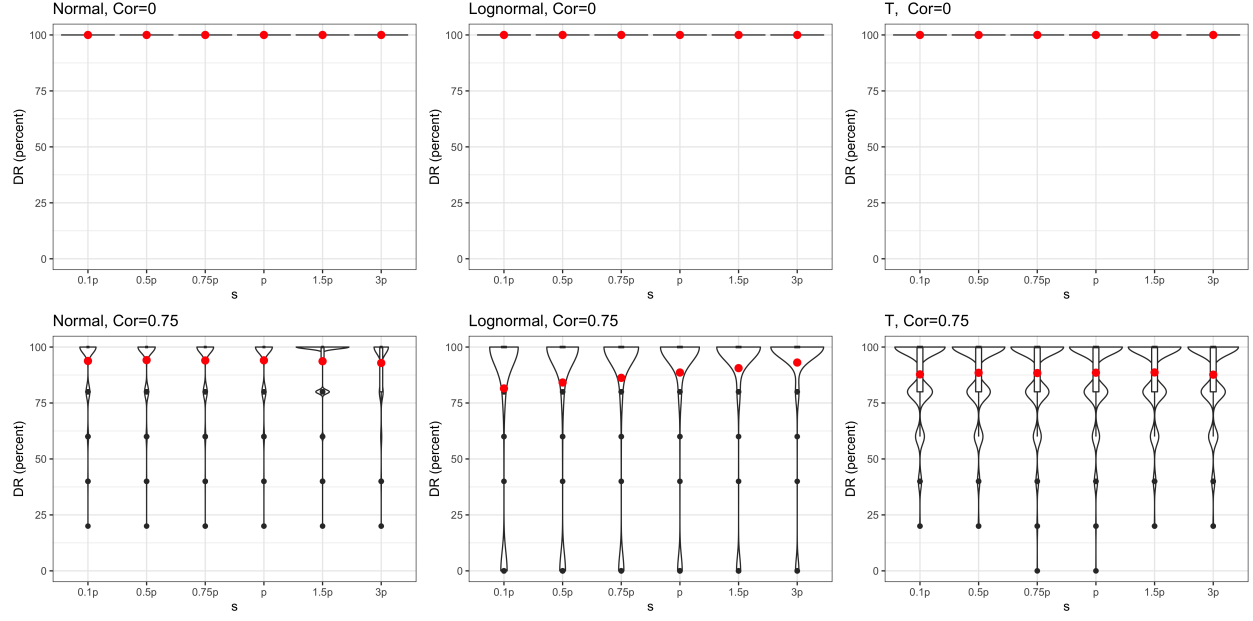


Figure 8: Detection Rate for samples with $N=50$, $p=50$, 10% outliers. The red dot indicates the mean.

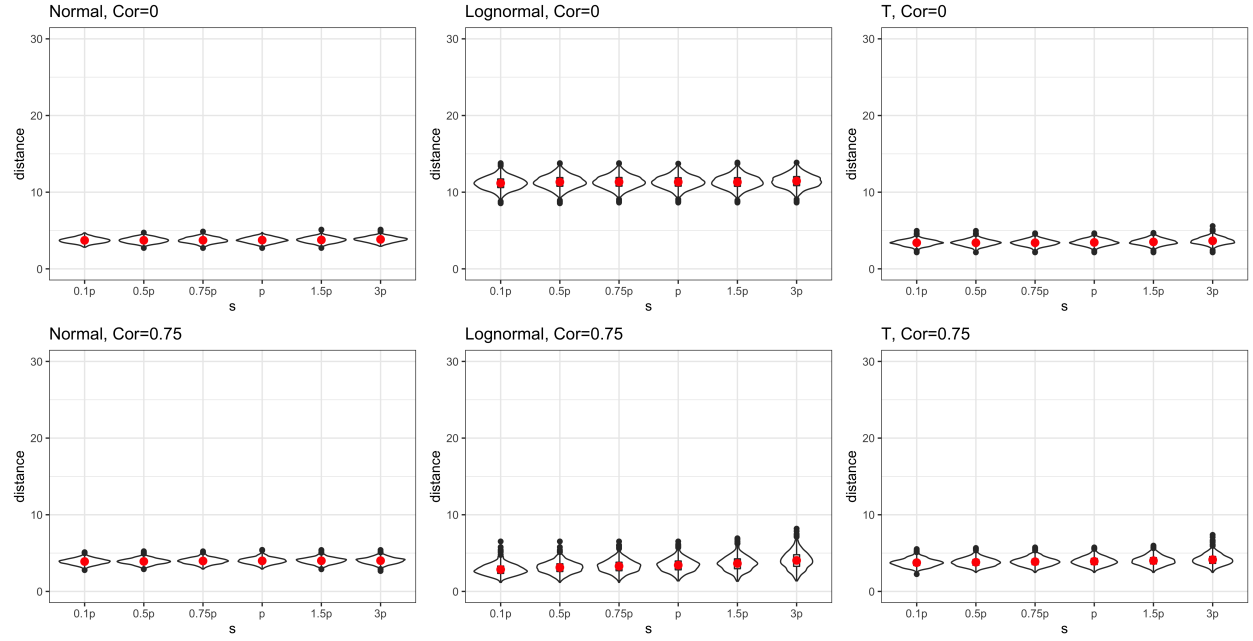


Figure 9: Statistical Distance between $\hat{\mu}_{OCP}$ and μ_0 for samples with $N=50$, $p=50$, 10% outliers. The red dot indicates the mean.

5 Example

Martinez et al.⁷ analyzed an example data set which contained Wikipedia search hits on pages associated with players, teams, and coaches in the NFL. The OCP method was applied to residuals from an additive Holt-Winters model lagged by 24 hours on the original data using the empirical threshold from the $t(df=10)$ - distribution. The data set is estimated to be contaminated with about 29% outliers which are defined by unusual events of interest. We applied the OCP method with different values of the bandwidth, s , to evaluate the detection rate and classification rate. Figure 10 displays the detection rate and classification rate for this example. Similar to the simulation results, the bandwidth parameter, s has little to no effect on the performance of the OCP method on this example data.

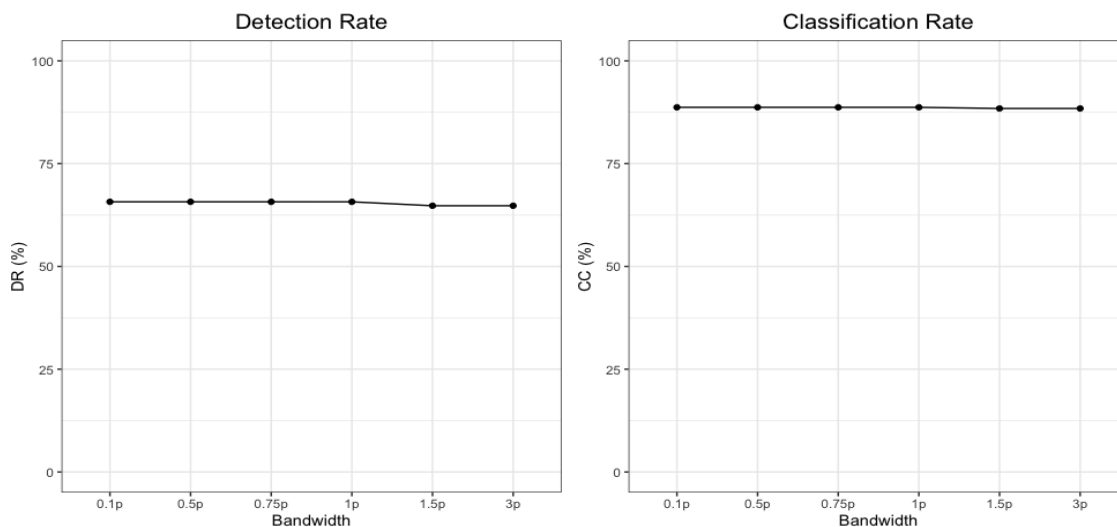


Figure 10: Detection Rate and Classification Rate over changing bandwidth for the NFL Example data from Martinez et al.⁷.

6 Discussion and Conclusion

The OCP method is a robust outlier detection method that works well for high-dimensional data and can be applied in Phase I of process control to identify a baseline sample. Code implementation of the OCP method can be found at <https://github.com/martinwg/OCP.git>. The OCP method

requires selection of a Gaussian kernel bandwidth parameter, s that controls the shape of the kernel used to define the perimeter of the data. Using extensive simulations, we have shown that the choice of the bandwidth parameter s , ranging from $s = 0.1p$ to $s = 3p$ has little effect on the performance of the OCP method in in-control and out-of-control situations. The primary advantage of this finding for the OCP method is the practitioner need not be concerned with the effect of s on the performance of the OCP method.

Our findings suggest the estimator, $\hat{\mu}_{\text{OCP}}$ is slightly biased in the cases of symmetric data and shows larger bias in the case of skewed data. The presence of high correlation in the case of the lognormal distribution mitigated this bias to some degree. Interestingly our results show that regardless of the bias in $\hat{\mu}_{\text{OCP}}$, the average in- and out-of-control performance as measured by empirical false positive, classification, and detection rates were unchanged, regardless of the choice of the bandwidth parameter, s .

While the bandwidth parameter, s does not affect the performance of the OCP method, it does have an affect on the number of observations peeled at each iteration when computing the estimator, $\hat{\mu}_{\text{OCP}}$. Figure 11 illustrates the number of support vectors assigned at each peel of the OCP method for in-control data. The support vectors are the observations which are removed each time a SVDD boundary is constructed. For the majority of the distributions studied, a smaller value of s will lead to more observations removed at each peel therefore requiring fewer SVDD boundaries to be constructed in application of the OCP method. The pattern of smaller s leading to more support vectors is observed across all sample sizes and for in- and out-of-control data (see Appendix D). Thus, we recommend the practitioner select a smaller value of s , which creates boundaries that more tightly estimate the data in order gain computational efficiencies.

As an additional note, the parameter C from Equation (1) also controls the number of observations identified as support vectors for a given data description from SVDD. Modifying C might also increase the number of observations removed at each iteration of the peeling algorithm; however, we have not evaluated how this will effect the performance of the method. Evaluating the effect of C on the performance of the method will be left to future research.

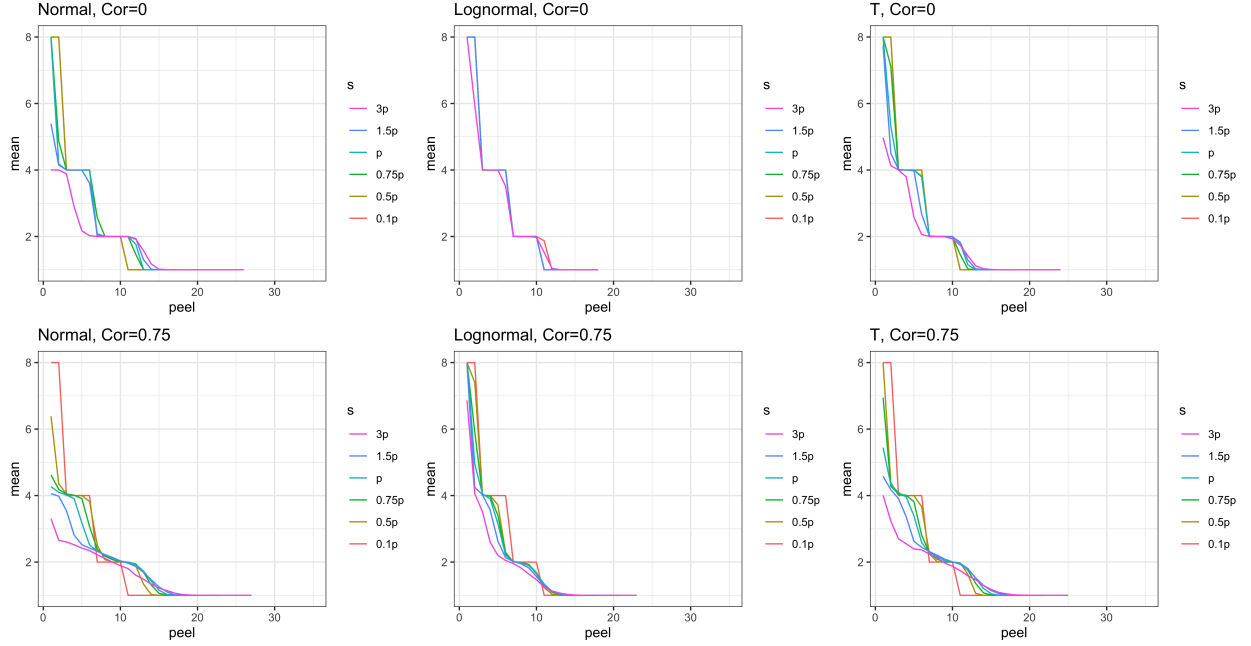


Figure 11: The Number of Support Vector Per Peel $N=50$, $p=50$

References

- [1] Anita Singh. Outliers and robust procedures in some chemometric applications. *Chemometrics and intelligent laboratory systems*, 33(2):75–100, 1996.
- [2] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John wiley & sons, 2005.
- [3] Irad Ben-Gal. Outlier detection. In *Data mining and knowledge discovery handbook*, pages 131–146. Springer, 2005.
- [4] Mia Hubert and Stephan Van der Veen. Outlier detection for skewed data. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 22(3-4):235–246, 2008.
- [5] Peter J Rousseeuw and Mia Hubert. Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):73–79, 2011.

- [6] Haizhou Du et al. Robust local outlier detection. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 116–123. IEEE, 2015.
- [7] Waldyn G. Martinez, Maria L. Weese, and L. Allison Jones-Farmer. A one-class peeling method for multivariate outlier detection with applications in phase i spc. *Quality and Reliability Engineering International*, 36(4):1272–1295, 2020. doi: 10.1002/qre.2629. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.2629>.
- [8] Kwangil Ro, Changliang Zou, Zhaojun Wang, Guosheng Yin, et al. Outlier detection for high-dimensional data. *Biometrika*, 102(3):589–599, 2015.
- [9] L Allison Jones-Farmer, William H Woodall, Stefan H Steiner, and Charles W Champ. An overview of Phase I analysis for process improvement and monitoring. *Journal of Quality Technology*, 46(3):265–280, 2014.
- [10] Vic Barnett. The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, pages 318–355, 1976.
- [11] David MJ Tax and Robert PW Duin. Support vector domain description. *Pattern recognition letters*, 20(11-13):1191–1199, 1999.
- [12] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004.
- [13] Maria L Weese, Waldyn G Martinez, and L Allison Jones-Farmer. On the selection of the bandwidth parameter for the k-chart. *Quality and Reliability Engineering International*, 33(7):1527–1547, 2017.
- [14] Ruixiang Sun and Fugee Tsung. A kernel-distance-based multivariate control chart using support vector methods. *International Journal of Production Research*, 41(13):2975–2989, 2003. doi: 10.1080/1352816031000075224. URL <https://doi.org/10.1080/1352816031000075224>.

- [15] Paulo Cortez, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553, 2009.
- [16] David L Donoho and Peter J Huber. The notion of breakdown point. *A festschrift for Erich L. Lehmann*, 157184, 1983.

A Additional FSRBP Simulation Results

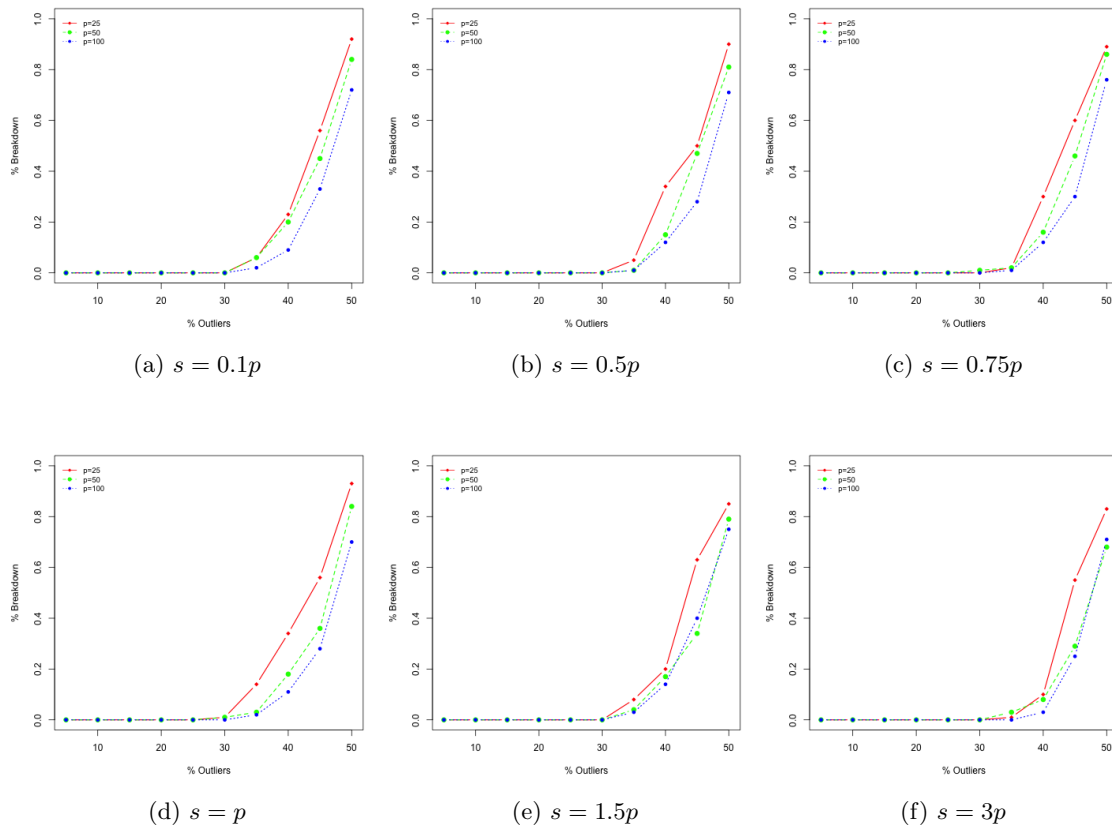


Figure 1: Breakdown Rate (Uncorrelated lognormal distribution)

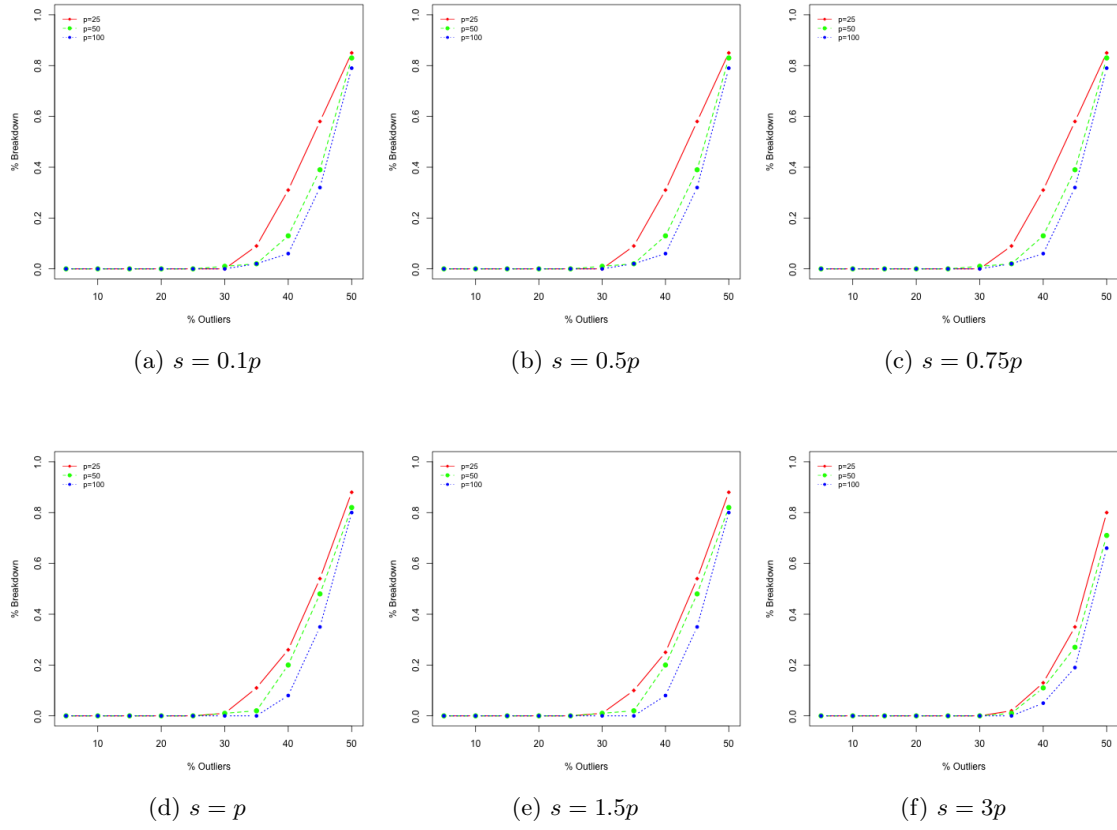
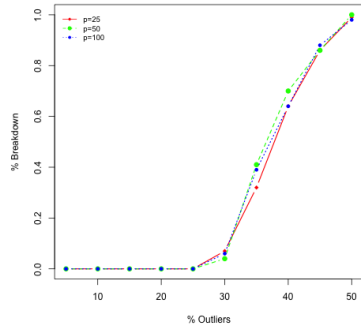
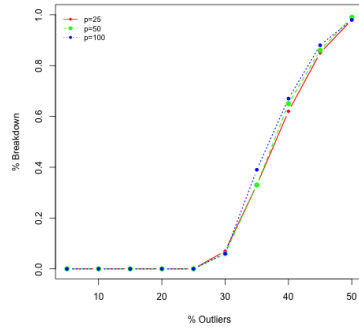


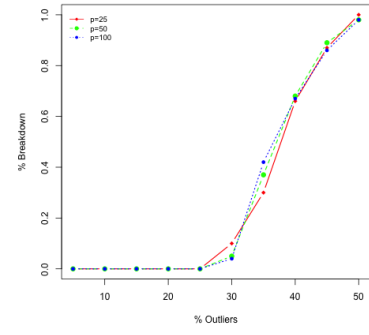
Figure 2: Breakdown Rate (Correlated lognormal distribution)



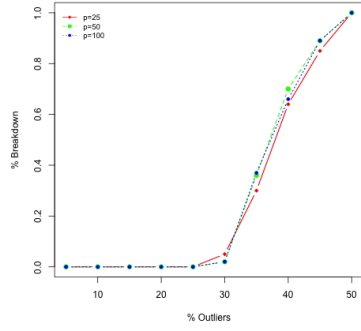
(a) $s = 0.1p$



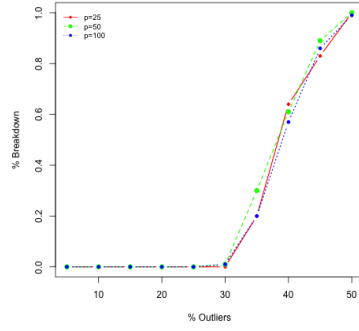
(b) $s = 0.5p$



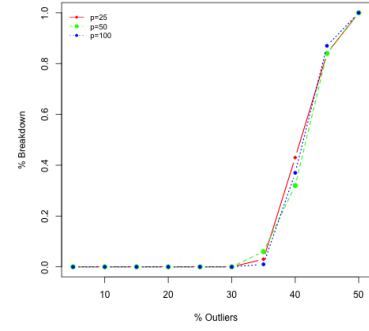
(c) $s = 0.75$



(d) $s = p$



(e) $s = 1.5p$



(f) $s = 3p$

Figure 3: Breakdown Rate (Uncorrelated t-distribution)

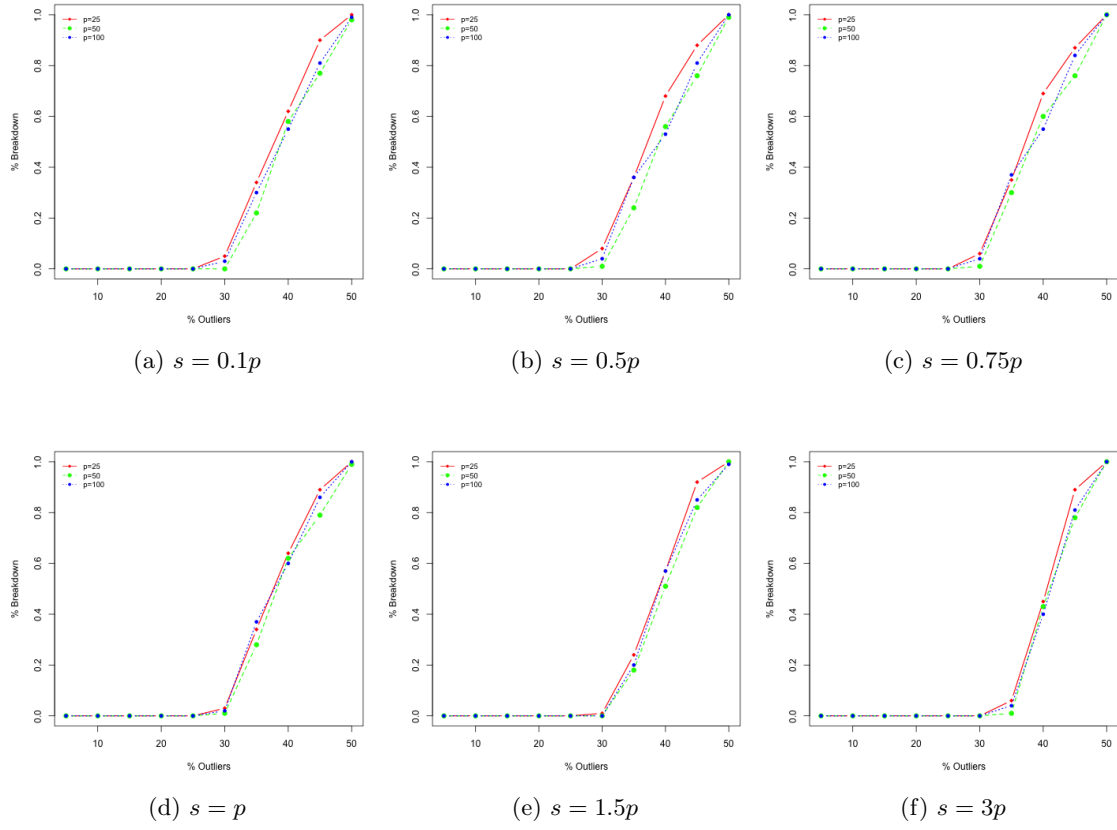


Figure 4: Breakdown Rate (Correlated t-distribution)

B Additional In-control Simulation Results

B.1 False Positive Rate

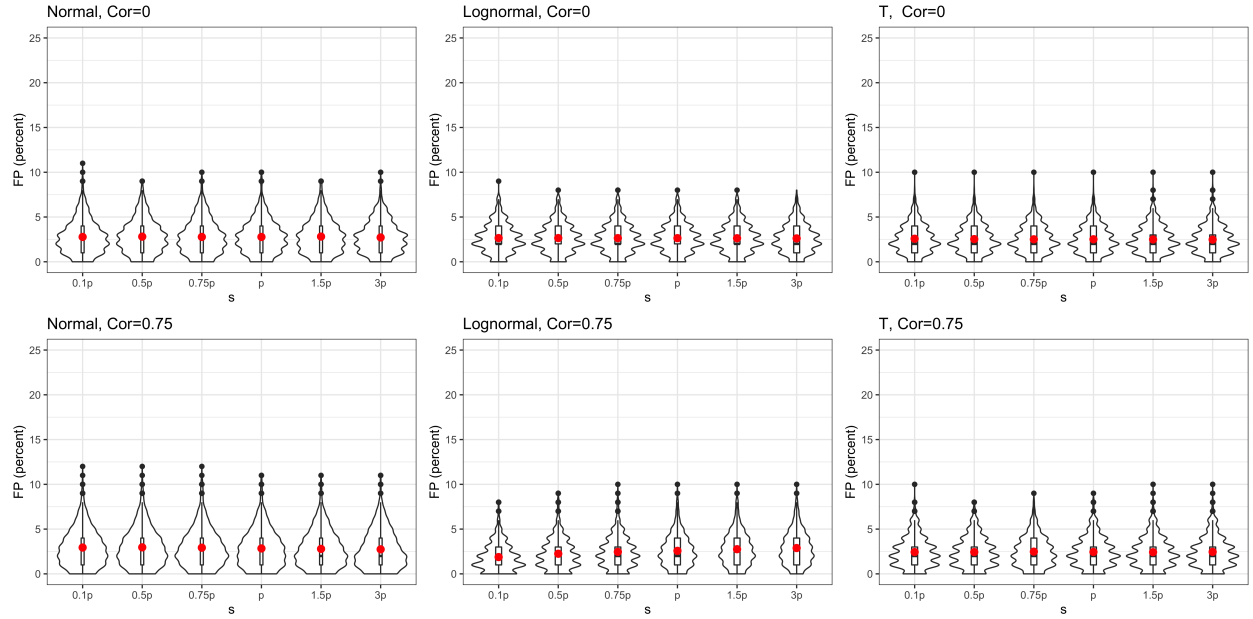


Figure 1: False Positive Rate $N=50$, $p=100$

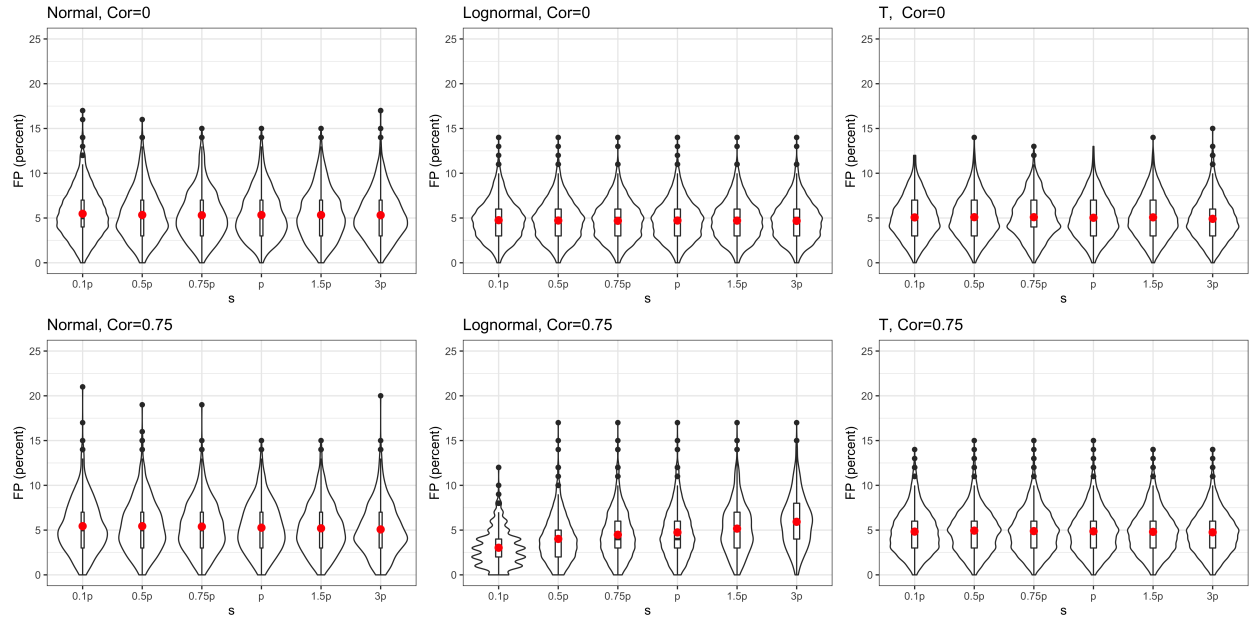


Figure 2: False Positive Rate $N=100$, $p=100$

B.2 Statistical distance

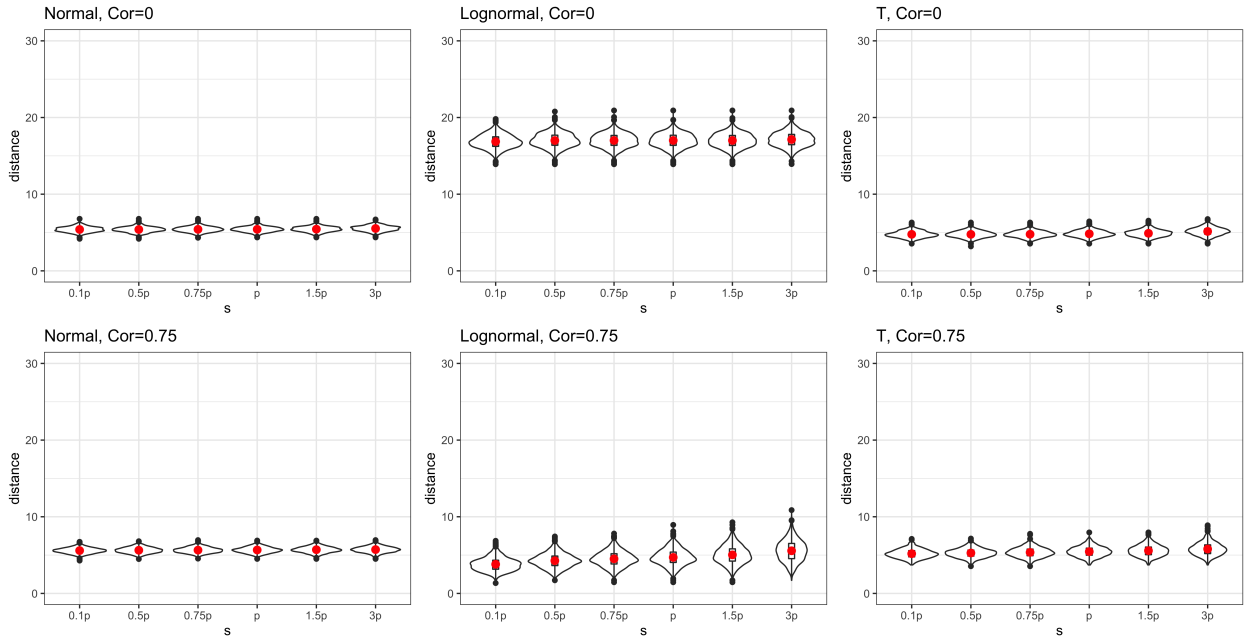


Figure 1: Distance between the true mean and the estimated mean, $N=50$, $p=100$

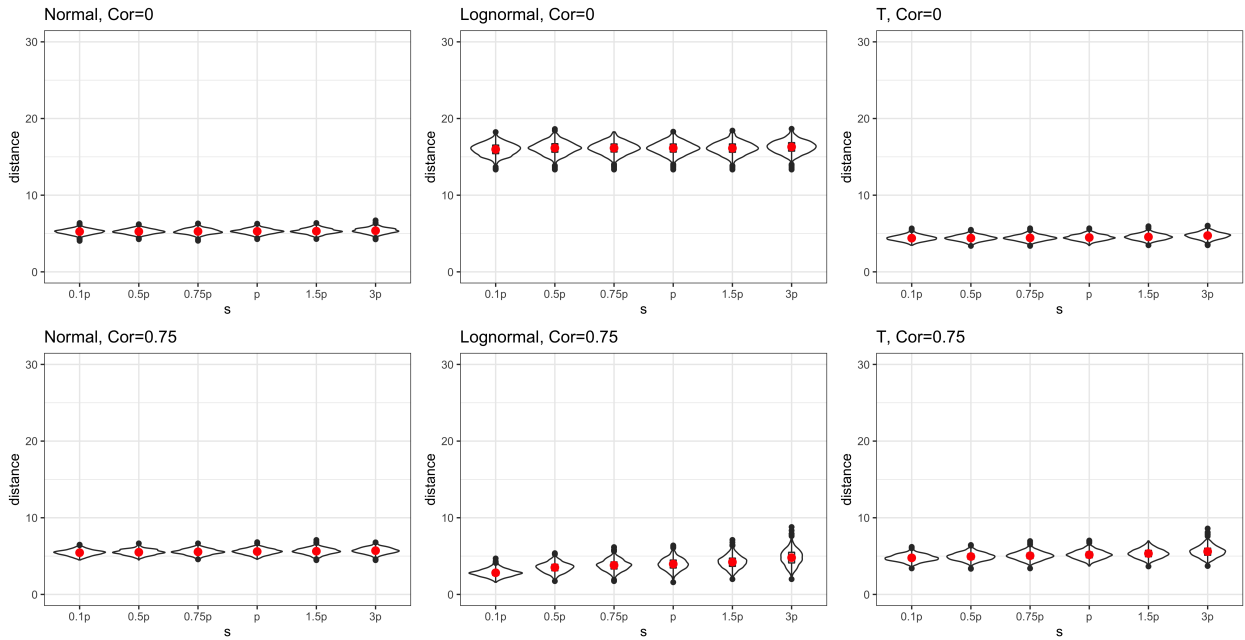


Figure 2: Distance between the true mean and the estimated mean, $N=100$, $p=100$

C Additional out-of-control Simulation Results

C.1 Detection Rate

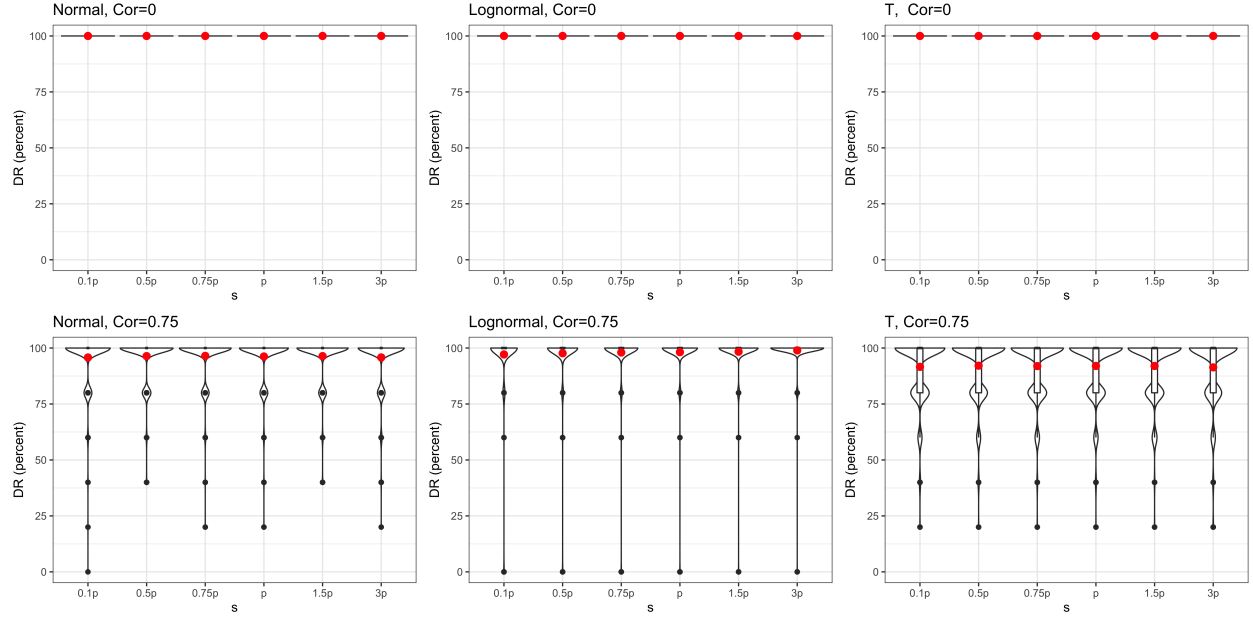


Figure 1: Detection Rate $N=50$, $p=100$

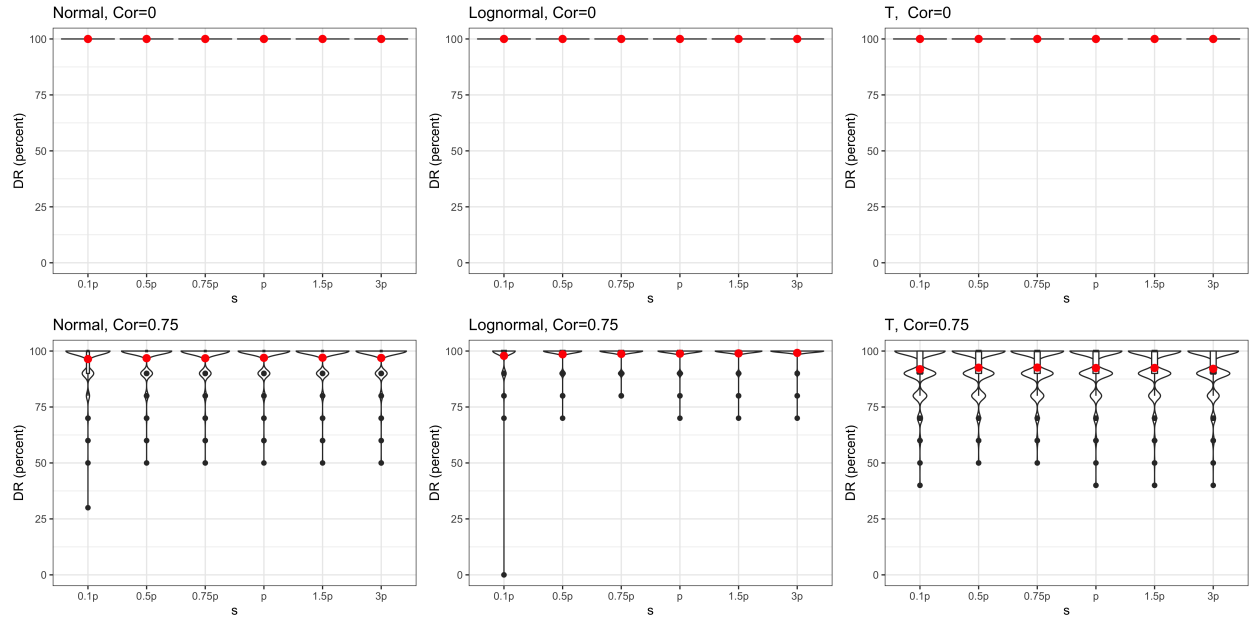


Figure 2: Detection Rate $N=100$, $p=100$

C.2 Classification Rate

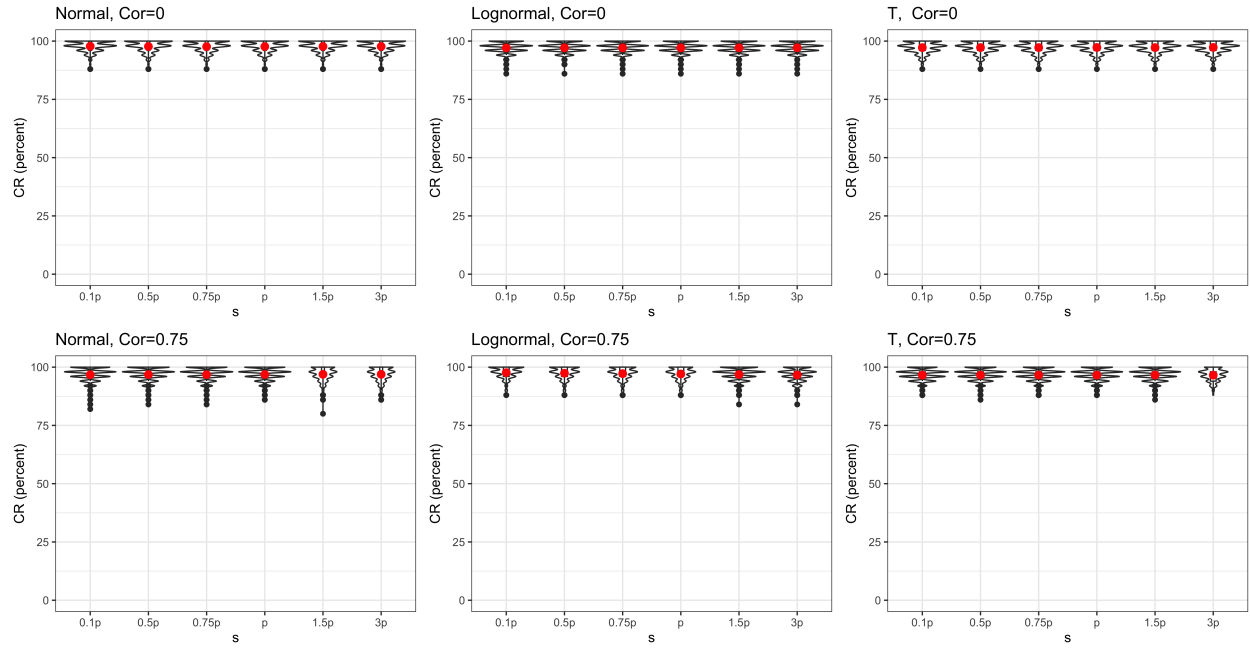


Figure 1: Classification Rate, $N=50$, $p=100$

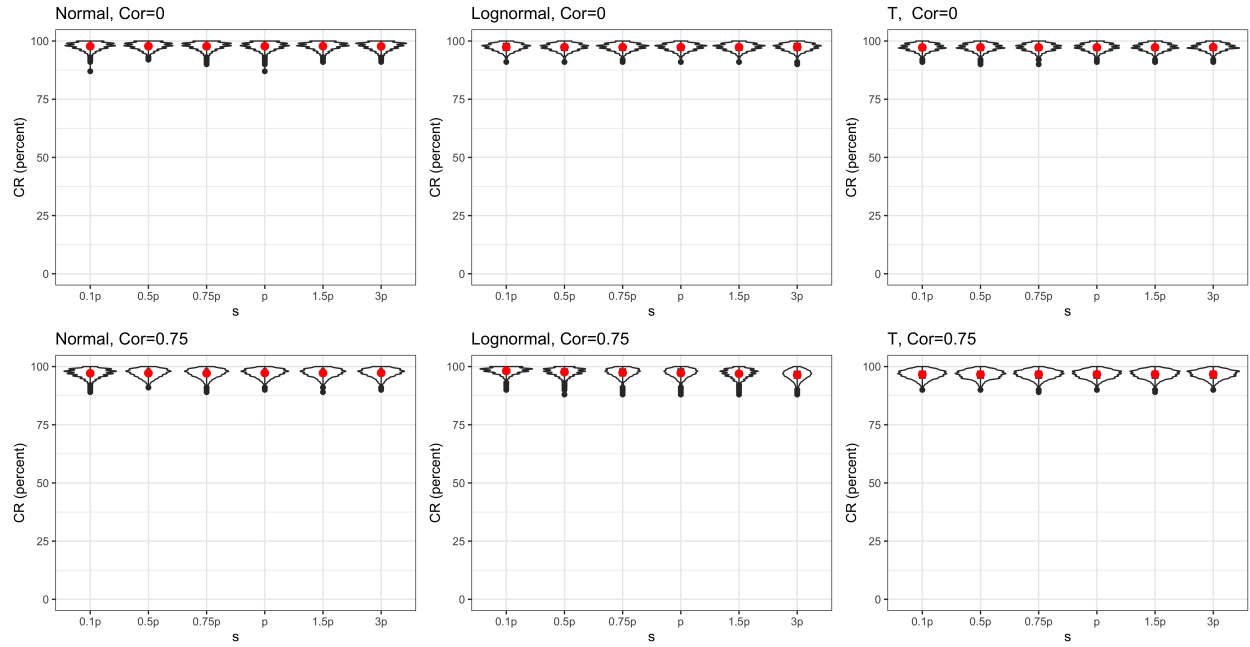


Figure 2: Classification Rate, $N=100$, $p=100$

C.3 Statistical Distance

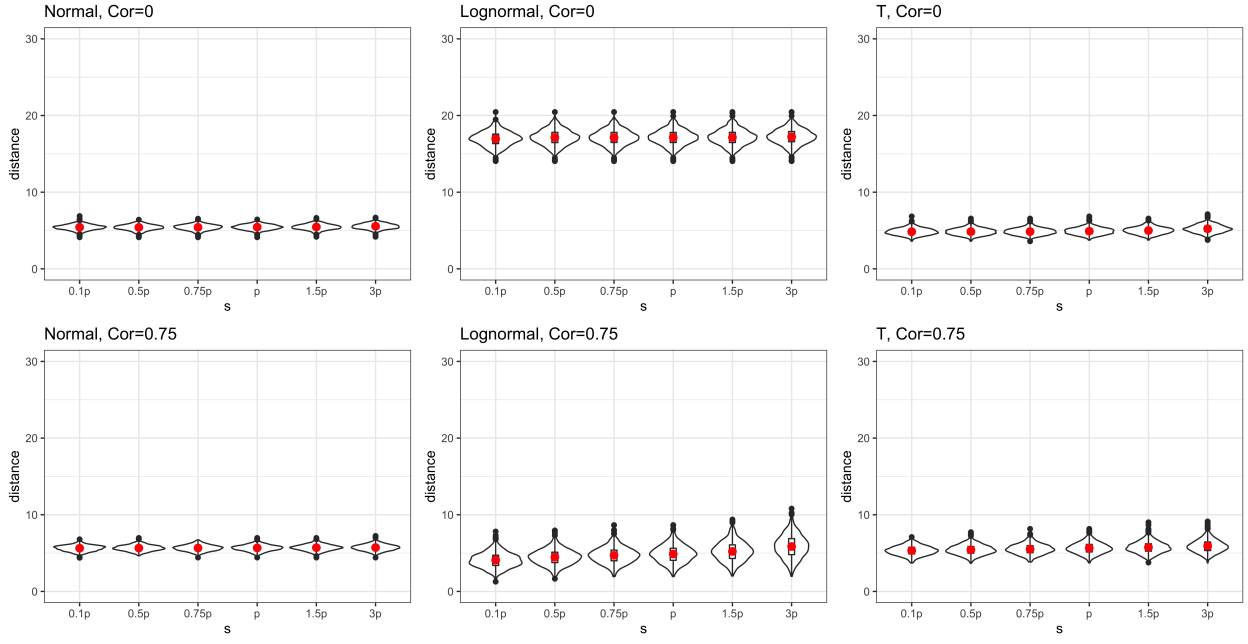


Figure 1: Distance between the true mean and the estimated mean ($N=50$, $p=100$)

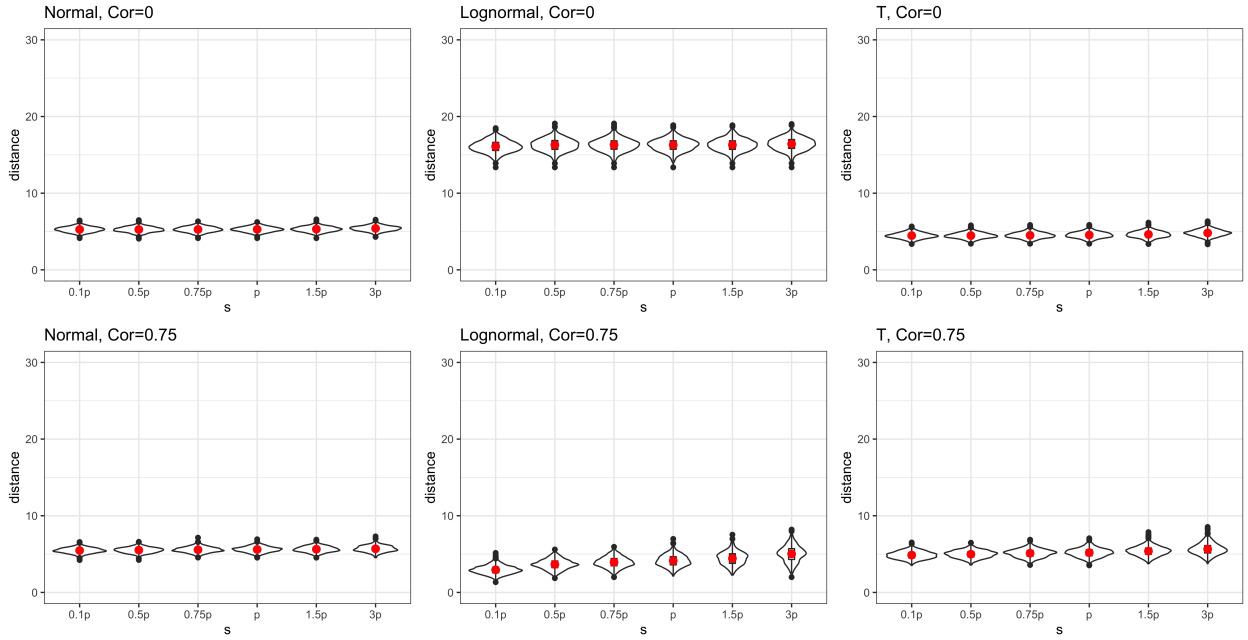


Figure 2: Distance between the true mean and the estimated mean ($N=100$, $p=100$)

D The Number of Support Vectors Per Peel

D.1 IC scenario

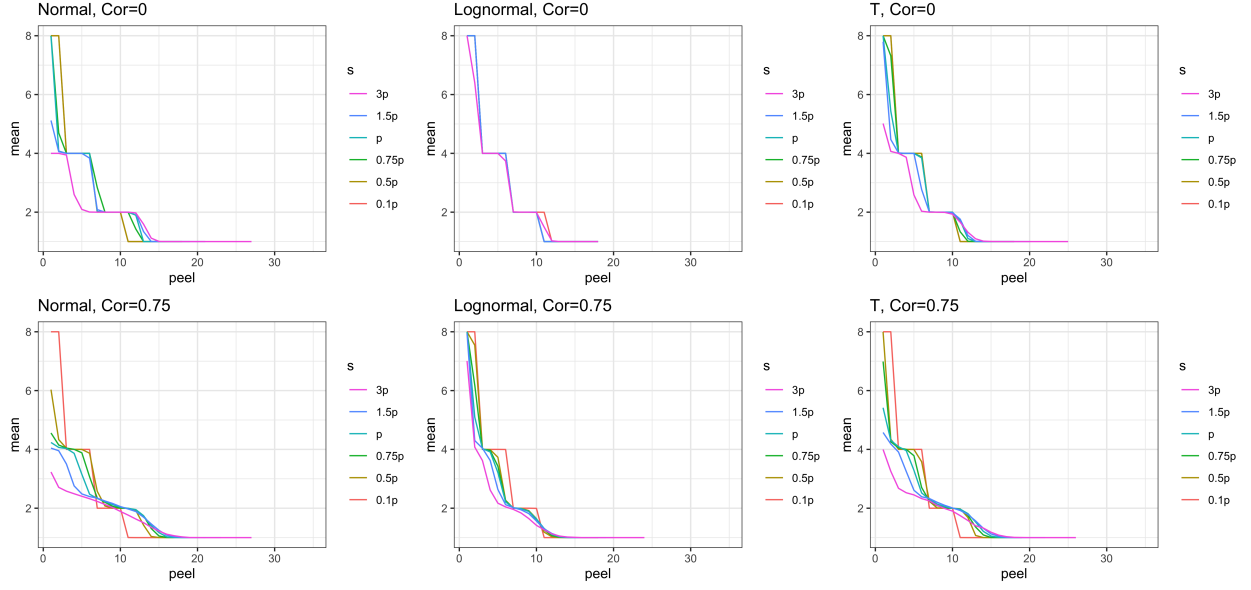


Figure 3: The Number of Support Vectors Per Peel, $N=50$, $p=100$

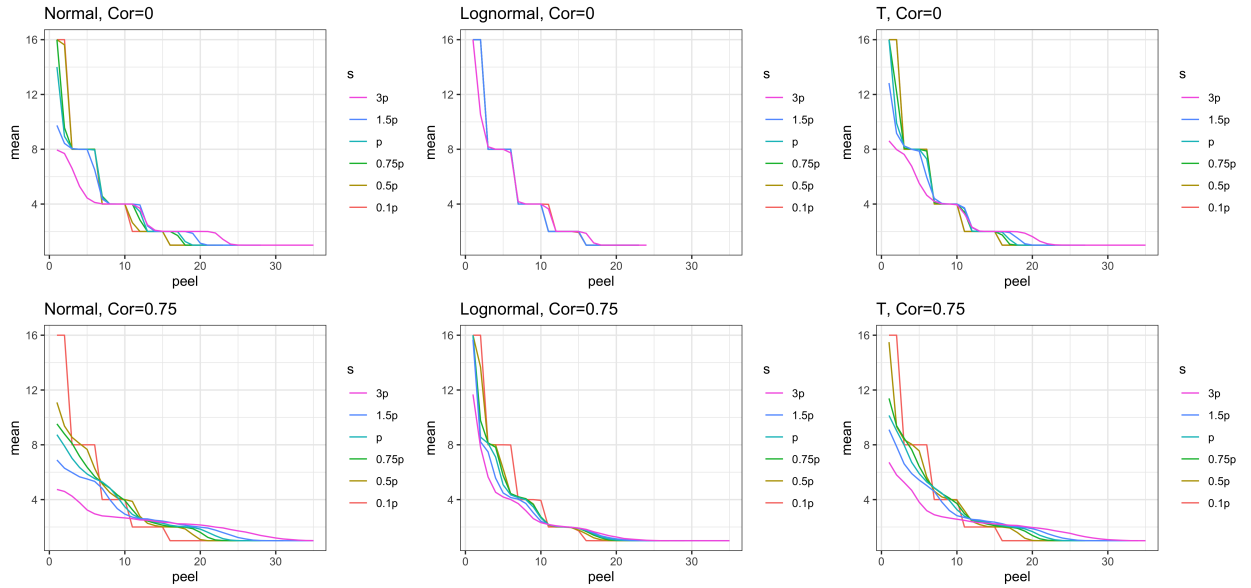


Figure 4: The Number of Support Vectors Per Peel, $N=100$, $p=100$

D.2 OC scenario

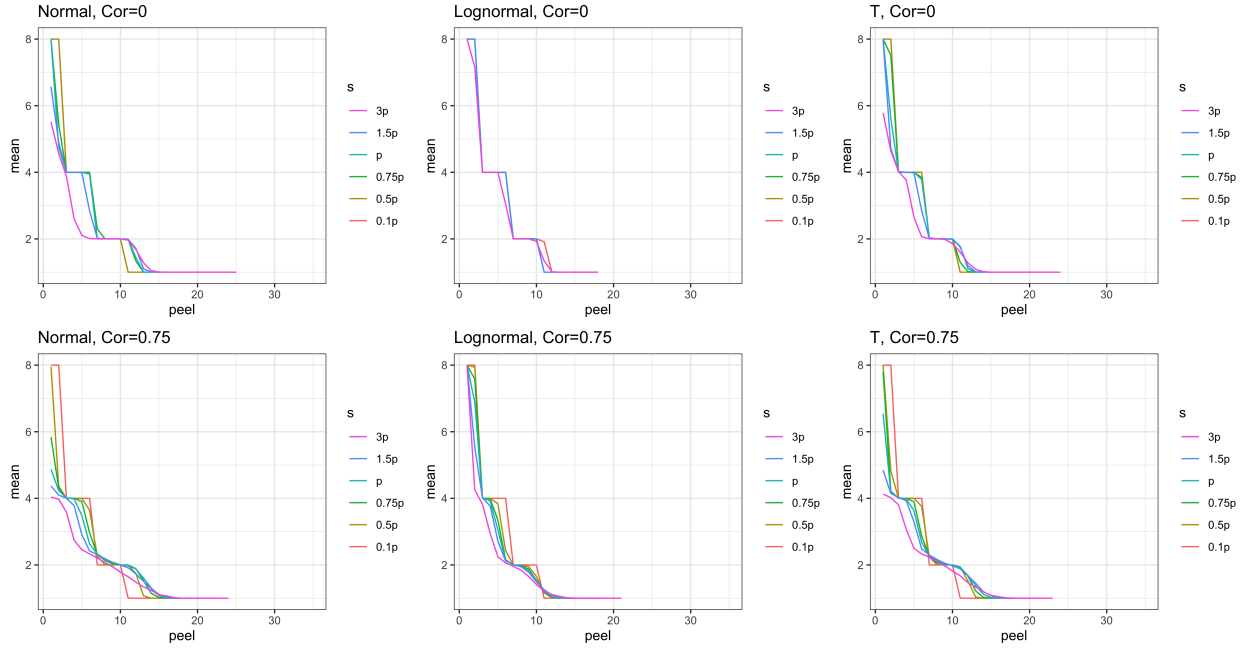


Figure 1: Number of Support Vectors Per Peel in the Simulation with Outliers, $N=50$, $p=50$

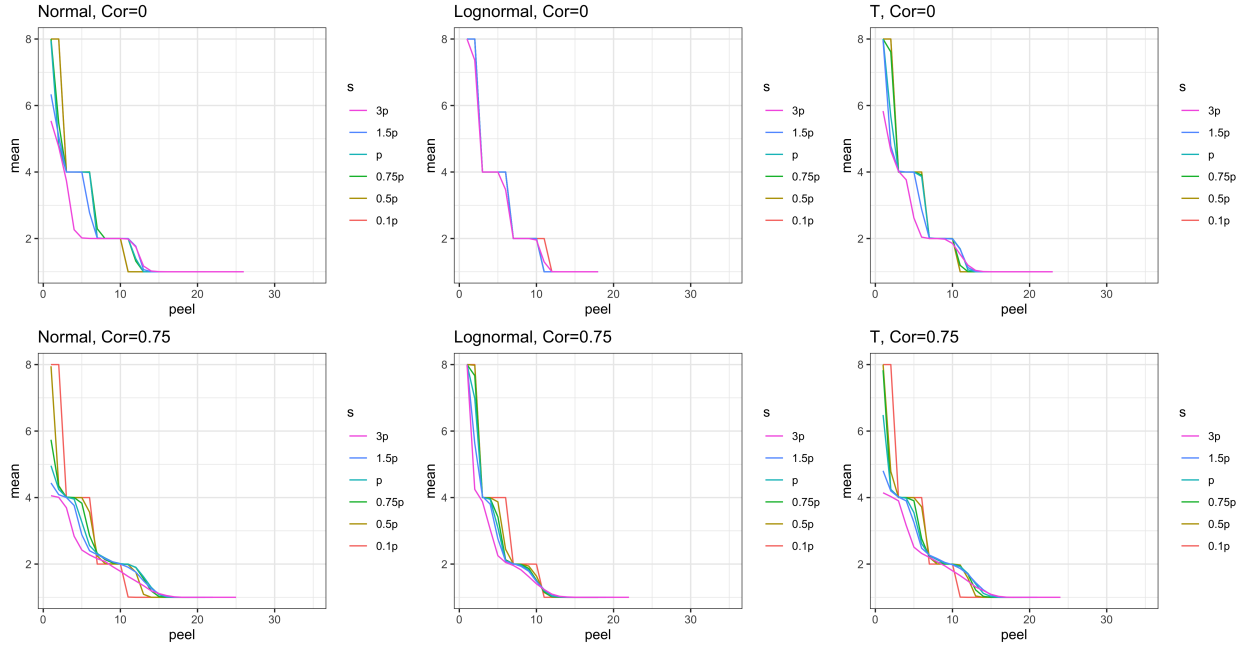


Figure 2: Number of Support Vectors Per Peel in the Simulation with Outliers, $N=50$, $p=100$

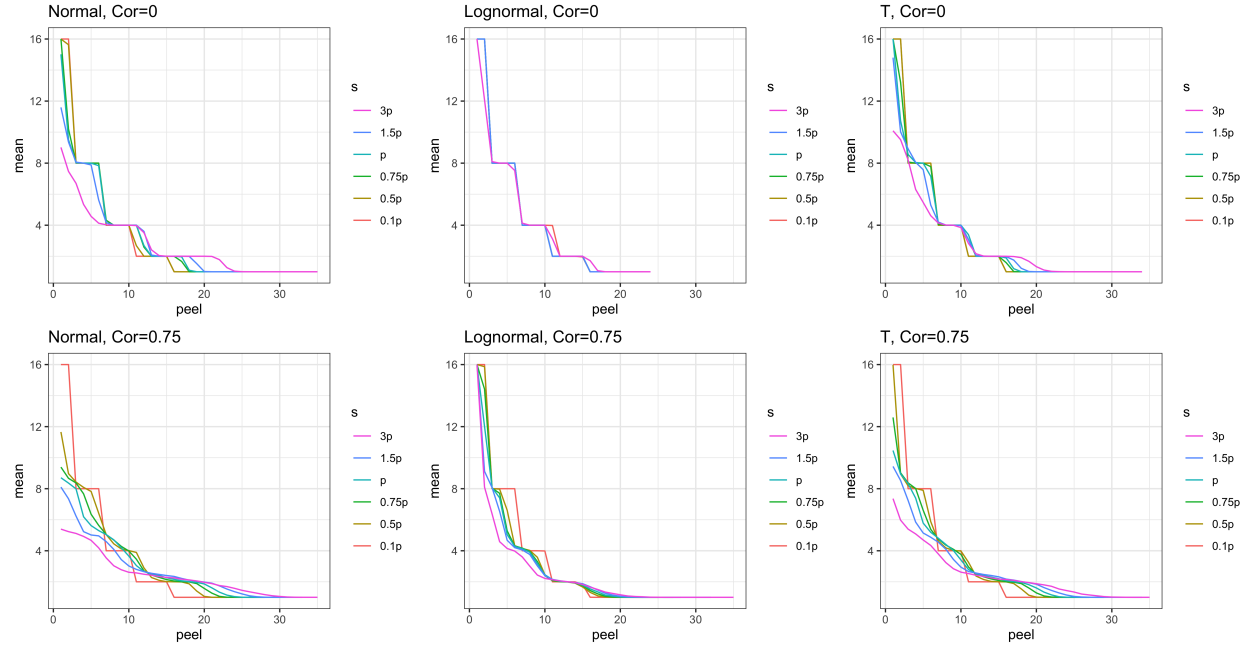


Figure 3: Number of Support Vectors Per Peel in the Simulation with Outliers, $N=100$, $p=100$