MULTIVARIATE ROBUST CONTROL CHARTS FOR LOCATION AND DISPERSION

Thesis submitted to the University of Calicut for
the award of the degree of

DOCTOR OF PHILOSOPHY
IN
STATISTICS
under the Faculty of Science

by

VIJAYALAKSHMI S

under the guidance of

Dr. NICY SEBASTIAN

DEPARTMENT OF STATISTICS
ST. THOMAS COLLEGE (AUTONOMOUS), THRISSUR
KERALA – 680001, INDIA

MAY 2026

CERTIFICATE

This is to certify that the adjudicators of the Ph.D thesis of Ms. Vijayalakshmi S., titled “Multivariate robust control charts for location and dispersion”, have not offered any directions for corrections or suggestions for changes in their reports. The content of the soft copy is identical to that of the hard copy.

Thrissur | Dr. Nicy Sebastian
Date: 06/05/2026 | Research Supervisor

CERTIFICATE

I hereby certify that, this thesis entitled "Multivariate robust control charts for location and dispersion" is a bonafide record of research work carried out by Ms. Vijayalakshmi S, Research Scholar, Department of Statistics, St. Thomas College (Autonomous), Thrissur, under my supervision and guidance for the award of the Degree of Doctor of Philosophy in Statistics, of the University of Calicut. The work reported herein does not form part of any other thesis or dissertation submitted previously for the award of any degree or diploma of any other university or institution.

Thrissur

Date: 06/05/2026

Dr. Nicy Sebastian

Research Supervisor

DECLARATION

I, Vijayalakshmi S, hereby declare that the work presented in the thesis entitled, "Multivariate robust control charts for location and dispersion" is based on the original work done by me under the guidance of Dr. Nicy Sebastian, Assistant Professor, Department of Statistics, St. Thomas College (Autonomous), Thrissur, Kerala and has not been included in any other thesis submitted previously for the award of any degree. The contents of the thesis are undergone plagiarism check using iThenticate software at C.H.M.K. Library, University of Calicut, and the similarity index found within the permissible limit. I also declare that the thesis is free from AI generated contents.

Vijayalakshmi S

Thrissur
Date: 06/05/2026

Dr. Nicy Sebastian
Research Supervisor

ACKNOWLEDGEMENT

First and foremost, I express my heartfelt gratitude to my research supervisor, Dr. Nicy Sebastian, Assistant Professor, Department of Statistics, St. Thomas College (Autonomous), Thrissur, for her invaluable guidance, constant encouragement, and unwavering support throughout the course of my research. Her insightful suggestions and patient mentorship have been instrumental in the successful completion of this work.

I am deeply grateful to Rev. Dr. Martin K. A., Principal in Charge, Rev. Fr. Biju Panengaden, Executive Manager, and the Management of St. Thomas College (Autonomous), Thrissur, for providing me with the opportunity and facilities to carry out my research in this esteemed institution.

My sincere thanks are due to Dr. Sajesh T. A., Associate Professor and Head, Department of Statistics, and to all the faculty members of the department for their valuable support and encouragement. I also extend my gratitude to Dr. Joyce Jose, Dean of Research, for her guidance and motivation. I would like to express my sincere gratitude to Mr. Sanjo Jose, the college librarian, for his prompt and kind assistance in providing the required research articles whenever I approached him, which greatly supported my research work.

I gratefully acknowledge the esteemed RAC members for their valuable comments, constructive feedback, and kind guidance throughout my research journey.

I express my sincere gratitude to Mr. Bobby Paul, Managing Director, Star Pipes and Fittings, Thrissur, for granting permission to collect data from their Athani factory. I am also thankful to Mr. Baiju Devassy, Manager, and Mr. Sreenath K. P. for their valuable support during the data collection process.

I extend my heartfelt thanks to Mr. Jaison P. Jose, Whole-time Director, Tholur Foods Private Limited, Thrissur, for kindly granting permission to collect data from their factory, and to Ms. Juhiya C. B., Food Quality Controller and Technologist, for her kind assistance and cooperation throughout the data collection.

I would also like to thank all my teachers who have guided and inspired me throughout my academic years, and my dear research friends for their companionship and encouragement.

Finally, with deepest love and gratitude, I thank my family members for their constant support, patience, and encouragement, as their care has been my greatest strength.

Vijayalakshmi S

ABSTRACT

Statistical quality control plays a crucial role in maintaining process stability and product reliability across industrial and service sectors. However, the presence of outliers, data contamination, and high-dimensional measurements poses serious challenges to conventional monitoring methods. Classical control charts, though widely used, are highly sensitive to outlying observations, often leading to false alarms or missed detections of genuine process shifts. This research focuses on developing robust estimation and control chart methodologies for effective monitoring of bivariate and multivariate processes under such complex conditions.

A major contribution of this study is the development of the Shrinkage Gnanadesikan-Kettenring (SGK) covariance estimator, a robust multivariate estimator derived from the Gnanadesikan–Kettenring approach. Simulation studies demonstrated that SGK efficiently identifies contaminated observations while maintaining low false detection rates across diverse data structures, including high-dimensional and low-sample-size cases. Comparative analysis with existing robust estimators such as Minimum Covariance Determinant (MCD), Robust Mahalanobis Distance-Shrinkage (RMD-S), Reweighted Orthogonalized Comedian (ROC), S_n, and Orthogonalized Gnanadesikan-Kettenring (OGK) established that SGK achieves superior robustness and computational efficiency.

Building upon robust estimation, novel bivariate and multivariate control charts were developed for both individual observations and rational subgroups. These charts exhibited strong in-control performance comparable to classical Hotelling's T^2 charts for clean data, while significantly outperforming them under contamination. Robust dispersion charts based on S_n Covariance, GK_S_n, GK_Q_n, Comedian, ROC, and SGK within Multivariate Exponentially Weighted Mean Squared deviation (MEWMS) and Multivariate Exponentially Weighted Moving Variance (MEWMV) frameworks further enhanced detection of variability shifts with low false alarm rates. Real data applications validated the proposed charts' effectiveness, robustness, and computational efficiency.

This research contributes a unified framework for robust process monitoring, offering practical, scalable, and reliable tools for modern quality management. Numerical studies and real-world applications confirm the models' capability to detect process shifts accurately, even under contamination and high dimensional conditions, ensuring enhanced decision making and operational reliability.

Keywords: robust estimation, multivariate control charts, SGK covariance estimator, process monitoring, data contamination.

പഠനസംഗ്രഹം

ഏതൊരു നിർമ്മാണ അല്ലെങ്കിൽ സേവന പ്രക്രിയയുടെയും ഒരു പ്രധാന ഭാഗമാണ് ഗുണനിലവാര നിയന്ത്രണം. ഇത് സ്ഥാപനങ്ങളെ അവരുടെ ഉൽപ്പന്നങ്ങളും സേവനങ്ങളും ആവശ്യമായ മാനദണ്ഡങ്ങൾ പാലിക്കുന്നുണ്ടെന്ന് ഉറപ്പാക്കാൻ സഹായിക്കുന്നു. എന്നിരുന്നാലും, ശേഖരിച്ച ഡാറ്റയിൽ അസാധാരണമോ തെറ്റായതോ ആയ മൂല്യങ്ങൾ (ഔട്ട്ലറുകൾ എന്ന് വിളിക്കപ്പെടുന്നു) അടങ്ങിയിരിക്കുമ്പോൾ, പരമ്പരാഗത ഗുണനിലവാര നിയന്ത്രണ രീതികൾ പലപ്പോഴും തെറ്റായ ഫലങ്ങൾ നൽകുന്നു. അവ തെറ്റായ സൂചനകൾ ഉയർത്തുകയോ പ്രക്രിയയിൽ യഥാർത്ഥ പ്രശ്നങ്ങൾ ശ്രദ്ധിക്കുന്നതിൽ പരാജയപ്പെടുകയോ ചെയ്തേക്കാം.

അത്തരം അസാധാരണമായ ഡാറ്റ ശരിയായി കൈകാര്യം ചെയ്യാൻ കഴിയുന്ന ശക്തമായ രീതികൾ വികസിപ്പിക്കുന്നതിലാണ് ഈ ഗവേഷണം ശ്രദ്ധ കേന്ദ്രീകരിക്കുന്നത്. ഡാറ്റ ഭാഗികമായി തെറ്റോ മലിനമായതോ ആണെങ്കിൽ പോലും വേരിയബിളുകൾ തമ്മിലുള്ള ബന്ധം അളക്കുന്നതിനായി SGK കോവേരിയൻസ് എസ്റ്റിമേറ്റർ എന്ന പുതിയ സ്റ്റാറ്റിസ്റ്റിക്കൽ രീതി വികസിപ്പിച്ചെടുത്തു. ഈ പുതിയ രീതി അറിയപ്പെടുന്ന നിരവധി എസ്റ്റിമേറ്ററുകളുമായി താരതമ്യം ചെയ്തതിൽ നിന്നും, SGK കൂടുതൽ കൃത്യവും സ്ഥിരതയുള്ളതുമായ ഫലങ്ങൾ നൽകുന്നുവെന്ന് കാണാവുന്നതാണ്, പ്രത്യേകിച്ച് ഡാറ്റ സങ്കീർണ്ണമോ നിരവധി വേരിയബിളുകളോ ഉള്ളപ്പോൾ.

ഈ ശക്തമായ രീതി ഉപയോഗിച്ച്, പ്രക്രിയകളുടെ പ്രകടനവും സ്ഥിരതയും നിരീക്ഷിക്കുന്നതിന് പുതിയ തരം നിയന്ത്രണ ചാർട്ടുകൾ സൃഷ്ടിച്ചു. ഈ ചാർട്ടുകൾക്ക് മാറ്റങ്ങൾ കൂടുതൽ ഫലപ്രദമായി കണ്ടെത്താനും ഡാറ്റയിലെ പിശകുകൾ ബാധിക്കാതിരിക്കാനും കഴിയും. വലിയ അളവിലുള്ള ഡാറ്റ കൈകാര്യം ചെയ്യുന്ന അല്ലെങ്കിൽ അളക്കലിൽ തെറ്റുകൾ സാധാരണമായ സാഹചര്യങ്ങളിൽ പ്രവർത്തിക്കുന്ന വ്യവസായങ്ങൾക്ക് അവ ഉപയോഗപ്രദമാണ്.

മൊത്തത്തിൽ, പ്രക്രിയയുടെ ഗുണനിലവാരം നിരീക്ഷിക്കുന്നതിനും മെച്ചപ്പെടുത്തുന്നതിനുമുള്ള മെച്ചപ്പെട്ട ഉപകരണങ്ങൾ ഈ ഗവേഷണം നൽകുന്നു. തെറ്റായ മുന്നറിയിപ്പുകൾ കുറയ്ക്കുന്നതിനും, യഥാർത്ഥ മാറ്റങ്ങൾ വേഗത്തിൽ കണ്ടെത്തുന്നതിനും, നിർമ്മാണം, ആരോഗ്യ സംരക്ഷണം, സേവനങ്ങൾ തുടങ്ങിയ വ്യവസായങ്ങളിൽ കൂടുതൽ വിശ്വസനീയമായ തീരുമാനമെടുക്കൽ ഉറപ്പാക്കുന്നതിനും ഈ രീതികൾ സഹായിക്കുന്നു.

സൂചക പദങ്ങൾ: ഗുണനിലവാര നിയന്ത്രണം, ശക്തമായ രീതികൾ, SGK എസ്റ്റിമേറ്റർ, പ്രക്രിയ നിരീക്ഷണം, ഔട്ട്ലയറുകൾ.

Contents

List of Figures

List of Tables

CHAPTER 1
INTRODUCTION AND OVERVIEW

1.1 Introduction

In both industrial and service-oriented sectors, quality plays a critical role in shaping organizational success, competitiveness, and long-term sustainability (Juran *et al*., 1999; Deming, 1986). In today's era of globalization and intense competition, quality has moved beyond being optional and has become a strategic imperative that influences business performance, customer loyalty, and market reputation (Oakland, 2014). Rapid technological progress in manufacturing, including automation, high-speed processes, and large-scale production, has further complicated the task of ensuring consistent quality. These developments necessitate the use of more advanced and adaptive methods of quality assurance (Montgomery, 2020). Moreover, under global competitive pressures, the philosophy of quality must extend across all organizational levels, embedding itself into culture rather than remaining restricted to isolated functions (Deming, 1986; Oakland, 2014). However, achieving and sustaining quality continues to be challenging, particularly in contexts where resources are scarce or constrained.

The concept of quality is inherently multidimensional. It encompasses attributes such as reliability, durability, maintainability, functionality, aesthetics, and conformity to standards and specifications. Among the many definitions of quality, one of the most widely accepted is "fitness for use," which emphasizes the product's ability to carry out its intended function effectively (Juran *et al*., 1999; Montgomery, 2020). The long-standing assumption that higher quality necessarily leads to higher costs has been shown to be misleading. As Taguchi *et al*. (1989) demonstrated, improving quality can actually reduce costs by lowering scrap, rework, and warranty claims. This perspective highlights the importance of minimizing process variation at every stage of production, since effective quality improvement benefits both producers and customers. Since the early twentieth century, statistical methods have been central to quality management, providing strong tools for achieving greater consistency, efficiency, and customer satisfaction.

Quality control is a systematic and ongoing process that involves assessing actual performance, comparing it with predefined standards, and taking corrective actions whenever deviations occur. It includes both engineering and managerial functions aimed at ensuring that products conform to specified requirements. More specifically, statistical quality control utilizes statistical methods across the stages of product design, production, maintenance, and service, with the primary objective of fulfilling customer needs in an efficient and cost-effective manner (Montgomery, 2020; Juran *et al*., 1999).

A fundamental objective of Statistical Quality Control is the systematic reduction of variation arising from controllable sources within production processes. Quality assurance within this framework is typically carried out in two complementary stages. The first stage involves Statistical Process Control (SPC), which comprises a set of analytical and diagnostic tools aimed at monitoring process behaviour, achieving operational stability, and enhancing capability by minimizing variability. Through the effective application of SPC, organizations can ensure that production processes remain consistent and within specified control limits (Montgomery, 2020). The second stage, referred to as acceptance sampling, is applied to assess the quality of finished product lots prior to shipment, providing a means to control the probability of accepting nonconforming items. This method offers a systematic and statistically rigorous framework for decision-making, ensuring that outgoing products meet specified quality standards and reinforcing both product reliability and customer confidence. SPC is considered superior to acceptance sampling because it focuses on monitoring and controlling variation within the process to prevent defects, whereas acceptance sampling is limited to detecting nonconformities after production has occurred (Shewhart, 1931).

1.2 Statistical Process Control

SPC is widely used in industrial environments, and this growing reliance has created a need for effective methods that can quickly detect shifts in process quality. The central objective of SPC is to enhance product quality while lowering production costs through the reduction of defective items. The main tools employed in SPC include histograms, check sheets, Pareto charts, cause-and-effect diagrams, defect concentration diagrams, scatter plots, and Shewhart control charts, each playing a distinct role in diagnosing and addressing quality-related problems.

In manufacturing systems, it is well recognized that not all units produced are identical with respect to their quality characteristics. Such variability is natural and may result from several factors, including fluctuations in raw material properties, differences in operator performance, variations in machine condition, changes in production methods, managerial practices, and environmental influences. Together, these elements determine the overall stability and performance of the process. Process variability is generally classified into two categories: common (or chance) causes and assignable causes. Common causes represent the inherent variability of a process, which tends to be stable and predictable over time. Assignable causes, in contrast, stem from unusual disturbances such as tool wear, operator mistakes, or irregularities in raw materials. The effective application of SPC requires the ability to distinguish between these two types of variation maintaining process stability in the presence of common causes while identifying and eliminating assignable causes when they occur.

1.3 Control Charts

The control chart is a foundational tool in quality engineering, providing a rigorous statistical framework for distinguishing between natural process variation and variation caused by assignable factors. Developed by Walter A. Shewhart at Bell Telephone Laboratories in the 1920s, the control chart transformed quality management from end-product inspection to proactive process regulation through statistical reasoning (Shewhart, 1931).

A control chart visualizes process performance over time, signalling deviations that may compromise product quality. Let w denote a sample statistic of a quality characteristic, with mean μ_w and standard deviation σ_w. The central line (CL) is set at $\mu_w$while the upper and lower control limits are set at μ_w+$d\sigma_w$ (UCL) and μ_w-$d\sigma_w$ (LCL), respectively, where d specifies the number of standard deviations from the mean (Ryan, 2011; Oakland, 2014). Observations within these limits reflect common-cause variation, whereas points outside or exhibiting systematic patterns indicate potential assignable causes requiring corrective action (Wheeler & Chambers, 1992).

Control charts are broadly categorized into variable charts, for continuously measured characteristics, and attribute charts, for discrete or conforming/nonconforming data. Variable charts, such as $\bar{X}$ and R charts, provide detailed insight into both process location and dispersion (Ryan, 2011). Attribute charts, including np, p, c, and u charts, rely on binomial or Poisson models and are well-suited for contexts where precise measurements are impractical.

While traditional Shewhart charts are effective for detecting large process shifts, their reliance on current-sample data limits sensitivity to small or gradual changes. Memory-type charts, such as the CUSUM (Page, 1954) and EWMA (Roberts, 2000), incorporate historical data to enhance responsiveness to subtle and sustained deviations (Woodall, 2000; Woodall & Montgomery, 1999). These approaches have become essential in modern quality control, particularly in high-precision or automated manufacturing environments.

In essence, control charts have evolved from simple diagnostic graphs into strategic instruments for continuous process improvement, integrating historical data, advanced monitoring techniques, and statistical rigor to maintain process stability and optimize product quality (Ryan, 2011; Woodall & Montgomery, 2014).

1.4 Univariate Control Charts

Univariate control charts are among the earliest and most extensively applied tools in Statistical Process Control, intended for monitoring a single quality characteristic at a time. The concept was pioneered by Shewhart (Shewhart, 1931), who proposed the use of statistical limits to distinguish between natural variability inherent in a process and unusual deviations attributable to special causes. These charts remain simple, practical, and highly effective in identifying substantial shifts in process variables, which accounts for their ongoing popularity in industrial practice (Montgomery, 2020).

Despite these advantages, a key limitation of univariate charts is the assumption that quality characteristics can be assessed independently. In situations where multiple process variables are correlated, monitoring them separately may yield incomplete insights or even misleading interpretations of process performance (Alt, 1985).

1.5 Multivariate Control Charts

With advances in real-time computing and modern data acquisition systems, it has become standard practice in many industries to evaluate multiple quality characteristics simultaneously rather than in isolation. For instance, in chemical production, process quality may depend jointly on variables such as temperature, pressure, and flow rate, making joint monitoring essential for ensuring consistency (Jackson, 1985). In such cases, multivariate control charts provide an advantage because they explicitly account for the correlations among process variables that univariate methods neglect.

Montgomery and Mastrangelo (1991) showed that applying separate univariate charts in a multivariate setting inflates the overall probability of false alarms (Type I errors) and weakens the ability to correctly identify in-control processes. Similarly, Lowery and Montgomery (1995) demonstrated that multivariate charts typically have greater sensitivity for detecting process shifts compared with their univariate counterparts. Additional studies (Woodall & Ncube, 1985; Mason *et al*., 1995) emphasized the importance of modelling interdependencies, noting that multivariate approaches not only improve detection capability but also provide a more realistic assessment of overall process performance. Collectively, these contributions underscore the essential role of multivariate control charts in contemporary quality control practice.

Beyond theoretical advantages, multivariate control charts have found applications across diverse domains. Early industrial applications, such as photographic processing, demonstrated the effectiveness of these methods in tracking several interrelated process variables, facilitating timely detection and correction of subtle process drifts (Jackson & Morris, 1957). In modern manufacturing, multivariate approaches support on-line monitoring of chemical processes, providing rapid identification of shifts and enabling root-cause analysis for process optimization (Chiang & Colegrove, 2007). Similarly, polymer production benefits from integrating infrared spectroscopy with multivariate quality-control methods, permitting real-time monitoring of chemical composition and process-induced variations to maintain product consistency (Bodecchi *et al.*, 2005).

Multivariate monitoring is also applied in healthcare, where tracking multiple service indicators in antenatal care helps identify systemic quality drops and guides timely interventions (Anggraini, 2021). In laboratory settings, multivariate monitoring allows simultaneous tracking of multiple correlated quality-control measurements, enabling early detection of systematic deviations that may not be evident with traditional univariate approaches (Ialongo, 2025). In consumer product evaluation, interdependent attributes such as taste, texture, and aroma are jointly monitored, improving product quality assessment (Ennis, 2000). Foundational principles, including design considerations, covariance estimation, and handling of correlated data, provide the theoretical framework necessary for successful implementation across industrial and service domains (Jackson, 1985). Collectively, these studies highlight the versatility, robustness, and practical impact of multivariate control charts in supporting comprehensive quality management and informed decision-making.

1.6 Effect of Outliers on Multivariate Control Charts

Outliers are statistically defined as values that belong to a different population because they originate from a different process or source, effectively being derived from one or more contaminating distributions (Hampel *et al*., 1986). Such observations often indicate potential errors, rare events, or phenomena governed by different processes. Identification of outliers in datasets plays a crucial role in statistical data analysis, driven by two primary motivations. Firstly, outliers themselves may carry significant information. Secondly, the presence of outliers can exert influence on the outcomes derived from the remaining data (Barnett & Lewis 1994). Detecting outliers in multivariate data presents more challenges compared to univariate data, as visually spotting multivariate outliers is difficult in most of the situations due to their lack of conspicuousness at the extremes (Gnanadesikan & Kettenring, 1972)
Outliers and departures from normality can severely distort the estimation of mean vectors and covariance matrices, resulting in unreliable control limits and misleading monitoring outcomes (Rocke & Woodruff, 1996; Rousseeuw & Hubert, 2011). Even a small proportion of atypical observations may inflate false alarms or mask genuine process shifts, compromising the effectiveness of classical multivariate control charts. This issue becomes particularly critical in high-dimensional settings, where correlations among variables amplify the impact of outliers on monitoring statistics (Hubert *et al*., 2015; Maronna *et al*., 2006).

To address these challenges, several strategies have been proposed, including nonparametric methods, data transformations, and resampling techniques. Among these, robust estimation-based methods have proven especially effective, as they mitigate the influence of outliers while preserving the accuracy of multivariate location and dispersion estimates. Robust estimators, such as high-breakdown alternatives for the mean vector and covariance matrix, ensure that monitoring performance remains reliable even in contaminated datasets (Maronna *et al*., 2006; Hubert *et al*., 2008).

1.7 Objectives of the Study

- To review available robust control charts for process location and dispersion
- To propose new robust control charts for monitoring the location parameter of bivariate process
- To propose an improved method for the estimation of location and dispersion.
- To propose robust multivariate control chart for process location
- To propose robust multivariate control chart for process dispersion
- To apply the proposed methods in real life data sets.

1.8 Organization of the Thesis

- This thesis is organized into eleven chapters, each addressing specific aspects of robust statistical process control and contributing new methodologies.
- Chapter 2 reviews existing robust estimators of multivariate location and scatter and proposes an improved robust covariance estimator with an accompanying outlier detection method. Its performance is evaluated through simulations and real-world datasets, demonstrating high accuracy, low false detection rates, and robustness across diverse data scenarios, including high-dimensional and small sample settings. The findings of this study have been published as:
  Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2025). Unveiling outliers with robust covariance matrix estimation: A shrinkage approach. *Statistical Papers, 66*(4).
- Chapter 3 proposes a robust bivariate control chart for process location based on individual observations, highlighting its advantages over classical methods. This work has been published as:
  Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2024). Bivariate robust control charts for individual observations. *Communications in Statistics – Simulation and Computation, 53*(5), 1234–1256.
- Chapter 4 proposes a robust multivariate control chart for process location based on individual observations, thereby extending the bivariate approach to higher dimensions. This research has been published as:
  Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2023). Enhancing multivariate control charts for individual observations using ROC estimates. *Stochastics and Quality Control, 38*(2), 109–118.

- Chapter 5 proposes a robust bivariate control chart for process location using rational subgroups, ensuring improved performance in grouped data settings. The findings of this research have been published as:
  Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2025). Robust bivariate control chart for process monitoring with rational subgroups. *Quality and Reliability Engineering International*. **https://doi.org/10.1002/qre.70022**
- Chapter 6 extends the idea of bivariate control chart to the multivariate setting by proposing a robust multivariate control chart for process location using rational subgroups.
- Chapter 7 focuses on process variability and proposes a robust bivariate dispersion control chart, designed to detect shifts in process spread.
- Chapter 8 generalizes the above to higher dimensions and proposes a robust multivariate dispersion control chart, ensuring reliable monitoring of multivariate process dispersion.
- Chapter 9 demonstrates the industrial application of the proposed robust control charts and evaluates their performance using real process data.
- Chapter 10 concludes the thesis by summarizing the major contributions, discussing practical implications.
- Chapter 11 provides recommendations based on the findings of this research and suggests directions for future studies in robust multivariate process monitoring. It highlights potential extensions to subgroup data, high-dimensional processes, mixed-type quality characteristics, and adaptive industrial applications.

CHAPTER 2

ROBUST ESTIMATION OF LOCATION VECTOR AND DISPERSION MATRIX

2.1 Introduction

Considering the location vector and the dispersion matrix are the cornerstones of multivariate statistical inference, reliable and effective techniques for estimating these quantities are indispensable for drawing valid conclusions from multivariate data. Under the presumption of multivariate normality, classical estimators such as, arithmetic mean vector and variance covariance matrix are optimal (Anderson, 2003). Nevertheless, the classical estimators are well-known for being sensitive to deviations from normality, especially in the presence of outliers (Maronna *et al.*, 2006). A single anomalous observation has the potential of significantly skew the mean vector or inflate the covariance matrix, making subsequent inferences invalid. This susceptibility emphasizes the importance of robust estimators that maintain efficiency under normality while showing resilience against contamination in the data.

Many methods have been proposed in the literature for the robust estimation of location and dispersion parameters. Some of the most important and recent estimators are reviewed in the following section. Despite being robust, many of these methods remain computationally expensive and often entail trade-offs between breakdown point and statistical efficiency. This paper proposes a robust shrinkage estimate for the dispersion matrix, based on the covariance estimator of Gnanadesikan and Kettenring (1972). An outlier identification process is built utilising this estimator. The suggested method's characteristics are empirically investigated, and its effectiveness is compared with established methods that have been discussed in the literature.

The remainder of this chapter is organized as follows: Section 2.2 presents the robustness concepts including breakdown point, influence function and affine equivariance. Section 2.3 provides a detailed review of robust estimators of multivariate location. Section 2.4 discusses robust estimators of dispersion, beginning with bivariate methods before extending to general multivariate methods. Section 2.5 describes the proposed shrinkage-based method for robust estimation of dispersion matrix, outlining its theoretical properties and computational procedure. Section 2.6 presents the results of a simulation study, while Section 2.7 demonstrates the application of the method to real life data. Finally, Section 2.8 summarizes the discussion.

2.2 Robustness Concepts

2.2.1 Breakdown point

The breakdown point is a key metric used to quantify the robustness of an estimator against the influence of outliers. Following Hodges (1967), Hampel (1968) and Hampel (1971), breakdown point of an estimator is defined as the maximum proportion of arbitrary contaminating observations a dataset can contain before the estimator's value becomes unbounded. Lopuhaä and Rousseeuw (1991) offered a formal and precise definition of the breakdown point for both location and covariance estimators. For a location estimator $\hat{\boldsymbol{\mu}}$ based on a set of observations $\boldsymbol{X}$, the breakdown point $\varepsilon_n^*(\hat{\boldsymbol{\mu}}, \boldsymbol{X})$ is defined as:

$$\varepsilon_n^*(\hat{\boldsymbol{\mu}}, \boldsymbol{X}) = \min_m \left\{ \frac{m}{n}; \sup_{\tilde{X}} \left\| \hat{\boldsymbol{\mu}}\left(\tilde{\boldsymbol{X}}\right) - \hat{\boldsymbol{\mu}}\left(\boldsymbol{X}\right) \right\| = \infty \right\}, \tag{2.1}$$

where $\tilde{\boldsymbol{X}}$ denotes the dataset obtained by replacing m observations in $\boldsymbol{X}$ with arbitrary values. From equation (2.1), it follows that the breakdown point of a location estimator corresponds to the smallest fraction of observations that must be contaminated for the estimator to deviate without bound from the true sample mean.

The breakdown point of a covariance estimator $\hat{\boldsymbol{\Sigma}}$ is formally defined as:

$$\varepsilon_n^*(\hat{\boldsymbol{\Sigma}}, \boldsymbol{X}) = \min_m \left\{ \frac{m}{n}; \sup_{\tilde{X}} D\left(\hat{\boldsymbol{\Sigma}}\left(\tilde{\boldsymbol{X}}\right) - \hat{\boldsymbol{\Sigma}}\left(\boldsymbol{X}\right) \right) = \infty \right\},$$

(2.2)

where $D(\mathbf{A}, \mathbf{B}) = \max\{|\lambda_1(\mathbf{A}) - \lambda_1(\mathbf{B})|, |\lambda_p(\mathbf{A})^{-1} - \lambda_p(\mathbf{B})^{-1}|\}$, and $\lambda_i(\mathbf{A})$ is the i^{th} ordered eigen value of $\mathbf{A}$. In other words, Equation (2.2) indicates that the breakdown point of a covariance estimator is the smallest fraction of observations that must be contaminated for the estimator to become unreliable either by causing the largest eigenvalue of the estimated covariance matrix to increase without bound or by reducing the smallest eigenvalue to a value arbitrarily close to zero. In the context of estimating the mean vector and covariance matrix, it is desirable to employ estimators with a high breakdown point, ideally approaching the theoretical maximum of 50% (Rousseeuw & Leroy 2003).

In contrast, the classical sample mean and covariance estimators have breakdown points of only $1/n$, where n is the sample size (Donoho & Huber, 1983). Consequently, even a single outlier can result in unbounded estimates, as described by Equations (2.1) and (2.2).

2.2.2 Influence function

The influence function (IF), introduced by Hampel (1974), is a key concept in robust statistics used to assess the sensitivity of an estimator to small changes or contaminations in the data. It measures how a slight perturbation at a single observation affects the value of an estimator, providing insight into its robustness.

For a multivariate estimator $\theta_\infty(F)$ of a p dimensional parameter based on an underlying distribution F, the influence function at a point $x_0 \in R^p$ is defined as

$$IF_\theta(x_0, F) = \lim_{\varepsilon \downarrow 0} \frac{\theta_\infty\left((1-\varepsilon)F + \varepsilon\delta_{x_0}\right) - \theta_\infty(F)}{\varepsilon},$$

where δ_{x_0} denotes the point-mass distribution concentrated at x_0, and ε represents a small fraction of contamination added at that point.

In simpler terms, the influence function describes how the estimator $\theta_\infty(F)$ changes when the data distribution F is slightly "contaminated" by placing a very small proportion of observations at the point x_0. The larger the value of $IF_\theta(x_0, F)$, the more sensitive the estimator is to contamination at that point.

If the influence function of an estimator is bounded, then small perturbations in the data produce only limited changes in the estimate, which is a desirable property for robustness (Hubert *et al.*, 2008). On the other hand, unbounded influence functions indicate that even a single outlier can have an arbitrarily large effect on the estimator, as is the case for classical estimators like the sample mean and covariance matrix (Maronna *et al.*, 2006).

2.2.3 Affine equivariance

Another key property of a robust estimator is affine equivariance. A location estimator $\hat{\boldsymbol{\mu}} \in \boldsymbol{R}^p$ is affine equivariant if for any non-singular $p \times p$ matrix $\boldsymbol{A}$ and any vector $\boldsymbol{b} \in \boldsymbol{R}^p$,

$$\hat{\boldsymbol{\mu}}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\,\hat{\boldsymbol{\mu}}(\boldsymbol{X}) + \boldsymbol{b}. \tag{2.3}$$

A scale estimator $\hat{\boldsymbol{\Sigma}}$, taking values in the set of positive-definite symmetric $p \times p$ matrices, is said to be affine equivariant if, for any non-singular $p \times p$ matrix $\boldsymbol{A}$ and vector $\boldsymbol{b} \in \boldsymbol{R}^{\boldsymbol{p}}$ it satisfies

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{AX}+\boldsymbol{b}) = \boldsymbol{A}\hat{\boldsymbol{\Sigma}}(\boldsymbol{X})\boldsymbol{A}^T. \qquad (2.4)$$

For an affine equivariant estimator, linear transformations of the data, such as scaling, rotation, or translation, do not affect the resulting estimate. Removing this property substantially enlarges the class of available estimators, and in many practical cases, non-affine equivariant estimators can demonstrate superior robustness and efficiency compared to their affine equivariant counterparts.

Beyond considerations of estimator breakdown, the phenomenon of outlier masking further highlights the need for robust methods in detecting multidimensional outliers. Masking occurs when extreme outliers distort non-robust estimates of the mean and covariance to such an extent that less extreme outliers appear typical when evaluated using Mahalanobis distances. For instance, if there is one or more highly distant outliers and one or more moderately distant outliers in the same direction, the extreme outlier(s) can shift the mean substantially and inflate the standard deviation, causing the less extreme outlier(s) to fall within approximately 2 to 3 standard deviations from the mean and thereby remain undetected. The severity of masking is reflected in an increase in Type II errors, as truly outlying observations are mistakenly classified as part of the uncontaminated population.

2.3 Robust Estimators of Multivariate Location

Robust estimation of multivariate location is central to multivariate analysis, particularly when datasets contain outliers or deviate from Gaussian assumptions. Unlike the classical sample mean vector, which is highly sensitive to contamination, robust estimators provide measures of central tendency that retain validity under data contamination. Several alternatives exist, differing in robustness, affine equivariance, and computational feasibility. In this section, we review five key estimators: the component-wise median, spatial (L_1) median, Oja median, Tukey depth median, and projection depth median.

2.3.1 Component-wise median

The component-wise median is obtained by computing the univariate median for each component of the data vector independently. For a dataset $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n \in \mathbb{R}^p$, the component-wise median is defined as

$$\hat{\theta}_j = median(X_{1j}, X_{2j}, \dots, X_{nj}), j = 1,2, \dots, p,$$

and the estimator is given by

$$\widehat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p\right)^T.$$

This estimator is highly robust in each marginal distribution but is not affine equivariant, meaning it may not appropriately account for correlations among variables (Lopuhaä, & Rousseeuw, 1991). Nevertheless, its computational simplicity and strong marginal robustness make it attractive in practice.

2.3.2. Spatial (L_1) Median

The spatial median (also known as the geometric median) generalizes the univariate median by minimizing the sum of Euclidean distances:

$$\widehat{\boldsymbol{\theta}} = arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\theta}\|.$$

Equivalently, it satisfies the first-order condition:

$$\sum_{i=1}^{n} \frac{\boldsymbol{X}_i - \widehat{\boldsymbol{\theta}}}{\|\boldsymbol{X}_i - \widehat{\boldsymbol{\theta}}\|} = 0.$$

This estimator is affine equivariant and possesses a breakdown point of 50%, making it robust to extreme values (Kemperman, 1987). Compared to the coordinate wise median, it better captures multivariate dependence, though it is computationally more demanding in higher dimensions.

2.3.3 Oja median

The Oja median (Oja, 1983) is defined as the point minimizing the sum of simplex volumes formed with the data points. For $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n \in \mathbb{R}^p$, the Oja median is

$$\widehat{\boldsymbol{\theta}} = arg\,\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \sum_{1\le i_1\le\cdots\le i_p\le n} Vol\left(X_{i_1}, X_{i_2} \dots X_{i_p}, \boldsymbol{\theta}\right),$$

where *Vol* denotes the volume of the simplex spanned by p data points and $\boldsymbol{\theta}$.

The Oja median is affine equivariant and provides a natural multivariate extension of the univariate median. However, computational complexity grows rapidly with sample size and dimension, limiting its applicability to small-to-moderate datasets.

2.3.4 Tukey depth median

The Tukey depth median (also called the half space median) (Tukey, 1975) maximizes the Tukey depth function:

$$D(\boldsymbol{\theta}, \boldsymbol{X}) = \inf_{\boldsymbol{u}\in\mathbb{S}^{p-1}} P(\boldsymbol{u}^T\boldsymbol{X} \le \boldsymbol{u}^T\boldsymbol{\theta}),$$

where $\mathbb{S}^{p-1}$ is the unit sphere in $\mathbb{R}^p$. The Tukey median is defined as the point with maximum depth:

$$\widehat{\boldsymbol{\theta}} = arg\,\max_{\boldsymbol{\theta}\in\mathbb{R}^p} D(\boldsymbol{\theta}, \boldsymbol{X}).$$

It is affine equivariant and achieves the highest possible breakdown point (50%) (Donoho & Gasko, 1992). However, exact computation is only feasible in low dimensions ($p \le 3$).

2.3.5 Projection depth median

The projection depth median (Zuo, 2003) is based on projection pursuit. For any point $\boldsymbol{\theta} \in \mathbb{R}^p$, its projection depth is

$$PD(\boldsymbol{\theta}, \boldsymbol{X}) = \left[1 + \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \frac{|\boldsymbol{u}^T\boldsymbol{\theta} - med(\boldsymbol{u}^T\boldsymbol{X})|}{MAD(\boldsymbol{u}^T\boldsymbol{X})}\right]^{-1},$$

where *med* and *MAD* denote the univariate median and median absolute deviation, respectively. The projection depth median is then

$$\widehat{\boldsymbol{\theta}} = arg\,\max_{\boldsymbol{\theta}\in\mathbb{R}^p} PD(\boldsymbol{\theta}, \boldsymbol{X}).$$

This approach balances robustness and computational tractability, offering a practical alternative to Tukey's median in moderate dimensions.

Each of these estimators has distinct strengths. The component-wise median is simple and robust but lacks affine equivariance, while the spatial median improves on this at a moderate computational cost. The Oja, Tukey, and projection depth medians are theoretically elegant and robust but computationally intensive, particularly in higher dimensions. In applied research, the component-wise median and spatial median often provide the best trade-off between robustness and feasibility. This justifies their adoption in the present study as reliable measures of multivariate location.

2.4 Robust Estimators for Dispersion

2.4.1 Robust bivariate estimators for dispersion

2.4.1.1 Comedian estimator

Falk (1997) introduced a robust measure of scatter known as *Comedian*, for bivariate data. Let X and Y be two random variables then the comedian of X and Y is defined as

$$\mathrm{COM}(X,Y) = med\{(X - med(X))(Y - med(Y))\}, \tag{2.5}$$

where '*med*' denotes median. It generalizes the median absolute deviation (MAD) as it equals MAD^2 when $X = Y$ and also has the highest possible breakdown point. COM (X, Y) parallels COV(X, Y), but COV(X, Y) requires the existence of the first two moments of X and Y, whereas COM(X, Y) always exists. The comedian is symmetric, location invariant and scale equivariant. Hall and Welsh (1985) discussed about the strong consistency and asymptotic normality of MAD and Falk (1997) established similar results for comedian.

2.4.1.2 S_n covariance

More efficient alternative for MAD with 50% breakdown point is discussed by Rousseeuw and Croux (1993). A pairwise distance estimator

$$S_n(X) = 1.1926\, \underset{l}{med}\, \underset{s}{med}\, |x_l - x_s|; \text{ for } l, s = 1, \ldots, n. \tag{2.6}$$

S_n estimator of scale assures bounded influence function and optimal breakdown point. Sajana and Sajesh (2020) proposed a bivariate extension of S_n estimator as a robust alternative for covariance estimator, S_n Covariance, as:

$$S_nCov(X,Y) = 1.4223\, \underset{l}{med} \left\{ \underset{s \neq l}{med} [(x_l - x_s)(y_l - y_s)] \right\}; \text{ for } l, s = 1, \ldots, n. \tag{2.7}$$

The S_n Covariance estimator satisfies the characteristics of sample covariance in independent and symmetric random variables. This estimator assures location invariant and scale equivariant properties as well.

2.4.1.3 Gnanadesikan – Kettenring (GK) estimator

Gnanadesikan and Kettenring (1972) proposed a covariance estimator defined as:

$$\mathrm{Cov}(X, Y) = \frac{1}{4}(\sigma(X + Y)^2 - \sigma(X - Y)^2),$$

where σ is the standard deviation, and X and Y are two random variables. To enhance its robustness, the GK estimator can be modified by substituting σ with a robust scale estimator. This robust version, known as the robust GK estimator, is expressed as:

$$\mathrm{Cov}_{\mathrm{R}}(X, Y) = \frac{1}{4}(\sigma_R(X + Y)^2 - \sigma_R(X - Y)^2), \tag{2.8}$$

where σ_R represents a robust scale estimator.

2.4.2 Robust multivariate estimators for dispersion

Maronna (1976) delved into affine equivariant M estimators for covariance matrices, while Campbell (1980) suggested employing M estimators for both mean and covariance matrix to study Mahalanobis distance. Stahel (1981) raised concerns about the low breakdown point of multivariate M-estimators for large p, prompting the proposal of the Stahel-Donoho estimator, a weighted mean and covariance based on outlyingness proposed by Stahel (1981) and Donoho (1982) independently. Although robust, computational intensity of Stahel-Donoho estimator limits its practicality (Tyler 1994; Maronna and Yohai, 1995). Rousseeuw (1984) introduced the Minimum Covariance Determinant (MCD) estimator with a high breakdown value, albeit computationally expensive for large p. Rousseeuw and Van Driessen (1999) addressed this by proposing FAST-MCD, an efficient algorithm for MCD computation. However, even FAST-MCD requires substantial running time for large p.

Pena and Prieto (2001) introduced a fast outlier-detection method and a robust estimator for the covariance matrix based on projections maximizing and minimizing the kurtosis coefficient. Unlike some methods, this approach does not necessitate subsampling. Maronna and Zamar (2002) proposed an Orthogonalized Gnanadesikan-Kettenring (OGK) estimator by a general method to obtain positive-definite and approximately affine-equivariant robust scatter matrices starting from any pair-wise robust scatter matrix. In a similar way, Sajesh and Srinivasan (2012) proposed the Reweighted Orthogonalized Comedian estimator (ROC), which is constructed based on the Comedian estimate of Falk (1997). More recently Sajana and Sajesh (2022) proposed multivariate S_n covariance estimator, developed from S_n Covariance originally proposed by Sajana and Sajesh (2020). Although these estimates are not affine equivariant, they perform very well under high collinearity. The reweighing step can be used to identify the outliers, where outliers get weight 0 and normal observations get 1. Cabana *et al.* (2021) proposed a multivariate outlier detection approach using shrinkage comedian estimator (RMD-S). The author demonstrated that the shrinkage estimator has a high breakdown value even in high dimension and a high success rate in outlier detection.

This chapter introduces a robust shrinkage estimate for the dispersion matrix, based on the covariance estimator of Gnanadesikan and Kettenring (1972). An outlier identification process is built utilising this estimator. The suggested method's characteristics are empirically investigated, and its effectiveness is compared with established methods that have been discussed in the literature.

2.5 Robust Shrinkage Estimator for Dispersion Matrix

As described in Section 2.4.1.3, covariance estimation between two random variables Y_1 and Y_2 has traditionally relied on location-based methods. To overcome the sensitivity of such approaches, Gnanadesikan and Kettenring (1972) proposed a scale-based formulation, which employs measures of scale and provides improved robustness in capturing variability compared to the classical covariance estimator. A robust version of this estimator can be obtained by replacing the standard deviation with a robust scale estimator σ_R, yielding

$$C_{\mathrm{R}}(Y_1, Y_2) = \frac{1}{4}(\sigma_R(Y_1 + Y_2)^2 - \sigma_R(Y_1 - Y_2)^2). \quad (2.9)$$

The key advantage of the scale-based approach is that the influence function of the covariance estimator depends solely on that of the scale estimator. Additionally, the breakdown point of this covariance estimator matches that of the robust scale estimator used in Equation (2.9), ensuring strong robustness properties (Genton & Ma., 1999).

The proposed approach uses the multivariate version of the robust C_{R} estimator for the detection of outliers. Let $\mathbf{Y}$ be a $n \times p$ data matrix with rows $\boldsymbol{y}_i^{\mathrm{T}}(i = 1, 2, \ldots, n)$ and columns $\boldsymbol{Y}_j$ ($j = 1, 2, \ldots, p$). Then the robust Gnanadesikan-Kettenring (GK) estimator for dispersion matrix is defined as:

$$\mathbf{C}_{GK}(\mathbf{Y}) = \left(C_R\left(\boldsymbol{Y}_i, \boldsymbol{Y}_j\right)\right), \; i, j = 1,2, \ldots p \tag{2.10}$$

Note that $\boldsymbol{C}_{GK}$ is a robust alternative for the covariance matrix, in which the highly robust dispersion estimator stands out as a better choice for a robust scale estimator, featuring a high breakdown point and a bounded influence function but in general, it is not positive semi definite (Maronna & Zamar 2002). To overcome the non-positive semi-definiteness of the estimator we adopt the shrinkage approach as follows:

1) Compute the eigenvalues τ_i and the eigenvectors e_i of $\mathbf{C}_{GK}(\mathbf{Y})$, $i = 1,2, \ldots p$, and call $\mathbf{E}$ the matrix whose columns are the e_i's, so that $\mathbf{C}_{GK}(\mathbf{Y}) = \mathbf{EDE}^{\mathrm{T}}$, where $\mathbf{D} = \mathrm{diag}(\tau_1, \tau_2, \ldots \tau_p)$.
2) Compute $\mathbf{D}^* = (1 - \lambda)\mathbf{D} + \lambda\mathbf{T}$; where $\mathbf{T}$ is the shrinkage target and λ is the shrinkage intensity. Several choices have been proposed in the literature for the shrinkage target $\mathbf{T}$ (Ledoit & Wolf, 2003a; Ledoit & Wolf, 2003b; Ledoit & Wolf, 2004; Schäfer & Strimmer, 2005; DeMiguel *et al*., 2013). This study makes use of the shrinkage target, $\vartheta_\Sigma \mathbf{I}$, proposed by Ledoit and Wolf (2004).
3) Compute an improved GK dispersion estimator by shrinkage as,

$$\mathbf{C}_{SGK} = \mathbf{ED}^*\mathbf{E}^{\mathrm{T}}, \tag{2.11}$$

where $\mathbf{D}^* = (1 - \lambda)\mathbf{D} + \lambda\vartheta_\Sigma\mathbf{I}$. This estimator is named as Shrinkage Gnanadesikan-Kettenring (SGK) estimator. Ledoit and Wolf (2004) suggested to estimate ϑ_Σ and λ by minimizing the risk (expected quadratic loss) $\mathrm{E}[\| \mathbf{C}_{SGK} - \boldsymbol{\Sigma} \|^2]$, where $\boldsymbol{\Sigma}$ is the variance covariance matrix of $\mathbf{Y}$ and $\| \mathbf{B} \|^2 = \frac{trace(\mathbf{BB}^{\mathrm{T}})}{p}$.

Theorem: The solution of λ and ϑ_Σ which minimizes $\mathrm{E}[\| \mathbf{C}_{SGK} - \boldsymbol{\Sigma} \|^2]$ are

$$\vartheta_\Sigma = \frac{Trace(\boldsymbol{\Sigma})}{p} \text{ and } \lambda = \frac{\mathrm{E}\|\boldsymbol{C}_{GK} - \boldsymbol{\Sigma}\|^2}{\mathrm{E}[\|\boldsymbol{C}_{GK} - \boldsymbol{\Sigma}\|^2] \; + \|\vartheta_\Sigma \boldsymbol{I} - \boldsymbol{\Sigma}\|^2}.$$

Proof

The Solution of λ and ϑ_Σ which minimizes $\mathrm{E}[\| \mathbf{C}_{SGK} - \boldsymbol{\Sigma} \|^2]$ are

$\vartheta_{\Sigma} = \frac{Trace(\boldsymbol{\Sigma})}{p}$ and $\lambda = \frac{\mathrm{E}\|\boldsymbol{C}_{GK}-\boldsymbol{\Sigma}\|^2}{\mathrm{E}[\|\boldsymbol{C}_{GK}-\boldsymbol{\Sigma}\|^2] \ +\|\vartheta_{\Sigma}\boldsymbol{I}-\boldsymbol{\Sigma}\|^2}$.

$$min_{\vartheta_{\Sigma},\lambda}\, \mathrm{E}[\| \boldsymbol{C}_{SGK} - \boldsymbol{\Sigma} \|^2] = min_{\vartheta_{\Sigma},\lambda} \{\mathrm{E}[\| (1-\lambda)\boldsymbol{C}_{GK} + \lambda\vartheta_{\Sigma}\boldsymbol{I} - \boldsymbol{\Sigma} \|^2]\}$$

$$= min_{\vartheta_{\Sigma},\lambda} \{(1-\lambda)^2\mathrm{E}[\| \boldsymbol{C}_{GK} - \boldsymbol{\Sigma} \|^2] + \lambda^2 \| \vartheta_{\Sigma}\boldsymbol{I} - \boldsymbol{\Sigma} \|^2$$

$$+2\mathrm{E}[\langle(1-\lambda)(\boldsymbol{C}_{GK} - \boldsymbol{\Sigma}), \lambda\langle\vartheta_{\Sigma}\boldsymbol{I}, \boldsymbol{\Sigma}\rangle \ \rangle\}$$

$$= min_{\vartheta_{\Sigma},\lambda} \{(1-\lambda)^2\mathrm{E}[\| \boldsymbol{C}_{GK} - \boldsymbol{\Sigma} \|^2] + \lambda^2 \| \vartheta_{\Sigma}\boldsymbol{I} - \boldsymbol{\Sigma} \|^2\}.$$

Note that $\mathrm{E}(\mathbf{C}_{GK}) = \boldsymbol{\Sigma}$ obtained by using robust unbiased scale estimator. To obtain the optimal value ϑ_{Σ} it is sufficient to minimize the term $\| \vartheta_{\Sigma}\mathbf{I} - \boldsymbol{\Sigma} \|^2$ of the above expression.

$$\| \vartheta_{\Sigma}\mathbf{I} - \boldsymbol{\Sigma} \|^2 = \vartheta_{\Sigma}^2 \| \mathbf{I} \|^2 + \| \boldsymbol{\Sigma} \|^2 - 2\vartheta_{\Sigma}\langle \boldsymbol{I}, \boldsymbol{\Sigma}\rangle.$$

The first order optimality condition for obtaining the scaling factor ϑ_{Σ} is

$$2\vartheta_{\Sigma} \| \mathbf{I} \|^2 - 2\langle \mathbf{I}, \boldsymbol{\Sigma} \rangle = 0,$$

$$=> \ \vartheta_{\Sigma} = \frac{\langle \mathbf{I}, \boldsymbol{\Sigma}\rangle}{\langle \mathbf{I}, \mathbf{I}\rangle} = Trace(\boldsymbol{\Sigma}\, \mathbf{I}^{\mathrm{T}})/p.$$

Therefore, the optimal value ϑ_{Σ} is $\frac{Trace(\boldsymbol{\Sigma})}{p}$.

The first order optimality condition for obtain the optimal shrinkage intensity is

$$2(1-\lambda)\mathrm{E}(\| \mathbf{C}_{GK} - \boldsymbol{\Sigma} \|^2)(-1) + \ 2\lambda \| \vartheta_{\Sigma}\mathbf{I} - \boldsymbol{\Sigma} \|^2 = 0.$$

This implies,

$$\lambda = \frac{\mathrm{E}\|\mathbf{C}_{GK}-\boldsymbol{\Sigma}\|^2}{\mathrm{E}[\|\mathbf{C}_{GK}-\boldsymbol{\Sigma}\|^2] \ +\|\vartheta_{\Sigma}\mathbf{I}-\boldsymbol{\Sigma}\|^2}.$$

The shrinkage intensity of the covariance matrix refers to the proportional expected reduction in loss for the $\mathbf{C}_{GK}$ covariance matrix compared to the combined expected loss for both the $\mathbf{C}_{GK}$ covariance matrix and the scaled identity matrix.

Since the direct computation of $\hat{\vartheta}_{\Sigma}$ and $\hat{\lambda}$ depends on the true covariance matrix, Ledoit and Wolf (2004) suggested to compute these estimates as follows:

$$\hat{\vartheta}_{\Sigma} = \frac{Trace(\mathbf{C}_{GK})}{p} \text{ and } \hat{\lambda} = \frac{b_n^2}{d_n^2},$$

where $d_n^2 = \left\|\mathbf{C}_{GK} - \hat{\vartheta}_\Sigma \mathbf{I}\right\|^2$ and $b_n^2 = min\left(\sum_{i=1}^{n}\left\|\boldsymbol{y}_i\boldsymbol{y}_i^{\mathrm{T}} - \mathbf{C}_{GK}\right\|^2, d_n^2\right)$.

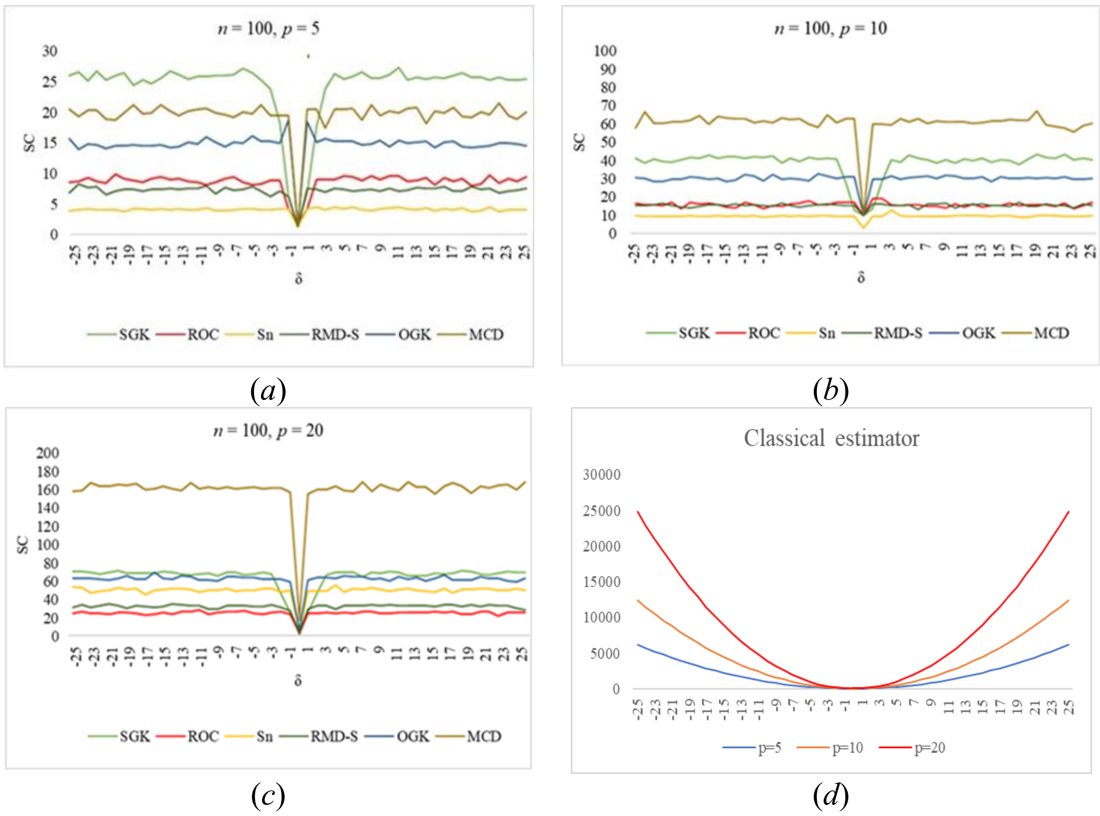
2.5.1 Influence function property of SGK estimator

The influence function of an estimator is an essential tool for evaluating its robustness properties. It serves as an asymptotic representation of its Sensitivity Curve (SC) which illustrates how minor changes in the data can impact the estimator. Let $\mathbf{Y}$ be a $n \times p$ data matrix with rows $\boldsymbol{y}_i^{\mathrm{T}}$(i = 1, 2, …, n) and $\mathbf{Y}_0$ be a $(n+1) \times p$ obtained by adding an outlying observation $\boldsymbol{y}_0$ to $\mathbf{Y}$. We define the sensitivity curve for the SGK estimator as

$$\mathrm{SC} = \|\mathbf{C}_{SGK}(\mathbf{Y}_0) - \mathbf{C}_{SGK}(\mathbf{Y})\|_F,$$

where $\| \, . \, \|_F$ is the Frobenius norm.

In order to draw the sensitivity curve, we have generated data from $N_p(\mathbf{0}, \boldsymbol{I})$ and added the extra outlier $\boldsymbol{y}_0 = \delta\boldsymbol{u}$, where $\boldsymbol{u}$ is a p dimensional vector of ones. Sensitivity curves for various combinations of n and p drawn for SGK, FAST-MCD, OGK, ROC, S_n, RMD-S and classical estimators for varying values of p are presented here.



(*a*) (*b*)

(*c*) (*d*)

Figure 2.1: Sensitivity curve for varying values of p

Figure 2.1 presents the sensitivity curve for n = 100 and p = 5, 10, 20. The sensitivity curve for robust estimators remains bounded across all dimensions, even as the contamination point (δ) moves far from the central region of the data. This indicates that these estimators are resistant to the influence of outliers. The sensitivity curves for classical estimator (shown in the bottom-right panel) grow unbounded as the δ moves farther from the central region of the data.

2.5.2 SGK method for detecting multiple outliers

The proposed robust covariance matrix can be used for the detection of multiple outliers from multidimensional data by defining a robust Mahalanobis distance as,

$$RD(\boldsymbol{y}_i, \boldsymbol{m}) = D_i = (\boldsymbol{y}_i - \boldsymbol{m})^{\mathrm{T}} \mathbf{C}_{SGK}^{-1} (\boldsymbol{y}_i - \boldsymbol{m}), i = 1, 2, \dots, n, \tag{2.12}$$

where $\mathbf{C}_{SGK}$ is defined in (3) and $\boldsymbol{m}$ is a robust location estimator. This study makes use of the shrinkage multivariate L_1 median discussed in Cabana *et al.* (2021) as location estimator which is defined as

$$\boldsymbol{m} = (1 - \alpha)\boldsymbol{L}_M + \alpha \nu_{\boldsymbol{\mu}} \boldsymbol{u},$$

where $\boldsymbol{L}_M$ is the L_1 median of $\boldsymbol{y}_i$ and $\boldsymbol{u}$ is p dimensional vector of ones. The scaling factor $\nu_{\boldsymbol{\mu}}$ and the intensity α should minimize the expected quadratic loss $\mathrm{E}[\| \boldsymbol{m} - \boldsymbol{\mu} \|_2^2]$ where $\| \boldsymbol{A} \|_2^2 = \sum_{j=1}^{p} a_j^2$. Cabana *et al.* (2021) obtained the estimates of $\nu_{\boldsymbol{\mu}}$ and α as

$$\hat{\nu}_{\boldsymbol{\mu}} = \frac{\boldsymbol{L}_M \boldsymbol{e}^{\mathrm{T}}}{p}; \ \hat{\alpha} = \frac{\mathrm{E}[\| \boldsymbol{L}_M - \boldsymbol{\mu} \|_2^2]}{\mathrm{E}[\| \boldsymbol{L}_M - \nu_{\boldsymbol{\mu}} \boldsymbol{e} \|_2^2]}.$$

To make these estimators more statistically efficient a weight function

$$w_i = \begin{cases} 1 \ if \ D_i \le cv \\ 0 \ if \ D_i > cv \end{cases}$$

is defined using the robust Mahalanobis distance in Equation (2.12) and a cut-off value

$$cv = \frac{\chi_p^2(0.95)\mathrm{median}\,(D_1, \dots, D_n)}{\chi_p^2(0.5)}$$

adopted from Maronna and Zamar (2002). Then the reweighted SGK estimators for location and scatter are defined as

$$\boldsymbol{m}_w = \frac{1}{N} \sum_i w_i \boldsymbol{y}_i \ \text{ and } \ \mathbf{C}_w = \frac{1}{N-1} \sum_i w_i (\boldsymbol{y}_i - \boldsymbol{m}_w)(\boldsymbol{y}_i - \boldsymbol{m}_w)^{\mathrm{T}},$$

where N is the sum of weights, $N = \sum_i w_i$. These re-weighted estimates have the same breakdown point of the raw estimates and better finite sample efficiency (Rousseeuw & Hubert, 2011). In order to identify the outliers, Robust Mahalanobis distance using these estimates are computed as

$$RD_w(\boldsymbol{y}_i, \boldsymbol{m}_w) = (\boldsymbol{y}_i - \boldsymbol{m}_w)^{\mathrm{T}} \mathbf{C}_w^{-1} (\boldsymbol{y}_i - \boldsymbol{m}_w), i = 1, 2, \ldots, n,$$

and if $RD_w(\boldsymbol{y}_i, \boldsymbol{m}_w) > \chi_p^2(0.975)$, the observation $\boldsymbol{y}_i$ may be considered as outlier.

2.6 Simulation Study

Efficiency of the SGK method to detect outliers is initially assessed through a simulation study by generating contaminated random samples from multivariate normal distribution and heavy tailed distributions, such as multivariate 't' and multivariate exponential distributions. Subsequently, the methods are evaluated using real life data sets. For the empirical comparison, data are generated by drawing a fraction (1 - ε) of the observations from the uncontaminated distribution, denoted as $F(\theta_1\boldsymbol{u}, \xi_1\mathbf{I}_p)$, and a fraction ε from the contaminated distribution, denoted as $F(\theta_2\boldsymbol{u}, \xi_2\mathbf{I}_p)$. Here $\boldsymbol{u}$ is a p dimensional vector of ones.

Without loss of generality, the parameters of uncontaminated distribution are taken as $\theta_1 = 0$ and $\xi_1 = 1$ and for the contaminated distribution the parametric values are $\theta_2 = 5, 10$ and $\xi_2 = 0.1, 1$. $M = 1000$ samples are generated for each combination of values of sample size n (n = 100, 500, 1000), dimension p (p = 5, 10, 20) and the contamination level ε (ε = 0.1, 0.2, 0.3).

Performance of the proposed method is compared with FAST-MCD (reweighted version), OGK, ROC, S_n, and RMD-S methods, focusing on the parameters success rate (SR) and false detection rate (FDR). SR gauges the detection of true outliers, while FDR measures the false detection of normal observations as outliers. The scale estimate Q_n proposed by Rousseeuw and Croux (1993) has been used as initial scale estimates for OGK. Since the proposed $\mathbf{C}_{SGK}$ estimator can be computed using any unbiased robust scale estimator, for the comparative study we have considered the robust scale estimators, MAD (Hampel, 1974), S_n, and Q_n which are unbiased under normality condition. $\mathbf{C}_{SGK}$ estimators based on these scale estimators are obtained using Equation (2.11) and the SGK approach using these estimates are denoted as $\mathbf{C}_{MAD}$, $\mathbf{C}_{Sn}$, and $\mathbf{C}_{Qn}$, respectively. Since the results are showing similar pattern, the results corresponding to multivariate normal and t distributions with $\theta_2 = 5$ and $\xi_2 = 1$ are presented here and the rest of the results are given in Appendix A. Also, to limit the size of the tables of success rates, only those cases, in which one of the methods scored less than 95% SR, have been included.

Table 2.1: SRs of outlier detection methods for Normal distribution

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	20	0.3	1	1	1	1	1	1	0.376	1
	30	0.2	1	1	1	1	1	1	0.491	1
		0.3	1	1	1	1	1	1	0.320	1
500	10	0.3	1	1	1	1	1	1	1	0.904
	20	0.3	1	1	1	1	1	1	0.434	0.494
	30	0.2	1	1	1	1	1	1	0.901	1
		0.3	1	1	1	1	1	1	0.084	0.152
1000	5	0.3	1	1	1	1	0.996	0.995	1	0.905
	10	0.3	1	1	1	1	1	1	1	0.71
	20	0.3	1	1	1	1	1	1	0.454	0.183
	30	0.3	0.794	1	1	1	1	1	0.066	0.035

Table 2.1 illustrates the success rates acquired for the normal distribution across various methodologies. Consistently, the SGK, S_n, ROC and OGK approaches demonstrate the highest success rates across all conditions. The RMD-S method exhibits commendable success rates across all scenarios, with the sample size set to 100. However, at $\varepsilon = 0.3$ there is a notable decline in success rates for data with $n = 500$, 1000, and $p = 20$, 30, indicating a deterioration in performance with increasing sample size and dimensionality. Notably, the MCD method performs comparably to other methods for dimensions $p = 5$ and 10, but its success rates experience a drastic decrease as both the dimensions and contamination levels increase.

Table 2.2: FDRs of outlier detection methods for Normal distribution

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0	0.000	0.000	0.000	0.110	0.000	0.045	0.073	0.003
		0.1	0.000	0.000	0.000	0.114	0.006	0.080	0.049	0.004
		0.2	0.000	0.000	0.000	0.101	0.005	0.056	0.037	0.000
		0.3	0.000	0.000	0.000	0.102	0.001	0.035	0.020	0.000
	10	0	0.000	0.000	0.000	0.000	0.000	0.047	0.117	0.000
		0.1	0.000	0.000	0.000	0.001	0.004	0.083	0.093	0.003
		0.2	0.000	0.000	0.000	0.000	0.002	0.061	0.063	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.033	0.040	0.000
	20	0	0.000	0.000	0.000	0.024	0.000	0.064	0.266	0.000
		0.1	0.000	0.000	0.000	0.046	0.002	0.090	0.197	0.001

		0.2	0.000	0.000	0.000	0.032	0.001	0.066	0.129	0.000
		0.3	0.000	0.000	0.000	0.014	0.000	0.038	0.231	0.000
	30	0	0.000	0.001	0.000	0.028	0.000	0.070	0.287	0.000
		0.1	0.000	0.000	0.000	0.047	0.001	0.094	0.202	0.000
		0.2	0.000	0.000	0.000	0.034	0.000	0.073	0.254	0.000
		0.3	0.000	0.000	0.000	0.015	0.000	0.038	0.274	0.000
500	5	0	0.000	0.000	0.000	0.087	0.000	0.028	0.028	0.001
		0.1	0.000	0.000	0.000	0.131	0.004	0.077	0.025	0.005
		0.2	0.000	0.000	0.000	0.120	0.002	0.052	0.021	0.001
		0.3	0.000	0.000	0.000	0.111	0.001	0.028	0.020	0.000
	10	0	0.000	0.000	0.000	0.000	0.000	0.031	0.032	0.001
		0.1	0.000	0.000	0.000	0.000	0.002	0.078	0.028	0.003
		0.2	0.000	0.000	0.000	0.000	0.001	0.054	0.024	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.028	0.020	0.000
	20	0	0.000	0.000	0.000	0.009	0.000	0.036	0.040	0.001
		0.1	0.000	0.000	0.000	0.037	0.000	0.080	0.035	0.001
		0.2	0.000	0.000	0.000	0.024	0.000	0.057	0.030	0.000
		0.3	0.000	0.000	0.000	0.012	0.000	0.030	0.028	0.000
	30	0	0.000	0.000	0.000	0.011	0.000	0.037	0.058	0.001
		0.1	0.000	0.000	0.000	0.038	0.000	0.080	0.053	0.001
		0.2	0.000	0.000	0.000	0.025	0.000	0.054	0.045	0.000
		0.3	0.000	0.000	0.000	0.012	0.000	0.029	0.050	0.000
1000	5	0	0.000	0.000	0.000	0.087	0.000	0.026	0.026	0.001
		0.1	0.000	0.000	0.000	0.137	0.003	0.075	0.025	0.006
		0.2	0.000	0.000	0.000	0.123	0.002	0.053	0.022	0.001
		0.3	0.000	0.000	0.000	0.115	0.001	0.029	0.021	0.000
	10	0	0.000	0.000	0.000	0.000	0.000	0.027	0.028	0.000
		0.1	0.000	0.000	0.000	0.000	0.002	0.077	0.025	0.004
		0.2	0.000	0.000	0.000	0.000	0.001	0.054	0.022	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.028	0.021	0.000
	20	0	0.000	0.000	0.000	0.007	0.000	0.030	0.030	0.001
		0.1	0.000	0.000	0.000	0.036	0.000	0.078	0.028	0.001
		0.2	0.000	0.000	0.000	0.023	0.000	0.054	0.024	0.000
		0.3	0.000	0.000	0.000	0.012	0.000	0.030	0.024	0.000
	30	0	0.000	0.000	0.000	0.008	0.000	0.034	0.037	0.001
		0.1	0.000	0.000	0.000	0.037	0.000	0.079	0.033	0.001
		0.2	0.000	0.000	0.000	0.024	0.000	0.056	0.028	0.000
		0.3	0.000	0.000	0.000	0.011	0.000	0.029	0.032	0.000

Table 2.2 displays the FDRs associated with different outlier detection techniques applied to normal data. It is evident that, across all cases, the SGK, ROC and RMD-S methods exhibit significantly low FDRs, while the OGK and MCD methods demonstrate higher FDRs. The FDR associated with the MCD method demonstrates an upward trend with increasing dimensionality, indicating a rise in false detections as dimensions expand. However, there is an improvement in FDR with larger sample sizes, suggesting that increased sample size mitigates false discoveries despite the dimensionality effect.

The S_n technique exhibits a declining tendency with increasing dimensionality, even while it presents a sizable FDR for low-dimensional data.

Table 2.3: SRs of outlier detection methods for multivariate *t* distribution with 3 d.f.

n	***p***	***ε***	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$	$\mathbf{S_n}$	**ROC**	**OGK**	**MCD**	**RMD-S**
100	10	0.3	1	1	1	1	1	1	0.650	1
	20	0.2	1	1	1	1	1	1	0.528	1
		0.3	1	1	1	1	1	1	0.399	1
	30	0.1	1	1	1	1	1	1	0.687	1
		0.2	1	1	1	1	1	1	0.370	1
		0.3	1	1	1	1	1	1	0.343	1
500	20	0.3	1	1	1	1	1	1	0.442	1
	30	0.2	1	1	1	1	1	1	0.504	1
		0.3	1	1	1	1	1	1	0.489	1
1000	20	0.3	1	1	1	1	1	1	0.439	1
	30	0.2	1	1	1	1	1	1	0.525	1
		0.3	1	1	1	1	1	1	0.496	1

The findings outlined in Table 2.3, derived from multivariate *t* distribution with 3 degrees of freedom data, indicate that all methods examined in this study, with the exception of MCD, achieve optimal success rates across datasets characterized by diverse combinations of size, dimension, and contamination levels. Specifically, for high-dimensional datasets with moderate to high degrees of contamination, the MCD technique exhibits lower success rates.

Examining the false detection rates presented in Table 2.4 for the multivariate *t* distribution, it is observed that the FDRs for all methodologies are higher compared to their corresponding values obtained for normal data. In this case as well, SGK demonstrates comparable performance to RMD-S and ROC, with similar FDRs except for 30% of outliers. However, the FDRs of S_n, OGK, and MCD methods are notably higher compared to SGK, RMD-S, and ROC.

Table 2.4: FDRs of outlier detection methods for multivariate *t* distribution with 3 d.f.

n	***p***	***ε***	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$	$\mathbf{S_n}$	**ROC**	**OGK**	**MCD**	**RMD-S**
100	5	0	0.000	0.000	0.000	0.097	0.001	0.038	0.243	0.001
		0.1	0.060	0.060	0.049	0.216	0.112	0.213	0.214	0.123
		0.2	0.047	0.047	0.043	0.203	0.086	0.166	0.190	0.072
		0.3	0.030	0.030	0.037	0.188	0.060	0.120	0.159	0.029
	10	0	0.000	0.000	0.000	0.000	0.001	0.044	0.344	0.002
		0.1	0.087	0.087	0.083	0.092	0.141	0.264	0.310	0.159
		0.2	0.076	0.076	0.078	0.085	0.115	0.207	0.265	0.101

		0.3	0.055	0.055	0.062	0.079	0.072	0.144	0.274	0.042
	20	0	0.000	0.000	0.000	0.000	0.001	0.044	0.381	0.002
		0.1	0.164	0.164	0.160	0.268	0.173	0.302	0.323	0.198
		0.2	0.145	0.145	0.148	0.212	0.137	0.240	0.349	0.123
		0.3	0.123	0.123	0.121	0.151	0.093	0.175	0.371	0.055
	30	0	0.000	0.000	0.000	0.024	0.000	0.064	0.339	0.001
		0.1	0.250	0.250	0.253	0.291	0.189	0.319	0.307	0.223
		0.2	0.228	0.228	0.219	0.242	0.153	0.259	0.330	0.140
		0.3	0.161	0.161	0.156	0.168	0.101	0.192	0.337	0.063
500	5	0	0.000	0.000	0.000	0.089	0.000	0.027	0.231	0.000
		0.1	0.045	0.045	0.045	0.234	0.111	0.214	0.212	0.124
		0.2	0.045	0.045	0.037	0.225	0.085	0.168	0.194	0.075
		0.3	0.031	0.031	0.034	0.215	0.062	0.122	0.181	0.037
	10	0	0.000	0.000	0.000	0.000	0.000	0.033	0.328	0.001
		0.1	0.069	0.069	0.069	0.104	0.141	0.261	0.313	0.163
		0.2	0.059	0.059	0.059	0.102	0.111	0.210	0.284	0.102
		0.3	0.053	0.051	0.050	0.099	0.078	0.149	0.254	0.048
	20	0	0.000	0.000	0.000	0.009	0.000	0.037	0.436	0.001
		0.1	0.096	0.096	0.101	0.272	0.173	0.302	0.403	0.204
		0.2	0.094	0.094	0.089	0.223	0.138	0.247	0.383	0.127
		0.3	0.072	0.071	0.077	0.160	0.099	0.180	0.410	0.062
	30	0	0.000	0.000	0.000	0.013	0.000	0.039	0.485	0.002
		0.1	0.116	0.116	0.124	0.296	0.189	0.323	0.446	0.227
		0.2	0.116	0.116	0.110	0.239	0.151	0.263	0.470	0.142
		0.3	0.094	0.092	0.093	0.176	0.107	0.194	0.468	0.067
1000	5	0	0.000	0.000	0.000	0.088	0.000	0.025	0.229	0.000
		0.1	0.044	0.044	0.044	0.235	0.109	0.214	0.213	0.124
		0.2	0.039	0.039	0.039	0.233	0.088	0.171	0.193	0.078
		0.3	0.033	0.033	0.030	0.220	0.057	0.123	0.177	0.033
	10	0	0.000	0.000	0.000	0.000	0.000	0.028	0.325	0.000
		0.1	0.064	0.064	0.065	0.107	0.142	0.265	0.311	0.163
		0.2	0.055	0.055	0.055	0.101	0.108	0.205	0.287	0.097
		0.3	0.049	0.049	0.045	0.102	0.076	0.146	0.265	0.044
	20	0	0.000	0.000	0.000	0.008	0.000	0.031	0.434	0.001
		0.1	0.097	0.097	0.093	0.274	0.172	0.303	0.416	0.206
		0.2	0.078	0.078	0.082	0.219	0.136	0.243	0.391	0.128
		0.3	0.072	0.072	0.072	0.163	0.099	0.183	0.413	0.062
	30	0	0.000	0.000	0.000	0.009	0.000	0.033	0.495	0.001
		0.1	0.115	0.115	0.118	0.295	0.188	0.324	0.470	0.226
		0.2	0.102	0.102	0.094	0.241	0.143	0.262	0.487	0.134
		0.3	0.086	0.086	0.086	0.176	0.107	0.194	0.487	0.069

2.6.1 Affine equivariance

This section investigates the potential consequences of employing SGK method lacking equivariance for the precise identification of outliers within the data. Given a dataset $\mathbf{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n\}$ and a non-singular $p \times p$ matrix $\mathbf{A}$, a transformed data is obtained as $\mathbf{Y_A} = \{\mathbf{A}\boldsymbol{y}_1, \mathbf{A}\boldsymbol{y}_2, \dots, \mathbf{A}\boldsymbol{y}_n\}$. When the estimates exhibit affine equivariance, then

$$\boldsymbol{m}_\mathbf{A} = \boldsymbol{m}(\mathbf{Y_A}) = \mathbf{A}\boldsymbol{m}(\mathbf{Y}) \; and \; \mathbf{S_A} = \mathbf{S}(\mathbf{Y_A}) = \mathbf{A}\mathbf{S}(\mathbf{Y})\mathbf{A}^\mathrm{T}.$$

Furthermore, the robust Mahalanobis distance of $\mathbf{A}\boldsymbol{y}_i$ from $\boldsymbol{m}_\mathbf{A}$ based on $\boldsymbol{S}_\mathbf{A}$ is defined as

$$RD(\mathbf{A}\boldsymbol{y}_i, \boldsymbol{m}_\mathbf{A}) = (\mathbf{A}\boldsymbol{y}_i - \boldsymbol{m}_\mathbf{A})^\mathrm{T}\boldsymbol{S}_\mathbf{A}^{-1}(\mathbf{A}\boldsymbol{y}_i - \boldsymbol{m}_\mathbf{A}), \quad i = 1,2,\dots,n.$$

If location and scatter estimates are affine-equivariant, then $RD(\mathbf{A}\boldsymbol{y}_i, \boldsymbol{m}_\mathbf{A}) = RD(\boldsymbol{y}_i, \boldsymbol{m})$ where $\boldsymbol{m}$ is the robust location estimate of $\mathbf{Y}$. To examine the effects of the lack of equivariance in estimates, Maronna and Zamar (2002), Sajesh and Srinivasan (2012), Cabana *et al.* (2021) generated random matrices as $\mathbf{A} = \mathbf{JD}$, where $\mathbf{J}$ is a random orthogonal matrix and $\mathbf{D} = \mathrm{diag}(v_1, \dots, v_p)$,, with v_j's being independent and uniformly distributed in (0, 1). A similar approach is adopted to investigate the lack of equivariance effects of the SGK methods. The simulation for untransformed data is repeated to evaluate the performance of SGK methods under transformation. Each generated data matrix $\mathbf{Y}$ is affinely transformed by multiplying it by a random matrix $\mathbf{A}$ generated as described above. The SGK methods are then applied to the transformed data matrix $\mathbf{Y_A}$ to detect outliers. The experiment is conducted across various sample space dimensions (p = 5, 10, 20, 30) and contamination levels ε (ε = 0.1, 0.2, 0.3).

Table 2.5: SRs of outlier detection methods for affinely transformed data

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.3	0.897	0.997	0.945	1	0.972	0.993	1	0.902
	20	0.3	0.900	0.900	0.990	1	1	1	0.351	0.986
	30	0.2	1	1	1	1	1	1	0.456	1
		0.3	1	1	1	1	1	1	0.312	0.968
500	5	0.3	0.901	0.901	0.941	1	0.988	0.997	1	0.847
	10	0.3	0.800	0.900	0.940	0.997	1	1	1	0.759
	20	0.3	0.743	1	1	1	1	1	0.356	0.634
	30	0.2	1	1	0.990	1	1	1	0.930	1
		0.3	0.505	0.7	0.89	1	1	1	0.150	0.496
1000	5	0.3	1	1	0.887	1	0.988	0.997	1	0.81
	10	0.3	0.889	1	1	0.998	1	1	1	0.722
	20	0.3	0.900	0.900	0.910	1	1	1	0.539	0.49
	30	0.3	0.900	0.800	0.662	1	1	1	0.087	0.385

Table 2.5 presents the success rates obtained for affinely transformed data using various outlier detection methods, mirroring the trends observed in the results obtained for normal data as presented in Table 2.1. Across all conditions, the SGK, S_n, Comedian, and OGK approaches consistently demonstrate the highest success rates. The RMD-S method exhibits commendable success rates across all scenarios when the sample size is set to 100. However, at an ε = 0.3, a notable decline in success rates is observed for datasets with larger sample and dimensions. It is noteworthy that the MCD method performs comparably to other methods for dimensions p = 5 and 10; however, its success rates experience a drastic decrease as both the dimensions and contamination levels increase.

Table 2.6: FDRs of outlier detection methods for affinely transformed data

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0	0.000	0.000	0.000	0.170	0.005	0.105	0.065	0.007
		0.1	0.000	0.000	0.000	0.097	0.007	0.081	0.048	0.021
		0.2	0.000	0.000	0.000	0.063	0.004	0.062	0.040	0.007
		0.3	0.000	0.000	0.000	0.054	0.002	0.034	0.018	0.002
	10	0	0.000	0.000	0.000	0.001	0.006	0.112	0.125	0.009
		0.1	0.000	0.000	0.000	0.000	0.003	0.091	0.092	0.023
		0.2	0.000	0.000	0.000	0.000	0.003	0.069	0.066	0.006
		0.3	0.000	0.000	0.000	0.000	0.001	0.037	0.040	0.001
	20	0	0.000	0.000	0.000	0.066	0.004	0.112	0.276	0.007
		0.1	0.000	0.000	0.000	0.047	0.002	0.094	0.199	0.016
		0.2	0.000	0.000	0.000	0.032	0.002	0.074	0.123	0.004
		0.3	0.006	0.000	0.000	0.019	0.001	0.041	0.230	0.000
	30	0	0.000	0.020	0.000	0.058	0.001	0.112	0.288	0.002
		0.1	0.000	0.000	0.000	0.053	0.001	0.107	0.205	0.014
		0.2	0.000	0.000	0.000	0.040	0.001	0.070	0.246	0.001
		0.3	0.000	0.000	0.001	0.020	0.000	0.039	0.283	0.000
500	5	0	0.000	0.000	0.000	0.170	0.007	0.100	0.030	0.019
		0.1	0.000	0.000	0.000	0.093	0.004	0.076	0.026	0.013
		0.2	0.000	0.000	0.000	0.060	0.002	0.053	0.020	0.003
		0.3	0.000	0.000	0.000	0.052	0.001	0.029	0.021	0.000
	10	0	0.000	0.000	0.000	0.000	0.003	0.100	0.032	0.016
		0.1	0.000	0.000	0.000	0.000	0.002	0.082	0.026	0.015
		0.2	0.000	0.000	0.000	0.000	0.001	0.056	0.023	0.004
		0.3	0.000	0.000	0.000	0.000	0.000	0.033	0.017	0.000
	20	0	0.000	0.000	0.000	0.055	0.001	0.104	0.040	0.010
		0.1	0.000	0.000	0.000	0.040	0.001	0.083	0.037	0.016
		0.2	0.000	0.000	0.000	0.028	0.000	0.059	0.033	0.002
		0.3	0.000	0.000	0.000	0.017	0.000	0.033	0.028	0.000
	30	0	0.000	0.000	0.000	0.055	0.000	0.104	0.057	0.006
		0.1	0.000	0.000	0.000	0.041	0.000	0.085	0.051	0.015
		0.2	0.000	0.000	0.000	0.028	0.000	0.060	0.043	0.001
		0.3	0.000	0.000	0.000	0.017	0.000	0.034	0.052	0.000

1000	5	0	0.000	0.000	0.000	0.171	0.006	0.103	0.027	0.020
		0.1	0.000	0.000	0.000	0.091	0.004	0.076	0.023	0.006
		0.2	0.000	0.000	0.000	0.059	0.002	0.054	0.021	0.001
		0.3	0.000	0.000	0.000	0.056	0.001	0.029	0.020	0.001
	10	0	0.000	0.000	0.000	0.000	0.003	0.101	0.029	0.015
		0.1	0.000	0.000	0.000	0.000	0.002	0.081	0.026	0.008
		0.2	0.000	0.000	0.000	0.000	0.001	0.058	0.021	0.001
		0.3	0.000	0.000	0.000	0.000	0.000	0.032	0.022	0.000
	20	0	0.000	0.000	0.000	0.052	0.001	0.101	0.030	0.010
		0.1	0.000	0.000	0.000	0.038	0.000	0.080	0.027	0.010
		0.2	0.000	0.000	0.000	0.028	0.000	0.058	0.025	0.001
		0.3	0.000	0.000	0.000	0.016	0.000	0.033	0.021	0.000
	30	0	0.000	0.000	0.000	0.051	0.000	0.104	0.035	0.006
		0.1	0.000	0.000	0.000	0.039	0.000	0.081	0.031	0.013
		0.2	0.000	0.000	0.000	0.028	0.000	0.059	0.028	0.001
		0.3	0.000	0.000	0.000	0.016	0.000	0.033	0.030	0.000

Similar to the success rate findings, the False Detection Rates corresponding to the transformed data as presented in Table 2.6 exhibit a similar pattern to that of the normal data. It is clear that, in every situation, the proposed SGK methods display significantly low FDRs, while the OGK and MCD methods show higher FDRs. This suggests that the affine transformation has minimal effect on the performance of the proposed SGK method, maintaining its effectiveness in outlier detection.

2.6.2 Correlated data

As the SGK method lacks affine equivariance, it becomes imperative to conduct a study utilizing correlated data, as its performance is contingent upon the covariance structure of the data. Devlin *et al*. (1981) employed a correlation matrix **P** in their investigation, generating Monte Carlo data from various distributions with moderate dimensionality (p = 6). The correlation matrix **P** takes the form

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 \end{bmatrix} \text{ with } \mathbf{P}_1 = \begin{bmatrix} 1 & 0.95 & 0.3 \\ 0.95 & 1 & 0.1 \\ 0.3 & 0.1 & 1 \end{bmatrix} \text{ and } \mathbf{P}_2 = \begin{bmatrix} 1 & -0.499 & -0.499 \\ -0.499 & 1 & -0.499 \\ -0.499 & -0.499 & 1 \end{bmatrix}$$

This matrix possesses advantageous attributes: its dimensions are sufficiently large for the study of multivariate estimators, and it encompasses a broad range of correlation values. The dataset, consisting of 100 observations, was generated from an asymmetric contaminated normal distribution, a mixture of $100(1-\varepsilon)$ observations from $\mathbf{N}_p(\mathbf{0}, \mathbf{P})$ and 100ε observations from $\mathbf{N}_p(5\boldsymbol{u}, \mathbf{P})$ where $\boldsymbol{u}^{\mathrm{T}} = (1, \ldots, 1)$. This setup allows for a comprehensive assessment of the methods' abilities in detecting outliers within highly correlated datasets.

Table 2.7: SRs of outlier detection methods for correlated data

n	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	0.1	1	1	1	1	1	1	1	1
	0.2	1	1	1	1	1	1	1	1
	0.3	1	1	1	1	1	1	1	1
500	0.1	1	1	1	1	1	1	1	1
	0.2	1	1	1	1	1	1	1	1
	0.3	1	1	1	1	1	1	1	1
1000	0.1	1	1	1	1	1	1	1	1
	0.2	1	1	1	1	1	1	1	1
	0.3	1	1	1	1	1	1	1	1

Tables 2.7 and Table 2.8 depict the success rates and false detection rates of various outlier detection methods for correlated data, respectively. The results demonstrate that all methods successfully identify outliers with optimal success rates while maintaining minimal FDRs. Among the methods, OGK exhibits slightly higher FDR compared to others. This underscores the applicability and effectiveness of the SGK method even in the presence of correlated data.

Table 2.8: FDRs of outlier detection methods for correlated data

n	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	0	0.000	0.000	0.000	0.004	0.028	0.112	0.072	0.016
	0.1	0.000	0.000	0.000	0.000	0.000	0.027	0.025	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.013	0.011	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.014	0.003	0.000
500	0	0.000	0.000	0.000	0.003	0.015	0.101	0.027	0.013
	0.1	0.000	0.000	0.000	0.000	0.000	0.021	0.002	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.013	0.001	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.010	0.002	0.000
1000	0	0.000	0.000	0.000	0.003	0.012	0.099	0.028	0.013
	0.1	0.000	0.000	0.000	0.000	0.000	0.020	0.001	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.020	0.001	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.009	0.001	0.000

2.6.3 Breakdown value of SGK method

The breakdown value of an outlier detection method, signifies the maximum fraction of contaminated data the method can handle while still maintaining reliable outlier identification. While this represents the worst-case robustness of the method, our experiments empirically evaluate the behaviour of outlier detection methods under controlled settings. These results are not intended to provide a breakdown theory but rather illustrate the robustness properties observed in practice under the specified conditions.

To examine the breakdown behaviour of the SGK method, we have followed the method used by Sajesh and Srinivasan (2012) and Cabana *et al*. (2021). The study incorporates two forms of contamination: ε percent symmetric, where the i^{th} observation is multiplied by $100i$, and ε percent asymmetric, where the i^{th} observation is replaced by $(100i)\boldsymbol{u}$, $i = 1, \ldots, 100\varepsilon$. In the first case, outliers are symmetrically distributed, while in the second case, they exhibit asymmetrical distribution. Various combinations of n (100, 500, 1000), p (10, 30, 50, 80, 100), and ε (0.1, 0.2, 0.3, 0.4, 0.45) were systematically selected to determine the empirical breakdown values of the SGK methods. This comprehensive exploration provides insights into the resilience and performance of the SGK method under diverse contamination scenarios and dimensionalities.

Table 2.9: Simulation results for breakdown value

		Symmetric						***Asymmetric***					
		SR			**FDR**			**SR**			**FDR**		
p	***ε***	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$
10	0.1	1	1	1	0	0	0	1	1	1	0	0	0
	0.2	1	1	1	0	0	0	1	1	1	0	0	0
	0.3	1	1	1	0	0	0	1	1	1	0	0	0
	0.4	1	1	1	0	0	0	1	1	1	0	0	0
	0.45	1	1	1	0	0	0	1	1	1	0	0	0
30	0.1	1	1	1	0	0	0	1	1	1	0	0	0
	0.2	1	1	1	0	0	0	1	1	1	0	0	0
	0.3	1	1	1	0	0	0	1	1	1	0	0	0
	0.4	1	1	1	0	0	0	1	1	1	0	0	0
	0.45	1	1	1	0	0	0	1	1	1	0	0	0
50	0.1	1	1	1	0	0	0	1	1	1	0	0	0
	0.2	1	1	1	0	0	0	1	1	1	0	0	0
	0.3	1	1	1	0	0	0	1	1	1	0	0	0
	0.4	1	1	1	0	0	0	1	1	1	0	0	0
	0.45	1	1	1	0	0	0	1	1	1	0	0	0
80	0.1	1	1	1	0	0	0	1	1	1	0	0	0
	0.2	1	1	1	0	0	0	1	1	1	0	0	0
	0.3	1	1	1	0	0	0	1	1	1	0	0	0
	0.4	1	1	1	0	0	0	1	1	1	0	0	0
	0.45	1	1	1	0	0	0	1	1	1	0	0	0
100	0.1	1	1	1	0	0	0	1	1	1	0	0	0
	0.2	1	1	1	0	0	0	1	1	1	0	0	0
	0.3	1	1	1	0	0	0	1	1	1	0	0	0
	0.4	1	1	1	0	0	0	1	1	1	0	0	0
	0.45	1	1	1	0	0	0	1	1	1	0	0	0

Table 2.9 presents the results for both ε percent symmetric and asymmetric cases with a sample size of n = 1000. It is evident from the table that the SGK method consistently achieves maximum success rates and minimum false detection rate for both symmetric and asymmetric data scenarios. Similar trends are observed for other values of n as well. These findings suggest that the proposed SGK method is adept at detecting a large amount of contamination and the estimates derived through this method are robust in the presence of a higher number of outliers.

2.7 Application to Real Life Data

To explore the efficacy of the outlier detection methods in real-world scenarios, we utilized several benchmark datasets, including *Bushfire* data (Campbell, 1989), *Milk* data (Daudin *et al*., 1988), *Stack Loss* data (Brownlee, 1965), *USJudgeRatings* data (available in "*datasets*" library of R) and the results obtained for *Bushfire* data and *Milk* data are presented here.

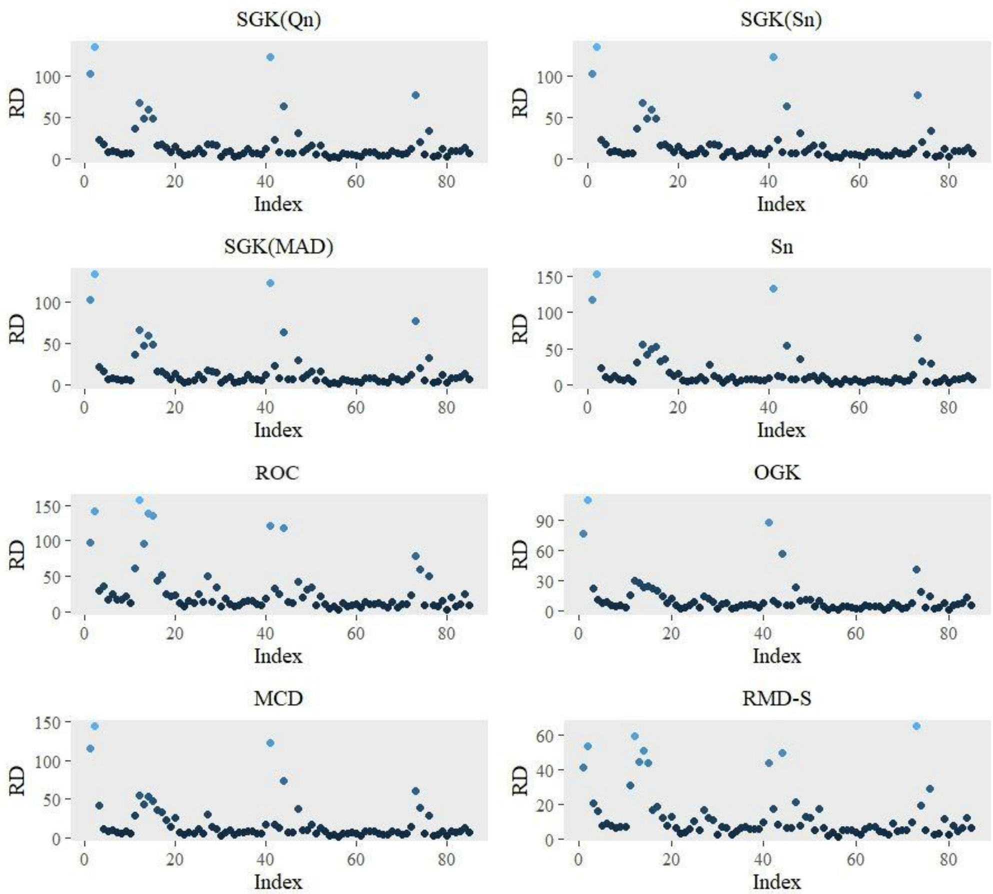


Figure 2.2: Robust Mahalano bis distance plots for Milk data

The *Milk* data set consists of 8 measurements on 86 samples of milk analysed by Atkinson (1994). He has suggested to remove the observation 70 from the analysis as it shows some weirdness in the preliminary analysis. Also, two of the observations (63 and 64) are exact duplicates. The duplicate point and observation 70 are omitted from the analysis. Atkinson identifies the observations 1, 2, 11, 12, 13, 14, 15, 41, 44 and 74 as potential outliers. Figure 2.2 presents the robust distance plots obtained for this data using various methods, and it is evident that all methods including SGK could successfully detect the known outliers.

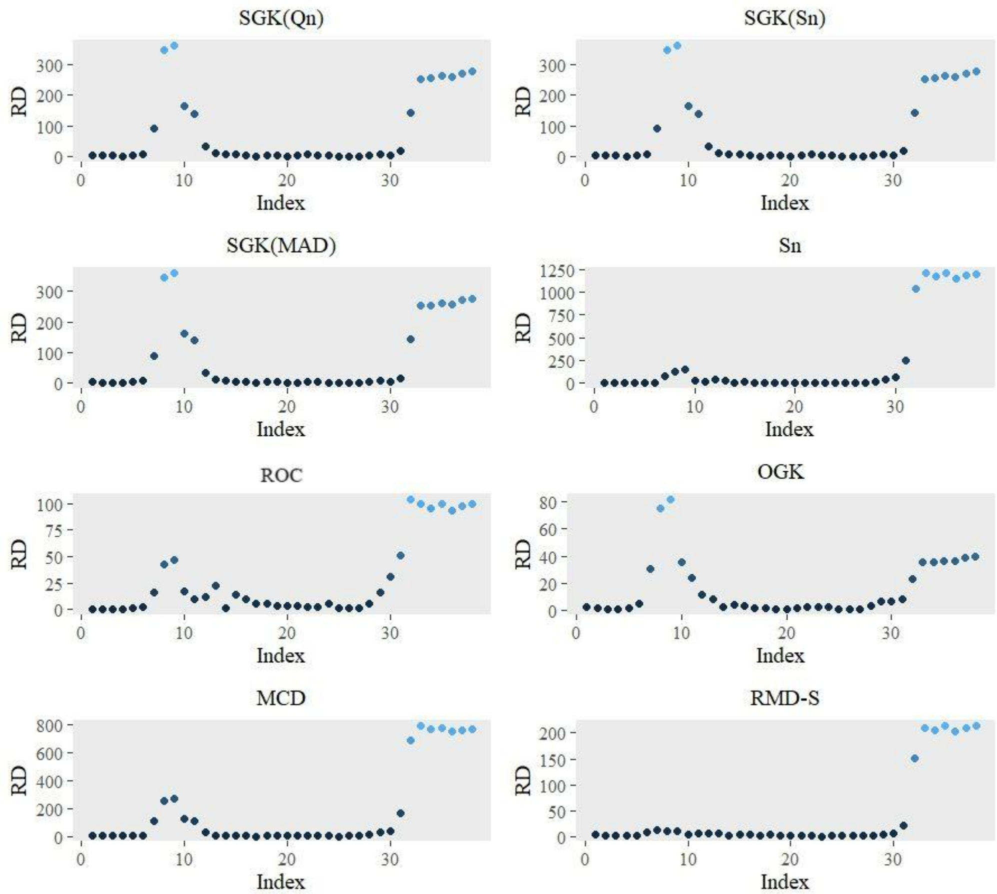


Figure 2.3: Robust Mahalanobis distance plots for Bushfire data

Bushfire data collected by Campbell (1989) consist of satellite measurement on five frequency bands each corresponding to 38 pixels. Maronna and Yohai (1995) analyzed this data and identified the observations 7-11 and 31-38 are outliers. Figure 2.3 displays the robust distance plots obtained for the Bushfire data employing different outlier detection methods. In contrast to the previous example, only the plots generated using SGK estimates exhibit a clear distinction between outliers and normal data points. While both the ROC and MCD methods detect the majority of outliers, the RMD-S method falls short in detecting outliers 7–11.

2.8 Summary

Covariance analysis goes beyond individual variable analysis by revealing the relationships between variables, which is essential for techniques like principal component analysis, factor analysis, and regression. However, traditional covariance matrix estimation can be influenced by outliers. To address this, we propose a robust estimator based on the Gnanadesikan-Kettenring approach for calculating covariance matrices. Additionally, an outlier detection method, SGK method, is proposed alongside this estimator. The effectiveness of both the estimator and the outlier detection method is compared to existing methods, such as ROC, S_n, OGK, MCD and RMD-S, through empirical evaluation.

The effectiveness of the SGK method is comprehensively evaluated using various metrics and scenarios. The simulation study utilizes tools success rate and false detection rate to compare its performance with existing methods. To assess the method's versatility in diverse settings, data with varying characteristics is employed. This includes correlated and uncorrelated data generated from multivariate normal, multivariate *t*, and multivariate exponential distributions. Across all evaluated scenarios, the proposed SGK method demonstrated a high success rate in identifying contaminated observations, accompanied by a very low false detection rate compared to other methods. While S_n, ROC, and OGK achieved similar success rates, the SGK method maintained a significantly lower FDR. RMD-S, a shrinkage-based method, showcased an optimal SR under the multivariate *t* distribution but exhibited a comparatively lower SR for large, heavily contaminated normal datasets. Additionally, MCD displayed a substantial decrease in SR as dimension and contamination sizes increased across most of the scenarios.

Despite not being affine equivariant, the SGK method maintained similar performance even under affinely transformed data, demonstrating its robustness to such transformations. In the simulation study, we computed the F-score for each contamination scheme. The result shows that the method exhibited exceptional resistance to contamination in datasets with varying sizes and dimensions, successfully identifying a high proportion of outliers with high accuracy even under heavy contamination.

Notably, the SGK method thrived in high-dimensional, low-sample size scenarios, a setting where shrinkage-based methods typically excel. This confirms its suitability for a wider range of data structures. Simulation results corresponding to small sample with high dimension cases and F Score are provided Appendix A. Finally, real-world applicability was validated by evaluating performance on benchmark datasets. The SGK method successfully identified all known outliers, corroborating the findings from the empirical study.

Table 2.10: Computation time

p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
5	0.1	0.0044	0.0096	0.0076	0.0502	0.0039	0.0060	0.0100	0.0038
	0.2	0.0049	0.0090	0.0068	0.0474	0.0037	0.0055	0.0097	0.0036
	0.3	0.0042	0.0091	0.0068	0.0470	0.0035	0.0061	0.0100	0.0043
10	0.1	0.0089	0.0274	0.0199	0.1520	0.0113	0.0200	0.0252	0.0063
	0.2	0.0090	0.0280	0.0198	0.1517	0.0110	0.0204	0.0252	0.0062
	0.3	0.0088	0.0286	0.0195	0.1519	0.0109	0.0209	0.0255	0.0062
20	0.1	0.0291	0.1103	0.0761	0.5715	0.0430	0.0832	0.1034	0.0174
	0.2	0.0326	0.1259	0.0830	0.6478	0.0467	0.0954	0.1150	0.0192
	0.3	0.0302	0.1159	0.0796	0.6054	0.0457	0.0865	0.1109	0.0191
30	0.1	0.0670	0.2596	0.1800	1.2971	0.0980	0.1946	0.2659	0.0360
	0.2	0.0683	0.2678	0.1817	1.3560	0.0985	0.2013	0.2705	0.0374
	0.3	0.0662	0.2601	0.1769	1.2919	0.0969	0.1960	0.2655	0.0363

Table 2.10 displays the computational times in seconds for the multivariate normal distribution with $\theta_2 = 5$ and $\xi_2 = 1$. The experiments were conducted using R software on a PC equipped with a 11^{th} Gen Intel(R) Core (TM) i3-1115G4 @ 3.00GHz processor and 8 GB of RAM. Among the SGK methods, C_{Sn} demonstrates faster performance compared to C_{MAD} and C_{Qn}. Additionally, the RMD-S method is slightly quicker than the proposed SGK method. Overall, the SGK method outperforms S_n, OGK, and MCD significantly. These findings suggest that the SGK method offers a superior balance between accurate outlier detection and minimizing false positives, particularly in high-contamination settings. Also, we have empirically compared three estimators - Q_n, S_n and MAD – as potential choices for the robust scale parameter σ_R. Based on our results, we recommend Q_n as the default choice for σ_R in most applications due to its superior performance in terms of success rate and false detection rate. However, S_n may be preferable in situations where computational speed is a key consideration.

CHAPTER 3

BIVARIATE ROBUST CONTROL CHARTS FOR INDIVIDUAL OBSERVATIONS

3.1 Introduction

Bivariate control charts for individual observations are a type of statistical process control tool, employed to monitor and assess process quality. These charts are specifically designed for situations where two variables are observed simultaneously, enabling the detection of patterns or trends that signal a shift or alteration in the process. Alt (1985) defined the concept of two distinct phases, Phases I and Phase II, in constructing control charts. In Phase I, historical observations are utilized to assess the process's stability and estimate the parameters of the in-control process, as well as to establish control limits. These control limits serve as reference boundaries for subsequent monitoring. In Phase II, the estimated parameters and control limits obtained from Phase I are applied to analyze the data collected during the actual manufacturing process, enabling the detection of any deviations or out-of-control signals. The purpose of Phase II is to monitor and maintain the process's stability based on the established control limits.

One commonly used approach in bivariate process monitoring and control is the Hotelling's T^2 control chart, which focuses on monitoring the mean vector of the process. This method serves as a direct counterpart to the univariate Shewhart chart (Montgomery, 2020). For the sample $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, . . ., \boldsymbol{x}_n\}$, where $\boldsymbol{x}_i$ represents a two-dimensional vector of measurements made on a process at the time period i, Hotelling's T^2 statistic is defined as

$$T^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}). \tag{3.1}$$

When the process is in statistical control, it is assumed that $\boldsymbol{x}_i$ is independent and follows a bivariate normal distribution $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ represents the mean vector and $\boldsymbol{\Sigma}$ represents the covariance matrix. If both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, T^2 follows a Chi-square distribution with 2 degrees of freedom. However, in practice, when dealing with individual observations, the true values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are generally unknown. In such cases, these parameters are usually substituted with classical estimators of mean vector ($\overline{\boldsymbol{x}}$) and covariance matrix ($\boldsymbol{S}$). Thus, the T^2 statistic for $\boldsymbol{x}_i$ can be constructed in the following way:

$$T^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}), \tag{3.2}$$

where $\overline{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2)^T$ and $\boldsymbol{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, S_{12} = Cov(X_1,\, X_2)$. Unfortunately, the classical estimators are highly susceptible to the influence of outliers, which can lead to misleading results when constructing control charts using phase I data (known as the *masking problem*). Recognizing this issue, researchers have proposed alternative approaches in the literature to mitigate the adverse impact of outliers. These alternative methods involve replacing the classical estimators with robust estimators, which are more resilient to the presence of outliers. By utilizing robust estimators, the control charts become more robust and reliable in the presence of outliers, ensuring more accurate monitoring of the process.

Vargas (2003) introduced robust control charts for identifying outliers in Phase I multivariate individual observations. These charts rely on two robust estimators, namely the minimum covariance determinant (MCD) and the minimum volume ellipsoid (MVE), to estimate the mean vector and covariance matrix. However, since the exact distribution of the T^2 statistic with these robust estimators is unavailable, the control limits for Phase I data are determined empirically. Vargas (2003) and Jensen *et al*. (2007) estimated the control limits for the robust T^2 charts for Phase I data using simulations. Alfaro and Ortega (2009) and Chenouri *et al*. (2009) have proposed robust multivariate control charts based on the reweighted Minimum Covariance Determinant (RMCD) estimator. Alfaro and Ortega (2009) determined the upper control limit (UCL) of their robust control chart empirically as the 95% quantile of the robust Hotelling's T^2 statistics while Chenouri *et al*. (2009) obtained the approximate control limits by modelling the empirical quantiles of robust Hotelling's T^2 statistics. Even though these control charts can successfully defend outliers, they also detect many normal observations as outliers (*swamping problem*). Also, the computational complexity of these control charts comparatively very high. Sajesh (2023) addressed this limitation by proposing a bivariate control chart using Gnanadesikan-Kettenring estimator. To further enhance its robustness, various robust dispersion estimators, including S_n and Q_n estimators proposed by Rousseeuw and Croux (1993), were incorporated. Among these variants, Sajesh (2023) showed that the GK(Q_n) chart, which utilizes Q_n estimator, is the most efficient.

In this chapter we have proposed bivariate robust control charts based on Comedian estimator and S_n Covariance estimator. These control charts are found to be suitable for bivariate data with outliers. The significance of the proposed control charts is evaluated through a comparative study with the performances of well-known methods.

3.2 Proposed Bivariate Robust Control Charts

Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a random sample of size n drawn from bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, which is considered the Phase I data in what follows. Let $\boldsymbol{y} \notin \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a Phase II observation, then it is known that

$$T^2(\boldsymbol{y}) \sim \left[\frac{2(n^2-1)}{n(n-2)}\right] F_{(2,n-2)},$$

where $T^2(\boldsymbol{y})$ is as defined in Equation (3.1) and $F_{(v_1,v_2)}$ is F distribution with (v_1, v_2) degrees of freedom (Wilks, 1962). We propose to replace the classical estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by its robust counterparts to robustify the T^2 control chart based on Phase I data. In this study we have considered the robust estimators S_n Covariance proposed by Sajana and Sajesh (2020) and Comedian proposed by Falk (1997) to replace the classical estimates of covariance matrix and the component wise median has been used to replace the classical estimate of mean vector. Suppose $\boldsymbol{x}_{\boldsymbol{med}}$ and $\boldsymbol{S}_R$ represent the robust mean vector and covariance matrix estimators, respectively. We define a robust Hotelling's T^2 for $\boldsymbol{y}$ based on these estimates by

$$T^2_{Com}(\boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{med}})^T \boldsymbol{S}^{-1}_{Com}(\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{med}}), \tag{3.3}$$

and

$$T^2_{S_n}(\boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{med}})^T \boldsymbol{S}^{-1}_{S_n}(\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{med}}), \tag{3.4}$$

where

$$\boldsymbol{S}_{Com} = \begin{bmatrix} MAD(X_1)^2 & COM(X_1,X_2) \\ COM(X_1,X_2) & MAD(X_2)^2 \end{bmatrix} \text{ and } \boldsymbol{S}_{S_n} = \begin{bmatrix} S_n(X_1)^2 & S_nCov(X_1,X_2) \\ S_nCov(X_1,X_2) & S_n(X_2)^2 \end{bmatrix}.$$

The finite sample distributions of the robust estimators which we have considered and thus the distribution of the robust Hotelling's T^2 statistic, denoted by T^2_R are unknown. By applying Slutsky theorem (Serfling, 2009), the asymptotic distribution of T^2_R can be obtained as Chi square distribution with 2 degrees of freedom. As $n \to \infty$

$$(\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{med}})^T \boldsymbol{S}_{\boldsymbol{R}}^{-1}(\boldsymbol{y} - \boldsymbol{x}_{\boldsymbol{med}}) \xrightarrow{D} (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \sim \chi^2_{(2)}. \tag{3.5}$$

However, this asymptotic distribution only works for large sample sizes. To compute the control limits for the proposed control charts we apply Monte Carlo simulations to estimate quantiles for various sample sizes. Then, we fit a smooth curve between the sample size and quantiles of T_R^2.These fits can be used to estimate appropriate control limits of the proposed control charts for small Phase I sample sizes ($n \leq 200$).

3.2.1 Estimation of Control Limits for the Proposed Control Charts

In order to estimate the 95%, 99%, 99.73% and 99.9% quantiles of T_R^2 for a given Phase I sample size n, we generate N = 10,000 samples of size n from a standard bivariate normal distribution $N_2(\boldsymbol{0}, \boldsymbol{I})$. For each data set of size n, we compute median vector and robust covariance matrix estimates using S_n Covariance and Comedian methods. In addition, for each data set, we randomly generate a new observation $\boldsymbol{y}_i$ from $N_2(\boldsymbol{0}, \boldsymbol{I})$ (treated as a Phase II observation) and calculate the corresponding $T_R^2\ (\boldsymbol{y}_i)$ value using S_n Covariance and Comedian estimates. The empirical distribution function of $T_R^2\ (\boldsymbol{y}_i)$ is based on the simulated values $T_R^2\ (\boldsymbol{y}_1), T_R^2\ (\boldsymbol{y}_2), \ldots., T_R^2\ (\boldsymbol{y}_K)$. By inverting the empirical distribution function of $T_R^2\ (\boldsymbol{y}_i)$, we obtain Monte Carlo estimates of the 95%, 99%, 99.73% and 99.9% quantiles. We construct the empirical distribution of $T_R^2\ (\boldsymbol{y}_i)$ for n = 20, 21, . . . ,50, 55, 60, . . . , 200. Figure 3.1 shows scatter plots of the empirical 95%, 99%, 99.73% and 99.9% quantiles of $T_R^2\ (\boldsymbol{y}_i)$ versus the sample size n for S_n Covariance control chart. Figure 3.2 shows scatter plots of the empirical 95%, 99%, 99.73% and 99.9% quantiles of $T_R^2\ (\boldsymbol{y}_i)$ versus the sample size n for Comedian control chart. These scatter plots suggest that we could model the quantiles using a family of regression curves of the form $f(n) = a + \frac{b}{n^c}$. Since the statistics asymptotically follow $\chi^2_{(2)}$ distribution, following two parameter family of curves is used for both S_n Covariance and Comedian control charts:

$$f_{1-\alpha}(n) = \boldsymbol{\chi}^2_{(2,1-\alpha)} + \frac{b_{1-\alpha}}{n^{c_{1-\alpha}}}, \tag{3.6}$$

where $\boldsymbol{\chi}^2_{(2,1-\alpha)}$ is the $1-\alpha$ quantile of the χ^2 distribution with 2 degrees of freedom and $b_{1-\alpha}$ and $c_{1-\alpha}$ are constants. Fitting this curve to the data will help us to estimate the desired control limits of S_n Covariance and Comedian control charts for any Phase I sample of size n. Note that, as n increases, $f_{1-\alpha}(n)$ approaches $\chi^2_{(2,1-\alpha)}$. Table 3.1 gives the least-square estimates of the parameters $b_{1-\alpha}$ and $c_{1-\alpha}$ for S_n Covariance and Comedian estimates.

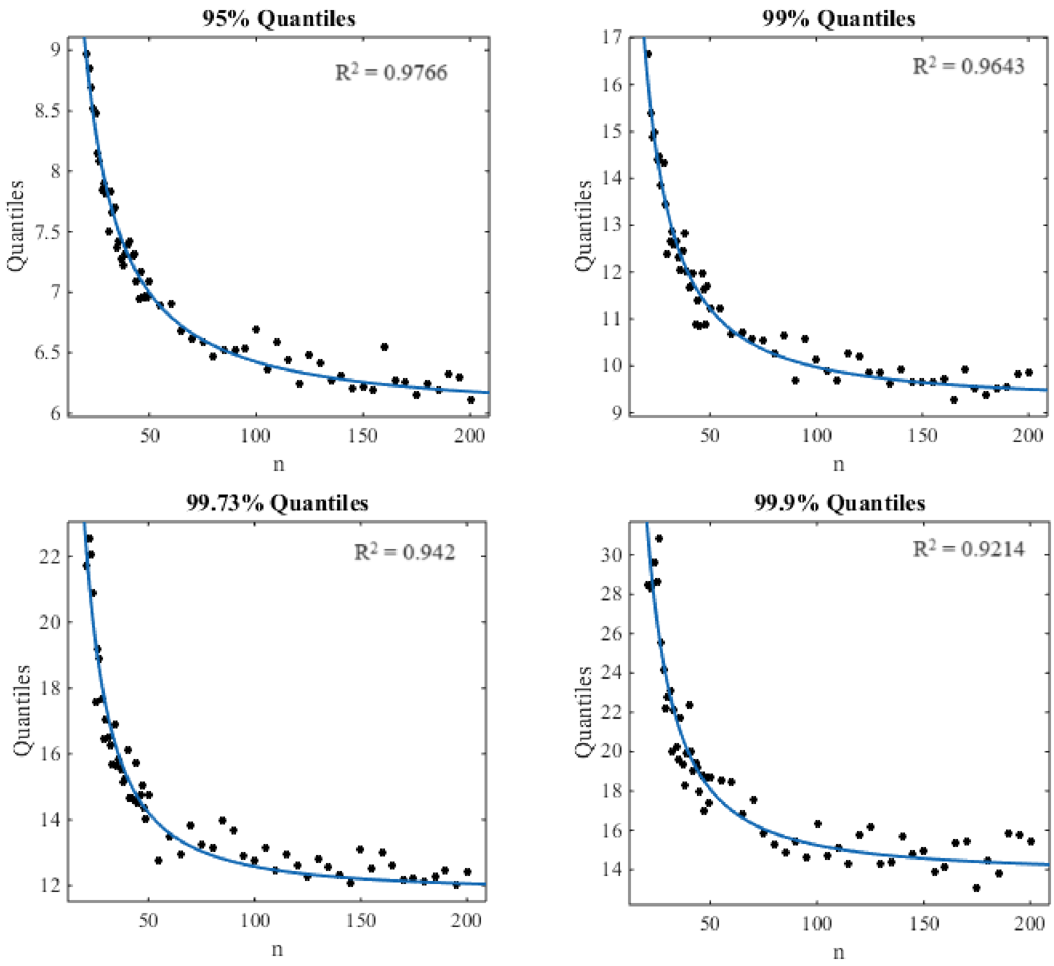


Figure 3.1: Simulated Quantiles of T_R^2 and the Fitted Curves for S_n Covariance

Using Table 3.1 and Equation (3.6), we can compute the 95%, 99%, 99.73% and 99.9% quantiles of $T_R^2\ (\boldsymbol{y}_i)$ for Phase I sample size n. The regression curves given by Equation (3.6) fit well to all the cases in Table 3.1, yielding R^2 values of at least 90%.

Table 3.1: The Least-Squares estimates of the regression parameters $b_{1-\alpha}$ and $c_{1-\alpha}$ for confidence levels $1-\alpha = 0.95, 0.99, 0.9973, 0.999$

Method	$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.0027$		$\alpha = 0.001$	
	$b_{1-\alpha}$	$c_{1-\alpha}$	$b_{1-\alpha}$	$c_{1-\alpha}$	$b_{1-\alpha}$	$c_{1-\alpha}$	$b_{1-\alpha}$	$c_{1-\alpha}$
S_n Covariance	116.5	1.214	467.2	1.393	1632	1.669	2069	1.581
Comedian	193.1	1.267	590.2	1.344	2178	1.556	4306	1.628

3.2.2 Implementation Procedures

A step-by-step approach for constructing proposed robust control chart is given as follows:

Phase I

1. Decide on the sample size n and confidence level $1 - \alpha$.

2. Collect the Phase I data $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ at well-defined periodic intervals.

3. Using the Phase I data, compute robust estimates of location and scale parameters.

4. For the desired α, choose the least square estimates of the parameters $b_{1-\alpha}$ and $c_{1-\alpha}$ using Table 3.1, then compute the control limit using Equation (3.6).

Phase II

5. For each new observation, compute the monitoring statistic T_R^2 according to the chosen robust estimation method, and plot the resulting values on the corresponding control chart using the Phase-I control limit obtained in step 4.

6. Interpret the chart and look for out-of-control points or patterns. Diagnose the process if needed.

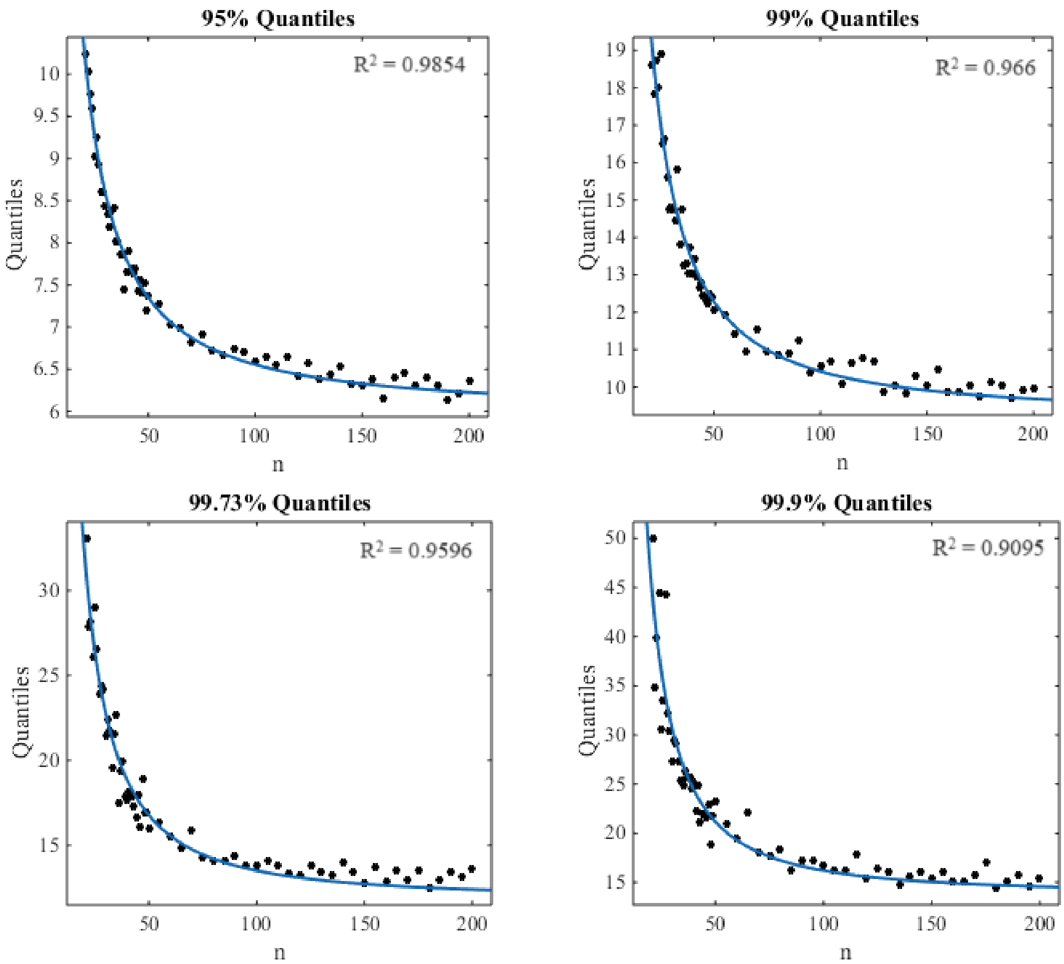


Figure 3.2: Simulated Quantiles of T_R^2 and the Fitted Curves for Comedian

3.3 Simulation Study

To evaluate the effectiveness of the suggested control charts, we carry out multiple simulation studies that vary in their Phase I data structures and the extent of change in the process mean vector and dispersion matrix in the Phase II data. The control chart's efficacy is assessed by analysing its ability to detect changes and the rate of false detections in the process behaviour using different estimators in Phase I. We measure the performance of the control chart by *success rate* (SR)- which is the proportion of statistic values that exceed the control limits across 1000 replications, which provides an estimate of the likelihood of detecting changes and *false alarm rate* (FAR) which is the false detections in the process behaviour based on the Phase II data. We have considered a contaminated model by using a mixture of bivariate normal distributions as

$$(1-\varepsilon)\ \mathrm{N}_2(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1),$$

where ε is the proportion (in as much by one) of outliers, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the in-control parameters (mean vector and covariance matrix respectively) and $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are the out-of-control parameters. Various scenarios are analysed with the following elements: number of observations(n), proportion of outliers (ε), and how to generate the outliers using different values of $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$. We can, without loss of generality, consider $\boldsymbol{\mu}_0$ to be a zero vector. We have considered that ε takes the values 0, 0.1 and 0.2, and the probability of detecting a change depends on the values of $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$. Using 1000 replications, we considered different sample sizes where n = 25, 50, 100 and 1000 with dimensions 2 and an overall false alarm probability of α = 0.001, 0.01, 0.0027 and 0.05. We generated 1000 samples of size n in Phase I and computed the robust and traditional location and scale estimates for each data set in this phase using MATLAB Program. Various cases are, hence, considered:

Case A: **Independent variables**

In this case, the two variables (Quality Characteristics) $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are assumed to be independent. The contaminated normal model considered is as follows:

$$(1-\varepsilon)\ \mathrm{N}_2(\mathbf{0}, \boldsymbol{I}_2) + \varepsilon\ \mathrm{N}_2(\boldsymbol{\mu}_1, \boldsymbol{I}_2),$$

where we consider $\boldsymbol{\mu}_1$ to be a vector of size 2 where the elements are all 0 (when there is no change), 5 or 10, and $\boldsymbol{I}_2$ to be the identity matrix of size 2. In this case, we want to compare the behaviour of different robust alternatives when there are different-sized changes in the average of all the variables if the variables are independent.

Case B: **Correlated variables**

In this case, two variables, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are assumed to be correlated. The contaminated normal model considered is as follows:

$$(1-\varepsilon)\ \mathrm{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_0),$$

where we consider $\boldsymbol{\mu}_1$ to be a vector of size 2 where the elements are all 0 (when there is no change), 5 or 10 (which is a good leverage point), and $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$. We have used this value of $\boldsymbol{\Sigma}_0$ to analyse whether the correlation level affects the detection probability of each alternative.

Case C: **Correlated variables and regression outliers**

Here, the two variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are assumed to be correlated and regression outliers are introduced. The contaminated normal model considered is as follows:

$$(1-\varepsilon)\ \mathrm{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1),$$

where we consider $\boldsymbol{\mu}_1$ to be a vector of size 2 where the elements are all 0 (when there is no change), 5 or 10 (which is a good leverage point), $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$. In this case, we analyse and compare the proposed robust methods in terms of the so-called good leverage and regression outliers.

The empirical performance of the proposed control charts is evaluated in comparison with classical control chart as well as the robust control charts of Alfaro and Ortega (2009), Chenouri *et al*. (2009) and Sajesh (2023). To facilitate comparison, we will refer to the methods introduced by Alfaro and Ortega (2009) and Chenouri *et al*. (2009) as MCD1 and MCD2, respectively. The control limits presented in these articles have been utilized to calculate the SR and false alarm rate FAR. It is worth noting that Chenouri *et al*. (2009) demonstrated the superiority of their method over the control charts proposed by Vargas (2003) and Jensen *et al*. (2007). Hence, for the purpose of comparison, we have excluded the methods of Vargas (2003) and Jensen *et al*. (2007) from the analysis.

We present only the result for α = 0.01 here and note that similar conclusions hold for other values of α.

Table 3.2: SR (FAR) in Phase II when there are case A outliers in Phase I data with different values of μ_1 and ε

n	μ_1	ε	S_n Covariance	Comedian	Classical	MCD1	MCD2	GK(Q_n)
25	(5,5)	0	1 (0.009)	0.998 (0.009)	1 (0.007)	1 (0.153)	1 (0.053)	0.998 (0.004)
		0.1	1 (0.003)	1 (0.004)	0.937 (0.002)	1 (0.113)	1 (0.053)	0.961 (0.001)
		0.2	0.999 (0.001)	1 (0.001)	0.461 (0.006)	1 (0.093)	1 (0.033)	0.607 (0.003)
	(10,10)	0	1 (0.018)	1 (0.014)	1 (0.016)	1 (0.18)	1 (0.067)	1 (0.014)
		0.1	1 (0.003)	1 (0.002)	0.017 (0.006)	1 (0.147)	1 (0.047)	1 (0.004)
		0.2	0.998 (0.001)	0.997 (0.004)	0.006 (0.006)	1 (0.087)	1 (0.013)	1 (0.002)
50	(5,5)	0	1 (0.013)	1 (0.011)	1 (0.013)	1 (0.147)	1 (0.033)	1 (0.02)
		0.1	1 (0.002)	1 (0.002)	0.999 (0.002)	1 (0.08)	1 (0.02)	0.988 (0.003)
		0.2	1 (0.001)	1 (0.003)	0.7 (0.004)	1 (0.047)	1 (0.013)	0.862 (0)
	(10,10)	0	1 (0.01)	1 (0.01)	1 (0.011)	1 (0.153)	1 (0.08)	1 (0.011)
		0.1	1 (0.001)	1 (0.002)	0.036 (0.001)	1 (0.093)	1 (0.013)	0.994 (0.003)
		0.2	1 (0)	1 (0)	0.009 (0.003)	1 (0.087)	1 (0.027)	0.883 (0)
100	(5,5)	0	1 (0.013)	1 (0.014)	1 (0.01)	1 (0.113)	1 (0.053)	1 (0.006)
		0.1	1 (0.003)	1 (0.004)	1 (0.005)	1 (0.047)	1 (0.033)	0.997 (0.002)
		0.2	1 (0)	1 (0)	0.794 (0.003)	1 (0.047)	1 (0)	0.925 (0)
	(10,10)	0	1 (0.008)	1 (0.008)	1 (0.007)	1 (0.1)	1 (0.047)	1 (0.015)
		0.1	1 (0)	1 (0)	0.048 (0.001)	1 (0.08)	1 (0.013)	0.998 (0.002)
		0.2	1 (0.001)	1 (0.001)	0.009 (0.004)	1 (0.067)	1 (0.007)	0.911 (0.001)
1000	(5,5)	0	1 (0.006)	1 (0.006)	1 (0.005)	1 (0.033)	1 (0)	1 (0.008)
		0.1	1 (0.001)	1 (0.002)	1 (0.001)	1 (0.053)	1 (0.007)	0.999 (0.002)
		0.2	1 (0.001)	1 (0.001)	0.869 (0.001)	1 (0.06)	1 (0.027)	0.949 (0)
	(10,10)	0	1 (0.015)	1 (0.015)	1 (0.014)	1 (0.053)	1 (0.013)	1 (0.018)

		0.1	1 (0)	1 (0)	0.059 (0.002)	1 (0.067)	1 (0.027)	0.998 (0.001)
		0.2	1 (0)	1 (0)	0.005 (0)	1 (0.067)	1 (0)	0.938 (0)

Table 3.2 presents the results obtained from simulations conducted for Case A samples. The findings clearly indicate that when the phase I data contains outliers, the S_n Covariance, Comedian, MCD1 and MCD2 control charts exhibit superior performance in terms of success rate compared to the classical control chart. The GK(Q_n) chart, although robust, displays significantly lower SR compared to other robust counterparts when contamination increases. In Conversely, when the phase I data is free from outliers, the robust control charts demonstrate similar performance to their classical counterpart.

Furthermore, it is crucial to consider the comparison in terms of FAR. Although the control charts based on the MCD show optimal success rates across all scenarios, their false alarm rates are significantly higher than those of the control charts based on the S_n Covariance and Comedian methods. Specifically, when comparing the two MCD-based control charts, the control chart based on MCD2 outperforms the one based on MCD1 in terms of FAR.

Table 3.3: SR (FAR) in Phase II when there are case B outliers in Phase I data with different values of $\boldsymbol{\mu}_1$ and ε

n	μ_1	ε	S_n Covariance	Comedian	Classical	MCD1	MCD2	GK(Q_n)
25	(5,5)	0	0.957 (0.034)	0.951 (0.025)	1 (0.008)	1 (0.167)	1 (0.06)	0.848 (0.031)
		0.1	0.924 (0.016)	0.931 (0.025)	0.816 (0.003)	1 (0.147)	0.9933 (0.053)	0.676 (0.026)
		0.2	0.839 (0.009)	0.854 (0.026)	0.358 (0.002)	1 (0.133)	1 (0.047)	0.338 (0.017)
	(10,10)	0	0.95 (0.031)	0.957 (0.027)	1 (0.013)	1 (0.16)	1 (0.067)	0.974 (0.029)
		0.1	0.922 (0.019)	0.928 (0.024)	0.027 (0.002)	1 (0.153)	1 (0.047)	0.964 (0.023)
		0.2	0.823 (0.011)	0.852 (0.026)	0.007 (0.002)	1 (0.113)	1 (0.033)	0.943 (0.016)
50	(5,5)	0	0.987 (0.028)	0.972 (0.018)	1 (0.009)	1 (0.147)	1 (0.06)	0.947 (0.045)
		0.1	0.963 (0.027)	0.962 (0.019)	0.961 (0.002)	1 (0.12)	1 (0.027)	0.847 (0.03)

		0.2	0.952 (0.016)	0.958 (0.01)	0.565 (0.003)	1 (0.04)	1 (0.013)	0.57 (0.023)
	(10,10)	0	0.99 (0.034)	0.985 (0.02)	1 (0.008)	1 (0.107)	1 (0.047)	0.953 (0.034)
		0.1	0.976 (0.023)	0.974 (0.021)	0.046 (0.003)	1 (0.087)	1 (0.033)	0.844 (0.039)
		0.2	0.943 (0.019)	0.95 (0.012)	0.95 (0.021)	1 (0.06)	1 (0.02)	0.571 (0.025)
100	(5,5)	0	1 (0.017)	0.997 (0.028)	1 (0.014)	1 (0.067)	1 (0.013)	0.985 (0.027)
		0.1	0.993 (0.017)	0.991 (0.018)	0.981 (0.005)	1 (0.12)	1 (0.033)	0.904 (0.015)
		0.2	0.974 (0.013)	0.989 (0.018)	0.66 (0.004)	1 (0.047)	1 (0.007)	0.633 (0.009)
	(10,10)	0	0.998 (0.013)	0.989 (0.025)	1 (0.005)	1 (0.1)	1 (0.047)	0.988 (0.021)
		0.1	0.986 (0.019)	0.984 (0.018)	0.047 (0.001)	1 (0.073)	1 (0.007)	0.912 (0.007)
		0.2	0.973 (0.015)	0.97 (0.013)	0.006 (0.003)	1 (0.08)	1 (0.007)	0.624 (0.007)
1000	(5,5)	0	1 (0.004)	1 (0.003)	1 (0.01)	1 (0.033)	1 (0.007)	0.988 (0.013)
		0.1	1 (0)	1 (0.003)	0.987 (0.002)	1 (0.067)	1 (0.013)	0.913 (0)
		0.2	1 (0)	1 (0.004)	0.694 (0.001)	1 (0.053)	1 (0.007)	0.647 (0)
	(10,10)	0	1 (0.003)	1 (0.002)	1 (0.004)	1 (0.053)	1 (0.007)	0.983 (0.016)
		0.1	1 (0.004)	1 (0.006)	0.051 (0.001)	1 (0.073)	1 (0.027)	0.901 (0.001)
		0.2	0.999 (0)	1 (0.001)	0.008 (0.004)	1 (0.067)	1 (0.02)	0.648 (0)

Table 3.3 and Table 3.4 provide the simulation results for Case B and Case C, respectively. These tables demonstrate that both the MCD methods and classical methods exhibit similar behavior as observed in the case of independent variables. However, the methods based on S_n Covariance and Comedian show slight decreases in success rates. However, under Case B contamination, GK(Q_n) shows a significant drop in SR,

sometimes falling below 65% when ε = 0.2**.** It is worth noting that the success rates of the proposed methods and GK(Q_n) improve as the sample size increases. Similarly, the failure rates of all the methods follow a similar pattern to that observed in the case of independent variables (Case A).

Furthermore, the proposed methods outperform the MCD-based control charts in terms of false alarm rate (FAR) in the simulations of Case B and Case C. The GK(Q_n) chart shows a favourable FAR pattern, staying low across Case B and Case C. Overall, S_n Covariance emerges as the most balanced and reliable control chart across all contamination types, with Comedian as a close competitor.

Table 3.4: SR (FAR) in Phase II when there are case C outliers in Phase I data with different values of $\boldsymbol{\mu}_1$ and ε

n	$\boldsymbol{\mu}_1$	ε	S_n Covariance	Comedian	Classical	MCD1	MCD2	GK(Q_n)
25	(5,5)	0	0.957 (0.034)	0.951 (0.025)	1 (0.008)	1 (0.167)	1 (0.06)	0.848 (0.031)
		0.1	0.924 (0.016)	0.931 (0.025)	0.816 (0.003)	1 (0.147)	0.9933 (0.053)	0.676 (0.026)
		0.2	0.839 (0.009)	0.854 (0.026)	0.358 (0.002)	1 (0.133)	1 (0.047)	0.338 (0.017)
	(10,10)	0	0.95 (0.031)	0.957 (0.027)	1 (0.013)	1 (0.16)	1 (0.067)	0.974 (0.029)
		0.1	0.922 (0.019)	0.928 (0.024)	0.027 (0.002)	1 (0.153)	1 (0.047)	0.964 (0.023)
		0.2	0.823 (0.011)	0.852 (0.026)	0.007 (0.002)	1 (0.113)	1 (0.033)	0.943 (0.016)
50	(5,5)	0	0.987 (0.028)	0.972 (0.018)	1 (0.009)	1 (0.147)	1 (0.06)	0.947 (0.045)
		0.1	0.963 (0.027)	0.962 (0.019)	0.961 (0.002)	1 (0.12)	1 (0.027)	0.847 (0.03)
		0.2	0.952 (0.016)	0.958 (0.01)	0.565 (0.003)	1 (0.04)	1 (0.013)	0.57 (0.023)
	(10,10)	0	0.99 (0.034)	0.985 (0.02)	1 (0.008)	1 (0.107)	1 (0.047)	0.953 (0.034)
		0.1	0.976 (0.023)	0.974 (0.021)	0.046 (0.003)	1 (0.087)	1 (0.033)	0.844 (0.039)
		0.2	0.943 (0.019)	0.95 (0.012)	0.95 (0.021)	1 (0.06)	1 (0.02)	0.571 (0.025)
100	(5,5)	0	1 (0.017)	0.997 (0.028)	1 (0.014)	1 (0.067)	1 (0.013)	0.985 (0.027)

		0.1	0.993 (0.017)	0.991 (0.018)	0.981 (0.005)	1 (0.12)	1 (0.033)	0.904 (0.015)
		0.2	0.974 (0.013)	0.989 (0.018)	0.66 (0.004)	1 (0.047)	1 (0.007)	0.633 (0.009)
	(10,10)	0	0.998 (0.013)	0.989 (0.025)	1 (0.005)	1 (0.1)	1 (0.047)	0.988 (0.021)
		0.1	0.986 (0.019)	0.984 (0.018)	0.047 (0.001)	1 (0.073)	1 (0.007)	0.912 (0.007)
		0.2	0.973 (0.015)	0.97 (0.013)	0.006 (0.003)	1 (0.08)	1 (0.007)	0.624 (0.007)
1000	(5,5)	0	1 (0.004)	1 (0.003)	1 (0.01)	1 (0.033)	1 (0.007)	0.988 (0.013)
		0.1	1 (0)	1 (0.003)	0.987 (0.002)	1 (0.067)	1 (0.013)	0.913 (0)
		0.2	1 (0)	1 (0.004)	0.694 (0.001)	1 (0.053)	1 (0.007)	0.647 (0)
	(10,10)	0	1 (0.003)	1 (0.002)	1 (0.004)	1 (0.053)	1 (0.007)	0.983 (0.016)
		0.1	1 (0.004)	1 (0.006)	0.051 (0.001)	1 (0.073)	1 (0.027)	0.901 (0.001)
		0.2	0.999 (0)	1 (0.001)	0.008 (0.004)	1 (0.067)	1 (0.02)	0.648 (0)

3.4 Time Complexity

A simulation study was conducted to compare the computation time of these methods across various sample sizes. The results, given in Table 3.5, reveal that the Comedian and Classical charts are multiple times faster than the MCD-based methods in all cases examined. GK(Q_n) is also remains computationally practical, showing only a slight increase as n grows. S_n Covariance prove computationally efficient for small to moderate sample sizes, but its computation time increases sharply as the sample size increases.

Table 3.5: Computation time (in seconds)

n	S_n Covariance	Comedian	Classical	MCD1	GK(Q_n)
25	0.00128	0.00054	0.00031	0.00342	0.00040
50	0.00214	0.00064	0.00032	0.00480	0.00050
100	0.00418	0.00067	0.00036	0.00769	0.00067
1000	0.06557	0.00073	0.00039	0.02768	0.00259

3.5 Application to Real Life Data

Further, to investigate the performance of the methods for real situation we have used the data given by Quesenberry (2001). The original data consists of 11 quality variables and measured on 30 products form a production process. For our comparison purposes we consider 25 observations from the first two variables since the restriction from MCD1 method. Figure 3.3 presents the control charts based on the robust and classical estimators. Control limits are calculated for $\alpha = 0.01$. All the control charts show that the second observation lying outside the upper control limits.

To examine the performance under contamination, an artificial outlying observation is created by adding a shift of 5 to the second variable of the 20^{th} observation. The resulting control charts, shown in figure 3.4, indicates that all the robust control charts successfully detected both the 2^{nd} and 20^{th} observations, while the classical chart fails to identify the 20^{th} observation.

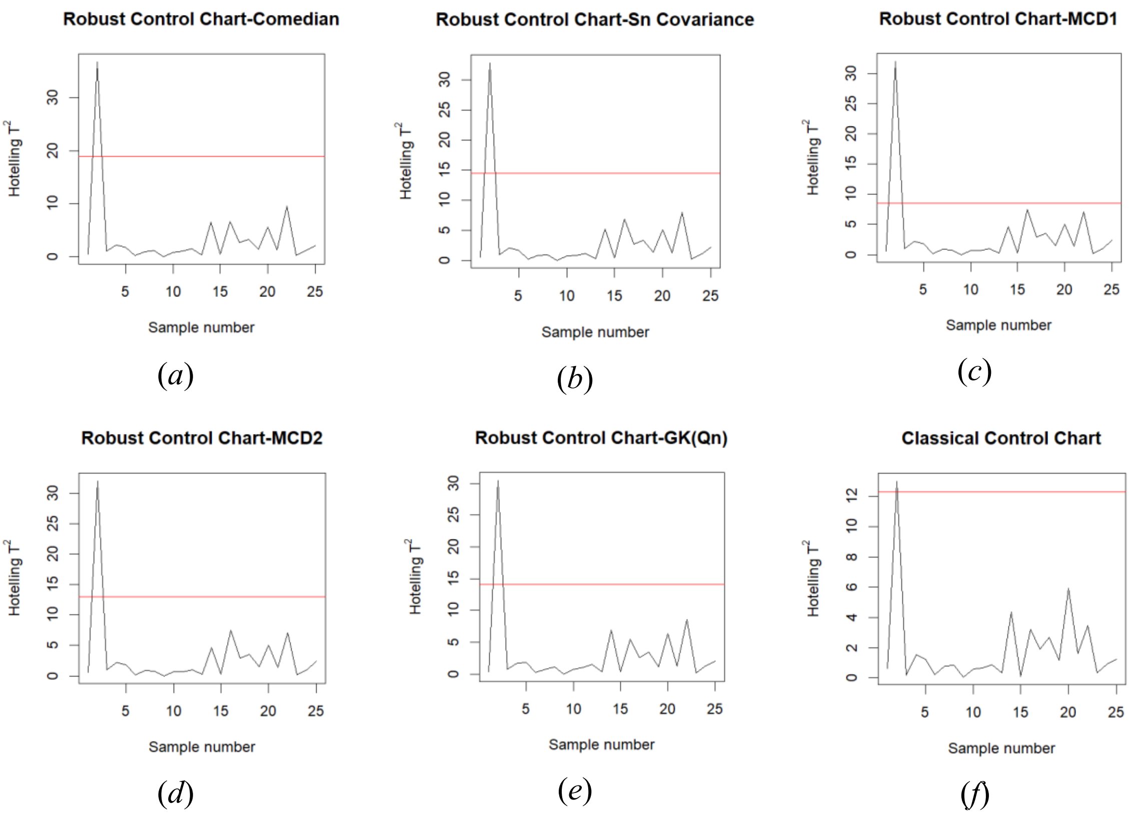


Figure 3.3: Control charts for the Quesenberry data. (*a*) S_n Covariance, (*b*) Comedian, (*c*) MCD1, (*d*) MCD2, (*e*) GK(Q_n), and (*f*) Classical

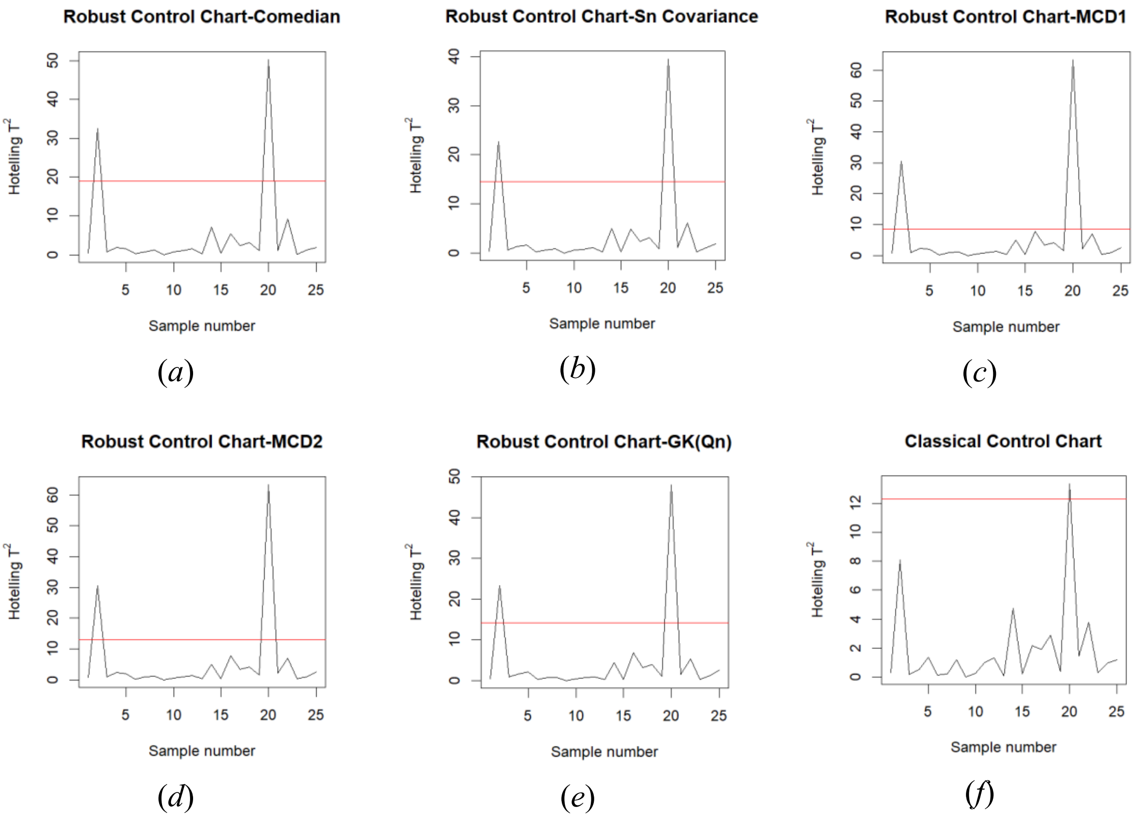


Figure 3.4: Control charts for the Quesenberry data (contaminated). (*a*) S_n Covariance, (*b*) Comedian, (*c*) MCD1, (*d*) MCD2, (*e*) GK(Q_n), and (*f*) Classical

3.6 Summary

In this chapter, we introduced bivariate robust Hotelling's T^2 control charts based on the Comedian and S_n Covariance estimators as alternatives to the classical bivariate T^2 control charts for Phase II data. To establish control limits for the proposed control charts, we conducted Monte Carlo simulations. The empirical quantiles obtained from these simulations were then modelled to approximate control limits for various sample sizes. Our simulation studies demonstrated that the proposed robust control charts perform similarly to the classical T^2 control chart when the process is in control. However, they outperformed the classical T^2 control charts when outliers were present in Phase I data. The MCD-based methods exhibited optimal success rates across all situations considered.

Furthermore, when comparing the methods based on false alarm rate (FAR), the proposed methods outperformed the MCD-based methods in the presence of outliers.

This indicates that the proposed control charts are more effective at controlling false alarms when the data contains outliers. Additionally, the results revealed that the proposed control charts had significantly shorter computation times compared to the MCD-based control charts. This computational efficiency contributes to faster analysis and decision-making processes. Overall, our findings highlight the effectiveness, efficiency, and robustness of the proposed control charts for monitoring bivariate processes in real-world applications. However, due to its nature as a variable control chart, the proposed control chart may not be applicable for quality characteristics that adhere to discrete distributions.

CHAPTER 4

MULTIVARIATE ROBUST CONTROL CHARTS FOR INDIVIDUAL OBSERVATIONS

4.1 Introduction

As discussed in Chapter 3, multivariate control charts monitor the joint behaviour of two or more related process variables (Alt, 1985; Montgomery, 2020). Historical data in Phase I are used to estimate the in-control mean vector and covariance matrix and to establish control limits, which serve as benchmarks for subsequent monitoring. In Phase II, these limits are applied to new observations to detect deviations and ensure the stability of the process.

The Hotelling's T^2 control chart is commonly used for monitoring shifts in the mean vector of multivariate processes, serving as the multivariate counterpart to the univariate Shewhart chart (Montgomery, 2020). For a sample $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$, with p dimensional observations, the Hotelling's T^2 statistic is defined as

$$T_i^2 = (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}). \tag{4.1}$$

When the process is in statistical control, the observations $\boldsymbol{x}_i$ is independent and follows a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. If both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, T_i^2 follows a Chi-square distribution with p degrees of freedom. In practice, however, these parameters are typically unknown and are estimated using the sample mean $\overline{\boldsymbol{x}}$ and the sample covariance matrix $\boldsymbol{S}$. By substituting the unknown parameters with their sample estimates, the statistic for an individual observation is given by

$$T_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}). \tag{4.2}$$

Classical estimators are highly sensitive to outliers, which may distort control limits and mask true process shifts. Robust estimators have therefore been developed to mitigate this issue, improving the reliability of multivariate control charts in the presence of contaminated data.

As mentioned in Chapter 3, Vargas (2003) introduced robust multivariate control charts using MCD and Minimum Volume Ellipsoid (MVE) estimators, while Alfaro and Ortega

(2009) and Chenouri *et al*. (2009) proposed charts based on the RMCD estimator. These methods handle outliers effectively but may incorrectly flag normal observations (swamping) and involve high computational cost. Cabana and Lillo (2022) introduced a robust multivariate robust control chart for individual observations based on the shrinkage estimators proposed by Cabana *et al*. (2021). This control chart is obtained by replacing these robust estimators of location and covariance matrix in the definition of the classical approach.

In this chapter, robust multivariate control charts are developed using two advanced estimators: the ROC estimator proposed by Sajesh and Srinivasan (2012) and the reweighted SGK estimator proposed by Vijayalakshmi *et al*. (2025).
Section 4.2 provides a detailed overview of the ROC estimator proposed by Sajesh and Srinivasan (2012). Section 4.3 presents the development of the proposed robust control charts based on both the ROC and SGK estimators. Section 4.4 describes the determination of control limits using Monte Carlo simulation, outlines the step-by-step implementation procedure for the proposed control charts, and presents the performance evaluation of the proposed methods. Their effectiveness is compared with several existing approaches, including the classical Hotelling's T^2 chart, the robust Mahalanobis distance-based reweighted Shrinkage estimator (RMDS) method of Cabana and Lillo (2022), and two RMCD-based approaches MCD1 by Alfaro and Ortega (2009) and MCD2 by Chenouri *et al*. (2009). The comparative analysis is conducted under various simulated scenarios to assess detection capability and robustness. Section 4.5 illustrates the performance of the proposed control charts using a real-world dataset, while Section 4.6 concludes the chapter by summarizing the major findings and insights.

4.2 Reweighted Orthogonalized Comedian (ROC)

Let $\mathbf{X}$ be an $n \times p$ data matrix with rows $\boldsymbol{x}_i^{\mathrm{T}}$ ($i = 1, 2, \ldots, n$) and columns $\boldsymbol{X}_j$ ($j = 1, 2, \ldots, p$). Then the comedian matrix $\mathbf{COM}(\mathbf{X})$ is defined as

$$\mathbf{COM}(\mathbf{X}) = (\mathrm{COM}\ (\boldsymbol{X}_i, \boldsymbol{X}_j)),\ i, j = 1, 2, \ldots, p, \tag{4.3}$$

where $\mathrm{COM}(X_i, X_j) = med\left\{\left(X_i - med(X_i)\right)\left(X_j - med(X_j)\right)\right\}$,

Similarly, multivariate correlation median matrix $\boldsymbol{\delta}$ is defined as,

$$\boldsymbol{\delta}\ (\mathbf{X}) = \boldsymbol{D}\mathbf{COM}(\mathbf{X})\boldsymbol{D}^{\mathrm{T}}, \tag{4.4}$$

where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements $1/MAD(\mathbf{X}_i)$) (i = 1, …, p). To overcome the non-positive semi-definiteness of comedian matrix and to obtain robust estimates for location and scatter, the following steps are adopted;

(i) Compute the eigen values λ_j and eigenvectors $\mathbf{e}_j$ of $\boldsymbol{\delta}(\mathbf{X})$ (j = 1, 2, …, p), and call $\mathbf{E}$ the matrix whose columns are the $\mathbf{e}_j$'s, so that $\boldsymbol{\delta}(\mathbf{X}) = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}'$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$.

(ii) Let $\mathbf{Q} = \mathbf{D}(\mathbf{X})^{-1}\mathbf{E}$, where $\mathbf{D}$ is defined as above and $\mathbf{z}_i = \mathbf{Q}^{-1}\mathbf{x}_i$, i = 1, 2, …, n.

(iii)The resulting robust estimates for location ($\boldsymbol{m}(\mathbf{X})$) and scatter ($\boldsymbol{S}(\mathbf{X})$) is then defined as

$$\boldsymbol{S}(\mathbf{X}) = \mathbf{Q}\boldsymbol{\Gamma}\mathbf{Q}^{\mathbf{T}} \text{ and } \boldsymbol{m}(\mathbf{X}) = \mathbf{Q}\boldsymbol{l},$$

where $\boldsymbol{\Gamma} = \text{diag}(\text{MAD}(\boldsymbol{Z}_1)^2, \ldots, \text{MAD}(\boldsymbol{Z}_p)^2)$ and $\boldsymbol{l} = (\text{med}(\boldsymbol{Z}_1), \ldots, \text{med}(\boldsymbol{Z}_1))^{\mathbf{T}}$.

The ROC estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the weighted mean vector,

$$\widehat{\boldsymbol{\mu}}_{ROC} = (\textstyle\sum_{i=1}^{n} w_i \boldsymbol{x}_i)/(\sum_{i=1}^{n} w_i), \tag{4.5}$$

and the weighted covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_{ROC} = (\textstyle\sum_{i=1}^{n} w_i (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{ROC})\,(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{ROC})^T)/(\sum_{i=1}^{n} w_i), \tag{4.6}$$

where the weights

$$w_i = \begin{cases} 1, if\ D(\boldsymbol{x}_i) \le cv \\ 0, \quad otherwise \end{cases}$$

are based on the robust distances

$$D(\boldsymbol{x}_i) = rd_i = (\boldsymbol{x}_i - \boldsymbol{m})^{\boldsymbol{T}}\boldsymbol{S}^{-\mathbf{1}}(\boldsymbol{x}_i - \boldsymbol{m}),$$

and

$$cv = 1.4826\ \frac{\chi_p^2(0.95)\text{median}\,(rd_1, \ldots, rd_n)}{\chi_p^2(0.5)}.$$

The ROC estimators of location and dispersion possess high breakdown value and they are approximately affine equivariant (Sajesh & Srinivasan, 2012).

In the statistical process control procedure, the use of ROC estimators and reweighted SGK estimators (discussed in section 2.5) eliminates the need to identify outliers in the

Phase I data, as these estimators are not influenced by outliers. Therefore, using them will provide us with robust Phase I estimates of location and covariance that will serve to compute a robust control chart for Phase II data. Next, the procedure for the main contribution of this paper, which defines a robust quality control approach based on the ROC and reweighted SGK estimators, is described in detail.

4.3 Proposed Multivariate Robust Control Charts

Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a random sample of size n drawn from multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, which is considered the Phase I data in what follows. Let $\boldsymbol{x}_j \notin \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a Phase II observation, then it is known that

$$T_j^2 \sim \frac{p(n+1)(n-1)}{n(n-p)} F_{(p,n-p)} ,$$

where T_j^2 is as defined in Equation (4.1) and $F_{(v_1,v_2)}$ is F distribution with (v_1, v_2) degrees of freedom (Wilks, 1962). In this chapter, we propose to robustify the definition of the Hotelling's T^2 control chart based on Phase I data, replacing the classical estimators with robust alternatives namely the ROC estimators $\widehat{\boldsymbol{\mu}}_{ROC}$ and $\widehat{\boldsymbol{\Sigma}}_{ROC}$ and the reweighted SGK estimators $\widehat{\boldsymbol{\mu}}_{SGK}$ and $\widehat{\boldsymbol{\Sigma}}_{SGK}$. The corresponding robust Hotelling's T^2 for the Phase II observation $\boldsymbol{x}_j$ are defined as

$$T_{ROC}^2(\boldsymbol{x}_j) = (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{ROC})^T \widehat{\boldsymbol{\Sigma}}_{ROC}^{-1}(\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}}_{ROC}), \tag{4.7}$$

$$T_{SGK}^2(\boldsymbol{x}_j) = (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{SGK})^T \widehat{\boldsymbol{\Sigma}}_{SGK}^{-1}(\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}}_{SGK}). \tag{4.8}$$

Note that this definition is similar to the definition of a squared robust Mahalanobis distance. It is well known that the distribution of the classical squared Mahalanobis distance based on sample mean and covariance matrix, is Chi-square with p degrees of freedom. When different estimators are used, this assumption does not have to be true. Although, in the literature the 97.5% quantile of the Chi-square distribution is usually used to determine a threshold for the distance to detect multivariate outliers (Gervini, 2003; Cabana & Lillo, 2022). In this study we apply Monte Carlo simulations to estimate the quantiles of the distribution of T_{ROC}^2 for several combinations of sample sizes n and dimensions p, similar as in Chenouri *et al.* (2009) and Cabana and Lillo (2022). For each dimension, a smooth curve between the sample size and quantiles of T_{ROC}^2 is fitted and used to estimate appropriate quantiles of T^2_{ROC} for different Phase I sample sizes. The same Monte Carlo procedure is applied to the T^2_{SGK} statistic to estimate its quantiles

4.3.1 Estimation of control limits for the proposed control charts

To approximate the 95%, 99%, and 99.9% quantiles of the robust T^2 statistic for a given Phase I sample size n, a Monte Carlo simulation study with M =10,000 replications was carried out. In each replication, a random sample of size n was drawn from a p variate standard normal distribution, $N_p(\mathbf{0}, \boldsymbol{I})$. For every simulated dataset, the ROC-based estimates of the location and scatter parameters, denoted by $\widehat{\boldsymbol{\mu}}_{ROC}(m)$ and $\widehat{\boldsymbol{\Sigma}}_{ROC}(m)$, and the reweighted SGK estimates, $\widehat{\boldsymbol{\mu}}_{SGK}$ and $\widehat{\boldsymbol{\Sigma}}_{SGK}$, are computed for m = 1, 2, …, M. Subsequently, a new observation $\boldsymbol{x}_{j,m}$ generated from $N_p(\mathbf{0}, \boldsymbol{I})$ to represent a Phase II sample, and its corresponding test statistics, $T^2_{ROC}\left(\boldsymbol{x}_{j,m}\right)$ and $T^2_{SGK}\left(\boldsymbol{x}_{j,m}\right)$ were evaluated. The empirical distributions of these statistics were then constructed using the simulated values

$$T^2_{ROC}\left(\boldsymbol{x}_{j,1}\right), T^2_{ROC}\left(\boldsymbol{x}_{j,2}\right), \ldots, T^2_{ROC}\left(\boldsymbol{x}_{j,M}\right),$$

and

$$T^2_{SGK}\left(\boldsymbol{x}_{j,1}\right), T^2_{SGK}\left(\boldsymbol{x}_{j,2}\right), \ldots, T^2_{SGK}\left(\boldsymbol{x}_{j,M}\right).$$

The quantiles of of $T^2_{ROC}\left(\boldsymbol{x}_j\right)$ and $T^2_{SGK}\left(\boldsymbol{x}_j\right)$, were derived by inverting their respective empirical distribution functions, resulting in Monte Carlo approximations of the 95%, 99%, and 99.9% critical values. These empirical distributions were generated for a broad set of dimensional settings, with p ranging from 2 to 15, and for Phase I sample sizes n = 20, 21, . . . ,50, 55, 60, . . . , 200. To illustrate the dependence of the quantiles on sample size, Figure 4.1 displays scatter plots of the estimated 95%, 99%, and 99.9% quantiles of $T^2_{ROC}\left(\boldsymbol{x}_j\right)$ and $T^2_{SGK}\left(\boldsymbol{x}_j\right)$ across selected dimensions (p = 3, 6 and 10). The overall trend in these plots indicates a smooth, monotonic relationship between the empirical quantiles and the Phase I sample size. The pattern observed in the empirical quantiles can be effectively modelled using the following regression relationship

$$f(n) = a_{p,1-\alpha} + \frac{b_{p,1-\alpha}}{n^{c_{p,1-\alpha}}}, \tag{4.9}$$

where $a = \chi^2_{(p,1-\alpha)}$ as suggested by Chenouri *et al.* (2009), and the parameters $b_{p,1-\alpha}$ and $c_{p,1-\alpha}$ are constants that depend on p and $1-\alpha$ as well. This functional form provides a flexible representation of the relationship between the empirical quantiles and the Phase I sample size n, allowing accurate prediction of the 95%, 99%, and 99.9% quantiles of both $T^2_{ROC}(\boldsymbol{x}_j)$ and $T^2_{SGK}(\boldsymbol{x}_j)$ for any given n.

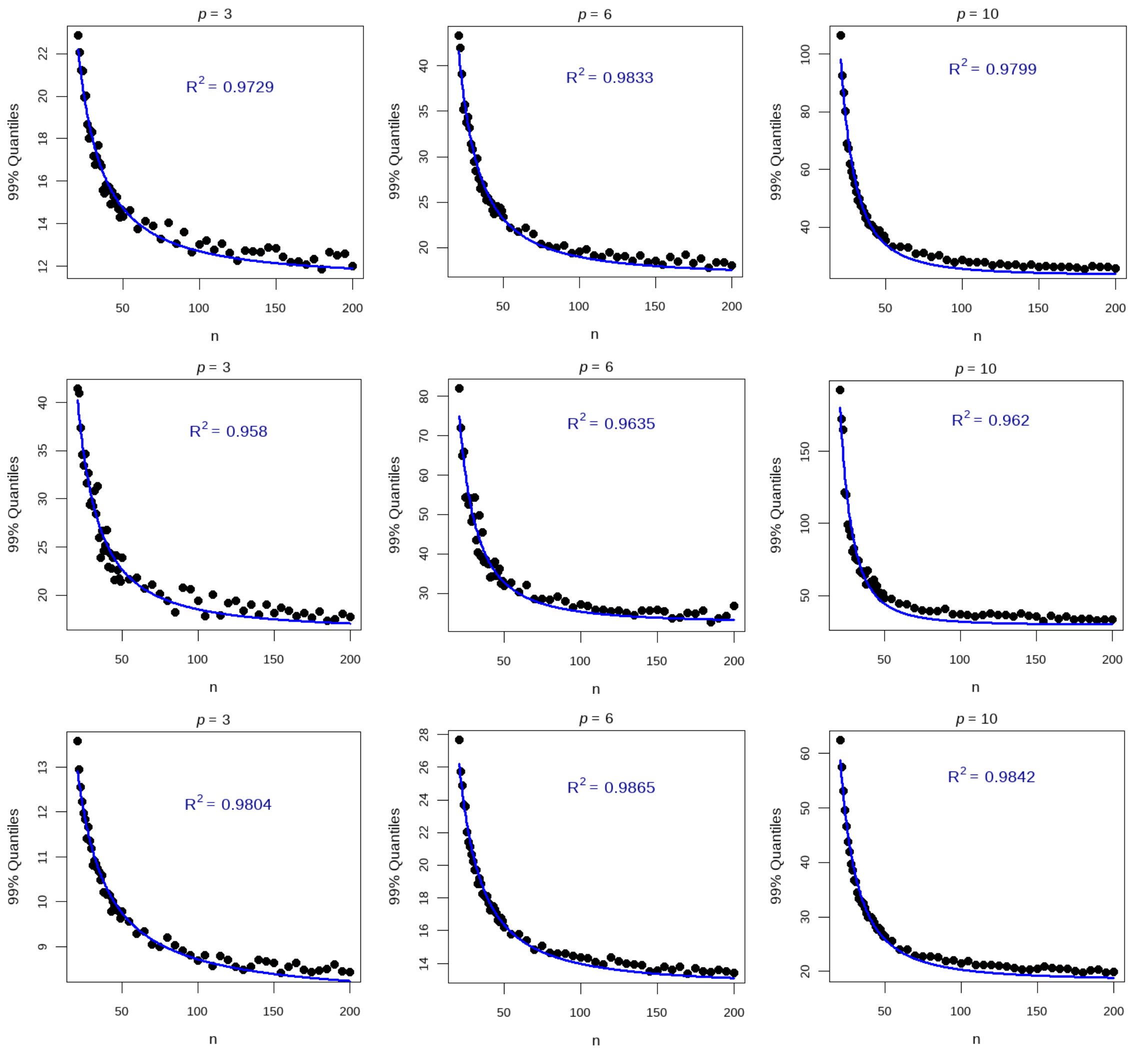


Figure 4.1: Simulated quantiles of T^2_{ROC} and the fitted curves for p = 3, 6, 10, α = 0.01 (Upper Panel), α = 0.001 (Middle Panel) and α = 0.05 (Lower Panel)

As the sample size increases, $f(n)$ approaches the theoretical quantile $\chi^2_{(p,1-\alpha)}$, which serves as the asymptotic reference in multivariate monitoring. The parameters $b_{p,1-\alpha}$ and $c_{p,1-\alpha}$, estimated via nonlinear least-squares, are listed in Table 4.1 and Table 4.2. Using these estimates in Equation (4.9) allows computation of the corresponding quantiles of $T^2_{ROC}(\boldsymbol{x}_j)$ and $T^2_{SGK}(\boldsymbol{x}_j)$ for any Phase I sample size. The fitted models show excellent agreement with the simulated data, with R^2 values consistently above 0.90, confirming their reliability for Phase II monitoring.

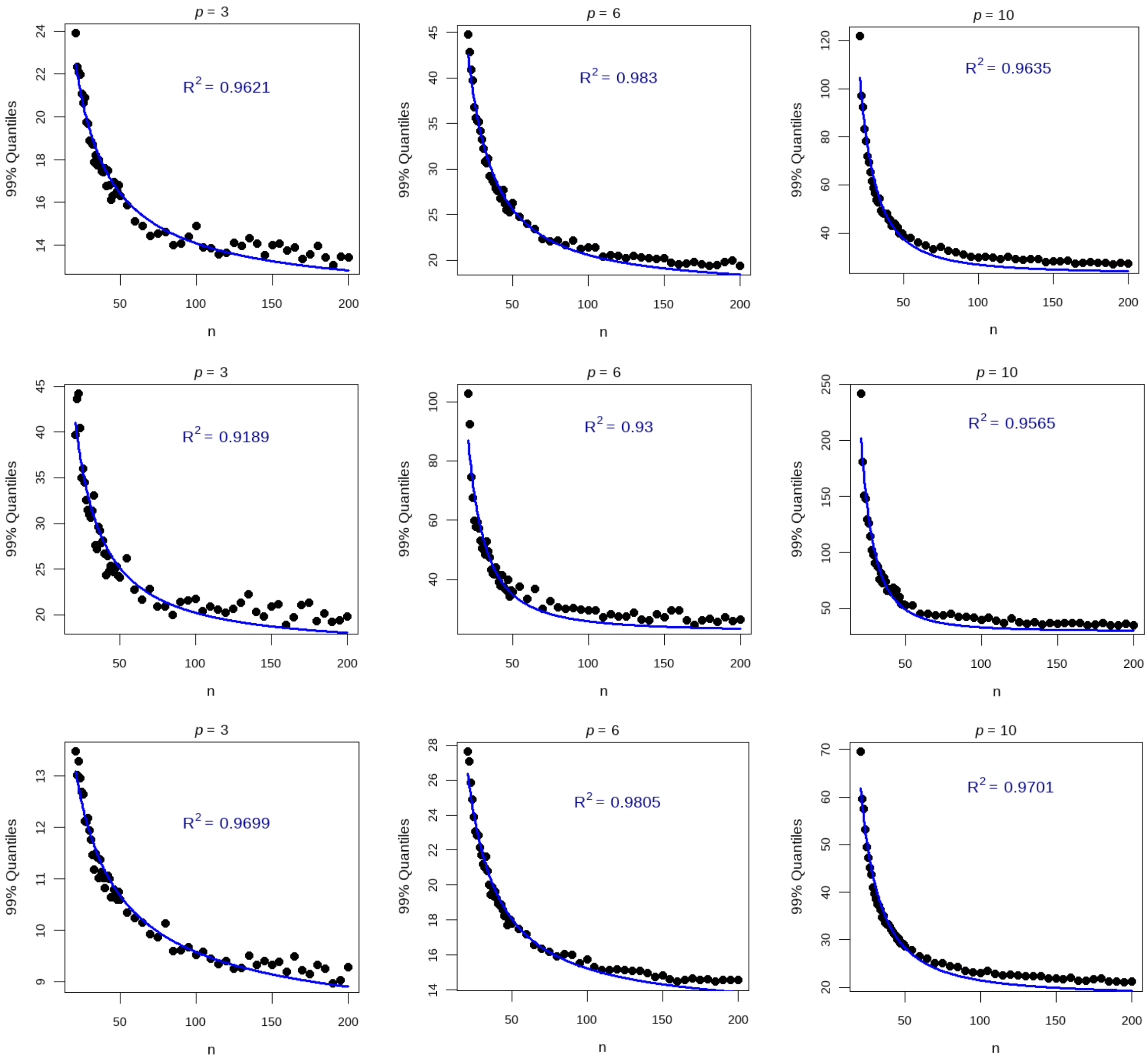


Figure 4.2: Simulated quantiles of T^2_{SGK} and the fitted curves for $p = 3, 6, 10$, $\alpha = 0.01$ (Upper Panel), $\alpha = 0.001$ (Middle Panel) and $\alpha = 0.05$ (Lower Panel)

For dimensions $p \geq 16$, a similar pattern is expected, although for a given situation, a practitioner may simulate the control limit directly. The simulations were done using Matlab software.

4.3.2 Implementation procedure

The implementation steps of the proposed ROC and SGK control charts were described in Chapter 3 for bivariate robust control chart. This procedure can be directly extended to the multivariate control chart, with the primary difference being that the estimates and control limits are computed for a general dimension p. To ensure completeness and clarity, the multivariate implementation steps are outlined below and then applied to Phase I and Phase II data.

Table 4.1: The Least-Squares estimates of the regression parameters $b_{p,1-\alpha}$, $c_{p,1-\alpha}$ for dimensions $p = 2, \ldots, 15$, and confidence levels $1 - \alpha = 0.95$, 0.99 and 0.999 (ROC)

p	**95%**		**99%**		**99.9%**	
	$b_{p,1-\alpha}$	$c_{p,1-\alpha}$	$b_{p,1-\alpha}$	$c_{p,1-\alpha}$	$b_{p,1-\alpha}$	$c_{p,1-\alpha}$
2	64.61	0.96	59.95	0.7893	6652.00	1.841
3	149.00	1.11	631.20	1.334	2356.00	1.507
4	288.60	1.21	937.10	1.374	9210.00	1.829
5	628.10	1.36	2052.00	1.528	10260.00	1.809
6	1175.00	1.47	2850.00	1.558	14780.00	1.853
7	1786.00	1.52	5728.00	1.695	48200.00	2.14
8	3680.00	1.66	10410.00	1.807	74390.00	2.207
9	5765.00	1.72	32030.00	2.077	124300.00	2.304
10	13750.00	1.92	62300.00	2.208	488100.00	2.655
11	29180.00	2.07	153600.00	2.417	3881000.00	3.208
12	97160.00	2.36	1022000.00	2.914	16770000.00	3.582
13	280500.00	2.62	1119000.00	2.952	801500000.00	4.681
14	1937000.00	3.13	2192000.00	3.168	810900000000.00	6.664
15	39630000.00	3.97	2009000.00	3.155	235900000000.00	6.166

Table 4.2: The Least-Squares estimates of the regression parameters $b_{p,1-\alpha}$, $c_{p,1-\alpha}$ for dimensions $p = 2, \ldots, 15$, and confidence levels $1 - \alpha = 0.95$, 0.99 and 0.999 (SGK)

p	**95%**		**99%**		**99.9%**	
	$b_{p,1-\alpha}$	$c_{p,1-\alpha}$	$b_{p,1-\alpha}$	$c_{p,1-\alpha}$	$b_{p,1-\alpha}$	$c_{p,1-\alpha}$
2	17.26	0.54	70.04	0.73	339.19	0.96
3	44.71	0.70	169.72	0.90	1204.20	1.26
4	96.17	0.84	329.51	1.02	1572.34	1.28
5	196.60	0.97	566.18	1.11	2307.83	1.33
6	350.62	1.06	1084.04	1.23	19600.29	1.88
7	728.25	1.20	2161.05	1.37	11769.97	1.68
8	1321.05	1.31	4507.58	1.52	18752.51	1.76
9	3084.73	1.49	1.16264E+04	1.73	176629.30	2.36
10	7158.02	1.68	3.41404E+04	1.98	3.508E+05	2.50
11	19419.78	1.91	1.75236E+05	2.40	7.454E+06	3.33
12	82355.04	2.27	1.25377E+06	2.92	1.102E+08	4.09
13	4.4466E+05	2.70	1.78246E+08	4.36	2.638E+15	9.31
14	1.1607E+07	3.61	4.69925E+11	6.75	5.779E+16	10.07
15	7.4293E+08	4.81	7.94465E+15	9.60	9.355E+20	11.26

Phase I

1. Decide on the sample size n and confidence level $1 - \alpha$.
2. Collect the Phase I data $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\}$ at well-defined periodic intervals.
3. Using the Phase I data, compute robust estimates for location and dispersion.
4. For the desired α and p values choose the least square estimates of the parameters $b_{p,1-\alpha}$ and $c_{p,1-\alpha}$ using Table 4.1 for ROC and Table 4.2 for SGK, then compute the control limit using Equation (4.9).

Phase II

5. Compute both T^2_{ROC} and T^2_{SGK} for each new observation and plot them on their respective control charts with the limit derived in Phase I (step 4).
6. Interpret the chart and look for out-of-control points or patterns. Diagnose the process if needed.

4.4 Simulation Study

To assess the performance of the proposed control charts, several simulation studies are conducted under varying Phase I data structures and different magnitudes of shifts in the process mean vector and dispersion matrix in Phase II, following the approach of Alfaro and Ortega (2009). The efficiency of the chart is evaluated based on its capability to identify process changes and its tendency to generate false alarms when different estimators are applied in Phase I. The performance metrics include the success rate defined as the proportion of statistic values that exceed the control limits across 1000 replications, representing the probability of detecting shifts and the false alarm rate representing the frequency of incorrect detections of process changes based on Phase II data. A contaminated model is employed, represented by a mixture of multivariate normal distributions as

$$(1-\varepsilon)\ \mathrm{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1),$$

where ε is the proportion (in as much by one) of outliers, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the in-control parameters (mean vector and covariance matrix respectively) and $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are the out-of-control parameters. Different simulation scenarios are examined by varying n, p, ε and method of generating the outliers using different values of $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$. . Without loss of generality, $\boldsymbol{\mu}_0$ is taken as a zero vector. The contamination proportion ε is considered at levels 0, 0.1, and 0.2, and the probability of detecting a process shift is influenced by the chosen values of $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$.

The simulation scenarios considered in this chapter, Cases A, Case B, and Case C, are the same as those described in Chapter 3. These cases are designed to evaluate the performance of the proposed robust control charts under different conditions. Case A examines mean shifts with independent variables, Case B incorporates correlated variables to assess the effect of correlation on detection probability, and Case C includes both mean and covariance changes to study the impact of good leverage and regression outliers. The full mathematical definitions and parameter settings are provided in Section 3.5.

Using 1000 replications, we considered different sample sizes where n = 50, 100 and 200 with dimensions 2, 5 and 10 and an overall false alarm probability of α = 0.001, 0.01and 0.05. We generated 1000 samples of size n in Phase I and computed the robust and traditional location and scale estimates for each data set. Performance of the proposed control charts is compared with that of the charts based on MCD1, MCD2, RMDS and the classical estimates using the parameters SR and FAR. We present only the result for α = 0.01 and u = 3 since similar pattern conclusions hold for other values of α and u.

Table 4.3: SR (FAR) in Phase II when there are ε proportion of outliers in Phase I data from Case A.

n	p	ε	ROC	SGK	MCD2	MCD1	RMDS	Classical
50	2	0	1 (0.012)	1 (0.015)	1 (0.038)	0.997 (0.01)	1 (0.009)	1 (0.011)
		0.1	0.995 (0.007)	1 (0.009)	0.995 (0.03)	0.997 (0.002)	1 (0.006)	0.929 (0.002)
		0.2	0.945 (0.008)	0.995 (0.001)	0.991 (0.023)	0.966 (0.001)	0.939 (0.001)	0.012 (0.006)
	5	0	1 (0.014)	1 (0.005)	1 (0.053)	1 (0.009)	1 (0.008)	1 (0.109)
		0.1	1 (0.009)	1 (0.011)	1 (0.038)	1 (0.006)	1 (0.012)	1 (0.088)
		0.2	1 (0.015)	1 (0.012)	1 (0.015)	0.988 (0.003)	0.999 (0.011)	0.247 (0.078)
	10	0	1 (0.012)	1 (0.009)	1 (0.062)	1 (0.013)	1 (0.018)	1 (0.605)
		0.1	1 (0.022)	1 (0.011)	1 (0.037)	1 (0.005)	1 (0.009)	1 (0.529)
		0.2	1 (0.019)	1 (0.026)	1 (0.007)	0.889 (0.001)	1 (0.012)	0.822 (0.501)
100	2	0	1 (0.011)	1 (0.02)	1 (0.03)	1 (0.011)	1 (0.005)	1 (0.006)
		0.1	1 (0.007)	1 (0.01)	1 (0.02)	1 (0.01)	1 (0.007)	0.944 (0.005)
		0.2	1 (0.012)	0.999 (0.008)	1 (0.02)	0.999 (0.009)	0.978 (0.005)	0.023 (0.004)
	5	0	1 (0.012)	1 (0.008)	1 (0.048)	1 (0.006)	1 (0.012)	1 (0.096)
		0.1	1 (0.009)	1 (0.014)	1 (0.03)	1 (0.003)	1 (0.01)	1 (0.073)
		0.2	1 (0.009)	1 (0.011)	1 (0.018)	1 (0.011)	1 (0.005)	0.269 (0.085)
	10	0	1 (0.019)	1 (0.007)	1 (0.055)	1 (0.008)	1 (0.012)	1 (0.537)

		0.1	1 (0.022)	1 (0.015)	1 (0.036)	1 (0.012)	1 (0.006)	1 (0.464)
		0.2	1 (0.025)	1 (0.01)	1 (0.039)	1 (0.016)	1 (0.01)	0.816 (0.496)
200	2	0	1 (0.007)	1 (0.006)	1 (0.014)	1 (0.011)	1 (0.007)	1 (0.01)
		0.1	1 (0.006)	1 (0.008)	1 (0.014)	1 (0.005)	1 (0.008)	0.971 (0.004)
		0.2	1 (0.005)	1 (0.004)	1 (0.012)	1 (0.002)	0.984 (0.006)	0.03 (0.004)
	5	0	1 (0.013)	1 (0.009)	1 (0.022)	1 (0.018)	1 (0.008)	1 (0.08)
		0.1	1 (0.014)	1 (0.006)	1 (0.019)	1 (0.011)	1 (0.008)	1 (0.08)
		0.2	1 (0.012)	1 (0.011)	1 (0.021)	1 (0.004)	1 (0.011)	0.247 (0.059)
	10	0	1 (0.013)	1 (0.015)	1 (0.033)	1 (0.01)	1 (0.013)	1 (0.498)
		0.1	1 (0.016)	1 (0.008)	1 (0.025)	1 (0.008)	1 (0.01)	1 (0.445)
		0.2	1 (0.017)	1 (0.008)	1 (0.028)	1 (0.01)	1 (0.009)	0.799 (0.491)

Table 4.3 presents the results obtained for Case A contamination. It is clear from the table that, in Case A, the main factor affecting performance is the proportion of outliers: as ε increases, the classical Hotelling's T^2 chart deteriorates sharply with high false alarms, while robust charts maintain high success rates and low variability. Larger sample sizes n reduce variability for all methods but do not resolve the classical chart's sensitivity to contamination, whereas ROC and SGK remain stable across n. The effect of dimension p is less pronounced; at higher p, ROC, SGK, MCD1, and RMDS continue to perform reliably, while MCD2 shows greater inconsistency. Overall, contamination strongly impacts the classical chart, but the proposed ROC and SGK demonstrate robustness across different dimensions and sample sizes.

Table 4.4: SR (FAR) in Phase II when there are ε proportion of outliers in Phase I data from Case B.

n	p	ε	**ROC**	**SGK**	**MCD2**	**MCD1**	**RMDS**	**Classical**
50	2	0	0.976 (0.01)	1 (0.006)	0.987 (0.037)	0.953 (0.006)	0.972 (0.01)	0.97 (0.016)
		0.1	0.952 (0.007)	0.993 (0.008)	0.952 (0.021)	0.839 (0.004)	0.919 (0.009)	0.713 (0.005)
		0.2	0.9 (0.009)	0.902 (0.002)	0.941 (0.017)	0.545 (0.004)	0.676 (0.003)	0.018 (0.005)
	5	0	0.947 (0.028)	0.998 (0.013)	0.874 (0.057)	0.811 (0.011)	0.914 (0.01)	0.994 (0.095)

		0.1	0.888 (0.011)	0.947 (0.006)	0.775 (0.033)	0.543 (0.011)	0.729 (0.021)	0.899 (0.095)
		0.2	0.886 (0.014)	0.668 (0.008)	0.762 (0.021)	0.217 (0.011)	0.369 (0.009)	0.243 (0.078)
	10	0	0.931 (0.031)	0.984 (0.014)	0.556 (0.046)	0.523 (0.01)	0.748 (0.031)	0.999 (0.607)
		0.1	0.854 (0.029)	0.805 (0.011)	0.345 (0.046)	0.247 (0.007)	0.533 (0.017)	0.992 (0.572)
		0.2	0.83 (0.031)	0.448 (0.01)	0.258 (0.03)	0.081 (0.003)	0.223 (0.016)	0.79 (0.504)
100	2	0	0.985 (0.022)	1 (0.013)	0.988 (0.035)	0.982 (0.015)	0.973 (0.01)	0.977 (0.011)
		0.1	0.951 (0.005)	0.999 (0.005)	0.963 (0.012)	0.932 (0.005)	0.933 (0.006)	0.781 (0.006)
		0.2	0.909 (0.006)	0.96 (0.002)	0.952 (0.011)	0.773 (0.001)	0.663 (0.001)	0.454 (0.008)
	5	0	0.973 (0.022)	0.999 (0.015)	0.976 (0.035)	0.936 (0.011)	0.944 (0.027)	0.995 (0.119)
		0.1	0.926 (0.012)	0.98 (0.005)	0.878 (0.026)	0.771 (0.014)	0.752 (0.006)	0.921 (0.074)
		0.2	0.875 (0.016)	0.78 (0.005)	0.865 (0.021)	0.336 (0.007)	0.304 (0.012)	0.25 (0.066)
	10	0	0.964 (0.035)	0.996 (0.01)	0.908 (0.051)	0.788 (0.01)	0.839 (0.032)	0.999 (0.555)
		0.1	0.909 (0.029)	0.909 (0.005)	0.738 (0.047)	0.444 (0.006)	0.512 (0.021)	0.989 (0.526)
		0.2	0.887 (0.028)	0.54 (0.009)	0.727 (0.033)	0.15 (0.006)	0.181 (0.015)	0.771 (0.49)
200	2	0	0.993 (0.015)	1 (0.008)	0.997 (0.024)	0.978 (0.005)	0.985 (0.01)	0.977 (0.01)
		0.1	0.954 (0.01)	0.999 (0.005)	0.972 (0.016)	0.95 (0.008)	0.947 (0.01)	0.793 (0.001)
		0.2	0.925 (0.01)	0.962 (0.001)	0.965 (0.017)	0.79 (0.005)	0.686 (0.006)	0.029 (0.004)
	5	0	0.978 (0.016)	1 (0.009)	0.977 (0.026)	0.948 (0.012)	0.951 (0.034)	0.993 (0.104)
		0.1	0.923 (0.022)	0.992 (0.006)	0.897 (0.031)	0.802 (0.005)	0.78 (0.019)	0.917 (0.079)
		0.2	0.858 (0.014)	0.796 (0.005)	0.88 (0.019)	0.35 (0.007)	0.347 (0.012)	0.263 (0.077)
	10	0	0.954 (0.015)	0.999 (0.007)	0.944 (0.029)	0.876 (0.008)	0.904 (0.026)	0.999 (0.521)
		0.1	0.884 (0.021)	0.932 (0.007)	0.833 (0.036)	0.579 (0.13)	0.494 (0.027)	0.987 (0.504)
		0.2	0.84 (0.023)	0.611 (0.005)	0.822 (0.024)	0.184 (0.008)	0.187 (0.03)	0.753 (0.444)

Simulation results corresponding to Case B contamination is given in Table 4.4. In Case B, ROC and SGK consistently achieve higher success rates than the other methods across different dimensions, sample sizes, and contamination levels. MCD2 shows lower success rates for small sample sizes and high dimensions (p = 10) but improves as n increases. MCD1 and RMDS perform well for low-dimensional data (p = 2,5) with up to 10% outliers, but their performance declines as dimension and contamination increase. Across all settings, false alarm rates follow a pattern similar to Case A, robust methods maintain low false alarms under contamination, while the classical Hotelling's T^2 chart is highly sensitive to outliers, showing substantial deterioration in performance as the proportion of outliers rises.

Table 4.5: SR (FAR) in Phase II when there are ε proportion of outliers in Phase I data from Case C.

n	p	ε	ROC	SGK	MCD2	MCD1	RMDS	Classical
50	2	0	1 (0.021)	1 (0.009)	1 (0.037)	0.999 (0.007)	1 (0.007)	1 (0.005)
		0.1	0.98 (0.005)	1 (0.006)	0.981 (0.027)	0.884 (0.001)	0.926 (0.005)	0.97 (0.003)
		0.2	0.968 (0.015)	0.999 (0.008)	0.973 (0.023)	0.405 (0.006)	0.383 (0.003)	0.025 (0.004)
	5	0	0.997 (0.015)	1 (0.01)	0.956 (0.057)	0.941 (0.009)	0.994 (0.02)	1 (0.122)
		0.1	0.951 (0.015)	0.996 (0.005)	0.928 (0.038)	0.479 (0.002)	0.61 (0.018)	0.997 (0.089)
		0.2	0.975 (0.013)	0.735 (0.01)	0.9 (0.027)	0.115 (0.005)	0.177 (0.005)	0.215 (0.078)
	10	0	0.982 (0.034)	1 (0.019)	0.538 (0.055)	0.512 (0.01)	0.873 (0.036)	1 (0.569)
		0.1	0.948 (0.033)	0.811 (0.013)	0.359 (0.05)	0.177 (0.005)	0.342 (0.028)	1 (0.572)
		0.2	0.957 (0.035)	0.415 (0.011)	0.293 (0.033)	0.051 (0.008)	0.116 (0.015)	0.778 (0.532)
100	2	0	1 (0.016)	1 (0.017)	1 (0.028)	1 (0.01)	1 (0.013)	1 (0.012)
		0.1	0.997 (0.011)	1 (0.005)	0.999 (0.017)	0.995 (0.01)	0.983 (0.007)	0.995 (0.007)
		0.2	0.979 (0.005)	1 (0.003)	0.997 (0.01)	0.631 (0.008)	0.298 (0.004)	0.414 (0.004)
	5	0	1 (0.013)	1 (0.006)	1 (0.033)	1 (0.008)	1 (0.032)	1 (0.117)
		0.1	0.968 (0.017)	1 (0.007)	0.94 (0.029)	0.746 (0.007)	0.646 (0.009)	1 (0.07)
		0.2	0.977 (0.006)	0.969 (0.005)	0.894 (0.019)	0.156 (0.006)	0.146 (0.01)	0.217 (0.066)
	10	0	1 (0.025)	1 (0.014)	0.988 (0.054)	0.924 (0.01)	0.969 (0.023)	1 (0.533)

		0.1	0.966 (0.035)	0.987 (0.011)	0.902 (0.052)	0.32 (0.006)	0.346 (0.018)	1 (0.501)
		0.2	0.987 (0.036)	0.551 (0.007)	0.89 (0.041)	0.069 (0.01)	0.119 (0.021)	0.772 (0.499)
200	2	0	1 (0.015)	1 (0.008)	1 (0.022)	1 (0.016)	1 (0.012)	1 (0.009)
		0.1	0.998 (0.011)	1 (0.003)	1 (0.017)	1 (0.003)	0.999 (0.006)	1 (0.003)
		0.2	0.985 (0.007)	1 (0.005)	1 (0.012)	0.747 (0.005)	0.357 (0.004)	0.492 (0.005)
	5	0	1 (0.018)	1 (0.008)	1 (0.031)	1 (0.007)	1 (0.032)	1 (0.097)
		0.1	0.979 (0.01)	1 (0.006)	0.986 (0.014)	0.848 (0.011)	0.758 (0.027)	0.92 (0.072)
		0.2	0.958 (0.014)	0.996 (0.006)	0.997 (0.021)	0.15 (0.01)	0.195 (0.015)	0.746 (0.008)
	10	0	1 (0.021)	1 (0.006)	0.999 (0.032)	0.991 (0.014)	0.998 (0.043)	1 (0.527)
		0.1	0.985 (0.019)	1 (0.003)	0.988 (0.032)	0.401 (0.013)	0.391 (0.024)	1 (0.473)
		0.2	0.979 (0.016)	0.672 (0.005)	0.976 (0.029)	0.103 (0.01)	0.121 (0.022)	0.739 (0.473)

The results for Case C contamination, presented in Table 4.5, indicate that both ROC and SGK based control charts achieve consistently high success rates across different sample sizes, dimensions, and contamination levels, with ROC being slightly more stable overall. SGK performs nearly identically to ROC in low and moderate dimensions (p = 2, 5) and small to moderate contamination levels, but at higher dimension (p = 10) and contamination ($\varepsilon = 0.2$), its success rate declines somewhat more than ROC, though still outperforming the classical chart. False alarm rates for SGK remain low and stable, similar to ROC, indicating strong robustness. By contrast, MCD2 achieves high success rates but with higher and more variable false alarm rates, while MCD1 and RMDS perform adequately in low dimensions and mild contamination but deteriorate under more challenging settings. The classical Hotelling's T^2 chart shows the steepest decline under contamination, confirming the advantage of robust approaches.

4.5 Time Complexity

The computation times (in seconds) for six covariance estimators ROC, SGK, MCD2, MCD1, RMDS, and the classical method were evaluated under varying sample sizes (n = 50, 100, 200) and dimensions (p = 2, 5, 10). The results presented in Table 4.6 show that computation time increases with both sample size and dimensionality for all methods, as larger and higher-dimensional datasets require more intensive matrix operations. Among all methods, the classical estimator is consistently the fastest due to its straightforward computation without robust iterations.

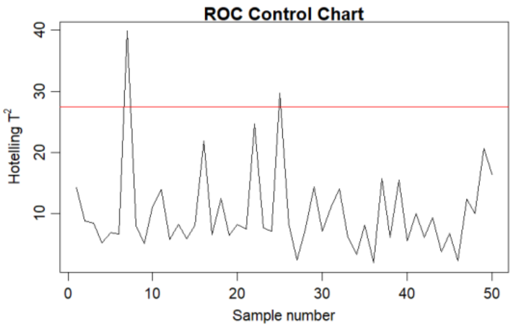
Table 4.6: Computation Time (in seconds)

n	p	ROC	SGK	MCD2	MCD1	RMDS	Classical
50	2	0.0011	0.0032	0.0039	0.0044	0.0030	0.0005
	5	0.0017	0.0036	0.0067	0.0073	0.0034	0.0005
	10	0.0026	0.0076	0.0147	0.0163	0.0066	0.0005
100	2	0.0016	0.0048	0.0069	0.0072	0.0072	0.0009
	5	0.0018	0.0065	0.0112	0.0116	0.0057	0.0009
	10	0.0032	0.0133	0.0272	0.0267	0.0078	0.0009
200	2	0.0018	0.0067	0.0131	0.0131	0.0080	0.0015
	5	0.0022	0.0116	0.0200	0.0197	0.0097	0.0015
	10	0.0045	0.0251	0.0521	0.0494	0.0120	0.0017

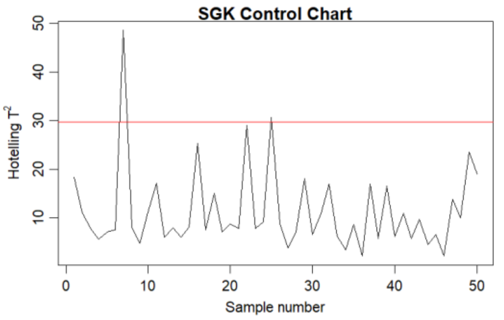
Among robust approaches, ROC exhibits the highest computational efficiency, with times only slightly greater than the classical method, followed by SGK, which remains reasonably fast and scalable. RMDS performs moderately, while MCD1 and MCD2 are the slowest, showing a rapid increase in time as the dimension grows. For example, at $n = 200$ and $p = 10$, MCD2 requires 0.0521 seconds compared to only 0.0045 seconds for ROC. Overall, the ranking of computational efficiency is: Classical < ROC < SGK < RMDS < MCD1 < MCD2. These results indicate that ROC and SGK provide an excellent balance between robustness and computational efficiency, making them suitable for practical, real-time process monitoring applications where both speed and reliability are essential.

4.6 Application to Real Life Data

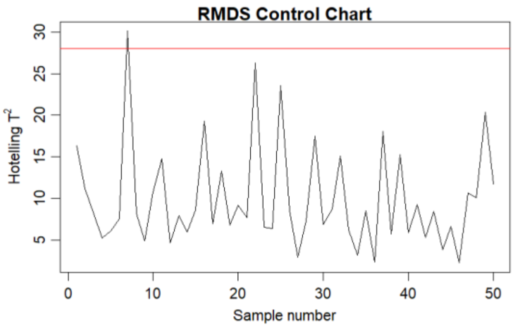
To demonstrate the application of the proposed robust multivariate control charts, we employ a mechanical process dataset consisting of seven interrelated quality characteristics reported by Santos-Fernández (2012). The dataset is organized into two phases: Phase I contains $m = 45$ samples for each characteristic, while Phase II includes 50 additional samples. Figure 4.3 presents the control charts constructed using ROC, SGK, RMDS, RMCD, and the Classical method. As shown, all robust estimators flagged observations 7 and 25 as out-of-control, consistent with the classical method, which performs best under uncontaminated data.



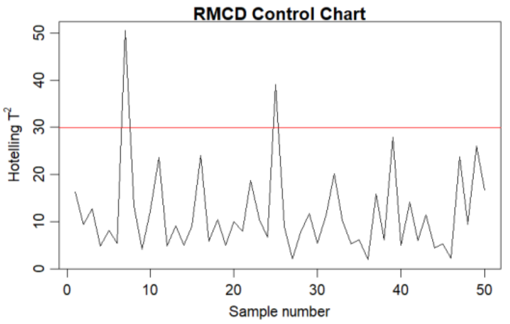
(*a*)



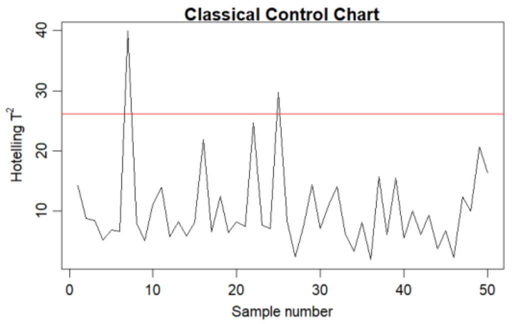
(*b*)



(*c*)



(*d*)



(*e*)

Figure 4.3: Control charts based on (*a*) ROC, (*b*) SGK, (*c*) RMDS, (*d*) RMCD and (*e*) Classical estimators under uncontaminated Phase I data.

To evaluate robustness under contamination, the Phase I data were deliberately altered by introducing a mean shift in the first three observations. Specifically, the mean of the length characteristic was increased by 3σ. The resulting control charts, displayed in Figure 4.4, reveal that the ROC, SGK, and RMCD methods continued to identify observations 7 and 25 as out-of-control, whereas the RMDS and classical charts incorrectly treated them as in-control. This outcome highlights the enhanced robustness of the proposed methods when data contamination is present.

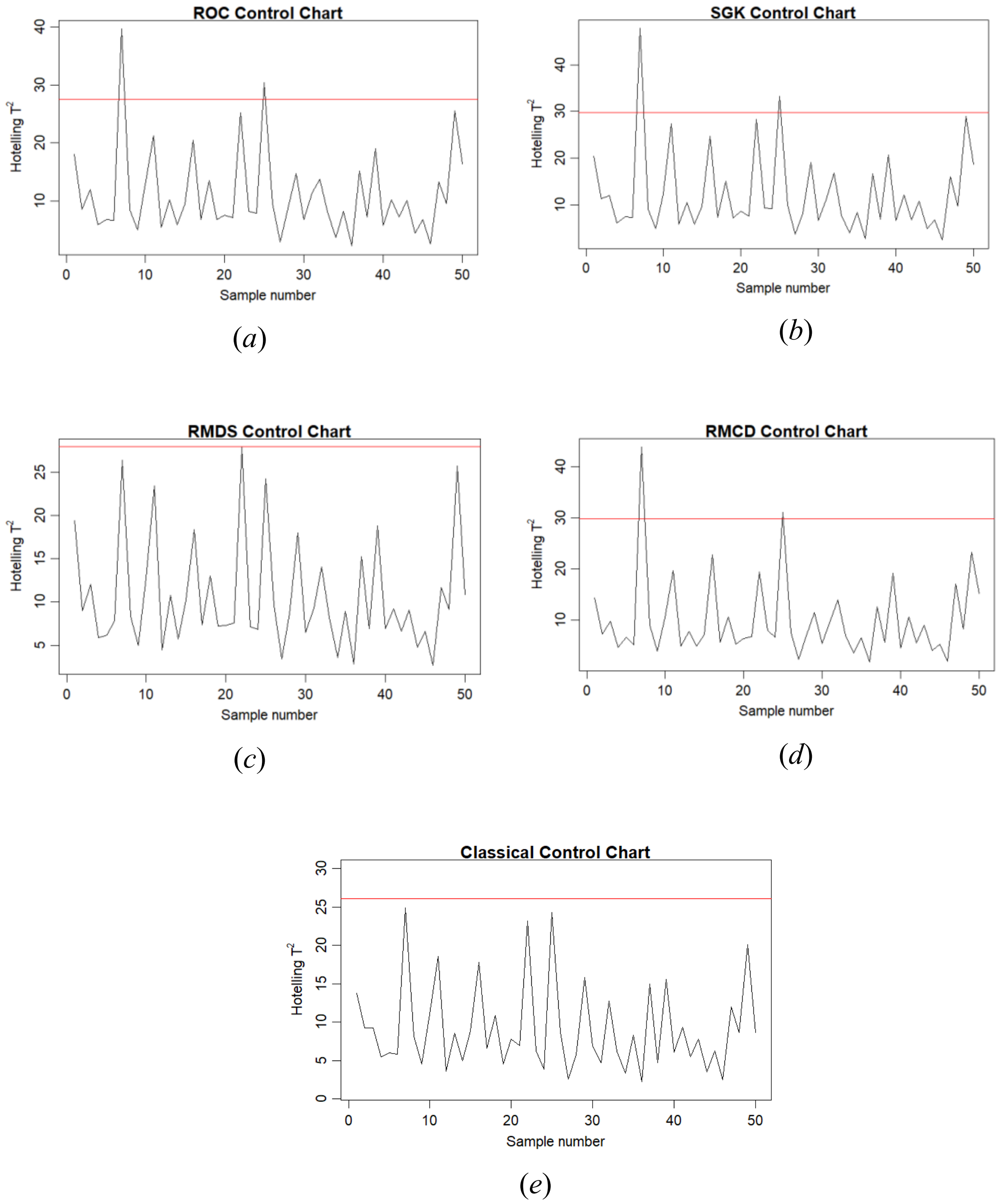


Figure 4.4: Control charts based on (*a*) ROC, (*b*) SGK, (*c*) RMDS, (*d*) RMCD, and (*e*) Classical estimators under contaminated Phase I data.

4.7 Summary

The findings in this chapter confirm that the proposed multivariate robust control charts based on ROC and SGK estimators provide reliable detection of Phase II process shifts while maintaining low false alarm rates. Across Cases A, B, and C, simulation studies show that both charts perform consistently well under varying sample sizes, dimensions, contamination levels, and correlation structures. ROC demonstrates slightly more stable performance in higher-dimensional settings and under heavier contamination, while SGK performs comparably in moderate dimensions and contamination levels. Compared with classical Hotelling's T^2 and other robust approaches such as MCD1, MCD2, and RMDS, the proposed charts achieve higher success rates and better control of false alarms. Application to real-world datasets further validates their practical utility, confirming that both ROC and SGK are effective and robust alternatives for multivariate process monitoring.

CHAPTER 5

BIVARIATE ROBUST CONTROL CHART FOR PROCESS MONITORING WITH RATIONAL SUBGROUPS

5.1 Introduction

The effective monitoring of process quality in industrial and service settings often involves assessing multiple interrelated quality characteristics. In such cases, multivariate control charts are indispensable tools for detecting shifts in the process mean vector, ensuring that deviations from the desired operational state are promptly identified. Among these, bivariate control charts serve as a focused yet powerful method for managing processes involving two correlated variables. For rational subgroups, where observations are logically grouped based on temporal or operational proximity, it is critical to develop control charts that not only account for the interdependence of variables but also provide precise monitoring of subgroup dynamics.

Alt (1985) defined the concept of two distinct phases, Phases I and Phase II, in constructing control charts. In Phase I, historical observations are utilized to assess the process's stability and estimate the parameters of the in-control process, as well as to establish control limits. These control limits serve as reference boundaries for subsequent monitoring. In Phase II, the estimated parameters and control limits obtained from Phase I are applied to analyse the data collected during the actual manufacturing process, enabling the detection of any deviations or out-of-control signals. The purpose of Phase II is to monitor and maintain the process's stability based on the established control limits. One commonly used approach in bivariate process monitoring and control is the Hotelling's T^2 control chart, which focuses on monitoring the mean vector of the process (Montgomery, 2020).

For a random sample $\boldsymbol{x}_i$ of size n of a bivariate quality characteristic, the Hotelling's T^2 statistic is defined as

$$T_i^2 = n(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}), \tag{5.1}$$

where $\overline{\boldsymbol{x}}_i$ is the sample mean vector of $\boldsymbol{x}_i$. When the process is in statistical control, it is assumed that $\boldsymbol{x}_i$ is independent and follows a bivariate normal distribution $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ represents the mean vector and $\boldsymbol{\Sigma}$ represents the covariance matrix. If both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, T^2 follows a Chi-square distribution with 2 degrees of freedom. However, in practice, the true values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are generally unknown. In such cases, these parameters are usually substituted with classical estimators of mean vector and covariance matrix.

5.2 Computational Algorithm for Bivariate Control Chart

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ be m random samples of a bivariate quality characteristic, where each sample consists of n observations. The following steps outline the procedure for constructing a Hotelling's T^2 control chart:

1. *Calculate the Mean Estimator*: For each subgroup i, calculate the sample mean vector $\overline{\boldsymbol{x}}_i = (\bar{x}_{i1}, \bar{x}_{i2})^{\boldsymbol{T}}$, where:

   $$\bar{x}_{ij} = \frac{1}{n}\sum_{k=1}^{n} x_{ijk} \text{ for } i = 1, 2, \ldots, m;\ j = 1, 2.$$

   The overall sample mean vector $\overline{\overline{\boldsymbol{x}}}$ is computed as the average of the subgroup means:

   $$\overline{\overline{\boldsymbol{x}}} = \frac{1}{m}\sum_{i=1}^{m} \overline{\boldsymbol{x}}_i. \tag{5.2}$$

2. *Calculate the Sample Covariance Matrix for Each Subgroup*: For each subgroup i, calculate the sample covariance matrix $\boldsymbol{S}_i$ of the two random variables. The covariance matrix is given by:

   $$\boldsymbol{S}_i = \begin{pmatrix} S_{i11} & S_{i12} \\ S_{i12} & S_{i22} \end{pmatrix}, \tag{5.3}$$

   where S_{ijj} represents the variance of the j^{th} variable in i^{th} subgroup, and S_{i12} represents the covariance between the two variables in the i^{th} subgroup.

3. *Compute the Average Covariance Matrix*: The sample covariance estimator $\boldsymbol{S}$ is obtained by averaging the subgroup covariance matrices:

   $$\boldsymbol{S} = \begin{pmatrix} \overline{\boldsymbol{S}}_1 & \overline{\boldsymbol{S}}_{12} \\ \overline{\boldsymbol{S}}_{12} & \overline{\boldsymbol{S}}_2 \end{pmatrix}, \tag{5.4}$$

   where $\overline{\boldsymbol{S}}_j = \frac{1}{m}\sum_{i=1}^{m} S_{ijj}$; j=1,2 and $\bar{S}_{12} = \frac{1}{m}\sum_{i=1}^{m} S_{i12}$.

4. *Compute the Hotelling's T^2 Statistic*: The Hotelling's T^2 statistic for each subgroup i is computed as follows:

$$T_i^2 = n(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})^T \boldsymbol{S}^{-1} (\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}}). \tag{5.5}$$

5. *Estimation of parameters for Phase II application*: Plot the calculated values of T^2 on the control chart. If any T^2 value falls outside the control limits, remove the corresponding samples and re-estimate the location and dispersion parameters using the remaining sample subgroups for the phase II applications. UCLs corresponding to Phase I and Phase II are available in Montgomery (2020).

However, the classical estimators of mean vector and covariance matrix are often sensitive to outliers and non-normal data, can produce inaccurate findings. Addressing these limitations, robust control charts have emerged as a powerful alternative, offering improved performance under non-ideal conditions. Several methods have been proposed to monitor mean shifts for individual samples (see Vargas, 2003; Chenouri *et al*., 2009; Jensen *et al*., 2007; Sajesh, 2023; Vijayalakshmi *et al*., 2024).

However, robust alternatives specifically designed for rational subgroups remain scarce. Abu-Shawiesh *et al*. (2014) proposed a robust bivariate control chart for rational subgroups as an alternative to the Hotelling's T^2 control chart. This method makes use of component-wise median and Comedian estimator proposed by Falk (1997) to compute robust Hotelling's T^2 statistic. In this procedure the robust estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are obtained by averaging the subgroup component-wise median vectors and the robust sample covariance matrices over the m subgroups. Main drawback of this method is that, the robust estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are affected by even a single locally contaminated outlier, that means all the observations in the subgroup are outliers. Consequently, the primary advantage of robust method will be lost.

Maleki *et al*. (2020) and Mahpouya *et al*. (2022) partially addressed this limitation by estimating $\boldsymbol{\mu}$ as the median of the component-wise medians across subgroups. Both studies utilized the Comedian estimator to compute the covariance matrix at the subgroup level. Similar to Abu Shawiesh *et al*. (2014), the overall robust estimate of the process covariance matrix $\boldsymbol{\Sigma}$ is obtained by averaging these robust sample covariance matrices across the m subgroups. Although these methods enhance robustness against outliers caused by mean shifts, their performance can still be compromised by even a single locally contaminated outlier that distorts the covariance structure, thereby affecting the reliability of the control chart.

This study proposes a procedure to construct robust bivariate robust control chart that can successfully defend all types of outliers and useful for Phase II application. The proposed control chart estimates $\boldsymbol{\mu}$ by computing the L_1 median of subgroup-wise L_1 median vectors while the $\boldsymbol{\Sigma}$ is robustly estimated from the combined dataset across all subgroups. Three robust covariance estimators are employed for this study: the Gnanadesikan–Kettenring (GK) estimator (Gnanadesikan & Kettenring, 1972), the Comedian estimator (Falk, 1997), and the S_n Covariance (Sajana & Sajesh, 2020), which are briefly described in Chapter 2, Section 2.4. For the robust GK estimator, the Q_n scale estimator of Rousseeuw and Croux (1993) is used. These estimators are chosen due to their high breakdown point, defined as the maximum proportion of outliers that the estimator can reliably withstand. The Lower control limits (LCL) of these control charts are set to zero and the upper control limits (UCL) of the control charts are estimated empirically through simulations and the results are compared with each other. These control charts are also compared with the Comedian based chart proposed by Maleki *et al.* (2020), hereafter referred to as the Avg-Comedian chart, using the Average Run Length as the performance metric.

5.3 Bivariate Robust Control Chart

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ represent m random samples of a bivariate quality characteristic, each of size n. The following steps describe the construction of a robust bivariate control chart:

1. *Compute the L_1 Median for Each Sample*: For each sample i, compute the L_1 median $\tilde{\boldsymbol{x}}_i$, defined as:

$$\tilde{\boldsymbol{x}}_i = \arg\min_{\boldsymbol{y}\in\mathbb{R}^2} \sum_{j=1}^{n} \left\| \boldsymbol{y} - \boldsymbol{x}_{ij} \right\|, \tag{5.6}$$

   where $\boldsymbol{x}_{ij}$ is the j^{th} observation of i^{th} sample.

2. *Compute the L_1 Median of the Sample Medians*: Compute the L_1 median of the sample medians $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots, \tilde{\boldsymbol{x}}_m$, denoted as $\widetilde{\boldsymbol{M}}_R$, which serves as a robust estimator of the population mean $\boldsymbol{\mu}$.

3. *Estimate the Covariance Matrix Using Robust Estimation*: The proposed chart uses the robust estimator of $\boldsymbol{\Sigma}$ based on the whole set of data, denoted as $\tilde{\boldsymbol{S}}_R$, instead of the $\boldsymbol{S}$ matrix given in the construction of Hotelling's T^2 chart.

4. *Compute the Robust Hotelling's T^2 Statistic*: For a new observation with data matrix $\boldsymbol{z}$, compute robust Hotelling's T^2 statistic using the robust estimators $\widetilde{\boldsymbol{M}}_R$ and $\tilde{\boldsymbol{S}}_R$ as follows:

$$T_{R_i}^{\,2} = n\left(\tilde{\boldsymbol{z}} - \widetilde{\boldsymbol{M}}_R\right)^T \tilde{\boldsymbol{S}}_R^{\,-1}\left(\tilde{\boldsymbol{z}} - \widetilde{\boldsymbol{M}}_R\right), \tag{5.7}$$

where $\tilde{\boldsymbol{z}}$ is the L_1 median of the new observation $\boldsymbol{z}$.

5. *Plot the Statistic on the Control Chart*: Plot the computed statistic on the control chart with the suitable UCL empirically obtained through simulation for a given α risk.

Table 5.1: Simulated and exact Upper control limits

n	α	S_n Covariance	Comedian	GK	Exact
5	0.05	6.781	6.992	6.873	6.618
	0.01	10.629	11.015	10.756	10.387
	0.0027	13.642	14.697	13.926	13.568
	0.001	16.060	17.401	16.711	16.053
10	0.05	6.946	7.030	6.913	6.433
	0.01	10.739	11.194	10.736	9.979
	0.0027	13.482	14.230	13.842	12.912
	0.001	15.958	16.322	16.200	15.165
20	0.05	6.447	6.512	6.460	6.358
	0.01	9.858	9.845	9.923	9.815
	0.0027	12.517	12.655	12.494	12.650
	0.001	14.959	15.726	15.238	14.813
50	0.05	6.492	6.520	6.488	6.317
	0.01	10.042	10.083	10.169	9.726
	0.0027	12.521	12.183	12.586	12.509
	0.001	14.431	14.595	14.372	14.624
100	0.05	6.623	6.625	6.604	6.304
	0.01	10.261	10.300	10.371	9.698
	0.0027	12.783	12.639	12.972	12.464
	0.001	14.986	15.326	15.108	14.564

In order to estimate the UCL for a given α, we have generated m samples of n observations from a standard bivariate normal distribution $N_2(\boldsymbol{0}, \boldsymbol{I})$ as phase I sample and computed the robust estimators of mean vector and covariance matrix $\tilde{\boldsymbol{M}}_R$ and $\tilde{\boldsymbol{S}}_R$. In addition, for each data set, we randomly generate a new observation z_l from $N_2(\boldsymbol{0}, \boldsymbol{I})$ (treated as a Phase II observation) and calculate the corresponding $T^2_{R_l}$ statistic. This procedure is repeated for N=10,000 times and computed $T^2_{R_1}, T^2_{R_2}, \ldots, T^2_{R_N}$. Then the UCL corresponding to α is the $100(1 - \alpha)^{\text{th}}$ quantile of $T^2_{R_1}, T^2_{R_2}, \ldots, T^2_{R_N}$. UCLs for n = 5, 10, 20, 50 and 100 were generated using Comedian, S_n Covariance, and GK methods for m = 50. The UCLs for α = 0.05, 0.01, 0.0027, and 0.001 are presented in Table 5.1.

5.4 Simulation Study

This section evaluates and compares the performance of the control charts in the presence of outliers using a simulation study, similar as in Alfaro and Ortega (2009). The data are modelled using a contaminated mixture of bivariate normal distributions, expressed as:

$$(1 - \varepsilon)\, N_2(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \varepsilon\, N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \tag{5.8}$$

where ε is the proportion of outliers, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the in-control parameters and $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are the out-of-control parameters of location and dispersion respectively. Various scenarios are examined based on the key elements such as number of observations(n), proportion of outliers, and the method of generating outliers using different values of $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$. Without loss of generality, $\boldsymbol{\mu}_0$ is assumed to be a zero vector. In this study ε takes the values of 0, 0.1, 0.2 and 0.03, and the probability of detecting a change depends on the values of $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$. The probability of detecting a change is influence by the values of $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$. Following cases are, therefore, considered to analyze this influence.

Case A: **Independent variables**

In this case, the quality characteristics are assumed to be independent. The contaminated normal model considered is as follows:

$$(1 - \varepsilon)\, N_2(\boldsymbol{0}, \boldsymbol{I}_2) + \varepsilon\, N_2(\boldsymbol{\mu}_1, \boldsymbol{I}_2).$$

where $\mathbf{0} = (0, 0)^T$, $\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12})^T$, the mean vector of the contaminated samples and $\boldsymbol{I}_2$ to be the identity matrix of size 2. In this case, we want to compare the behaviour of different robust alternatives when there are varying mean shifts across all the variables, assuming the variables are independent.

Case B: **Correlated variables**

In this case, the quality characteristics are assumed to be correlated. The contaminated normal model considered is as follows:

$$(1-\varepsilon)\ \mathrm{N}_2(\boldsymbol{0}, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_0),$$

where $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$. We have used this value of $\boldsymbol{\Sigma}_0$ to analyse whether the correlation level affects the detection probability of each alternative.

Case C: **Correlated variables and regression outliers**

Here, the quality characteristics are assumed to be correlated and regression outliers are introduced. The contaminated normal model considered is as follows:

$$(1-\varepsilon)\ \mathrm{N}_2(\boldsymbol{0}, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1),$$

where $\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$. In this case, we analyse and compare the proposed robust methods in terms of the so-called good leverage and regression outliers.

5.4.1. Impact of outliers on classical control chart

The estimates of location and dispersion derived during Phase I are crucial for calculating the Hotelling's T^2 statistic to monitor new observations in Phase II. In Phase I, these estimators are computed after excluding out-of-control samples from the data. Consequently, it is imperative to evaluate the efficacy of classical control charts in accurately identifying out-of-control samples in the presence of outliers. However, in the context of statistical process control, robust estimators offer a significant advantage as they are unaffected by outliers, eliminating the need for their explicit identification in Phase I data. Utilizing robust estimators ensures reliable Phase I estimates of location and covariance, which in turn facilitate the development of a robust control chart for analysing Phase II data.

The impact of outliers and the magnitude of mean shifts on the performance of classical bivariate control charts during Phase I application is investigated in this study. Specifically, m random samples of size n are generated using the contaminated model outlined in (5.8), and the percentage of out-of-control samples is computed. This procedure is repeated 1000 times to calculate the average percentage of out-of-control samples across various sample sizes (n = 5, 10, 20, 50, 100), contamination levels (ε = 0.1, 0.2, 0.3), and mean shifts. The results for Case A and Case B contaminations at α = 0.05 are presented in Figure 5.1 and Figure 5.2, respectively.

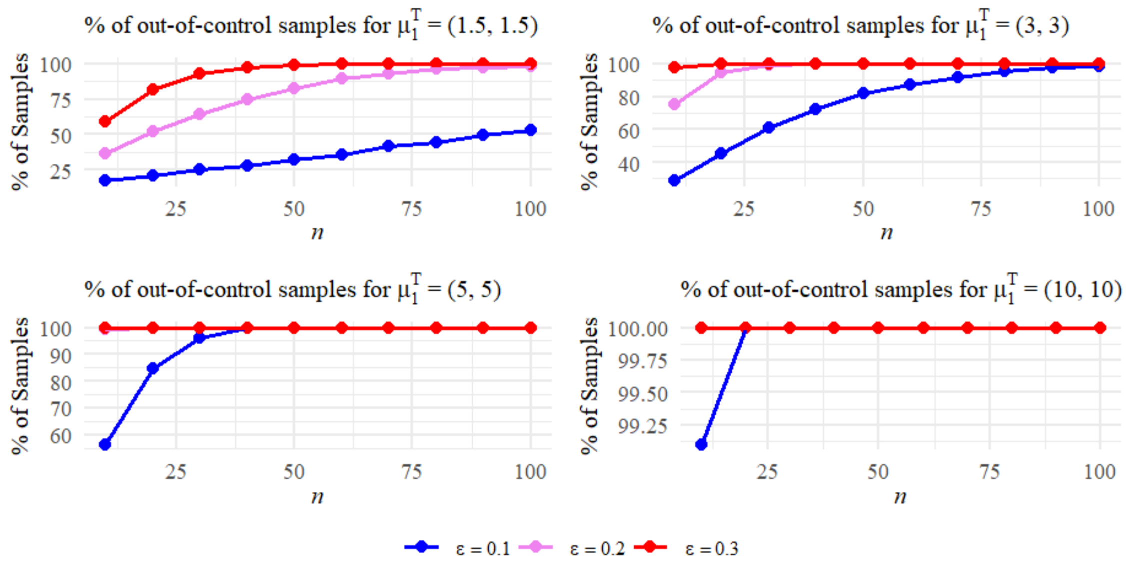


Figure 5.1: Average percentage of out-of-control samples under Case A contamination

Figure 5.1 illustrates the graphs showing the percentages of out-of-control samples under Case A contamination, with varying proportions of outliers and different mean shifts. The figure reveals that as the contamination level increases, the percentage of out-of-control samples rapidly approaches 100%. This process accelerates with larger sample sizes or greater mean shifts. Consequently, this reduces the number of valid samples available for the estimation of location and dispersion parameters in Phase II.

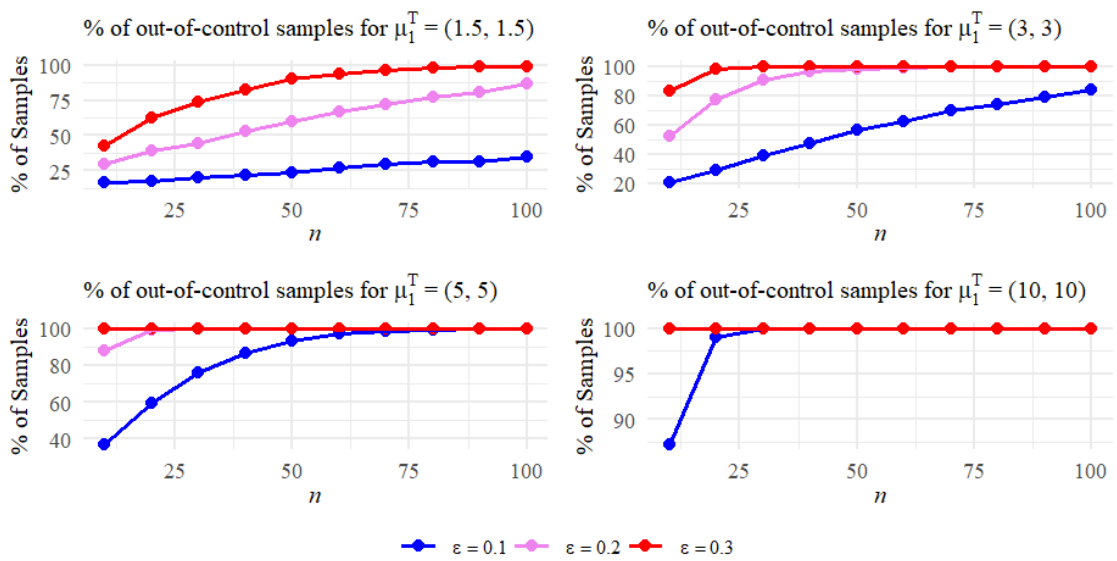


Figure 5.2: Average percentage of out-of-control samples under Case B contamination

[Figure 5.2](#) presents the percentages of out-of-control samples under Case B contamination, with varying proportions of outliers and different mean shifts. The results are similar to those obtained for Case A. This highlights that even small amounts of contamination can cause significant distortion in the control chart, making it difficult to accurately estimate location and dispersion parameters for Phase II. These findings demonstrate that classical control charts are ineffective in the presence of outliers, especially in the early stages of monitoring, emphasizing the critical impact that contamination has on both Phase I and Phase II applications.

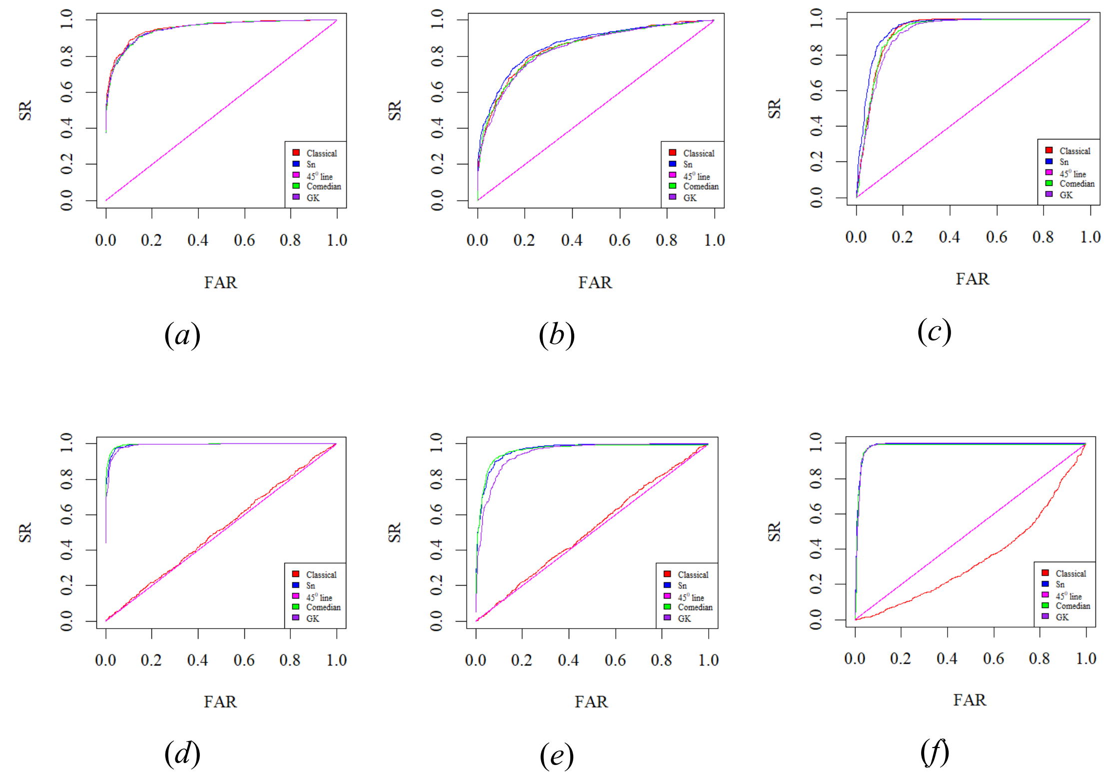


Figure 5.3: ROC curves of Bivariate control charts based on Classical, S_n Covariance, Comedian and GK estimators.

Figure 5.3 presents the Receiver Operating Characteristic (ROC) curves comparing the performance of four estimators — Classical, S_n Covariance, GK, and Comedian — under two scenarios: clean data (0% contamination) and contaminated data (20% contamination). The upper panel (a–c) illustrates the results without contamination for Cases A, B, and C, where all estimators, including the Classical method, show strong and comparable performance, with ROC curves closely approaching the ideal top-left corner, reflecting high success rates across different false alarm rates.

However, the lower panel (d–f) highlights the impact of 20% contamination. The robust estimators — S_n Covariance, GK, and Comedian — maintain their effectiveness, as shown by their curves remaining near the top-left corner. In contrast, the Classical estimator's performance declines considerably under contamination, especially in Case C (f), where its ROC curve drops below the diagonal random guess line. This indicates that the Classical method performs worse than random classification in this scenario, demonstrating its sensitivity to outliers. These findings clearly underline the robustness and superior reliability of S_n Covariance, GK, and Comedian estimators for outlier detection in contaminated datasets.

5.4.2. Impact of outliers on robust control charts

Using 1000 replications, we considered different sample sizes where n = 5, 10, 20, 50 and 100 with dimensions 2 and an overall false alarm probability of α = 0.05, 0.01, 0.0027, and 0.001. For Phase I, 1000 samples of size n were generated using the contaminated model (5.8) and computed the estimates of location and scatter for Phase II application. In Phase II, two types of observations were generated: one with in-control parameters and another with out-of-control parameters. The Hotelling's T^2 statistic was computed for both observations using Equation (5.7). The performance of the control charts is assessed using the parameters SR and FAR. These metrics were evaluated for different estimators in Phase I across 1000 replications, offering insights into the effectiveness and reliability of the proposed control charts under varying conditions.

In this study we have examined the performance of the proposed robust control charts for $\boldsymbol{\mu}_1 = (3, 3)^T$, $(3, -3)^T$, $(3, 0)^T$, $(5, 5)^T$, and $(10, 10)^T$. For brevity, we present results for α = 0.05 and $\boldsymbol{\mu}_1 = (5, 5)^T$, noting that similar trends and conclusions were observed for other values of α. (Results corresponding to the remaining values of $\boldsymbol{\mu}_1$ are included in the Appendix B).

The results obtained for Case A contamination are presented in Table 5.2. Under Case A contamination, where quality characteristics are independent, all methods consistently achieve a perfect success rate (SR = 100%) across all sample sizes and contamination levels. This indicates their effectiveness in detecting expected changes, regardless of sample size or contamination. However, false alarm rates (FAR) vary significantly, providing insights into the robustness of each method. S_n Covariance proves to be the most reliable, maintaining consistently low FAR even for small sample sizes (n) and high contamination levels (ε = 0.3), with further improvement as sample size increases. GK also shows strong robustness, achieving low FAR across all scenarios, particularly for $\varepsilon \leq 0.2$. Comedian, while reliable, exhibits slightly higher FAR, especially for smaller sample sizes (n = 5, 10) and moderate contamination levels (ε = 0.1, 0.2).

Table 5.2: SR (FAR) in Phase II when there are case A outliers in Phase I data

n	**ε**	**S_n Covariance**	**Comedian**	**GK**
5	0	1 (0.024)	1 (0.024)	1 (0.026)
	0.1	1 (0.009)	1 (0.016)	1 (0.006)
	0.2	1 (0)	1 (0)	1 (0)
	0.3	1 (0.001)	1 (0.001)	1 (0.002)
10	0	1 (0.034)	1 (0.036)	1 (0.033)
	0.1	1 (0.008)	1 (0.011)	1 (0.008)
	0.2	1 (0.003)	1 (0.006)	1 (0.003)
	0.3	1 (0)	1 (0.002)	1 (0)
20	0	1 (0.031)	1 (0.034)	1 (0.033)
	0.1	1 (0.011)	1 (0.015)	1 (0.012)
	0.2	1 (0.002)	1 (0.003)	1 (0.003)
	0.3	1 (0)	1 (0)	1 (0)
50	0	1 (0.033)	1 (0.031)	1 (0.032)
	0.1	1 (0.01)	1 (0.012)	1 (0.008)
	0.2	1 (0.001)	1 (0.002)	1 (0.001)
	0.3	1 (0)	1 (0)	1 (0)
100	0	1 (0.029)	1 (0.028)	1 (0.027)
	0.1	1 (0.006)	1 (0.011)	1 (0.007)
	0.2	1 (0.001)	1 (0.001)	1 (0.001)
	0.3	1 (0)	1 (0)	1 (0)

The effects of contamination and sample size reveal clear trends. False alarm rates generally decrease as the sample size increases, suggesting that larger sample sizes provide more reliable estimates and better performance in Phase II monitoring. Larger sample sizes (n = 50, 100) lead to reduced FAR across all methods. For smaller sample sizes (n = 5, 10), S_n Covariance and GK are the most resilient, maintaining low FAR even under challenging conditions.

Results obtained for Case B contamination are presented in Table 5.3. Under Case B contamination, where quality characteristics are assumed to be correlated, all methods maintain high success rates across most scenarios. Slight variations are observed with small sample sizes (n = 5) and higher contamination levels (ε). Both S_n Covariance and GK consistently exhibit low false alarm rates, demonstrating exceptional robustness to contamination, even with small sample sizes. Comedian also performs reasonably well but shows slightly higher FAR compared to S_n Covariance and GK, particularly at smaller sample sizes and moderate contamination levels (ε = 0.1, 0.2).

Table 5.3: SR (FAR) in Phase II when there are case B outliers in Phase I data

n	***ε***	**S_n Covariance**	**Comedian**	**GK**
5	0	0.994 (0.028)	0.916 (0.035)	1 (0.04)
	0.1	0.967 (0.017)	0.898 (0.035)	0.996 (0.023)
	0.2	0.926 (0.021)	0.848 (0.03)	0.996 (0.019)
	0.3	0.889 (0.014)	0.809 (0.014)	0.994 (0.008)
10	0	1 (0.027)	0.983 (0.053)	1 (0.04)
	0.1	0.998 (0.015)	0.964 (0.027)	1 (0.015)
	0.2	0.985 (0.005)	0.932 (0.025)	1 (0.001)
	0.3	0.958 (0.008)	0.882 (0.018)	1 (0.001)
20	0	1 (0.015)	0.997 (0.025)	1 (0.025)
	0.1	1 (0.007)	0.998 (0.021)	1 (0.009)
	0.2	0.997 (0.005)	0.981 (0.018)	1 (0.001)
	0.3	0.992 (0.009)	0.95 (0.017)	1 (0)
50	0	1 (0.013)	1 (0.017)	1 (0.032)
	0.1	1 (0.008)	1 (0.013)	1 (0.012)
	0.2	1 (0.001)	0.996 (0.01)	1 (0.003)
	0.3	0.999 (0)	0.994 (0.01)	1 (0.001)
100	0	1 (0.013)	1 (0.018)	1 (0.036)
	0.1	1 (0.01)	1 (0.015)	1 (0.013)
	0.2	1 (0.001)	1 (0.005)	1 (0.005)
	0.3	1 (0)	0.998 (0.004)	1 (0)

FAR trends indicate improvements with increasing sample size for all robust methods. S_n Covariance and GK achieve nearly perfect results with minimal FAR for n ≥ 20, even under severe contamination (ε = 0.3). Comedian shows improvement in FAR as sample size increases, but it is still somewhat more sensitive to contamination compared to S_n Covariance and GK.

The results for Case C contamination, which assumes correlated quality characteristics and introduces regression outliers, presented in Table 5.4, show that all methods generally achieve high success rates. Minor variations are observed for small sample sizes (n = 5) and higher contamination levels. Among the methods, GK demonstrates the most robust performance, maintaining near-perfect SR and negligible false alarm rates across all scenarios, including severe contamination. S_n Covariance also performs reliably, with high SR and low FAR, although FAR increases slightly at higher contamination levels (ε = 0.2, 0.3) for small n. Comedian shows moderate performance, with SR improving as n increases, but its FAR remains higher than that of GK and S_n Covariance.

Table 5.4: SR (FAR) in Phase II when there are case C outliers in Phase I data

n	ε	S_n Covariance	Comedian	GK
5	0	0.994 (0.029)	0.922 (0.037)	0.999 (0.041)
	0.1	0.975 (0.016)	0.899 (0.025)	0.997 (0.018)
	0.2	0.926 (0.02)	0.873 (0.024)	0.998 (0.006)
	0.3	0.895 (0.011)	0.816 (0.023)	1 (0.002)
10	0	1 (0.024)	0.976 (0.038)	1 (0.041)
	0.1	0.999 (0.011)	0.958 (0.028)	1 (0.014)
	0.2	0.981 (0.017)	0.934 (0.018)	1 (0.002)
	0.3	0.961 (0.012)	0.882 (0.019)	1 (0.003)
20	0	1 (0.021)	0.999 (0.032)	1 (0.03)
	0.1	1 (0.011)	0.992 (0.021)	1 (0.015)
	0.2	0.998 (0.004)	0.982 (0.02)	1 (0.002)
	0.3	0.994 (0)	0.93 (0.024)	1 (0.004)
50	0	1 (0.016)	1 (0.018)	1 (0.027)
	0.1	1 (0.005)	1 (0.019)	1 (0.011)
	0.2	1 (0.001)	1 (0.002)	1 (0.001)
	0.3	1 (0.001)	0.993 (0.008)	1 (0.001)
100	0	1 (0.014)	1 (0.016)	1 (0.032)
	0.1	1 (0.002)	1 (0.004)	1 (0.004)
	0.2	1 (0)	1 (0.003)	1 (0)
	0.3	1 (0)	0.999 (0.004)	1 (0.003)

5.4.3 ARL plots

Average Run Length (ARL) is a key metric used to evaluate control chart performance. It represents the expected number of samples taken before detecting an out-of-control signal. An effective chart minimizes false alarms (high in-control ARL) while quickly identifying actual shifts (low out-of-control ARL).

In Phase I, datasets with contaminated observations are generated to examine the impact on estimation methods. Both classical and robust estimators are used to calculate the process mean and covariance matrix, and corresponding control limits are established.

Out of control samples identified using Hotelling's T^2 in the classical method are excluded prior to estimation. The process shift is quantified using the squared non-centrality parameter:

$$\delta_I^2 = (\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0)$$

with values $\delta_I^2 = 75$ and 150.

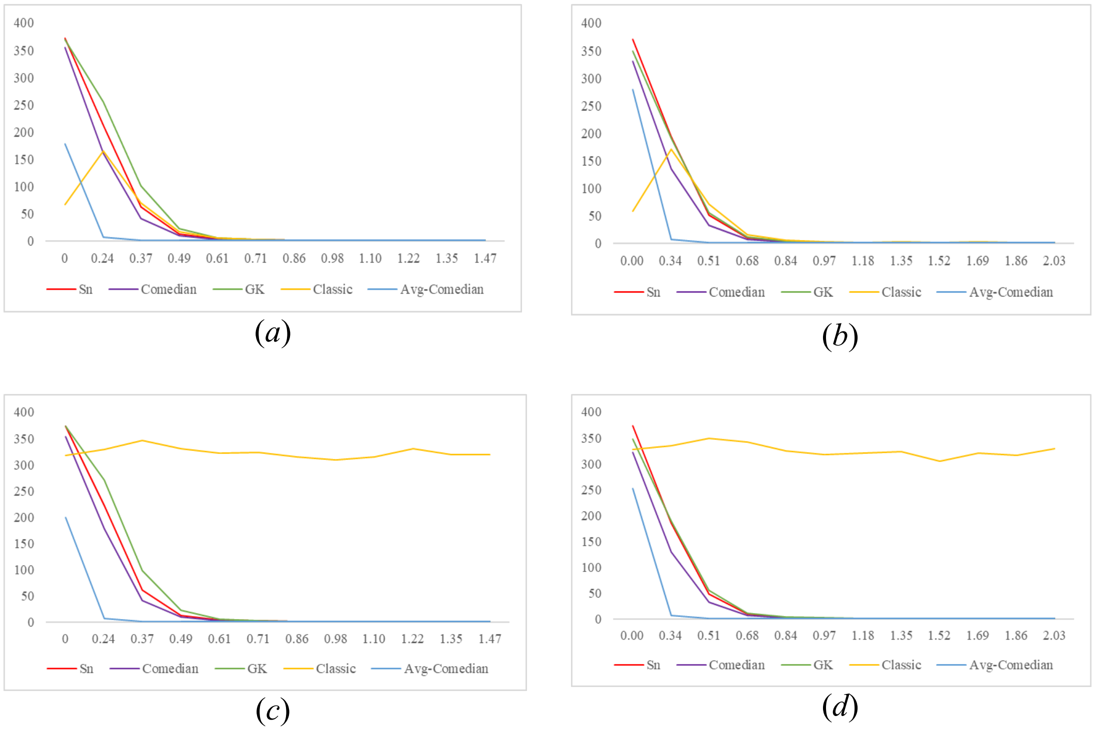


Figure 5.4: ARL plots of Bivariate control charts based on Classical, S_n Covariance, Comedian, GK and Avg-Comedian estimators under Case A and B contamination for $\delta^2 = 75$ and 150.

In Phase II, simulated samples are tested under varying shift magnitudes:

$$\delta_{II}^2 = (\mu_2 - \mu_0)^T \Sigma^{-1} (\mu_2 - \mu_0)$$

with δ_{II}^2 values ranging from 0.0 to 4.32. For each sample, the T^2 statistic is calculated and compared against control limits from both classical and robust estimation methods. The ARL is computed across $r = 1000$ iterations using:

$$\text{ARL} = \frac{1}{r} \sum_{i=1}^{r} \text{RL}_i.$$

The significance level is set at $\alpha = 0.0027$. Results comparing classical and robust charts under contamination are shown in Figure 5.4.

The ARL plots presented in Figure 5.4 illustrate the performance of various estimators under two contamination scenarios (Case A and Case B) and two levels of process shift

(δ^2 = 75 and 150). The comparison clearly highlights the distinction between classical and robust estimators when monitoring contaminated processes. The Classical and Avg-Comedian based control charts show relatively high ARL values under both contamination cases, indicating poor detection ability and a strong sensitivity to the influence of outliers. In contrast, the control charts based on robust estimators — S_n Covariance, GK, and Comedian — demonstrate significant resistance to contamination, as reflected by their consistently lower ARL values when the shift magnitude increases, which indicates reliable and timely detection of true process changes.

A closer inspection of the plots reveals that the Avg-Comedian based chart exhibits a steep decrease in ARL even for very small shifts when compared to the other robust methods. While this behavior might superficially suggest improved shift detection, it actually points to the estimator's vulnerability to contamination. Specifically, the early drop in ARL is often driven by the effect of outlying observations, which distort the location estimate and cause the chart to signal prematurely, even in the absence of meaningful process shifts. This leads to an increased likelihood of false detections, where the control chart incorrectly signals an out-of-control condition caused by outliers rather than an actual shift. Therefore, although Avg-Comedian based chart appears highly sensitive, its performance under contaminated data is unstable, making it prone to false alarms and undermining its reliability. On the other hand, S_n Covariance, GK, and Comedian demonstrate more controlled ARL behavior, reflecting their robustness against outliers and a lower risk of false detections, thereby ensuring that signals are more likely to reflect genuine process changes.

5.5 Time complexity

The Table 5.5 shows the running times for three methods - S_n Covariance, Comedian, Avg- Comedian and GK - for various sample sizes (n = 5, 10, 20, 50, 100). As the subgroup size increases, the computational time required by the S_n Covariance estimator grows significantly, indicating a relatively higher computational cost. In contrast, the Comedian and Avg-Comedian estimators maintain relatively stable and low running times across all sample sizes, ranging between 0.0037 and 0.0067 seconds. The GK estimator exhibits moderate computational complexity, with running times increasing gradually from 0.0042 to 0.0100 seconds as the sample size grows. In conclusion, Comedian is the fastest and most consistent, followed by GK and Avg-Comedian, while S_n Covariance is the slowest, particularly for larger sample sizes.

Table 5.5: Computation time (in seconds)

n	**S_n Covariance**	**Comedian**	**Avg-Comedian**	**GK**
5	0.0070	0.0037	0.0050	0.0042
10	0.0109	0.0041	0.0061	0.0048
20	0.0221	0.0043	0.0056	0.0054
50	0.0789	0.0047	0.0056	0.0070
100	0.2577	0.0051	0.0067	0.0100

5.6 Application to Real Life Data

To illustrate the implementation of the proposed robust bivariate control charts, we utilize the carbon fiber dataset presented by Santos-Fernández (2012). This dataset comprises three interrelated quality characteristics measured in inches: inner diameter, thickness, and length of carbon fiber tubes.

The data are structured into two phases. In Phase I, each quality characteristic includes $m = 30$ samples, with each sample containing $n = 8$ observations. Phase II consists of 25 additional samples, also of size $n = 8$, for each characteristic. For the purpose of this study, which focuses on bivariate control charts, we consider only two of the quality characteristics—thickness and length of the tubes.

Figure 5.5 displays the control charts constructed using various estimators— S_n Covariance, Comedian, Avg-Comedian, GK, and the Classical method. As evident from the figure, all robust estimators produce results comparable to the classical method, as both the Phase I and Phase II datasets do not contain any out-of-control observations. To assess the robustness of the control charts in the presence of process disturbances, the Phase I data were deliberately contaminated by introducing a mean shift in the *length of the fiber* for the first two subgroups. Specifically, the mean of the *length* was increased by a magnitude of 1σ in both subgroups. The resulting control charts, presented in Figure 5.6, clearly indicate that while the robust methods remain stable and unaffected by the induced shift, the classical chart responds by incorrectly identifying an in-control observation as out-of-control. This underscores the superior robustness of the proposed methods in the presence of data contamination.

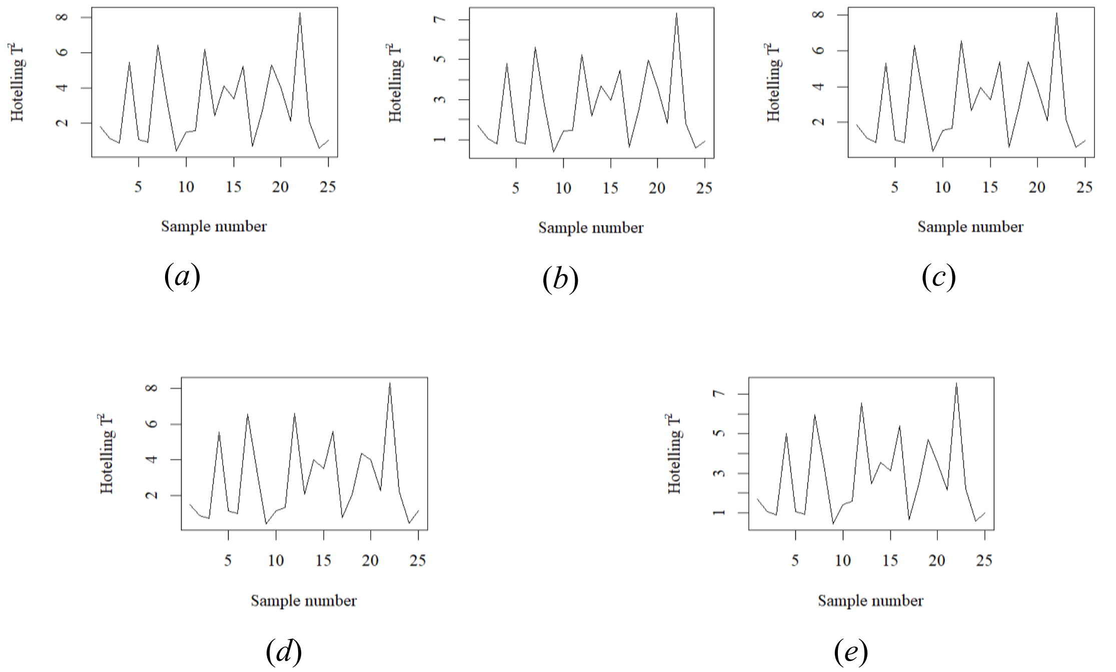


Figure 5.5: Bivariate control charts based on (*a*) S_n Covariance, (*b*) Comedian, (*c*) Avg-Comedian, (*d*) GK and (*e*) Classical estimators under uncontaminated Phase I data.

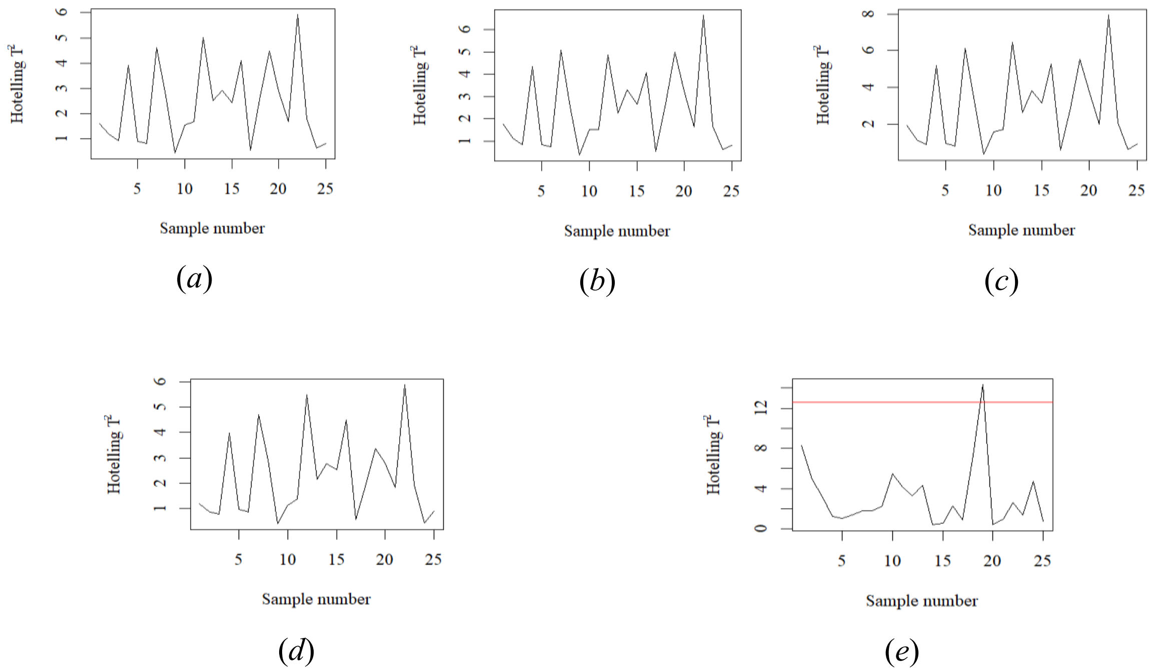


Figure 5.6: Bivariate control charts based on (*a*) S_n Covariance, (*b*) Comedian, (*c*) Avg-Comedian, (*d*) GK and (*e*) Classical estimators under contaminated Phase I data.

5.7 Summary

This chapter discusses the development of a robust bivariate control chart for process monitoring, particularly in the presence of outliers and non-normal data. The study aims to improve the performance of conventional bivariate control charts, like the Hotelling's T^2 chart, by introducing robust estimators for the mean vector and covariance matrix, which are less sensitive to outliers. The chapter highlights the importance of monitoring processes with bivariate quality characteristics, using rational subgroups, and constructing control charts that accurately capture shifts in the process mean vector.

The chapter begins by explaining the concept of Phase I and Phase II in control chart construction. Phase I involves estimating the in-control parameters and setting control limits using historical data, while Phase II involves using these control limits to detect deviations during actual process monitoring. The Hotelling's T^2 statistic is the foundation of the bivariate control chart, which is typically based on the assumption of normally distributed data. However, in practice, estimators of the mean and covariance matrix can be influenced by outliers, leading to the false alarming in process monitoring.

To address this limitation, the study proposes a new approach using robust estimators, such as the L_1 median vector, Comedian estimator, S_n Covariance estimator, and the robust GK estimator, which are known for their high breakdown points. These robust estimators are used to construct a bivariate control chart that can effectively detect shifts in the mean and covariance of a process, even in the presence of outliers. The control limits for the robust chart are derived empirically through simulations, and the results are compared with each other.

The performance of robust control charts was empirically assessed under different sample sizes, contamination levels, and contamination scenarios. Success rate and false alarm rate were assessed for different estimators, revealing that S_n Covariance and GK demonstrated the most reliable performance, maintaining low FAR even under high contamination and small sample sizes. Comedian, while performing well, showed slightly higher FAR, particularly for smaller sample sizes. The results indicated that as the sample size increased, FAR decreased, highlighting the importance of larger sample sizes for better performance in Phase II monitoring. Additionally, Comedian exhibited the fastest running times across all contamination scenarios, followed by GK, with S_n Covariance showing the highest computational costs, particularly for larger sample sizes. These findings suggest that robust control charts, especially S_n Covariance and GK, are highly effective in detecting out-of-control observations while minimizing false alarms, making them valuable tools for statistical process control.

CHAPTER 6

MULTIVARIATE ROBUST CONTROL CHART FOR PROCESS MONITORING WITH RATIONAL SUBGROUPS

6.1 Introduction

Monitoring process quality in industrial and service applications often requires the simultaneous assessment of multiple correlated quality characteristics. Multivariate control charts provide an effective means for this purpose, especially when data are organized into rational subgroups that capture natural temporal or operational dependencies. Alt (1985) distinguished between Phase I, where historical data are analysed to estimate in-control parameters and establish control limits, and Phase II, where these limits are applied to ongoing production for detecting deviations from stability. Among the established techniques, the Hotelling's T^2 chart (Montgomery, 2020) remains one of the most widely used methods for monitoring shifts in the process mean vector.

For a random sample $\boldsymbol{x}_i$ of size n of a multivariate quality characteristic, the Hotelling's T^2 statistic is defined as

$$T_i^2 = n(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}_i - \boldsymbol{\mu}), \tag{6.1}$$

where $\overline{\boldsymbol{x}}_i$ is the sample mean vector of $\boldsymbol{x}_i$. When the process is in statistical control, it is assumed that $\boldsymbol{x}_i$ is independent and follows a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ represents the mean vector and $\boldsymbol{\Sigma}$ represents the covariance matrix. If both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, T^2 follows a Chi-square distribution with p degrees of freedom. However, in practice, the true values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are generally unknown. In such cases, these parameters are usually substituted with classical estimators of mean vector and covariance matrix.

6.2 Computational algorithm for Multivariate control chart

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ be m random samples of a multivariate quality characteristic, where each sample consists of n observations. The following steps outline the procedure for constructing a Hotelling's T^2 control chart:

1. *Calculate the Mean Estimator*: For each subgroup i, calculate the sample mean vector $\bar{\boldsymbol{x}}_i = (\bar{x}_{i1}, \bar{x}_{i2} \ldots \bar{x}_{ip})^{\boldsymbol{T}}$, where:

   $$\bar{x}_{ij} = \frac{1}{n}\sum_{k=1}^{n} x_{ijk} \text{ for } i = 1, 2, \ldots, m;\ j = 1, 2, \ldots p$$

   The overall sample mean vector $\bar{\bar{\boldsymbol{x}}}$ is computed as the average of the subgroup means:

   $$\bar{\bar{\boldsymbol{x}}} = \frac{1}{m}\sum_{i=1}^{m} \bar{\boldsymbol{x}}_i. \quad (6.2)$$

2. *Calculate the Sample Covariance Matrix for Each Subgroup*: For each subgroup i, calculate the sample covariance betweentwo random variables. The covariance matrix $\boldsymbol{S}_i$ is given by:

   $$\boldsymbol{S}_i = \begin{pmatrix} S_{i11} & S_{i12} \cdots & S_{i1p} \\ \vdots & \ddots & \vdots \\ S_{i1p} & \cdots & S_{ipp} \end{pmatrix} \text{ for } i = 1, 2, \ldots, m, \quad (6.3)$$

   where S_{ijj} represents the variance of the j^{th} variable in i^{th} subgroup, and $S_{ijj'}$ represents the covariance between the two variables j and j' in the i^{th} subgroup.

3. *Compute the Average Covariance Matrix*: The sample covariance estimator $\boldsymbol{S}$ is obtained by averaging the subgroup covariance matrices:

   $$\boldsymbol{S} = \begin{pmatrix} \bar{\boldsymbol{S}}_1 & \bar{\boldsymbol{S}}_{12} \cdots & \bar{\boldsymbol{S}}_{1p} \\ \vdots & \ddots & \vdots \\ \bar{\boldsymbol{S}}_{1p} & \cdots & \bar{\boldsymbol{S}}_{pp} \end{pmatrix} \quad (6.4)$$

   where $\bar{\boldsymbol{S}}_j = \frac{1}{m}\sum_{i=1}^{m} S_{ijj}$ and $\bar{S}_{jj'} = \frac{1}{m}\sum_{i=1}^{m} S_{ijj'}\,; j, j' = 1,2, \ldots, p.$

4. *Compute the Hotelling's T^2 Statistic*: The Hotelling's T^2 statistic for each subgroup i is computed as follows:

   $$T_i^2 = n(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})^T \boldsymbol{S}^{-1} (\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}}) \text{ for } i = 1, 2, \ldots, m. \quad (6.5)$$

5. *Estimation of parameters for Phase II application*: Plot the calculated values of T^2 on the control chart. If any T^2 value falls outside the control limits, remove the corresponding samples and re-estimate the location and dispersion parameters using the remaining sample subgroups for the phase II applications. UCLs corresponding to Phase I and Phase II are available in Montgomery (2009).

However, the classical estimators of the mean vector and covariance matrix are often sensitive to outliers and deviations from normality, which can lead to inaccurate results. Addressing these limitations, robust control charts have emerged as a powerful alternative, offering improved performance under non-ideal conditions. Various approaches have been developed for detecting mean shifts in individual sample monitoring (see Vargas, 2003; Chenouri *et al*., 2009; Jensen *et al*., 2007; Sajesh, 2023; Vijayalakshmi *et al*., 2024). However, robust alternatives specifically designed for rational subgroups remain scarce. Several robust subgroup-based control charts have been proposed in the literature, such as those by Abu-Shawiesh *et al*. (2012), Maleki *et al*. (2020), and Mahpouya *et al*. (2021). These charts employ the component-wise median and the Comedian estimator (Falk, 1997) for robust estimation of the mean vector and covariance matrix. Although these methods improve resistance to outliers arising from mean shifts, their performance can still be compromised when a subgroup is locally contaminated, as the covariance structure becomes distorted. Building on these ideas, the present chapter extends the focus to robust techniques for multivariate control charts

6.3 Multivariate Robust Control Chart

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ represent m random samples of a multivariate quality characteristic, each of size n. The following steps describe the construction of a robust multivariate control chart:

1. *Compute robust location estimator for Each Sample*: For each sample i, a robust location estimator $\widetilde{\boldsymbol{\theta}}_i$ is computed.

2. *Estimate robust location estimator for population mean*: Compute the robust location estimator based on $\widetilde{\boldsymbol{\theta}}_1, \widetilde{\boldsymbol{\theta}}_2, \ldots, \widetilde{\boldsymbol{\theta}}_m$, denoted as $\widetilde{\boldsymbol{M}}_R$, which serves as a robust estimator of the population mean $\boldsymbol{\mu}$.

3. *Estimate the Covariance Matrix Using Robust Estimation*: The proposed chart uses the robust estimator of $\boldsymbol{\Sigma}$ based on the whole set of data, denoted as $\tilde{\boldsymbol{S}}_R$, instead of the $\boldsymbol{S}$ matrix given in the construction of Hotelling's T^2 chart.

4. *Compute the Robust Hotelling's T^2 Statistic*: For a new observation with data matrix $\boldsymbol{z}$, compute robust Hotelling's T^2 statistic using the robust estimators $\widetilde{\boldsymbol{M}}_R$ and $\tilde{\boldsymbol{S}}_R$ as follows:

$$T_{R_i}^2 = n\left(\tilde{\boldsymbol{z}} - \widetilde{\boldsymbol{M}}_R\right)^T \widetilde{\boldsymbol{S}}_R^{-1}\left(\tilde{\boldsymbol{z}} - \widetilde{\boldsymbol{M}}_R\right), \tag{6.6}$$

where $\tilde{\boldsymbol{z}}$ is the L_1 median of the new observation $\boldsymbol{z}$.

5. *Plot the Statistic on the Control Chart*: Plot the computed statistic on the control chart with the suitable UCL empirically obtained through simulation for a given α risk.

Table 6.1: Simulated Upper control limits

n	*p*	**ROC**			**SGK**		
		95%	99%	99.73%	95%	99%	99.73%
20	2	6.922	10.554	13.240	6.871	10.623	13.335
	5	11.898	16.587	19.672	12.019	16.548	19.571
	10	19.702	25.023	30.113	19.700	25.261	29.732
30	2	6.910	10.334	13.311	6.910	10.530	13.502
	5	11.912	16.589	20.628	12.174	16.242	19.238
	10	19.609	24.419	27.833	19.496	24.784	28.991
50	2	6.758	10.228	13.563	6.962	10.735	13.592
	5	11.988	16.899	21.803	11.999	16.482	20.340
	10	18.958	24.470	29.886	19.120	26.132	28.258
n	*p*	**OGK**			**RMCD**		
		95%	99%	99.73%	95%	99%	99.73%
20	2	8.596	13.164	16.856	6.233	9.588	12.355
	5	13.904	18.895	22.674	11.794	16.002	19.294
	10	21.418	27.397	32.602	19.215	24.465	28.487
30	2	8.454	12.683	15.624	6.315	9.915	12.673
	5	13.644	18.647	22.625	11.403	15.680	19.027
	10	21.212	26.517	30.722	19.394	24.428	28.083
50	2	8.097	12.922	16.717	6.259	9.853	13.441
	5	13.362	17.555	20.905	11.125	17.325	19.502
	10	21.782	28.482	34.476	19.270	24.602	29.552

To estimate the UCL corresponding to a given significance level α, we first generate m Phase I samples, each consisting of n observations drawn from the standard multivariate normal distribution $N_p(\boldsymbol{0}, \boldsymbol{I})$. From these samples, the robust estimators of the mean vector and covariance matrix, denoted as $\widetilde{\boldsymbol{M}}_R$ and $\widetilde{\boldsymbol{S}}_R$, are obtained. Next, for each Phase

I dataset, an additional observation z_l is independently generated from $N_p(\mathbf{0}, \boldsymbol{I})$ and treated as a Phase II observation. The corresponding $T^2_{R_l}$ statistic is then computed based on $\widetilde{\boldsymbol{M}}_R$ and $\widetilde{\boldsymbol{S}}_R$. This entire procedure is repeated N=10,000 times to obtain a distribution of $T^2_{R_1}, T^2_{R_2}, \dots, T^2_{R_N}$. The UCL for the specified significance level α is finally determined as the $100(1-\alpha)^{\text{th}}$ percentile of these simulated $T^2_{R_1}, T^2_{R_2}, \dots, T^2_{R_N}$ statistics. UCLs for n = 20, 30 and 50 were generated using ROC, SGK, OGK and RMCD methods for m = 50. The UCLs for α = 0.05, 0.01, 0.0027, and 0.001 are presented in Table 6.1.

6.4 Simulation Study

This section evaluates the performance of the proposed multivariate control charts in the presence of outliers through a simulation study, following the approach of Alfaro and Ortega (2009) and as discussed in Chapter 5. The data are generated from a contaminated multivariate normal mixture given by

$$(1-\varepsilon)\, N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \varepsilon\, N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1),$$

where ε represents the contamination proportion. The first component denotes the in-control process, while the second represents the out-of-control condition. Different scenarios are considered by varying the sample size, the proportion of contamination (ε = 0,0.1,0.2,0.3), and the method of generating outliers through changes in the mean vector and covariance matrix. The in-control mean is assumed to be a zero vector. The study investigates how these factors influence the ability of the chart to detect process shifts.

Case A: **Independent variables**

In the first scenario, it is assumed that all quality characteristics are statistically independent. The data are generated from a contaminated multivariate normal distribution expressed as:

$$(1-\varepsilon)\, N_p(\mathbf{0}, \boldsymbol{I}) + \varepsilon\, N_p(\boldsymbol{\mu}_1, \boldsymbol{I}).$$

where $\mathbf{0} = (0\ 0\ \dots 0)^T$, $\boldsymbol{\mu}_1 = (\mu_{11}\ \ \mu_{12} \dots \mu_{1p})^T$, the mean vector of the contaminated samples and $\boldsymbol{I}$ to be the identity matrix of size p. This case is designed to examine how various robust control charting methods respond to mean shifts occurring independently across different variables.

Case B: Correlated variables

In this case, the process variables are assumed to be correlated. The contaminated normal model considered is as follows:

$$(1-\varepsilon)\ \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_0),$$

Where $\boldsymbol{\Sigma}_0 = \begin{pmatrix} 1 & 0.9\cdots & 0.9 \\ \vdots & \ddots & \vdots \\ 0.9 & \cdots & 1 \end{pmatrix}$. This setting allows an assessment of how inter-variable correlation influences the sensitivity and reliability of the competing robust approaches in detecting process shifts.

Case C: **Correlated variables and regression Outliers**

Here, both correlation among variables and the presence of regression-type outliers are considered. The contaminated model is defined as

$$(1-\varepsilon)\ \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_0) + \varepsilon\ \mathrm{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1),$$

where

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 1 & 0.9\cdots & 0.9 \\ \vdots & \ddots & \vdots \\ 0.9 & \cdots & 1 \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_1 = \begin{pmatrix} 0.1 & 0\cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0.1 \end{pmatrix}.$$

This configuration is used to explore the ability of the proposed robust estimators to handle both good leverage points and regression outliers when the variables are correlated.

6.4.1 ARL Plots

The Average Run Length (ARL) serves as a key criterion for evaluating control chart effectiveness. It represents the expected number of samples taken before the chart signals a process shift. An optimal control chart is characterized by a high in-control ARL, minimizing false alarms, while maintaining a low out-of-control ARL to ensure rapid detection of genuine process deviations.

In Phase I, data sets containing a certain proportion of contaminated observations are generated to study their influence on parameter estimation. Both classical and robust approaches are applied to estimate the process mean vector and covariance matrix. In the classical approach, observations identified as out-of-control by Hotelling's T^2 statistic are removed prior to estimation. The magnitude of the process shift in this phase is quantified by

$$\delta_I^2 = (\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0),$$

with values δ_I^2 = 150 and 300.

In Phase II, simulated samples are evaluated under varying magnitudes of shift, expressed as

$$\delta_{II}^2 = (\mu_2 - \mu_0)^T \Sigma^{-1} (\mu_2 - \mu_0),$$

with δ_{II}^2 values ranging from 0.0 to 0.3. For each sample, the T^2 statistic is calculated and compared against control limits from both classical and robust estimation methods. The ARL is computed across r = 1000 iterations using:

$$\text{ARL} = \frac{1}{r} \sum_{i=1}^{r} \text{RL}_i.$$

The significance level is set at α = 0.0027, and the comparative performance of classical and robust control charts under different contamination scenarios is illustrated in the following figures..

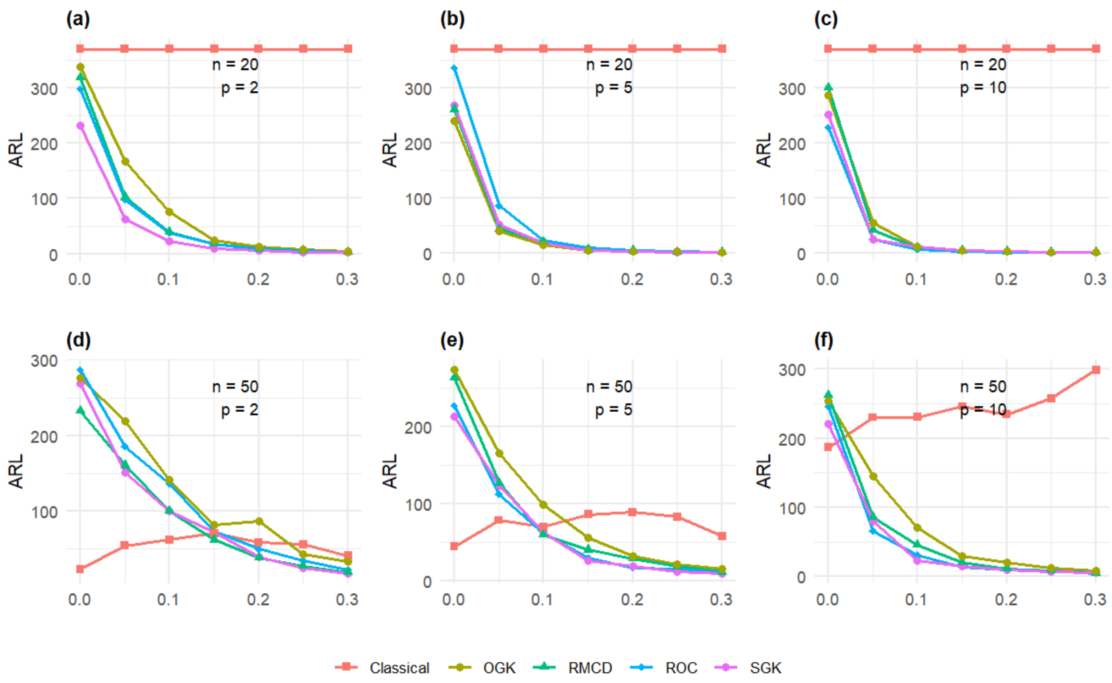


Figure 6.1: ARL plots of multivariate control charts based on ROC, SGK, RMCD, OGK and Classical estimators under Case A contamination for ε = 0.1 and δ^2 = 150.

Figure 6.1 and Figure 6.2 display ARL curves for classical and robust methods at different sample sizes (n = 20, 50) and dimensions (p = 2, 5, 10) under case A and case B contamination scenarios respectively with 10% contamination.

For the classical method, under both contamination scenarios, when n = 20, the ARL remains flat at the nominal in-control level across all shift magnitudes (panels a-c). This suggests that with small samples, the classical method is unable to detect shifts, even with contamination. However, for larger sample sizes (n = 50), the classical ARL curves become more erratic (panels d-f), showing unnatural decreases and increases as the shift grows. Especially in high dimensions (p = 10, panel f), the classical ARL stays high even under large shifts, clearly reflecting a breakdown under contamination.

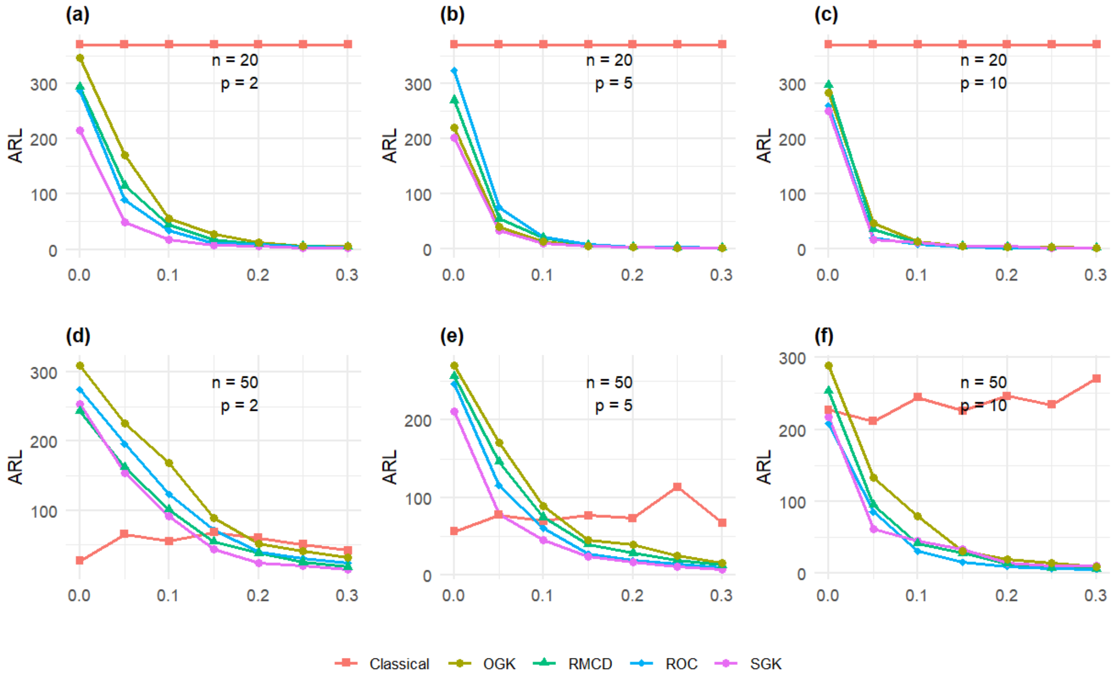


Figure 6.2: ARL plots of multivariate control charts based on ROC, SGK, RMCD, OGK and Classical estimators under Case B contamination for $\varepsilon = 0.1$ and $\delta^2 = 150$.

In contrast, the robust methods all show the desired pattern across the scenarios: high ARL at zero shift (good in-control performance) and a rapid decline as the shift magnitude increases (good out-of-control detection). Among them, under both contamination scenarios, ROC, SGK, and RMCD perform almost identically and efficiently. OGK also works well in this case, although in some panels (e.g., d and e), its detection slope is slightly less sharp than the others.

Sample size and dimension both influence the performance of the control charts. With a larger sample size (n = 50), the robust methods produce smoother and more consistent ARL curves compared to smaller samples (n = 20). As the dimension increases, the robust methods continue to perform reliably, whereas the classical method becomes increasingly unstable and unreliable.

Figure 6.3 and Figure 6.4 present the ARL curves of classical and robust methods across varying sample sizes (n = 20, 50) and dimensions (p = 2, 5, 10). Figure 6.3 corresponds to Case A contamination, while Figure 6.4 illustrates Case B contamination, both evaluated under a 20% contamination level.

Across all panels, the classical estimator consistently fails under contamination. Its ARL remains nearly constant at a high value, irrespective of the shift size, especially in the smaller-dimensional cases. This indicates its inability to detect mean shifts under contamination, highlighting its non-robustness. In some higher-dimensional cases (e.g., n = 50 and p = 10)), the classical method produces irregular ARL behaviour, further confirming its breakdown in contaminated environments.

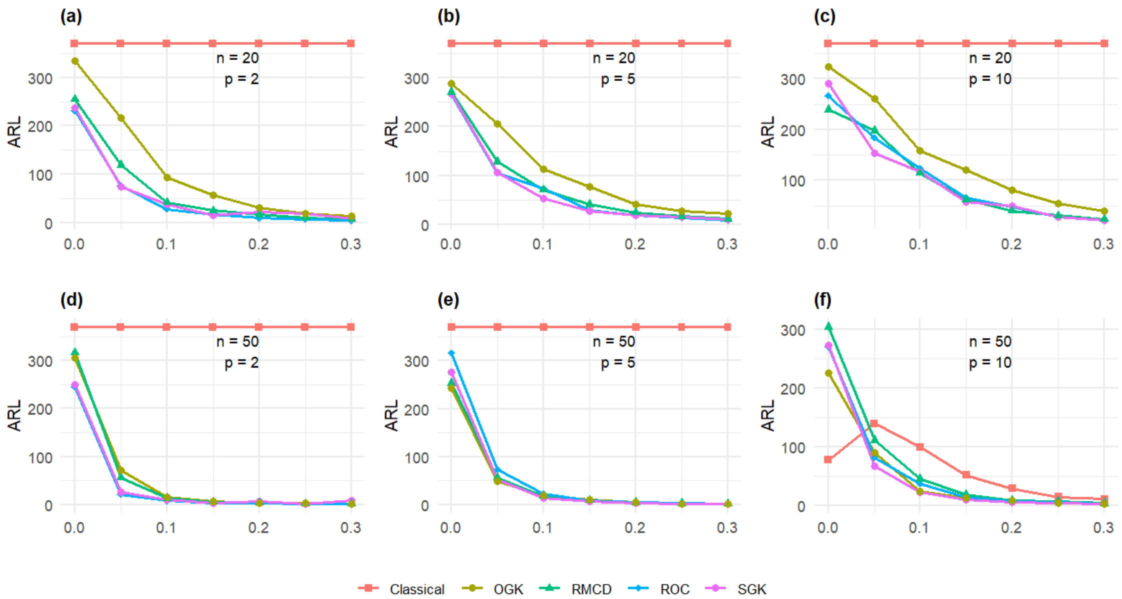


Figure 6.3: ARL plots of multivariate control charts based on ROC, SGK, RMCD, OGK and Classical estimators under Case A contamination for $\varepsilon = 0.2$ and $\delta^2 = 150$.

In contrast, the robust estimators demonstrate the expected ARL pattern: high ARL at zero shift (in-control) and steeply decreasing ARL as the shift increases under both contamination scenarios. Among them, ROC, SGK, and RMCD show very similar and strong performance across all settings. Their curves drop sharply with even small shifts, indicating high sensitivity to process changes while still maintaining control at zero shift. OGK, however, tends to be less efficient: its ARL decreases more slowly, particularly in lower sample size and higher dimension settings (e.g., n = 20 and p = 5, 10), which implies weaker sensitivity to shifts.

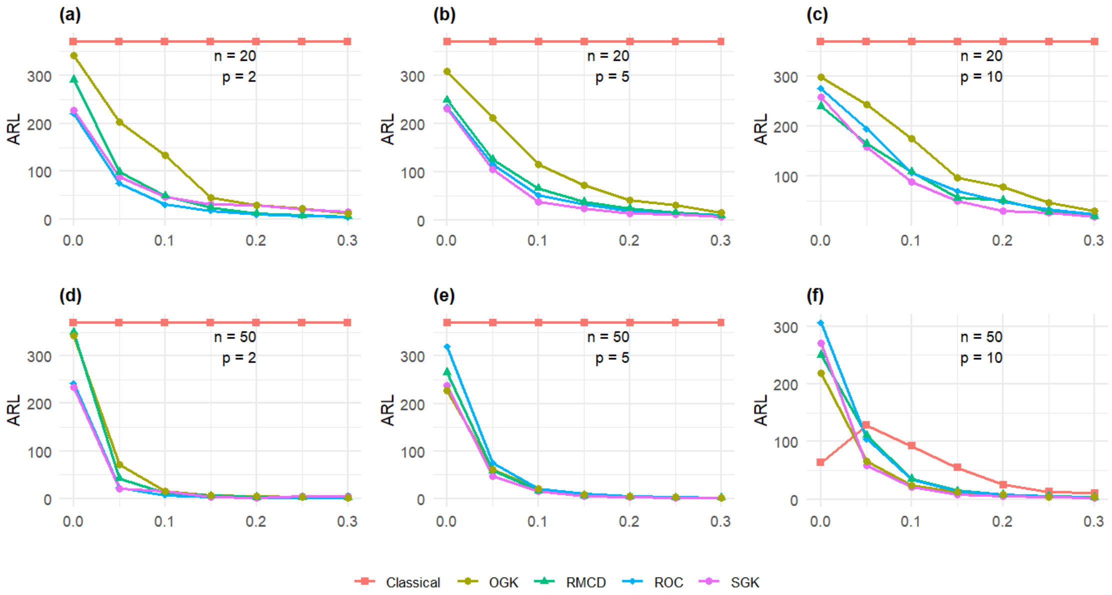


Figure 6.4: ARL plots of multivariate control charts based on ROC, SGK, RMCD, OGK and Classical estimators under Case B contamination for $\varepsilon = 0.2$ and $\delta^2 = 150$.

The effect of sample size is clearly evident. When $n = 50$, all robust methods produce tighter and more stable ARL curves, with faster detection compared to $n = 20$, confirming the stabilizing influence of larger samples on robust estimators. As the dimension increases (from $p = 2$ to $p = 10$), performance differences become more pronounced: OGK shows noticeable deterioration, while ROC, SGK, and RMCD maintain consistent and reliable performance.

At 10% contamination, all four robust methods performed very similarly, with their efficiency almost unaffected by the small level of contamination, demonstrating their stability and reliability even under mild contamination. In contrast, the classical estimator produced a flat, high ARL line at $n = 20$, indicating limited effectiveness but still functioning better than under higher contamination. At 20% contamination, the differences between robust estimators became more noticeable, with OGK performing relatively weaker, yet they still retained their desirable ARL pattern, highlighting their robustness under moderate contamination. Meanwhile, the classical estimator failed completely at this level, showing no detection ability or erratic behavior, especially in high-dimensional cases, confirming its high non-robustness and collapse in reliability as contamination increases. For Case C, similar patterns were observed, and the remaining results are presented in Appendix C.

6.5 Time Complexity

The computational times of the robust and classical multivariate control charts were examined across varying subgroup sizes (n = 20, 30, 50) and dimensions (p = 2, 5, 10). The results, presented in Table 6.2, reveal that the classical chart is by far the most efficient, with negligible computation times (0.002–0.004 seconds) regardless of subgroup size or dimension. Among the robust procedures, ROC is the fastest, followed by SGK, while OGK and RMCD are considerably slower. The effect of increasing subgroup size on computation time is relatively mild, whereas the impact of dimensionality is much more pronounced. For instance, at n = 30, the average running time of ROC increases from 0.028 seconds at p = 2 to 0.107 seconds at p = 10, while OGK rises sharply from 0.065 to 1.031 seconds over the same range. This pattern demonstrates that dimensionality plays a greater role than sample size in driving computational cost. Overall, the findings suggest that although robust charts inevitably require more computation than the classical approach, ROC and SGK offer a favorable balance between robustness and efficiency, whereas OGK and RMCD become computationally demanding in higher dimensions.

Table: 6.2 Computation time (in seconds)

n	p	ROC	SGK	OGK	RMCD	Classical
20	2	0.0299	0.0933	0.0595	0.2461	0.0022
	5	0.0468	0.1432	0.2477	0.2165	0.0025
	10	0.1061	0.3020	0.8348	0.5198	0.0041
30	2	0.0277	0.1110	0.0645	0.1852	0.0023
	5	0.0523	0.1947	0.2669	0.3083	0.0029
	10	0.1072	0.4167	1.0309	0.6587	0.0043
50	2	0.0334	0.2302	0.0748	0.2448	0.0023
	5	0.0590	0.3050	0.3041	0.3908	0.0027
	10	0.1349	0.8051	1.1517	0.9516	0.0042

6.6 Application to Real Life Data

To demonstrate the implementation of the proposed robust multivariate control charts, we analyze the carbon fiber dataset reported by Santos-Fernández (2012). This dataset comprises three correlated quality characteristics inner diameter, thickness, and length of carbon fiber tubes measured in inches. The data are organized into two phases: Phase I includes m = 30 samples of size n = 8, while Phase II consists of 25 additional samples, also of size $n = 8$. Figure 6.5 presents the control charts constructed using ROC, SGK, OGK, RMCD, and the Classical method. Under uncontaminated conditions, all robust charts exhibit patterns consistent with the classical chart, which performs optimally in such settings.

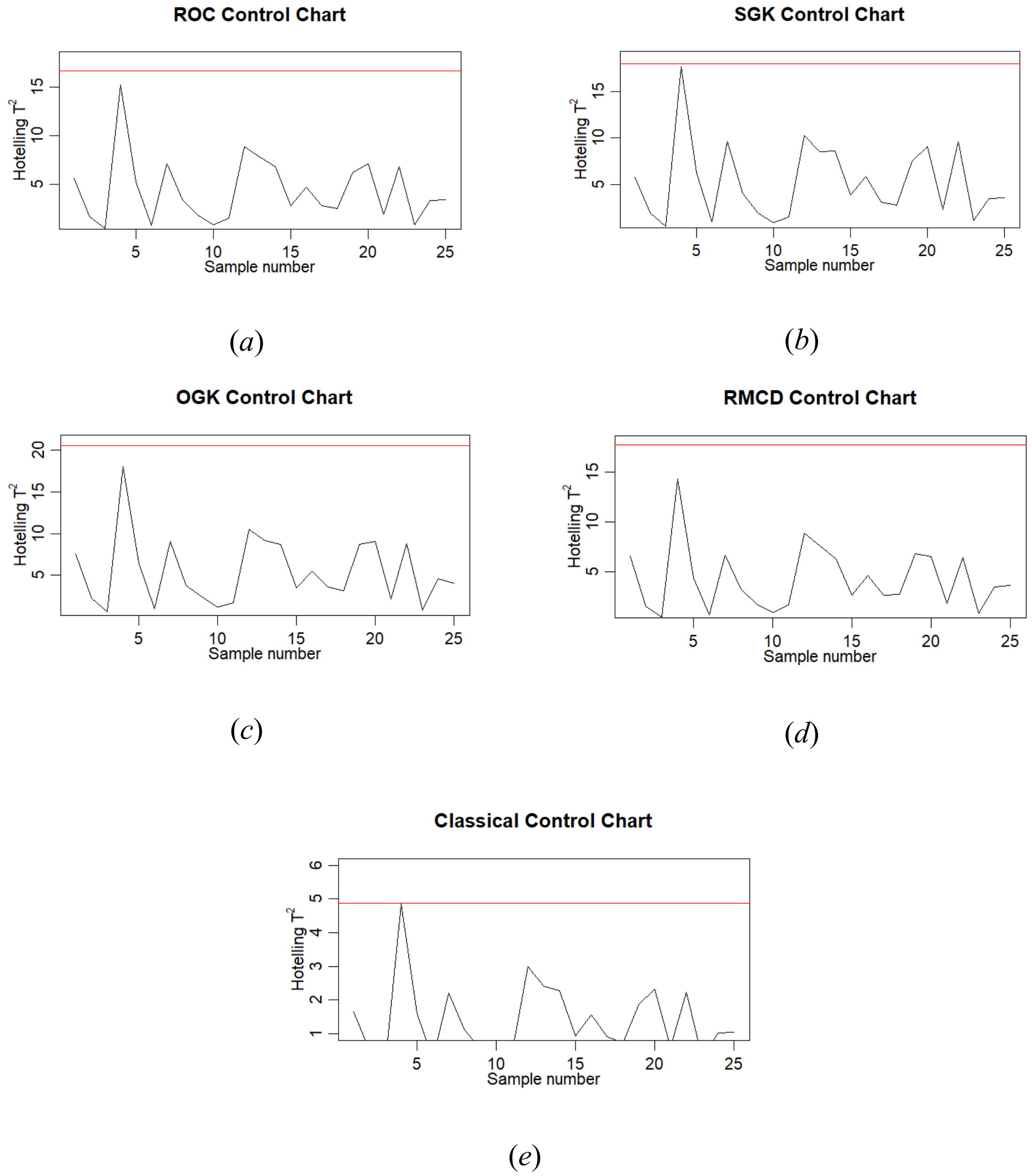
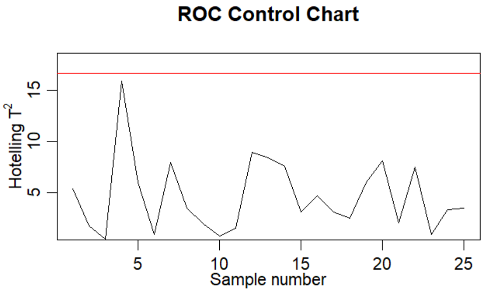
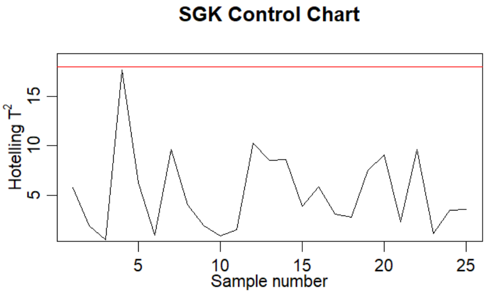


Figure 6.5: Control charts based on (*a*) ROC, (*b*) SGK, (*c*) OGK, (*d*) RMCD and (*e*) Classical estimators under uncontaminated Phase I data.

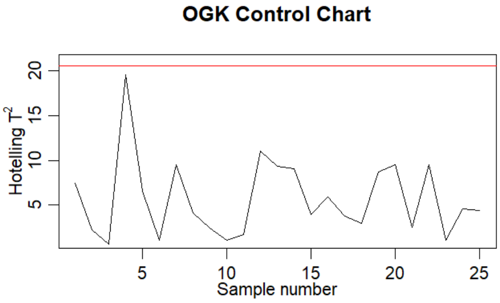
To assess performance under contamination, a mean shift of 2.5 was introduced to the first rational subgroup in Phase I. The resulting charts, shown in Figure 6.6, reveal that the robust methods remain unaffected, whereas the classical chart incorrectly signalled four subgroups (4, 12, 16, and 19) as out-of-control. These results highlight the improved robustness of the proposed approaches when data contamination is present.



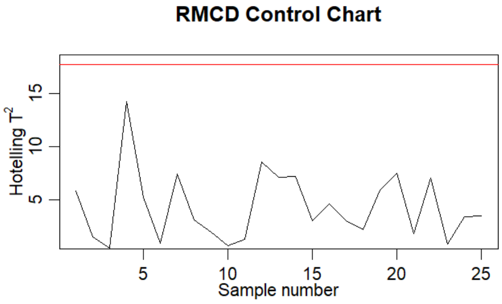
(*a*)



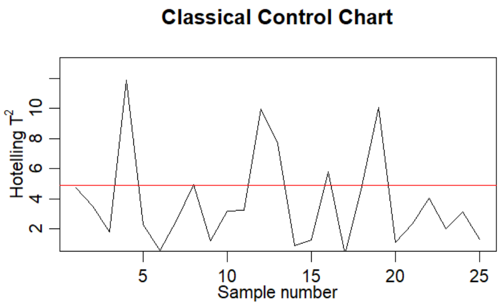
(*b*)



(*c*)



(*d*)



(*e*)

Figure 6.6: Control charts based on (*a*) ROC, (*b*) SGK, (*c*) OGK, (*d*) RMCD, and (*e*) Classical estimators under contaminated Phase I data.

6.7 Summary

The findings of this chapter demonstrate that the classical Hotelling's T^2 chart is highly sensitive to data contamination, often producing flat or erratic ARL patterns and failing to detect shifts reliably, particularly in small-sample and high-dimensional settings. In contrast, the proposed ROC and SGK based charts, along with RMCD, consistently maintained high in-control ARL values and showed rapid detection of process shifts

1under contamination. OGK also performed reasonably well but was less effective compared to the other robust alternatives.

The application to the carbon fiber dataset confirmed these results, while the classical chart generated false alarms under contamination, the proposed ROC and SGK charts, as well as RMCD, remained stable and identified only true out-of-control subgroups. These results highlight the practical value of the proposed approaches, as they combine robustness against outliers with efficiency under clean data, making them reliable tools for multivariate process monitoring in Phase II applications**.**

Chapter 7

MONITORING PROCESS VARIABILITY WITH BIVARIATE ROBUST CONTROL CHARTS

7.1 Introduction

Bivariate control charts serve as a fundamental statistical tool for simultaneous monitoring and detection of process deviations when two interrelated variables are of interest. Effective process monitoring necessitates the prompt detection of shifts in process parameters, whether in the mean, the variability, or both. All processes, whether in manufacturing, services, or other domains, inherently exhibit variation, which may lead to deviations from desired performance and reduced quality outcomes. As a result, the reduction of variability is a central focus in statistical quality control to ensure process stability and consistent quality. A number of multivariate procedures have been developed for simultaneous monitoring of such quality characteristics. Most of them have been developed to monitor the process mean vector. But monitoring the process variability is an important part of any control procedure (Woodall & Montgomery, 1999). Healy (1987) a cumulative sum (CUSUM) based dispersion chart for monitoring dispersion using individual observations, which considered shifts in both the mean vector and the covariance matrix. Tang and Barnett (1996) proposed methods to monitor multivariate process dispersion using subgroup data without needing prior covariance estimates, making them suitable for short-run production. To detect small shifts in process variability, Yeh *et al.* (2003) developed a Multivariate Exponentially Weighted Moving Average (MEWMA) control chart based on the probability integral transformation for rational subgroup samples. Subsequently, Yeh *et al.* (2004) proposed a likelihood ratio-based Exponentially Weighted Moving Average (EWMA) control chart for monitoring small changes in the variability of multivariate normal processes by applying EWMA to the log-likelihood ratio for testing equality of covariance matrices. Surtihadi *et al.* (2004) developed generalized likelihood ratio-based control charts for monitoring multivariate process dispersion, considering both Shewhart-type and CUSUM-type charts to handle specific cases of shifts in the covariance matrix. In order to monitor shifts in process variability for multivariate individual observations, Yeh *et* *al.* (2005) developed a MEWMA based control chart which shows strong performance, especially under high positive correlation. Following the work of Huwang *et al*. (2007), which established the effectiveness of Multivariate Exponentially Weighted Mean Square deviation (MEWMS) and Multivariate Exponentially Weighted Moving Variance (MEWMV) charts for monitoring variability in multivariate processes using individual observations, Memar and Niaki (2009) extended this approach by introducing control charts based on L_1 and L_2-norm distances. Rather than relying on the trace of the estimated covariance matrix, these charts measure deviations in diagonal elements from their expected values to improve sensitivity to process changes. Simulation studies demonstrated that these norm-based statistics enhance the detection of shifts in the covariance matrix, offering a more responsive alternative for multivariate dispersion monitoring.

7.1.1. MEWMS control charts

Let $\boldsymbol{x} = (x_1, x_2)'$ represent the process data vector consisting of a bivariate quality characteristic, assumed to follow a bivariate normal distribution with a mean vector $\boldsymbol{\mu_x}$ and covariance matrix $\boldsymbol{\Sigma_x}$. The parameters $\boldsymbol{\mu_x}$ and $\boldsymbol{\Sigma_x}$ are either known or estimated from the Phase I analysis of the in-control process conditions with $\boldsymbol{\mu_x} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma_x} = \boldsymbol{\Sigma}_0$. To facilitate further analysis, the process variable $\boldsymbol{x}$ is transformed into a new random vector $\boldsymbol{y}$, which follows a bivariate normal distribution with the mean $\boldsymbol{\mu_y}$ and covariance matrix $\boldsymbol{\Sigma_y}$ as defined below:

$$\boldsymbol{y} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}_0); \quad \boldsymbol{\mu_y} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\mu_x} - \boldsymbol{\mu}_0); \quad \boldsymbol{\Sigma_y} = \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma_x}\boldsymbol{\Sigma}_0^{-1/2}.$$

For an in-control process, the mean vector $\boldsymbol{\mu_y} = \boldsymbol{0}$ and $\boldsymbol{\Sigma_y} = \boldsymbol{I}_2$, where $\boldsymbol{0} = (0,0)'$ and $\boldsymbol{I}_2$ is a 2×2 identity matrix. In the case of individual observations, although the sample covariance matrix cannot be computed, the outer product $\boldsymbol{yy'}$ of each observation constitutes an unbiased estimator of $\boldsymbol{\Sigma_y}$ when the process mean $\boldsymbol{\mu_y} = \boldsymbol{0}$; however, $\boldsymbol{yy'}$ need not be a positive definite matrix. To address this, Huwang *et al*. (2007) proposed the exponentially weighted moving average for the t^{th} individual observation $\boldsymbol{y_t} = (y_1, y_2)'$ as follows:

$$\boldsymbol{S}_t = \omega \boldsymbol{y}_t \boldsymbol{y}_t' + (1 - \omega)\boldsymbol{S}_{t-1}, \quad t = 1,2,3 \ldots,$$

where ω is a smoothing constant, $0 < \omega < 1$, and $\boldsymbol{S}_0 = \boldsymbol{y}_t \boldsymbol{y}_t'$.

Huwang *et al*. (2007) established that, when there is no shift in the process mean, the matrix $\boldsymbol{S}_t$ becomes positive definite for $t \geq 2$ with probability 1 and its expectation satisfies $\mathrm{E}(\boldsymbol{S}_t) = \boldsymbol{\Sigma}_{\boldsymbol{y}}$. Since the trace of $\boldsymbol{S}_t$ quantifies the overall variability across the quality characteristics, the trace is used as monitoring statistic for detecting changes in the covariance matrix. The following expression for the trace:

$$tr(\boldsymbol{S}_t) = \sum_{i=1}^{t} c_i \, tr(\boldsymbol{y}_t \boldsymbol{y}_t') = \sum_{\boldsymbol{i=1}}^{\boldsymbol{t}} c_i \, (y_{i1}^2 + y_{i2}^2). \tag{7.1}$$

The mean and variance of $tr(\boldsymbol{S}_t)$ are given by 2 and $4\sum_{i=1}^{t} c_i^2$ respectively, where $\sum_{i=1}^{t} c_i^2 = \frac{\omega}{(2-\omega)} + \frac{2-2\omega}{2-\omega}(1-\omega)^{2(t-1)}$ which will converge to $\frac{\omega}{(2-\omega)}$ as $t \to \infty$. Based on this, the control limits for MEWMS control chart are defined as

$$E[tr(\boldsymbol{S}_t)] \pm \sqrt{\mathrm{Var}[tr(\boldsymbol{S}_t)]} = 2 \pm L\sqrt{4\sum_{i=1}^{t} c_i^2}\,,$$

where the value of L is typically determined by Monte Carlo simulation to achieve a desired in-control average run length (ARL_0).

7.1.2 MEWMV control charts

The MEWMS chart is primarily constructed to detect changes in the covariance structure under the assumption that the process mean remains stable. However, in practical scenarios, shifts in both the mean and the variability may occur simultaneously during the monitoring period. This highlights the need for a control chart capable of detecting changes in process variability as well as shift in process mean.

Huwang *et al*. (2007) proposed the MEWMV chart based on the statistic $\boldsymbol{V}_t$, which is constructed similarly to $\boldsymbol{S}_t$, with following modification:

$$\boldsymbol{V}_t = \omega(\boldsymbol{y}_t - \boldsymbol{z}_t)(\boldsymbol{y}_t - \boldsymbol{z}_t)' + (1-\omega)\boldsymbol{V}_{t-1}, \quad t = 1,2,3 \ldots,$$

where ω is a smoothing constant, $0 < \omega < 1$, and $\boldsymbol{V}_0 = (\boldsymbol{y}_1 - \boldsymbol{z}_1)(\boldsymbol{y}_1 - \boldsymbol{z}_1)'$. Here, a natural estimate $\boldsymbol{z}_t$ for the process mean at time t is the multivariate exponentially moving average of $\boldsymbol{y}_t$ (Lowry *et al.*, 1992) and its values are computed recursively as:

$$\boldsymbol{z}_t = \lambda \boldsymbol{y}_t + (1-\lambda)\boldsymbol{z}_{t-1}, \quad t = 1,2,3 \ldots,$$

with $\boldsymbol{z}_0 = \boldsymbol{0}$ and $\lambda \in (0,1)$ being another smoothing constant.

For $t \geq 2$, $\boldsymbol{V}_t$ is a positive definite matrix with probability 1 and, the expected value converges as:

$$E(\boldsymbol{V}_t) \to \frac{2(1-\lambda)^2}{(2-\lambda)} \boldsymbol{\Sigma}_{\boldsymbol{y}} \text{ as } t \to \infty.$$

Hence, $\frac{(2-\lambda)}{2(1-\lambda)^2} \boldsymbol{V}_t$ can be used as an unbiased estimator of $\boldsymbol{\Sigma}_{\boldsymbol{y}}$. The trace of $\boldsymbol{V}_t$ can be simplified to the following form:

$$tr(\boldsymbol{V}_t) = \sum_{i=1}^{t} \sum_{j=1}^{t} q_{ij} \, (x_{i1}x_{j1} + x_{i2}x_{j2}), \tag{7.2}$$

where $\boldsymbol{Q} = (q_{ij})$ is defined as:

$$\boldsymbol{Q} = (\boldsymbol{I}_t - \boldsymbol{M})'\boldsymbol{C}(\boldsymbol{I}_t - \boldsymbol{M}).$$

Here, $\boldsymbol{C} = diag((1-\omega)^{t-1}, \omega(1-\omega)^{t-2}, \dots, \omega(1-\omega), \omega))$, $\boldsymbol{I}_t$ is the $t \times t$ identity matrix and $\boldsymbol{M}$ is a lower triangular matrix determined by the smoothing parameter λ. The mean and variance of $tr(\boldsymbol{V}_t)$ are given by $2tr(\boldsymbol{Q}_t)$ and $4\sum_{i=1}^{t} \sum_{j=1}^{t} q_{ij}^2$ respectively. Hence, the control limits for the MEWMV chart are expressed as

$$E[tr(\boldsymbol{V}_t)] \pm \sqrt{\mathrm{Var}[tr(\boldsymbol{V}_t)]} = 2tr(\boldsymbol{Q}_t) \pm L \sqrt{4 \sum_{i=1}^{t} \sum_{j=1}^{t} q_{ij}^2},$$

where the value of L can be found using Monte Carlo simulation based on ARL_0.

7.1.3 Control charts based on $\mathrm{L_c}$-norm function

Memar and Niaki (2009) proposed the $\mathrm{MEWMSL_1}$ and $\mathrm{MEWMSL_2}$ control charts by employing the $\mathrm{L_1}$ norm and $\mathrm{L_2}$ norm distances between the vector of diagonal elements of $\boldsymbol{S}_t$. The respective monitoring statistics are defined as follows:

$$D_1(\boldsymbol{S}_t) = \left\| \left(S_{t(11)}, S_{t(22)}\right)' - \boldsymbol{1}_2 \right\| = \left|S_{t(11)} - 1\right| + \left|S_{t(22)} - 1\right|, \tag{7.3}$$

$$D_2(\boldsymbol{S}_t) = \left\| \left(S_{t(11)}, S_{t(22)}\right)' - \boldsymbol{1}_2 \right\|^2 = (S_{t(11)} - 1)^2 + (S_{t(22)} - 1)^2, \tag{7.4}$$

where $\boldsymbol{1}_2$ is a two-dimensional vector of ones, and $S_{t(ii)}$ represents the i^{th} diagonal element of the matrix $\boldsymbol{S}_t$. The control limits for these charts can be determined using Monte Carlo simulation based on the desired ARL_0.

Extending this concept, Memar and Niaki (2009) developed a related monitoring framework utilizing the $\boldsymbol{V}_t$ statistic. Since $E(\boldsymbol{V}_t) = \frac{2(1-\lambda)^2}{(2-\lambda)}\boldsymbol{\Sigma}_{\boldsymbol{y}}$ for large values of t, the transformed statistic $\boldsymbol{V}_t^* = \frac{(2-\lambda)}{2(1-\lambda)^2}\boldsymbol{V}_t$ can be used to monitor the process variability more effectively. Based on this transformed matrix, two monitoring statistic are defined as follows:

$$D_1(\boldsymbol{V}_t^*) = \left\|\left(V_{t(11)}^*, V_{t(22)}^*\right)' - \mathbf{1}_2\right\| = \left|V_{t(11)}^* - 1\right| + \left|V_{t(22)}^* - 1\right|, \quad (7.5)$$

$$D_2(\boldsymbol{V}_t^*) = \left\|\left(V_{t(11)}^*, V_{t(22)}^*\right)' - \mathbf{1}_2\right\|^2 = (V_{t(11)}^* - 1)^2 + (V_{t(22)}^* - 1)^2, \quad (7.6)$$

where $V_{t(ii)}^*$ denotes the i^{th} diagonal element of the matrix $\boldsymbol{V}_t^*$. These statistics, D_1 and D_2 respectively represent the L_1 and L_2 norm distances of the diagonal elements of $\boldsymbol{V}_t^*$ from their target values. The corresponding control limits can be determined via Monte Carlo simulation, depending on p, ω, λ and ARL_0.

Huwang *et al*. (2007) and Memar and Niaki (2009) constructed the control charts under the assumption that the in-control parameters are known. However, in practical applications, these parameters must be estimated from Phase I data. When Phase I data contain outliers, the estimates can be severely distorted, compromising chart performance.

Variyath and Vattathoor (2014) addressed this issue by introducing robust control charts for Phase I data based on the re-weighted minimum covariance determinant (RMCD) and re-weighted minimum volume ellipsoid (RMVE) proposed by Rousseeuw (1985). These estimators were integrated into MEWMS and MEWMV schemes to monitor process variability in multivariate individual observations. Simulation studies conducted for bivariate quality characteristics demonstrated that the robust charts outperform classical methods, especially in the presence of contamination in the phase I data.

A key limitation of both RMCD and RMVE estimators is their high computational cost, especially when applied to large or high-dimensional datasets. Additionally, these methods require the user to predefine the subset size, which reflects an assumed proportion of uncontaminated observations. Since the true level of contamination is typically unknown in practice, mis specifying this parameter may adversely affect the robustness and efficiency of the estimators. To address this drawback in the bivariate context, the present study proposes robust bivariate dispersion control charts based on robustified MEWMS and MEWMV statistics, utilizing high-breakdown robust covariance estimators such as the Comedian, S_n Covariance and Gnanadesikan-Kettenring estimators, which are discussed in Chapter 2.

7.2 Bivariate Robust Dispersion Control Chart for Individual Observations

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ represent n Phase I observations from a bivariate quality characteristic, based on which the robust bivariate dispersion control chart is constructed as follows:

1. *Estimation of robust location and dispersion parameters*: Using the Phase I data, compute the robust estimates of location and dispersion parameters, denoted as $\widehat{\boldsymbol{\mu}}_R$ and $\widehat{\boldsymbol{\Sigma}}_R$, respectively.

2. *Transformation of phase II observations*: For each new Phase II observation $\boldsymbol{y_t}$, obtain the standardized transformation $\boldsymbol{y}_t^*$ as:

$$\boldsymbol{y}_t^* = \widehat{\boldsymbol{\Sigma}}_R^{-\frac{1}{2}}(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_R)$$

3. *Computation of the monitoring statistic*: Select the appropriate monitoring statistic as defined in Equations (7.1)-(7.6), and evaluate it using the transformed observation $\boldsymbol{y_t^*}$.

4. *Charting and interpretation*: Plot the computed statistic on the control using empirically determined control limits for a given significance level α. Any observation exceeding the control limits indicates a potential shift in the process, warranting further investigation or corrective action.

As the monitoring statistics defined above are always positive, only upper control limits are required. To establish these limits, 10,000 Monte Carlo simulations were conducted for various combinations of the smoothing parameters ω and λ and for confidence levels corresponding to $\boldsymbol{\alpha}$ =0.05, 0.01, and 0.0027.

To estimate the upper control limit (UCL) for a given significance level α, we generate n individual observations from the standard bivariate normal distribution $N_2(\mathbf{0}, \boldsymbol{I}_2)$ as Phase I sample, and compute the robust estimators of the mean vector and covariance matrix. In addition, for each simulated dataset, we randomly generate a new n

observation from $N_2(\mathbf{0}, \boldsymbol{I}_2)$ (treated as Phase II observations) and calculate the corresponding statistic in Equation (7.1). This process is repeated 10,000 times, and the UCL for a given α is obtained as the empirical $100(1 - \alpha)^{th}$ percentile of the computed statistic. The same procedure is applied to obtain UCLs for the statistics defined in equations (7.2), (7.3), (7.4), (7.5) and (7.6). UCLs for n = 50 and 100 were computed using various estimation methods including Comedian, S_n Covariance, Classical estimator, RMCD, RMVE, GK_Q_n and GK_S_n for α = 0.05, 0.01 and 0.0027. The UCLs of $MEWMSL_1$, $MEWMSL_2$, MEWMS charts with ω = 0.4, and those of $MEWMVL_1$, $MEWMVL_2$ and MEWMV charts with ω = 0.4 and λ = 0.2, corresponding to α = 0.0027 for n = 50 and 100, are presented in Table 7.1 and Table 7.2.

Table 7.1: Simulated UCLs for n = 50 and α = 0.0027

Method	$MEWMSL_1$	$MEWMSL_2$	MEWMS	$MEWMVL_1$	$MEWMVL_2$	MEWMV
Comedian	6.91	38.12	8.53	11.75	96.27	6.01
S_n Covariance	5.77	26.37	7.34	9.29	57.55	5.15
Classical	4.44	14.72	5.95	7.46	33.4	4.29
RMCD	10.62	86.29	12.37	18.04	225.89	8.44
RMVE	7.21	40.22	8.78	11.07	89.56	6.03
GK_Q_n	5.57	24.18	7.18	8.91	50.73	4.96
GK_S_n	6.14	29.45	7.75	10.15	67.03	8.65

Table 7.2: Simulated UCLs for n = 100 and α = 0.0027

Method	$MEWMSL_1$	$MEWMSL_2$	MEWMS	$MEWMVL_1$	$MEWMVL_2$	MEWMV
Comedian	5.73	25.57	7.35	10.08	67.5	5.16
S_n Covariance	5.16	20.47	6.75	9	51.62	4.74
Classical	4.48	15.19	6.09	8.08	39.14	4.37
RMCD	6.83	36.17	8.54	11.9	94.27	5.94
RMVE	5.45	22.85	7.06	9.22	57.31	4.97
GK_Q_n	5.02	19.47	6.67	8.72	47.86	4.69
GK_S_n	5.24	21	6.9	9.25	54.06	4.87

7.3 Simulation Study

The performance of the proposed control charts is compared with that of control charts based on the classical, RMCD and RMVE estimators. We consider a bivariate process with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, defined as

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \delta\rho\sigma_1\sigma_2 \\ \delta\rho\sigma_1\sigma_2 & \delta^2\sigma_2^2 \end{pmatrix}.$$

For the in-control process, the parameters are set as $\mu_1 = \mu_2$=0, $\sigma_1^2 = \sigma_1^2 = 1, \delta = 1$ and $\rho = 0$. In Phase I analysis, the data were generated such that a proportion $(1 - \varepsilon)$ was drawn from the standard bivariate normal distribution $N_2(\mathbf{0}, \boldsymbol{I})$, while the remaining ε proportion was generated from a contaminated distribution $N_2(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I)$.

To evaluate the performance of the proposed control charts under out-of-control conditions, various shift scenarios were considered. The correlation coefficient ρ was set to 0, 0.5, and 0.9 to represent different levels of correlation between the variables. For MEWMV-based control charts, a shift in the process mean was introduced using $\boldsymbol{\mu}_I = (2\ 2)^t$ while a scale shift was incorporated through δ = 3 for both MEWMS and MEWMV based control charts. The study was conducted using sample sizes of n = 50 and 100, with contamination levels ε = 0.10 and 0.20, to assess the robustness and sensitivity of the control charts across various conditions using ARL plots.

7.3.1 ARL Plots

ARL is a fundamental performance metric for assessing the effectiveness of control charts. It denotes the expected number of samples required before an out-of-control condition is signalled. An efficient control chart is characterized by a high in-control ARL, which minimizes false alarms, and a low out-of-control ARL, which ensures prompt detection of actual process shifts.

In Phase II, simulated samples were tested to assess the sensitivity of the control charts to changes in process conditions. A mean shift of $\boldsymbol{\mu}_{II} = (1\ 1)^t$ was introduced specifically for the MEWMV based control charts. To evaluate sensitivity to dispersion changes, the scale shift parameter δ was gradually increased from 1 to 2 in increments of 0.1 (i.e., δ = 1.0, 1.1, 1.2, …, 2.0) and applied to both MEWMS and MEWMV based charts.

The significance level is set at α = 0.0027, corresponding to an ARL_0 of 370. In phase II, the number of observations required for the monitoring statistic to exceed the control limit is recorded as a run length observation RL_i. For simulations where no signal occurs within 400 observations, the run length is truncated and recorded as 400. The ARL is then computed across r =1000 iteration using the formula:

$$\text{ARL} = \frac{1}{r}\sum_{i=1}^{r} RL_i.$$

ARL plots corresponding to a contamination level of 20%, smoothing constant ω = 0.4 and λ=0.2 are presented here. The remaining results are provided in the Appendix D.

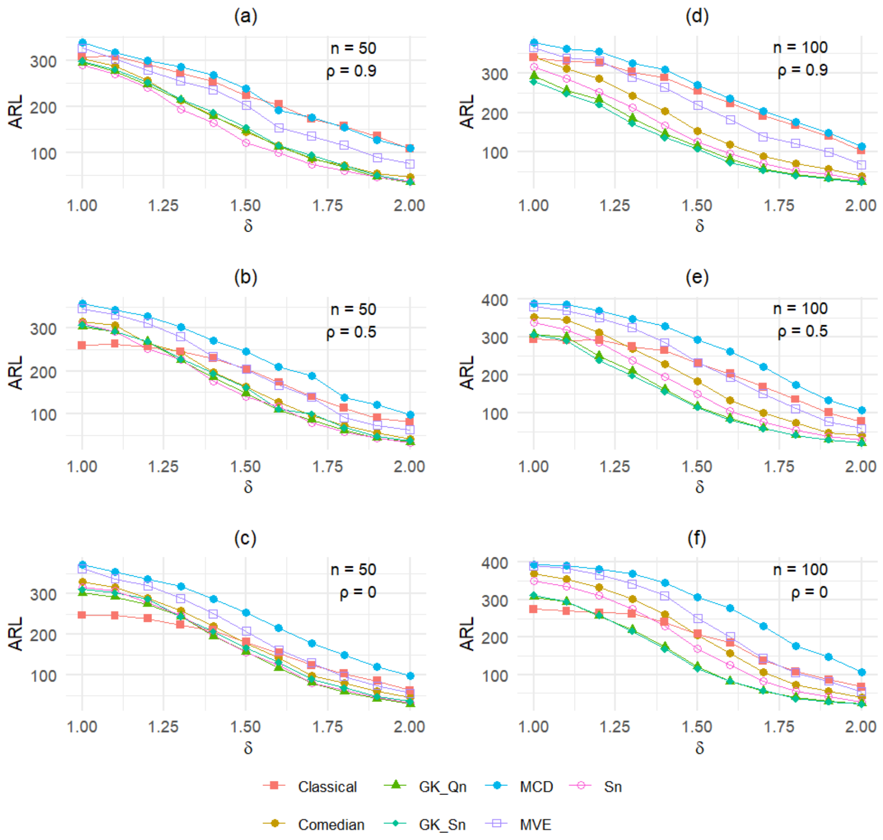


Figure 7.1: ARL plots for MEWMSL$_1$ statistic with ε = 0.2 and ω = 0.4

Figure 7.1 presents the ARL plots for the MEWMSL$_1$ statistic. As seen in this figure, robust estimators such as GK_Q$_n$, GK_S$_n$, S_n Covariance, and Comedian consistently outperform the Classical, RMCD, and RMVE estimators, particularly for larger shifts (δ ≥ 1.5). As the correlation coefficient ρ decreases or the sample size n increases, detection performance improves across all methods, while the relative ranking among estimators remains consistent.

Figure 7.2 displays the ARL plots for the MEWMSL$_2$ statistic. Similar to Figure 7.1, this plot shows that GK_Q$_n$, GK_S$_n$, S_n Covariance, and Comedian achieve faster shift detection, especially under high correlation (ρ = 0.9) and smaller sample sizes (n = 50). Classical, RMCD, and RMVE exhibit delayed detection, particularly in more challenging conditions.

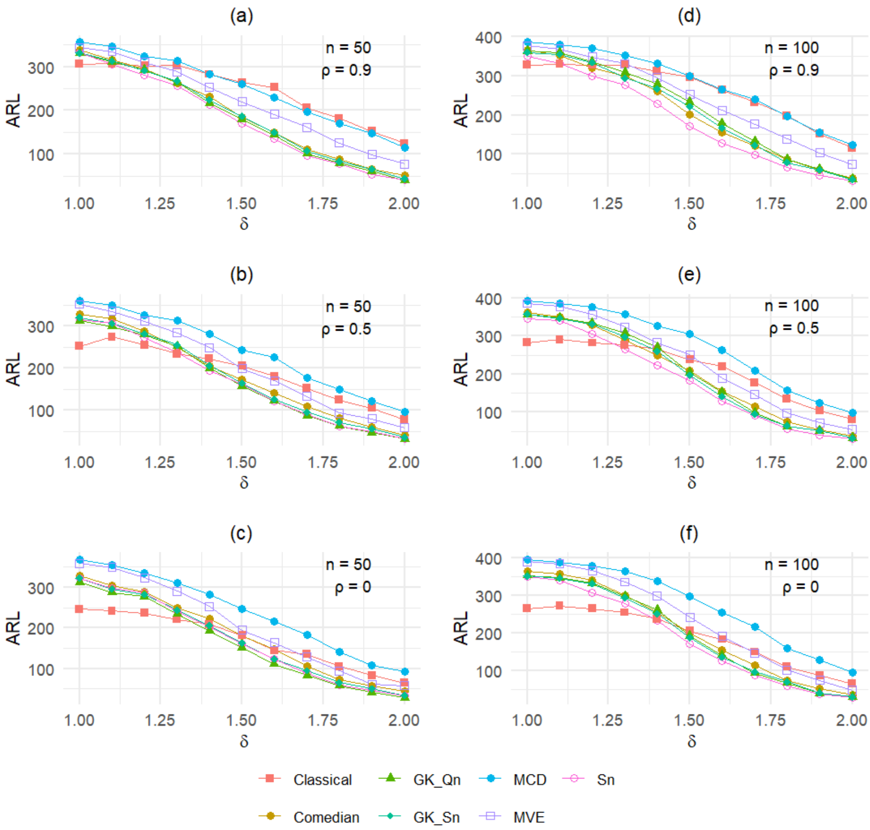


Figure 7.2: ARL plots for $MEWMSL_2$ statistic with $\varepsilon = 0.2$ and $\omega = 0.4$

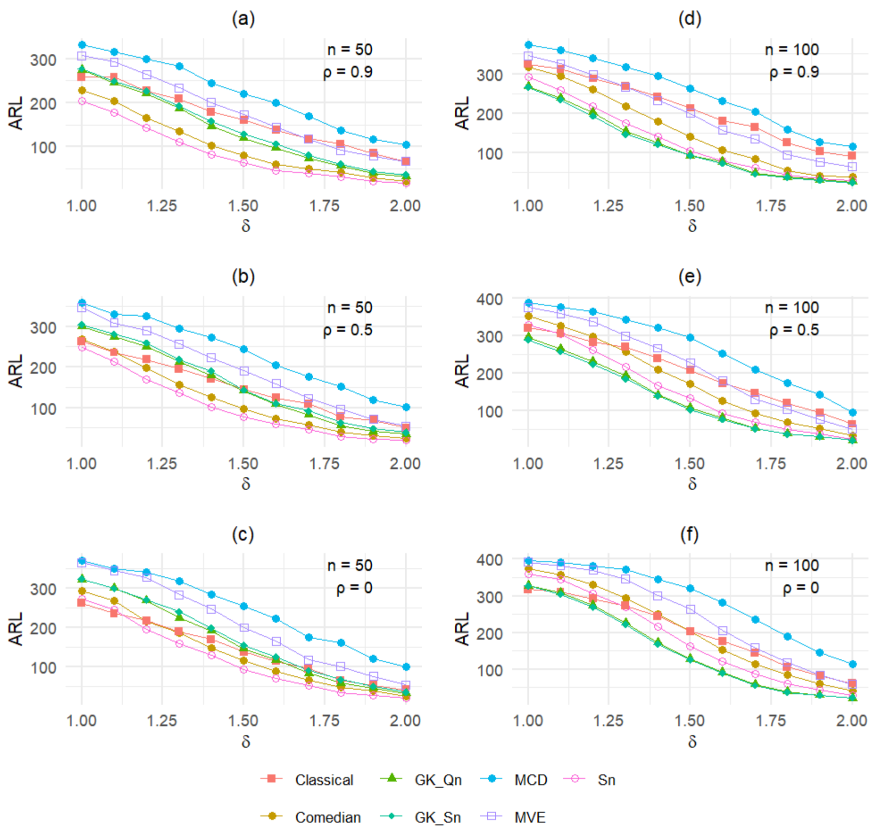


Figure 7.3: ARL plots for MEWMS statistic with $\varepsilon = 0.2$ and $\omega = 0.4$.

Figure 7.3 presents the ARL plots for the MEWMS statistic. As illustrated, GK_Q_n and GK_S_n continue to achieve the lowest ARLs across all scenarios. S_n Covariance emerges as the most effective under small-to-moderate shifts, while Comedian performs particularly well under low correlation and larger sample sizes. In contrast, Classical, RMCD, and RMVE estimators show significantly higher ARLs, especially under high correlation and small shifts. Increasing the sample size from 50 to 100 results in improved performance for all estimators, consistent with the trend observed in Figure 7.1 and Figure 7.2.

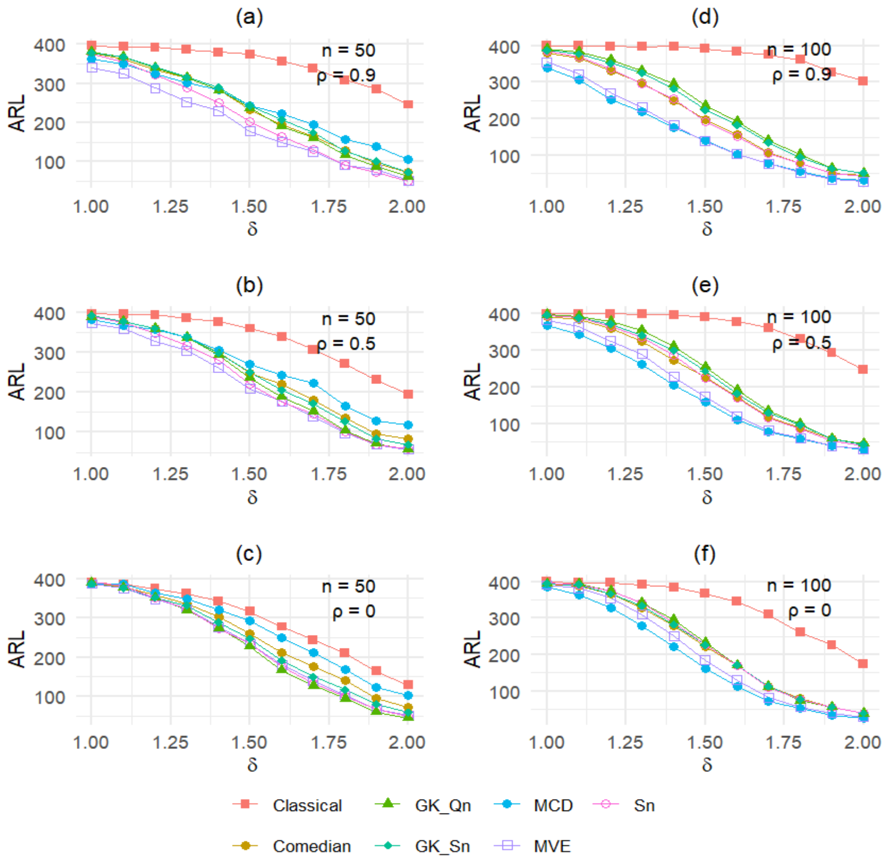


Figure 7.4: ARL plots for $MEWMVL_1$ statistic with $\varepsilon = 0.2$, $\omega = 0.4$ and $\lambda = 0.2$

Figure 7.4 illustrates the ARL plots for the $MEWMVL_1$ statistic. As seen here, the Classical estimator consistently shows the highest ARLs, indicating weak detection performance particularly under high correlation ($\rho = 0.9$). In contrast, robust estimators, particularly S_n Covariance, RMVE, and the GK-based methods, yield much lower ARLs, reflecting enhanced sensitivity. These results align with those in Figures 7.1 through 7.3, reaffirming the advantage of robust methods.

Figure 7.5 displays the ARL plots for MEWMVL$_2$. As with previous figures, the Classical estimator performs the worst across all conditions. Robust estimators—especially S_n Covariance, GK_S$_n$, and Comedian demonstrate superior performance across all shift levels and correlation structures. Their robustness remains evident even under strong correlation (ρ = 0.9). RMVE performs moderately well, outperforming RMCD and GK_Q$_n$ as correlation decreases. Increasing the sample size continues to enhance detection capability.

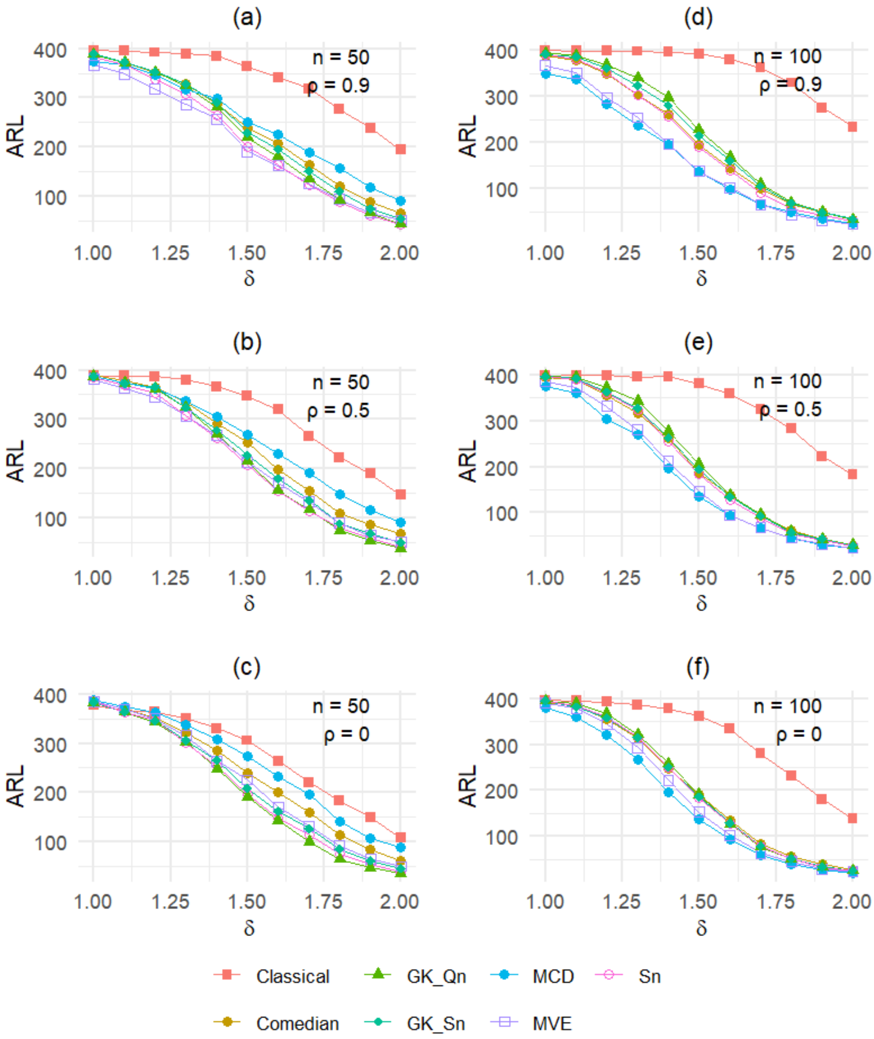


Figure 7.5: ARL plots for MEWMVL$_2$ statistic with $\varepsilon = 0.2$, $\omega = 0.4$ and $\lambda = 0.2$

Figure 7.6 shows the ARL plots for the MEWMV statistic. As in earlier figures, the Classical estimator consistently yields the highest ARLs, confirming its poor performance in the presence of contamination. In contrast, all six robust estimators show substantially lower ARLs. Among them, S_n Covariance, GK_S$_n$, and Comedian emerge as the most effective, particularly for larger shifts ($\delta \geq 1.5$) and under all correlation and sample size settings.

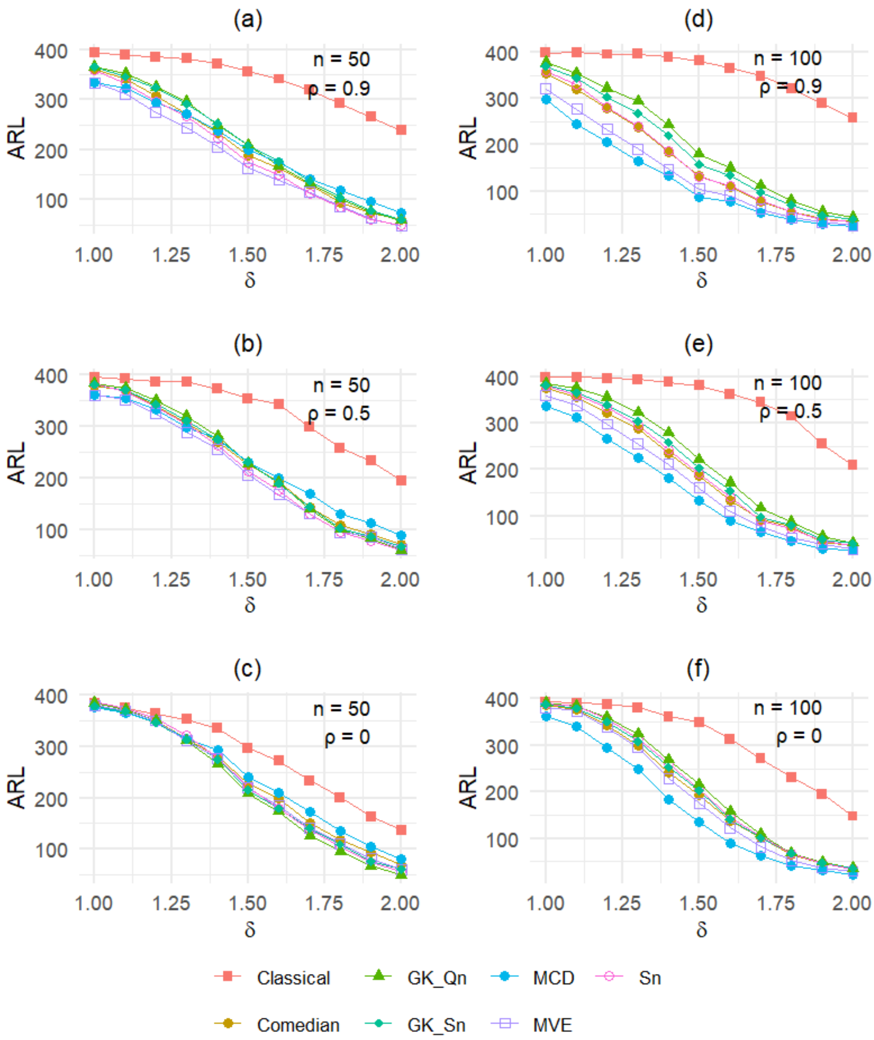


Figure 7.6: ARL plots for MEWMV statistic with $\varepsilon = 0.2$, $\omega = 0.4$ and $\lambda = 0.2$

Across all MEWMS and MEWMV statistics (Figures 7.1–7.6), robust estimators GK_S_n, S_n Covariance, and Comedian consistently achieve lower ARL values, showing strong detection performance under contamination. In contrast, the Classical estimator performs the worst, particularly under high correlation ($\rho = 0.9$) and small sample sizes ($n = 50$). GK_Q_n shows excellent performance for MEWMS statistics but is slightly less effective under MEWMV statistics, especially compared to S_n Covariance, GK_S_n, and Comedian. RMVE performs moderately well, while RMCD shows varying effectiveness.

For the remaining values of sample size (n), contamination level (ε), and smoothing constants (ω and λ), the ARL plots (presented in Appendix C) exhibit a consistent pattern.

This reaffirms the robustness and reliability of estimators like S_n Covariance, GK_S_n, and Comedian across a wide range of conditions.

7.4 Time Complexity

Table 7.3 presents the computation time (in seconds) for constructing bivariate control charts using various estimators across three statistics MEWMVL$_1$, MEWMVL$_2$, and MEWMV for sample sizes of 50 and 100. As expected, running times increase slightly with larger sample sizes. The Classical estimator is the fastest overall, with computation times as low as 0.0010 seconds, but it lacks robustness under contamination. Among robust methods, the Comedian, GK_Q_n, and GK_S_n estimators exhibit excellent computational efficiency, with running times generally between 0.0011 and 0.0026 seconds, making them practical for real-time or large-scale applications. The S_n Covariance estimator, while robust, takes slightly longer, especially for larger sample sizes. In contrast, RMCD and RMVE are the most computationally intensive, with RMCD reaching up to 0.0123 seconds and RMVE up to 0.0078 seconds, which may limit their use in time-sensitive settings.

Table 7.3: Computation time (in seconds)

Statistic	n	Comedian	S_n Covariance	GK_Q_n	GK_S_n	RMCD	RMVE	Classical
MEWMVL$_1$	50	0.0017	0.0028	0.0019	0.0017	0.0086	0.0070	0.0016
	100	0.0023	0.0049	0.0026	0.0025	0.0123	0.0077	0.0021
MEWMVL$_2$	50	0.0014	0.0022	0.0013	0.0012	0.0081	0.0053	0.0010
	100	0.0023	0.0045	0.0022	0.0020	0.0116	0.0078	0.0019
MEWMV	50	0.0014	0.0023	0.0013	0.0011	0.0077	0.0053	0.0011
	100	0.0022	0.0044	0.0022	0.0020	0.0109	0.0078	0.0020

7.5 Application to Real Life Data

To illustrate the application and comparative performance of robust bivariate control charts, Figure 7.7 presents six charts constructed using the MEWMVL$_1$ statistic for the first and second quality characteristics of the Chemical Process data from Montgomery (2020). This dataset includes 20 Phase I (in-control) and 10 Phase II (monitoring) observations. Based on prior analysis, a shift in the process is expected to occur around observation 24 or 25. The control limits for these charts were simulated using the parameters $n = 20$, $\omega = 0.4$, and $\lambda = 0.2$, as previously described.

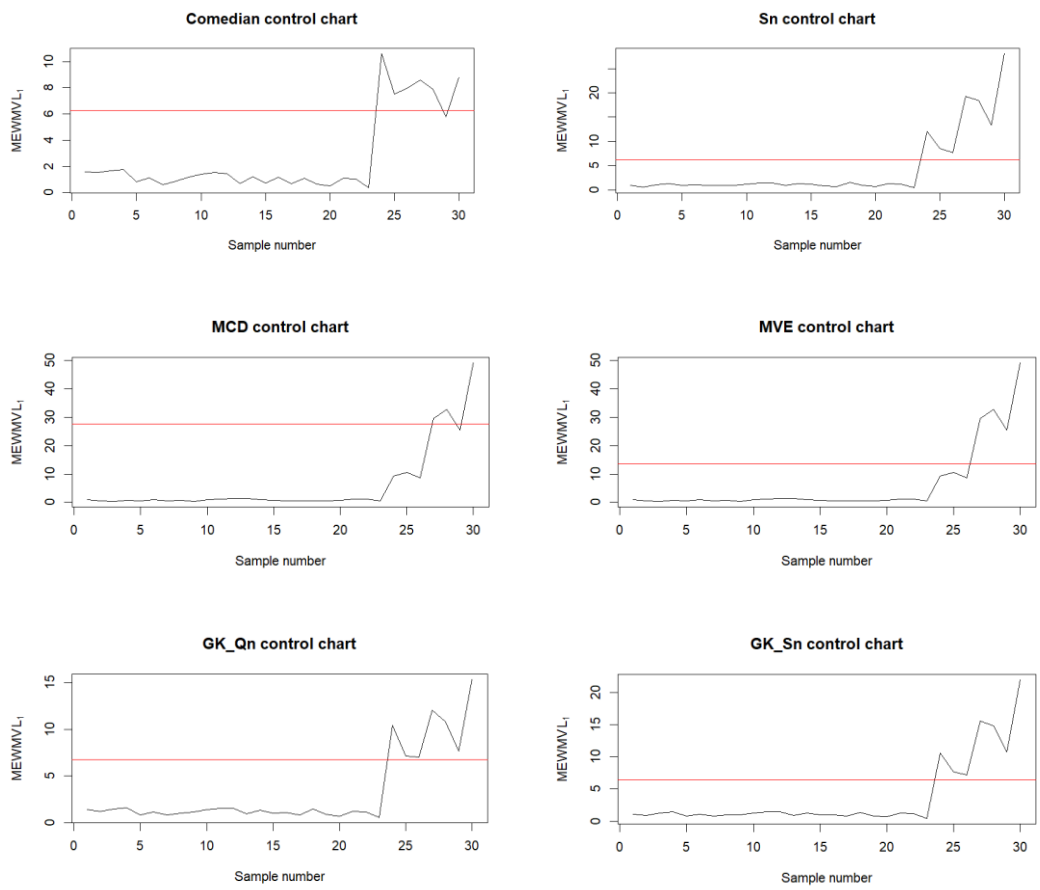


Figure 7.7: Robust bivariate control charts for Chemical Process data using MEWMVL₁ statistic

The charts constructed using the S_n Covariance, GK_S$_n$, and Comedian estimators show a clear and early detection of the process shift, with the control statistic rising sharply above the control limit starting at observation 24. GK_Q$_n$ also detects the shift, although its response under the MEWMVL$_1$ framework is slightly less pronounced compared to S_n Covariance and GK_S$_n$, which aligns with simulation results. RMCD and RMVE estimators identify the shift but with more variability or delayed response. These findings are consistent with the simulation study, where S_n Covariance, GK_S$_n$, and Comedian consistently achieved lower ARL values across various scenarios, confirming their robustness and reliability under contamination and high correlation.

7.6 Summary

This study develops robust bivariate control charts for individual observations by integrating high-breakdown estimators (S_n Covariance, GK_S$_n$, Comedian, and GK_Q$_n$)

into MEWMS and MEWMV frameworks. These charts address the limitations of classical methods, particularly their vulnerability to outliers in Phase I data.

Simulation results across a range of conditions contamination levels, correlation structures, and shift magnitudes demonstrate that the robust charts, especially those using S_n Covariance, GK_S_n, and Comedian, consistently outperform classical and RMCD/RMVE-based charts in terms of detection speed (lower ARLs) and reliability. These robust estimators also offer strong computational efficiency, avoiding the heavy processing times associated with RMCD and RMVE. The application to a real-world chemical process dataset confirms the simulation findings, with robust charts detecting shifts earlier and more clearly. Among all the methods studied, S_n Covariance, GK_S_n, and Comedian achieve the best balance between robustness, sensitivity, and efficiency.

In conclusion, the proposed charts provide a robust and practical approach for monitoring bivariate dispersion, particularly in environments where data contamination and correlation are prevalent.

CHAPTER 8

MONITORING PROCESS VARIABILITY WITH MULTIVARIATE ROBUST CONTROL CHARTS

8.1 Introduction

Monitoring multiple quality characteristics simultaneously is a fundamental requirement in modern production and service systems where maintaining process stability and product consistency is critical. Multivariate control charts have become the primary statistical tools for this purpose because they enable the timely detection of shifts in process parameters, whether in the mean, the variability, or both. All processes naturally exhibit variation, and such variability can lead to deviations from desired performance and compromise quality outcomes. Consequently, reducing variability is a central objective in statistical quality control. Although numerous multivariate procedures have been developed to monitor quality characteristics simultaneously, most focus on shifts in the mean vector, while monitoring process variability remains equally essential (Woodall & Montgomery, 1999). Identifying shifts in the covariance structure is crucial because such changes affect both the variability of individual quality characteristics and the correlations among them, potentially leading to hidden instabilities and degraded process performance. Monitoring covariance shifts is challenging because the sample covariance matrix requires large sample sizes for stable estimation, can become unreliable in high-dimensional settings, and may be affected by concurrent mean shifts (Lowry & Montgomery, 1995). These challenges highlight the need for more effective methods to detect covariance shifts in multivariate process monitoring.

A variety of approaches for monitoring process dispersion have been proposed, including CUSUM and EWMA type procedures, likelihood ratio–based methods, and norm-based techniques (Healy, 1987; Tang & Barnett, 1996; Yeh *et al*., 2003, 2004, 2005; Surtihadi *et al*., 2004; Huwang *et al*., 2007; Memar & Niaki, 2009). Building on the bivariate study, this Chapter extends the analysis to multivariate dispersion control, addressing the additional challenges posed by higher-dimensional processes.

8.1.1. MEWMS control charts

Consider the process data vector $\boldsymbol{x} = (x_1, x_2, \dots x_p)'$, consisting of p quality characteristics that follow a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$. These parameters are either known or estimated from Phase I analysis of in-control process conditions, denoted by $\boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{\Sigma}_0$. To facilitate further analysis, the process variable $\boldsymbol{x}$ is transformed into a new random vector $\boldsymbol{y}$, which follows a multivariate normal distribution with the mean $\boldsymbol{\mu}_{\boldsymbol{y}}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ as defined below:

$$\boldsymbol{y} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}_0); \quad \boldsymbol{\mu}_{\boldsymbol{y}} = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\mu}_{\boldsymbol{x}} - \boldsymbol{\mu}_0); \quad \boldsymbol{\Sigma}_{\boldsymbol{y}} = \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\Sigma}_0^{-1/2}.$$

Under in-control conditions, $\boldsymbol{y}$ follows a multivariate normal distribution $N_p(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is a $p \times p$ identity matrix. For individual observations, the sample covariance matrix cannot be computed; however, the outer product $\boldsymbol{y}\boldsymbol{y}'$ serves as an unbiased estimator of $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ provided that the process mean remains stable (i.e. $\boldsymbol{\mu}_{\boldsymbol{y}} = \boldsymbol{0}$). The drawback is that $\boldsymbol{y}\boldsymbol{y}'$ is not guaranteed to be positive definite. To overcome this limitation, Huwang *et al*. (2007) proposed the multivariate exponentially weighted moving average statistic for the t-th individual observation $\boldsymbol{y}_{\boldsymbol{t}} = (y_1, y_2, \dots y_p)'$:

$$\boldsymbol{S}_t = \omega \boldsymbol{y}_{\boldsymbol{t}}\boldsymbol{y}_{\boldsymbol{t}}' + (1 - \omega)\boldsymbol{S}_{t-1}, \quad t = 1,2,3 \dots,$$

where ω is a smoothing constant, $0 < \omega < 1$, and $\boldsymbol{S}_0 = \boldsymbol{y}_{\boldsymbol{t}}\boldsymbol{y}_{\boldsymbol{t}}'$.

Huwang *et al*. (2007) showed that when the process mean is in control, the matrix $\boldsymbol{S}_t$ becomes positive definite with probability 1 for $t \geq p$, and its expectation is $E(\boldsymbol{S}_t) = \boldsymbol{\Sigma}_{\boldsymbol{y}}$. Since the trace of $\boldsymbol{S}_t$ reflects the total variability across all p quality characteristics, it was proposed as the monitoring statistic for detecting covariance shifts. The trace is expressed as

$$tr(\boldsymbol{S}_t) = \sum_{i=1}^{t} c_i \, tr(\boldsymbol{y}_{\boldsymbol{t}}\boldsymbol{y}_{\boldsymbol{t}}') = \sum_{i=1}^{t} c_i \sum_{j=1}^{p}((y_{ij}^2). \tag{8.1}$$

The mean and variance of $tr(\boldsymbol{S}_t)$ are given by p and $2p\sum_{i=1}^{t} c_i^2$ respectively, where $\sum_{i=1}^{t} c_i^2 = \frac{\omega}{(2-\omega)} + \frac{2-2\omega}{2-\omega}(1-\omega)^{2(t-1)}$ which will converge to $\frac{\omega}{(2-\omega)}$ as $t \to \infty$. Based on this, the control limits for MEWMS control chart are defined as

$$E[tr(\boldsymbol{S}_t)] \pm \sqrt{\mathrm{Var}[tr(\boldsymbol{S}_t)]} = p \pm L\sqrt{2p\sum_{i=1}^{t} c_i^2},$$

where the value of L is typically determined by Monte Carlo simulation to achieve a specified in-control average run length (ARL_0).

8.1.2. MEWMV Control Charts

The MEWMS chart was originally developed to monitor changes in the covariance structure under the assumption of a stable process mean. In real applications, however, both the mean and the variability may shift simultaneously during process monitoring. This practical limitation motivated the development of a control chart that is sensitive to changes in variability as well as shifts in the mean.

To address this issue, Huwang *et al*. (2007) introduced the MEWMV chart, which is based on the statistic $\boldsymbol{V}_t$. The statistic is defined in a recursive manner similar to $\boldsymbol{S}_t$, with the following modification:

$$\boldsymbol{V}_t = \omega(\boldsymbol{y}_t - \boldsymbol{z}_t)(\boldsymbol{y}_t - \boldsymbol{z}_t)' + (1-\omega)\boldsymbol{V}_{t-1}, \quad t = 1,2,3 \ldots,$$

where ω is a smoothing constant, $0 < \omega < 1$, and $\boldsymbol{V}_0 = (\boldsymbol{y}_1 - \boldsymbol{z}_1)(\boldsymbol{y}_1 - \boldsymbol{z}_1)'$. Here, a natural estimate $\boldsymbol{z}_t$ for the process mean at time t is the multivariate exponentially moving average of $\boldsymbol{y}_t$ (Lowry *et al*., 1992) and its values are computed recursively as:

$$\boldsymbol{z}_t = \lambda\boldsymbol{y}_t + (1-\lambda)\boldsymbol{z}_{t-1}, \quad t = 1,2,3 \ldots,$$

with $\boldsymbol{z}_0 = \boldsymbol{0}$ and $\lambda \in (0,1)$ being another smoothing constant.

For $t \geq p$, $\boldsymbol{V}_t$ is a positive definite matrix with probability 1 and, its expectation converges as:

$$E(\boldsymbol{V}_t) \to \frac{2(1-\lambda)^2}{(2-\lambda)}\boldsymbol{\Sigma}_{\boldsymbol{y}} \text{ as } t \to \infty.$$

Thus, an unbiased estimator of $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ is obtained as $\frac{(2-\lambda)}{2(1-\lambda)^2}\boldsymbol{V}_t$. The trace of $\boldsymbol{V}_t$ admits the following simplified representation:

$$tr(\boldsymbol{V}_t) = \sum_{i=1}^{t} \sum_{j=1}^{t} q_{ij} \sum_{k=1}^{p} x_{ik}x_{jk}, \tag{8.2}$$

where the weight matrix $\boldsymbol{Q} = (q_{ij})$ is defined as:

$$\boldsymbol{Q} = (\boldsymbol{I}_t - \boldsymbol{M})'\boldsymbol{C}(\boldsymbol{I}_t - \boldsymbol{M}).$$

Here, $\boldsymbol{C} = diag((1-\omega)^{t-1}, \omega(1-\omega)^{t-2}, \ldots, \omega(1-\omega), \omega))$, $\boldsymbol{I}_t$ is the $t \times t$ identity matrix and $\boldsymbol{M}$ is a lower triangular matrix determined by the smoothing parameter λ. The mean and variance of $tr(\boldsymbol{V}_t)$ are given by $p \times tr(\boldsymbol{Q}_t)$ and $2p \times \sum_{i=1}^{t} \sum_{j=1}^{t} q_{ij}^2$ respectively. Hence, the control limits for the MEWMV chart are expressed as

$$E[tr(\boldsymbol{V}_t)] \pm \sqrt{\mathrm{Var}[tr(\boldsymbol{V}_t)]} = p \times tr(\boldsymbol{Q}_t) \pm L\sqrt{2p\ \sum_{i=1}^{t}\ \sum_{j=1}^{t} q_{ij}^2}\ ,$$

where the value of L can be determined through Monte Carlo simulation based on ARL_0.

8.1.3. Control charts based on L_c-norm function

Memar and Niaki (2009) proposed the $MEWMSL_1$ and $MEWMSL_2$ control charts, which utilize the L_1 norm and L_2 norm distances of the diagonal elements of $\boldsymbol{S}_t$. The monitoring statistics are defined as

$$D_1(\boldsymbol{S}_t) = \left\|\left(\boldsymbol{S}_{t(11)}, \boldsymbol{S}_{t(22)}, \ldots, \boldsymbol{S}_{t(pp)}\right)' - \mathbf{1}_p\right\| = \sum_{i=1}^{p}\left|\boldsymbol{S}_{t(ii)} - 1\right|, \quad (8.3)$$

$$D_2(\boldsymbol{S}_t) = \left\|\left(\boldsymbol{S}_{t(11)}, \boldsymbol{S}_{t(22)}, \ldots, \boldsymbol{S}_{t(pp)}\right)' - \mathbf{1}_p\right\|^2 = \sum_{i=1}^{p}(\boldsymbol{S}_{t(ii)} - 1)^2, \quad (8.4)$$

where $\mathbf{1}_p$ is a p dimensional vector of ones, and $\boldsymbol{S}_{t(ii)}$ represents the i^{th} diagonal element of the matrix $\boldsymbol{S}_t$. The control limits for these charts are typically obtained via Monte Carlo simulation for a specified ARL_0.

Building on this idea, Memar and Niaki (2009) further proposed a monitoring scheme based on the statistic $\boldsymbol{V}_t$ Since for large t, $E(\boldsymbol{V}_t) = \frac{2(1-\lambda)^2}{(2-\lambda)}\Sigma$, a transformed statistic,

$$\boldsymbol{V}_t^* = \frac{(2-\lambda)}{2(1-\lambda)^2}\boldsymbol{V}_t,$$

can be employed to provide an unbiased estimator of process variability. Using $\boldsymbol{V}_t^*$, two corresponding monitoring statistics are defined as follows:

$$D_1(\boldsymbol{V}_t^*) = \left\|\left(\boldsymbol{V}_{t(11)}^*, \boldsymbol{V}_{t(22)}^*, \ldots, \boldsymbol{V}_{t(pp)}^*\right)' - \mathbf{1}_p\right\| = \sum_{i=1}^{p}\left|\boldsymbol{V}_{t(ii)}^* - 1\right|, \quad (8.5)$$

$$D_2(\boldsymbol{V}_t^*) = \left\|\left(\boldsymbol{V}_{t(11)}^*, \boldsymbol{V}_{t(22)}^*, \ldots, \boldsymbol{V}_{t(pp)}^*\right)' - \mathbf{1}_p\right\|^2 = \sum_{i=1}^{p}(\boldsymbol{V}_{t(ii)}^* - 1)^2, \quad (8.6)$$

where $\boldsymbol{V}_{t(ii)}^*$ denotes the i^{th} diagonal element of $\boldsymbol{V}_t^*$. These statistics, D_1 and D_2 respectively represent the L_1 and L_2 norm distances of the diagonal elements of $\boldsymbol{V}_t^*$ from their nominal values. The corresponding control limits are determined through Monte Carlo simulation, with values depending p, ω, λ and ARL_0.

Huwang *et al*. (2007) and Memar and Niaki (2009) developed their control charts under the assumption that the in-control process parameters are known. In practice, however, these parameters must be estimated from Phase I data. When the Phase I dataset contains outliers, such estimates can be heavily biased, leading to a significant deterioration in chart performance.

To mitigate this issue, Variyath and Vattathoor (2014) proposed robust control charts for Phase I analysis by incorporating the re-weighted minimum covariance determinant (RMCD) and re-weighted minimum volume ellipsoid (RMVE) estimators originally introduced by Rousseeuw (1985). These robust estimators were embedded within the MEWMS and MEWMV to monitor process variability in multivariate individual observations. Simulation results for bivariate quality characteristics revealed that the robust versions substantially outperform their classical counterparts, particularly in the presence of contamination in Phase I data.

Despite these advantages, both the RMCD and RMVE estimators suffer from notable drawbacks. Their computational complexity is considerable, especially when applied to large-scale or high-dimensional datasets. Moreover, they require the specification of a subset size that implicitly assumes a known proportion of uncontaminated observations. Since the actual contamination level is typically unknown in practice, incorrect specification of this parameter can reduce both the robustness and efficiency of the resulting estimators.

In this study, we propose a multivariate dispersion control chart constructed from the MEWMS and MEWMV charts with L_c-norm formulations, employing the high-breakdown robust covariance estimator, specifically the ROC and SGK estimator.

8.2 Multivariate Robust Dispersion Control Chart

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ denote the n Phase I observations from a p variate quality characteristic. A robust multivariate dispersion control chart can be constructed through the following steps:

1. *Estimation of robust location and dispersion parameters*: Using the Phase I data, obtain robust estimates of the process location and dispersion, denoted by $\widehat{\boldsymbol{\mu}}_R$ and $\widehat{\boldsymbol{\Sigma}}_R$, respectively.

2. *Transformation of Phase II observations*: For each new Phase II observation $\boldsymbol{y}_t$, compute the standardized vector

$$\boldsymbol{y}_t^* = \widehat{\boldsymbol{\Sigma}}_R^{-\frac{1}{2}}(\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}}_R).$$

3. *Calculation of the monitoring statistic*: Select an appropriate monitoring statistic from Equations (8.1)–(8.6) and evaluate it using the transformed observation $\boldsymbol{y}_t^*$.
4. *Charting and interpretation*: Plot the computed statistics against empirically determined control limits corresponding to a specified significance level α. Any point falling outside the limits signals a potential shift in process dispersion, prompting further investigation or corrective action.

The monitoring of process variability is carried out using multivariate quality characteristics, with the dimension parameter considered at p = 2, 5 and 10. As the monitoring statistics defined above are always positive, only upper control limits are required. To establish these limits, 10,000 Monte Carlo simulations were conducted for various combinations of the smoothing parameters ω and λ and for confidence levels corresponding to $\boldsymbol{\alpha}$ =0.05, 0.01, and 0.0027

In order to estimate the UCL for a given α, we have generated n individual observations from a standard multivariate normal distribution $N_p(\mathbf{0}, \boldsymbol{I})$ as Phase I sample and computed the robust estimators of mean vector and covariance matrix $\widehat{\boldsymbol{\mu}}_R$ and $\widehat{\boldsymbol{\Sigma}}_R$. In addition, for each data set, we randomly generate a new observation from $N_p(\mathbf{0}, \boldsymbol{I})$ (treated as a Phase II observation) and calculate the corresponding statistic in Equation (8.1). This procedure is repeated for 10,000 times and then the UCL corresponding to α is the $100(1 - \alpha)^{th}$ quantile of above calculated statistic. The same procedure is followed to obtain UCL for the other statistic represented in Equation (8.2), (8.3), (8.4), (8.5) and (8.6). UCLs for n = 50 and 100 were generated using ROC, Classical estimator, RMCD and RMVE for α = 0.05, 0.01 and 0.0027. The UCLs of MEWMSL$_1$, MEWMSL$_2$, MEWMS charts with ω = 0.4, and those of MEWMVL$_1$, MEWMVL$_2$ and MEWMV charts with ω = 0.4 and λ = 0.2, corresponding to α = 0.0027 for n = 50 and 100, are presented in Table 8.1 and Table 8.2.

Table 8.1: Simulated UCLs for MEWMSL$_1$, MEWMSL$_2$, and MEWMS

Method	n	p	MEWMSL$_1$			MEWMSL$_2$			MEWMS		
			95%	99%	99.73	95%	99%	99.73	95%	99%	99.73
SGK	50	2	5.16	8.36	11.40	19.06	51.63	98.08	6.98	10.22	13.27

		5	8.89	12.88	16.67	31.35	70.62	123.36	12.95	17.02	20.80
		10	16.79	23.20	29.10	63.80	130.21	219.01	24.80	31.23	37.10
	100	2	4.53	7.02	9.24	14.62	36.25	64.30	6.34	8.88	11.09
		5	7.20	9.99	12.45	20.39	42.20	69.18	11.14	14.06	16.50
		10	12.14	15.83	19.07	33.07	61.03	94.44	19.60	23.42	26.53
ROC	50	2	3.65	5.99	8.20	9.55	26.84	51.51	5.34	7.73	9.95
		5	6.93	9.99	12.94	18.93	42.49	75.49	10.65	13.83	16.72
		10	13.67	17.91	21.91	46.50	91.36	148.34	20.91	25.07	28.66
	100	2	3.12	4.89	6.46	6.89	17.64	31.90	4.79	6.61	8.21
		5	5.77	7.96	9.87	12.75	26.64	43.16	9.33	11.67	13.59
		10	13.27	17.10	20.51	43.77	83.36	131.97	20.61	24.54	27.78
RMCD	50	2	4.59	8.04	11.76	15.38	48.67	103.86	6.30	9.82	13.59
		5	12.69	22.20	33.11	66.28	213.08	485.45	16.85	26.56	37.44
		10	47.58	87.97	134.13	541.08	1956.10	4697.95	56.43	97.12	143.41
	100	2	3.26	5.27	7.13	7.60	20.69	39.23	4.92	7.00	8.88
		5	6.71	9.62	12.33	17.82	39.98	68.85	10.45	13.51	16.32
		10	13.00	18.00	23.07	38.70	81.23	140.26	20.44	25.71	30.79
RMVE	50	2	3.32	5.58	7.89	7.86	23.75	48.80	4.87	7.17	9.52
		5	9.20	15.27	22.09	35.52	104.32	220.95	12.94	19.18	26.21
		10	34.11	59.54	88.05	281.33	884.98	2001.15	42.63	68.33	97.08
	100	2	2.71	4.32	5.75	5.10	13.89	25.60	4.26	5.93	7.41
		5	5.77	8.19	10.45	12.88	28.78	50.07	9.20	11.77	14.09
		10	11.43	15.46	19.34	29.69	60.30	99.78	18.40	22.66	26.53
Classical	50	2	2.89	4.63	6.27	5.83	15.99	29.97	4.48	6.28	7.93
		5	6.21	8.80	11.30	15.00	33.20	57.19	9.75	12.48	14.88
		10	12.59	16.87	20.84	35.99	70.92	115.68	19.80	24.17	27.84
	100	2	2.60	4.06	5.33	4.65	12.11	21.74	4.17	5.71	7.01
		5	5.33	7.30	9.05	10.72	22.46	36.68	8.73	10.87	12.61
		10	9.83	12.63	15.18	20.98	38.69	59.80	16.43	19.38	21.76

Table 8.2: Simulated UCLs $MEWMVL_1$, $MEWMVL_2$, and MEWMV

Method	n	p	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
			95 %	99 %	99.73 %	95 %	99 %	99.73 %	95 %	99%	99.73 %
SGK	50	2	8.64	11.95	14.81	45.52	89.31	141.88	3.78	5.43	6.97
		5	17.77	22.28	25.82	88.73	145.23	203.18	7.70	9.95	11.87
		10	36.69	44.03	49.76	195.42	292.75	389.01	15.25	18.80	21.85
	100	2	7.94	10.43	12.39	37.55	66.46	95.60	3.51	4.86	5.99
		5	15.48	18.75	21.19	65.18	98.44	129.60	6.83	8.56	9.98
		10	29.04	33.54	36.88	119.32	164.08	204.03	12.52	14.86	16.69
ROC	50	2	8.46	11.81	14.72	44.10	88.04	143.42	3.71	5.39	6.91
		5	16.84	21.18	24.73	80.92	134.67	190.91	7.39	9.58	11.53
		10	34.42	41.15	46.28	173.70	258.70	343.68	14.54	17.91	20.77
	100	2	7.48	9.83	11.66	33.68	59.42	85.86	3.37	4.65	5.73
		5	14.71	17.83	20.24	59.50	90.30	119.79	6.59	8.25	9.58
		10	27.54	31.82	35.05	108.31	149.20	185.45	12.07	14.30	16.09
RMCD	50	2	10.17	15.37	21.03	66.81	161.81	311.56	4.25	6.56	9.13
		5	29.68	47.21	66.37	293.24	841.63	1777.61	11.47	18.26	26.29
		10	106.56	170.90	243.05	2277.23	6489.72	14244.31	37.81	63.65	93.12
	100	2	7.82	10.61	12.97	37.84	72.22	112.89	3.45	4.90	6.19
		5	17.01	21.33	25.11	82.28	138.83	204.44	7.31	9.46	11.40
		10	34.54	42.59	49.71	178.13	289.87	424.55	14.24	17.87	21.29
RMVE	50	2	7.51	10.68	13.40	37.41	78.68	129.44	3.41	5.00	6.55
		5	21.95	32.08	42.83	160.86	373.18	728.38	9.08	13.42	18.15
		10	80.71	124.09	172.80	1264.81	3328.67	6979.87	29.84	47.52	67.73
	100	2	6.47	8.59	10.32	26.28	48.09	71.96	3.02	4.20	5.22
		5	14.37	17.80	20.69	60.43	99.11	142.53	6.49	8.31	9.89
		10	30.07	35.92	40.72	137.05	208.97	286.88	12.95	15.87	18.49
Classical	50	2	6.64	8.88	10.69	26.99	49.87	74.64	3.11	4.37	5.45
		5	14.89	18.36	21.06	63.02	99.75	136.19	6.77	8.63	10.23
		10	31.94	37.69	42.24	150.08	218.52	285.06	13.77	16.72	19.23
	100	2	6.26	8.17	9.63	23.72	41.35	58.48	2.94	4.03	4.95
		5	13.36	16.07	18.10	49.34	73.84	96.70	6.15	7.64	8.81
		10	25.93	29.87	32.74	96.32	131.66	162.62	11.55	13.66	15.29

8.3 Simulation Study

This section presents a simulation study to evaluate and compare the performance of the control charts in the presence of outliers. The data are generated from a contaminated mixture of multivariate normal distributions, defined as

$$(1-\varepsilon)\, \mathrm{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \varepsilon\, \mathrm{N}_p(\boldsymbol{\mu}_{\mathrm{I}}, \boldsymbol{\Sigma}_{\mathrm{I}}),$$

where ε denotes the proportion of outliers. Here, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ represent the in-control mean vector and covariance matrix, while $\boldsymbol{\mu}_{\mathrm{I}}$ and $\boldsymbol{\Sigma}_{\mathrm{I}}$ correspond to the Phase I out-of-control parameters for location and dispersion, respectively. Specifically, $\boldsymbol{\mu}_{\mathrm{I}} = (\mu_{\mathrm{I}1}, \mu_{\mathrm{I}2}, \ldots, \mu_{\mathrm{I}p})^T$ and $\boldsymbol{\Sigma}_{\mathrm{I}} = \boldsymbol{\delta}\boldsymbol{\Sigma}_0\boldsymbol{\delta}^{\boldsymbol{T}}$, where $\boldsymbol{\delta} = diag(\delta\ \delta \ldots \delta)$ represents a diagonal matrix with the scale shift parameter δ, taking the values 2 and 3.

The simulation study considered two contamination scenarios, with $\boldsymbol{\mu}_{\mathrm{I}} = (0, 0,\ldots,0)^t$ representing scale shift only, and $\boldsymbol{\mu}_{\mathrm{I}} = (2, 2,\ldots,2)^t$ corresponding to simultaneous shifts in both location and scale. The simulation further investigates the effects of key factors, including the number of observations (n), the proportion of outliers, and the method used to generate outliers through different values of $\boldsymbol{\mu}_{\mathrm{I}}$ and $\boldsymbol{\Sigma}_{\mathrm{I}}$. Without loss of generality, $\boldsymbol{\mu}_0$ is assumed to be a zero vector and ε takes the values of 0.1 and 0.2. The following cases are analyzed to examine how these factors influence the performance of the control charts.

Case A: **Independent variables**

In this case, the quality characteristics are assumed to be independent. The contaminated normal model is expressed as

$$(1-\varepsilon)\, \mathrm{N}_p\, (\mathbf{0}, \boldsymbol{I}) + \varepsilon\, \mathrm{N}_p(\boldsymbol{\mu}_{\mathrm{I}}, \delta\boldsymbol{I}\,).$$

Case B: **Correlated variables**

In this case, the quality characteristics are assumed to be correlated. The contaminated normal model considered is as follows:

$$(1-\varepsilon)\, \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_0) + \varepsilon\, \mathrm{N}_p(\boldsymbol{\mu}_{\mathrm{I}}, \delta\boldsymbol{\Sigma}_0)$$

where $\boldsymbol{\Sigma}_0$ is randomly generated by using R software. We have used this value of $\boldsymbol{\Sigma}_0$ to analyse whether the correlation level affects the detection probability of each alternative.

8.3.1 ARL Plots

ARL measures the expected number of samples until a control chart signals a process shift. An effective control chart achieves a high in-control ARL to reduce false alarms, while maintaining a low out-of-control ARL for rapid detection of process shifts.

In Phase II, simulated samples were generated with the mean vector $\boldsymbol{\mu}_{\text{II}}=(0,0,\ldots,0)^t$, , and $\boldsymbol{\Sigma}_{\text{II}} = \boldsymbol{\delta}_j \boldsymbol{\Sigma} \boldsymbol{\delta}_j^T{}_0$ for some $j = 1,2,..p$, where $\boldsymbol{\delta}_j$ is the transformation matrix corresponding to a scale change in the j^{th} component of the vector with the scale shift parameter δ varied from 1.0 to 3.0 (i.e., δ = 1.0, 1.2, 1.4, …, 3.0) and the corresponding ARL values were computed using Equation (8.1), (8.2) and (8.3). Similarly, for $\boldsymbol{\mu}_{\text{II}} = (1,1,\ldots,1)^t$, the parameter δ varied from 1.0 to 4.0 and the ARL values were obtained from Equation (8.4), (8.5) and (8.6) thereby enabling a systematic evaluation of the sensitivity of the proposed charts to different levels of dispersion change. The scale shift was introduced through the covariance matrix Σ_{II}, where δ was applied to one variable when p = 2, and to two variables when p=5 and p=10.

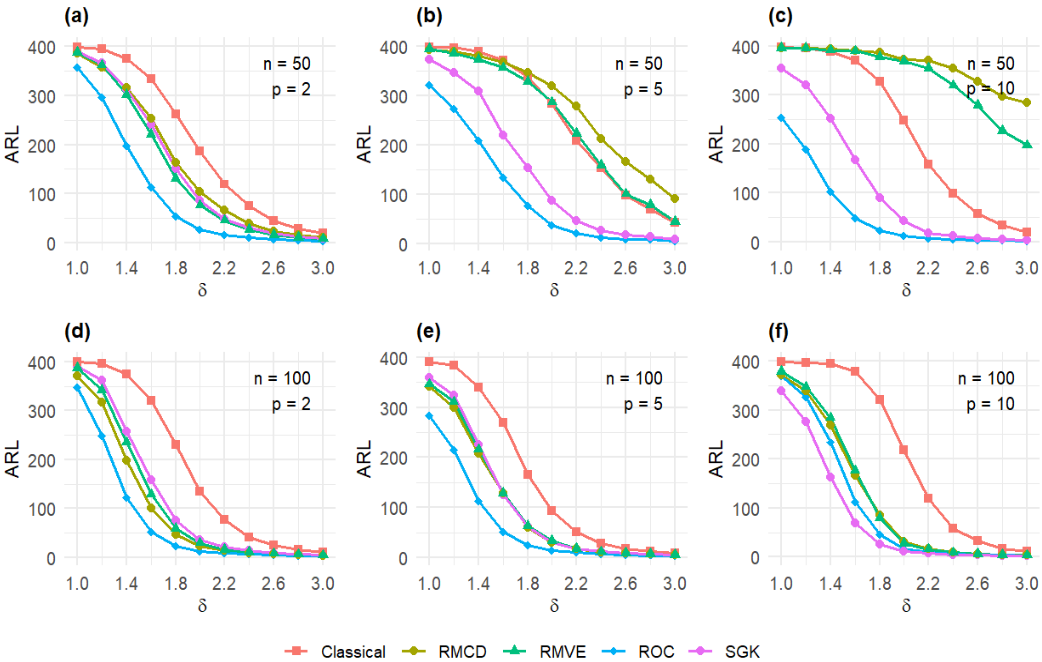


Figure 8.1: ARL plots of the MEWMSL$_1$ statistic with ε = 0.2 and ω = 0.4 under Case A.

The significance level is set at α = 0.0027, corresponding to an ARL$_0$ of 370. For simulations where no signal occurs within 400 observations, the run length RL_i is truncated and recorded as 400. The ARL is then computed across r =1000 iteration using the formula:

$$\text{ARL} = \frac{1}{r}\sum_{i=1}^{r} RL_i.$$

Figure 8.1 presents the ARL plots of the MEWMSL$_1$ statistic under Case A. Clear differences in performance are visible among the five methods. For $p = 2$, ARLs decrease as the shift size increases, with ROC and SGK detecting shifts more quickly than Classical, RMCD, and RMVE. At larger n (100), the gap becomes wider, with Classical showing the slowest detection. As p increases to 5 and 10, ROC and SGK consistently dominate, RMCD and RMVE perform moderately, and Classical becomes ineffective.

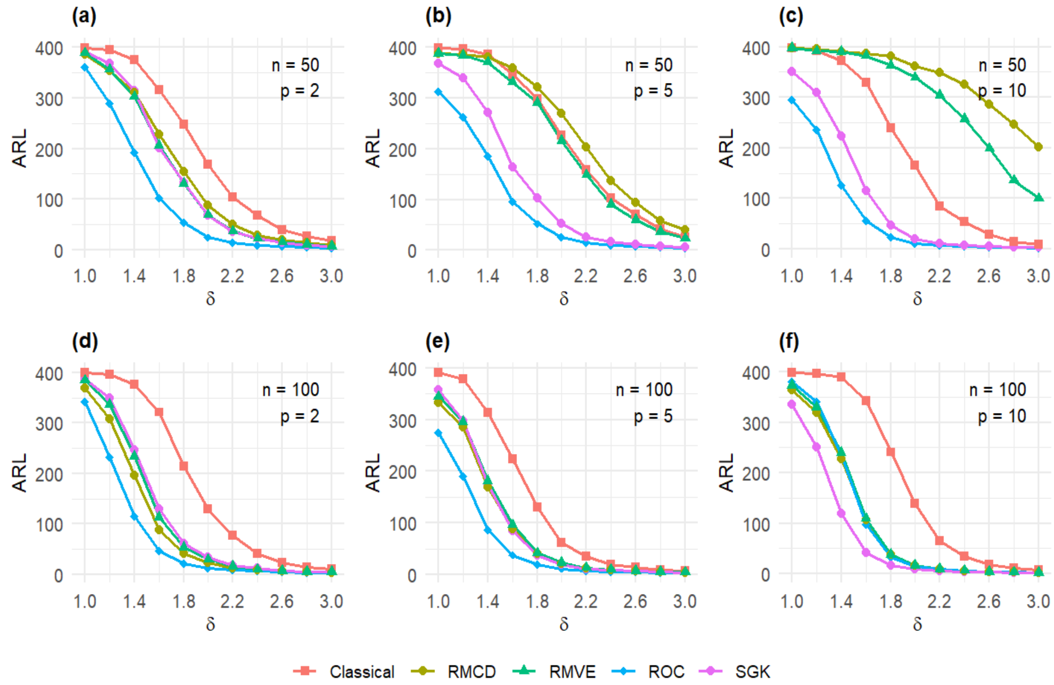


Figure 8.2: ARL plots of the MEWMSL$_2$ statistic with $\varepsilon = 0.2$ and $\omega = 0.4$ under Case A.

ARL plots of the MEWMSL$_2$ statistic under Case A are shown in Figure 8.2. The general trends mirror those in Figure 8.1, where ARL decreases as δ increases and larger n improves detection. Again, ROC and SGK perform best, RMCD and RMVE are moderate, and Classical lags behind, with the performance gap widening in higher dimensions.

In Figure 8.3, the ARL plots of the MEWMSL$_1$ statistic under Case B are displayed. The patterns remain consistent with Case A, where ARL values decline as δ increases, larger n accelerates detection, and higher dimensions amplify the differences among methods. ROC and SGK dominate, RMCD and RMVE follow, and Classical shows the poorest detection.

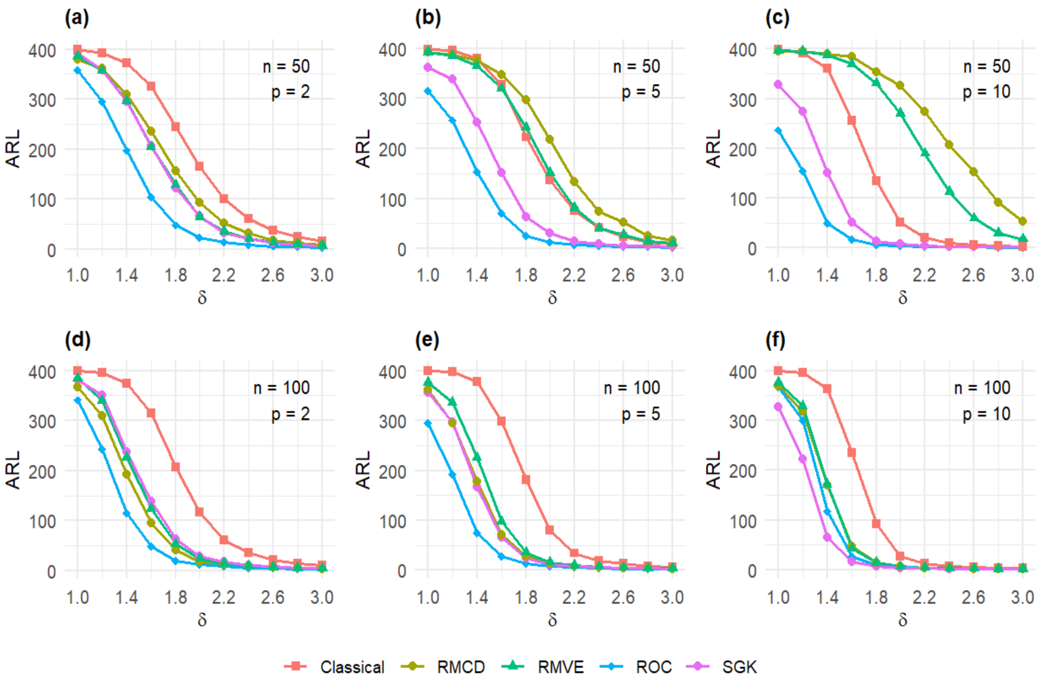


Figure 8.3: ARL plots of the $MEWMSL_1$ statistic with $\varepsilon = 0.2$ and $\omega = 0.4$ under Case B.

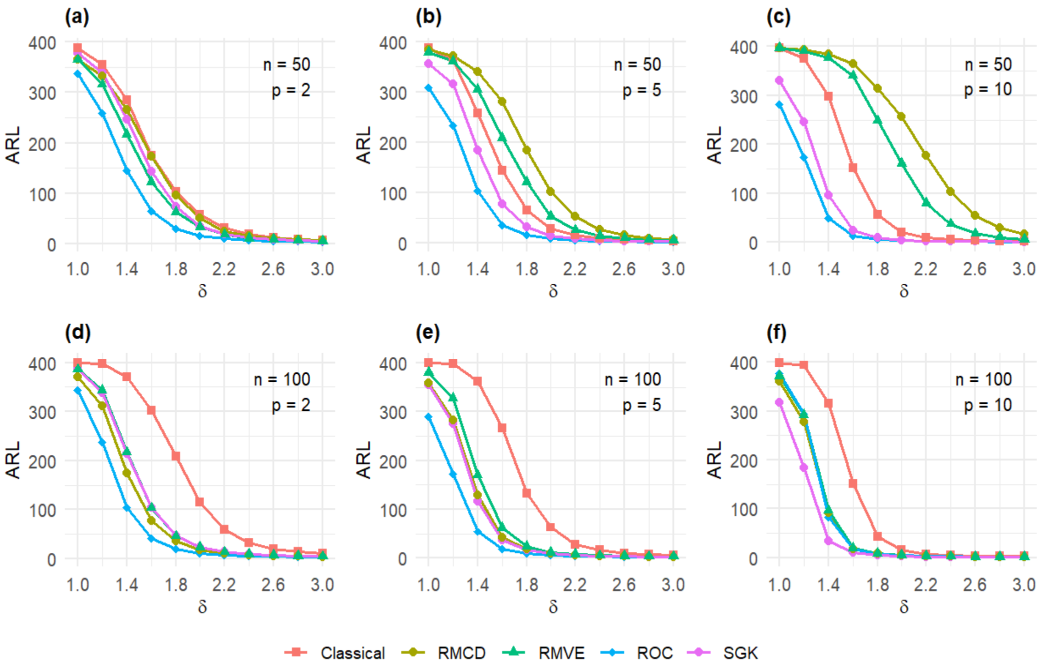


Figure 8.4: ARL plots of the $MEWMSL_2$ statistic with $\varepsilon = 0.2$ and $\omega = 0.4$ under Case B.

Figure 8.4 provides the ARL plots of the MEWMSL2 statistic under Case B. The results are similar to those in Figure 8.3, with ROC and SGK maintaining clear superiority, RMCD and RMVE offering balanced performance, and Classical trailing far behind, particularly in higher-dimensional settings.

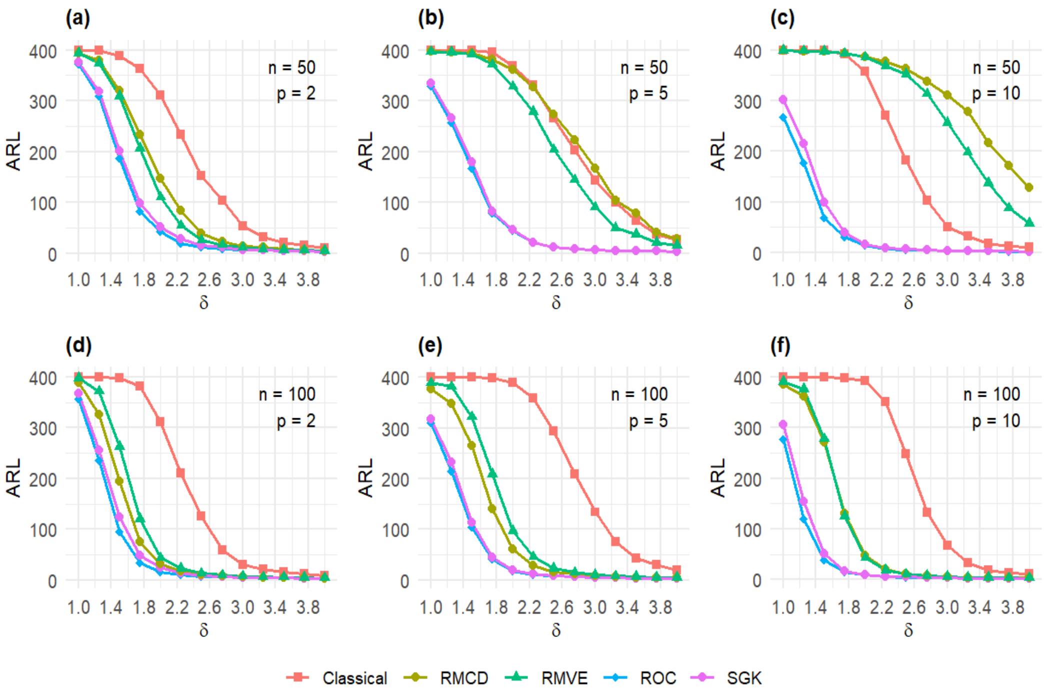


Figure 8.5: ARL plots of the $MEWMVL_1$ statistic with $\varepsilon = 0.2$, $\omega = 0.4$ and $\lambda = 0.2$ under Case A.

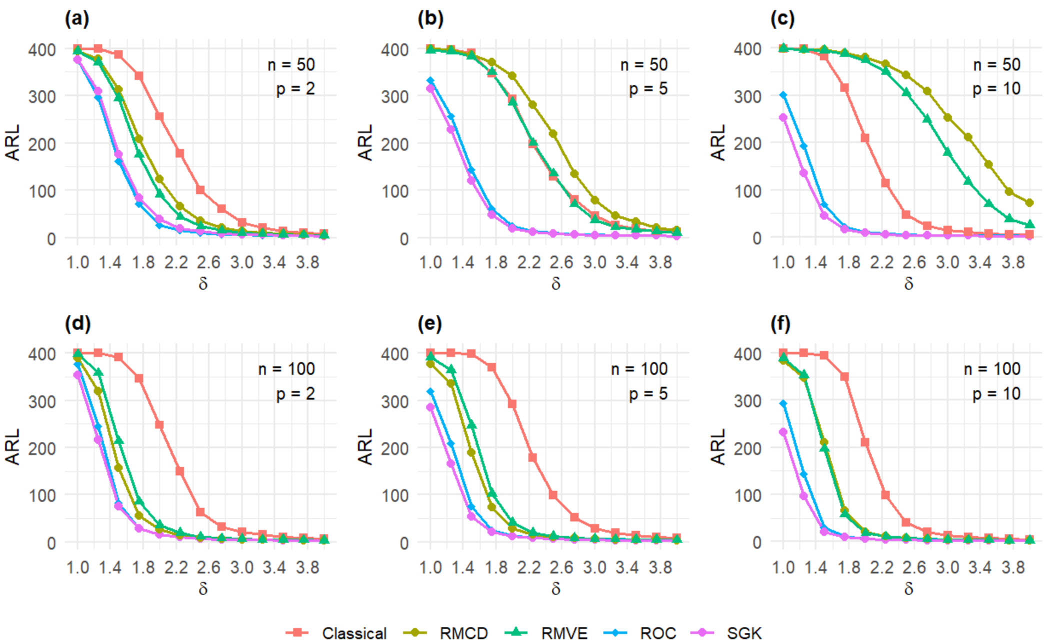


Figure 8.6: ARL plots of the $MEWMVL_2$ statistic with $\varepsilon = 0.2$, $\omega = 0.4$ and $\lambda = 0.2$ under Case A.

ARL plots of the MEWMVL1 statistic under Case A are presented in Figure 8.5. The results confirm that ARL decreases with δ and improves when the sample size increases.

ROC and SGK consistently achieve the steepest declines, RMCD and RMVE provide moderate detection, and Classical shows weak performance, especially when $p = 10$.

The ARL plots of the MEWMVL$_2$ statistic under Case A are shown in Figure 8.6. The findings are broadly consistent with those in Figure 8.5. Larger sample sizes enhance detection efficiency, especially for moderate shifts, while ROC and SGK remain superior, RMCD and RMVE achieve intermediate performance, and Classical is the least effective.

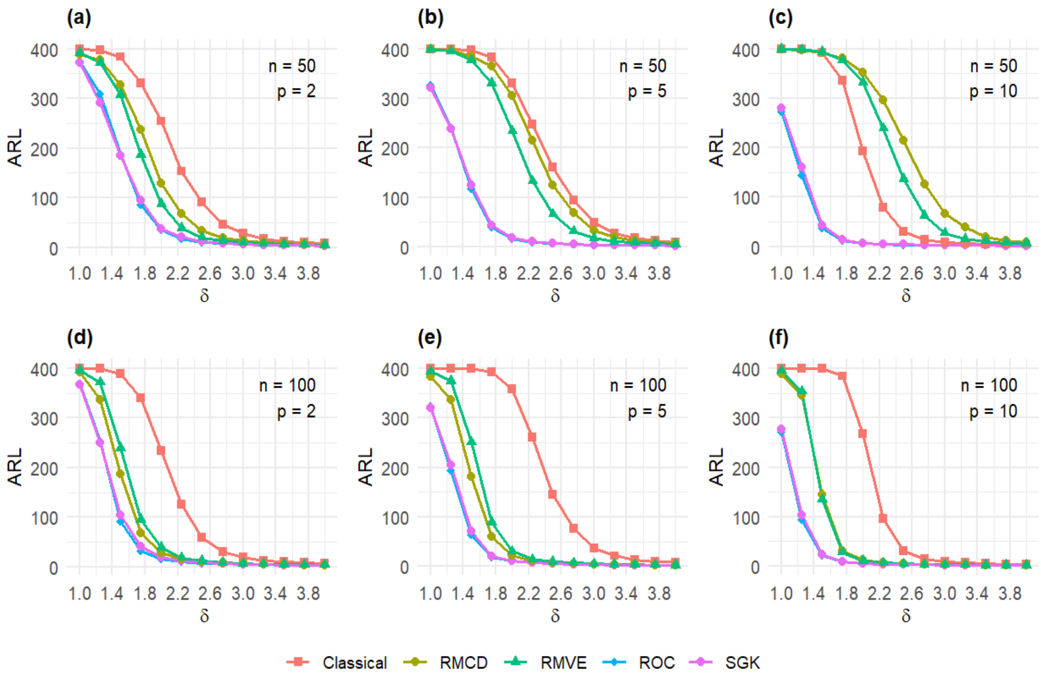


Figure 8.7: ARL plots of the MEWMVL$_1$ statistic with $\varepsilon = 0.2$, $\omega = 0.4$ and $\lambda = 0.2$ under Case B.

In Figure 8.7, the ARL plots of the MEWMVL$_1$ statistic under Case B are illustrated. The results echo earlier figures, with ROC and SGK consistently outperforming the other methods. RMCD and RMVE show moderate efficiency, while Classical remains the weakest, particularly as the dimension increases.

Finally, Figure 8.8 presents the ARL plots of the MEWMVL$_2$ statistic under Case B. The trends are consistent with Figure 8.7, where ARL decreases steadily as δ increases and larger n improves sensitivity. ROC and SGK dominate across all conditions, RMCD and RMVE remain moderate, and the Classical method performs worst, particularly in higher dimensions. Overall, across all figures, ROC and SGK emerge as the most effective methods, combining robustness with fast detection. RMCD and RMVE are conservative but acceptable alternatives. The Classical chart performs consistently poorly, especially in higher dimensions.

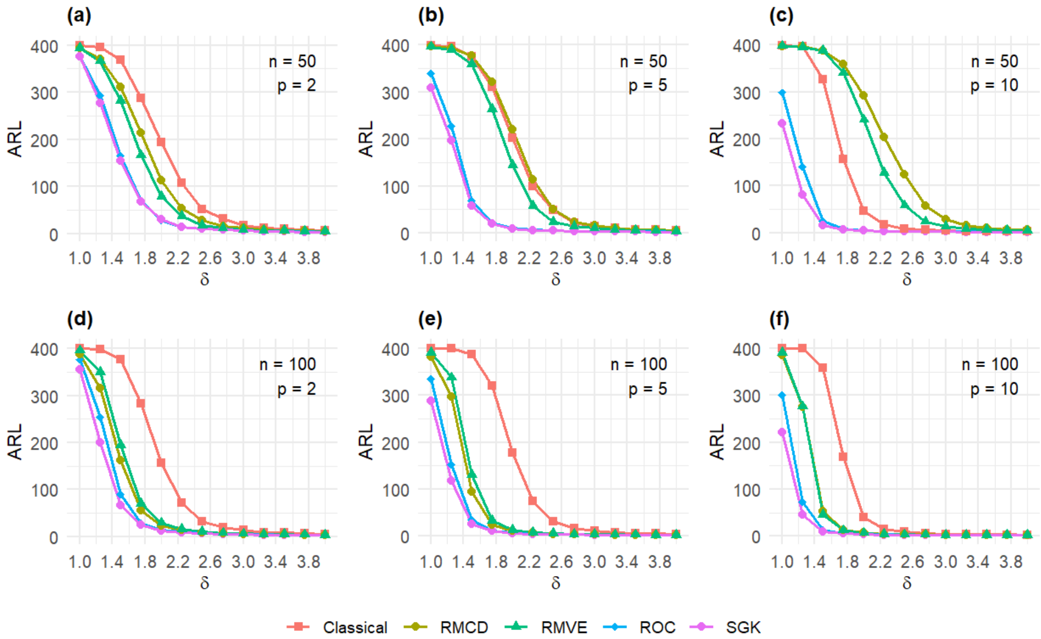


Figure 8.8: ARL plots of the MEWMVL$_2$ statistic with $\varepsilon = 0.2$, $\omega = 0.4$ and $\lambda = 0.2$ under Case B.

8.4 Time Complexity

A simulation study was carried out to evaluate the computation times of both classical and robust control charts under varying sample sizes. The results, averaged over 1000 trials, are presented in Table 8.3.

Table 8.3: Computation time (in seconds)

n	p	ROC	SGK	Classical	RMCD	RMVE
50	2	0.0036	0.0052	0.0016	0.0080	0.0053
	5	0.0043	0.0069	0.0023	0.0104	0.0070
	10	0.0095	0.0158	0.0020	0.0196	0.0224
100	2	0.0041	0.0071	0.0027	0.0116	0.0075
	5	0.0057	0.0119	0.0031	0.0144	0.0099
	10	0.0098	0.0025	0.0034	0.0312	0.0202

The table reports running times of five control chart methods for $n = 50$, 100 and $p = 2$, 5, 10. While computation time increases with both factors, the rise is most pronounced for the robust methods RMCD and RMVE, which are considerably slower than the others because they rely on intensive procedures for robust covariance estimation. In contrast, the Classical chart is always the fastest due to its simpler calculations, and ROC and SGK provide intermediate performance.

8.5 Application to Real Life Data

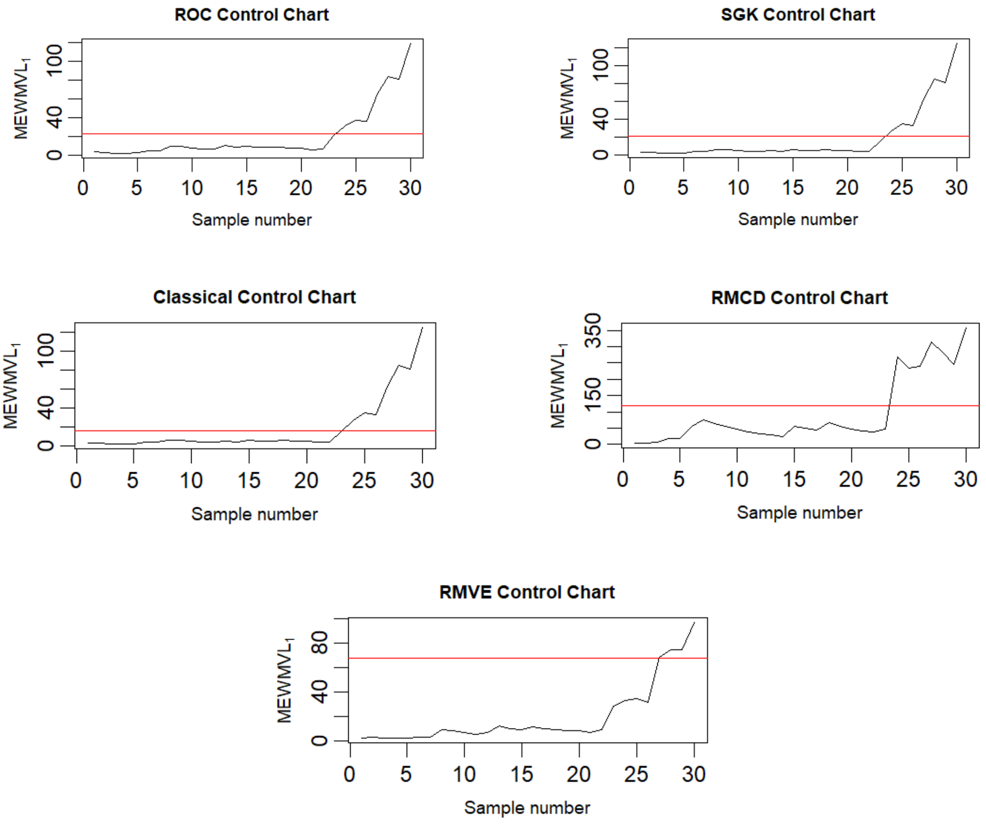


Figure 8.9: Control charts for Chemical Process data using $MEWMVL_1$ statistic

The performance of the robust control charts under contaminated scenarios was evaluated and compared through simulation. To assess their behavior in an uncontaminated setting, the Chemical Process data from Montgomery (2020) was employed. This dataset consists of 20 Phase I (in-control) observations and 10 Phase II (monitoring) observations. Prior analyses indicate that the Phase I data are in statistical control, whereas a shift is observed around observation 24 or 25 in the Phase II sample. Both robust and classical control charts were constructed using the $MEWMVL_1$ statistic, with control limits simulated based on the parameters $n = 20$, $\omega = 0.4$, and $\lambda = 0.2$, as previously described.

Figure 8.9 shows that all control charts remained stable during the Phase I (in-control) region and displayed a clear upward shift around observations 24–25, confirming earlier analyses of this dataset. As expected, the Classical control chart performed effectively, with tight in-control behaviour and prompt detection of the shift. Importantly, the robust charts exhibited performance comparable to the Classical chart in this uncontaminated setting: ROC and SGK most closely mirrored the Classical chart, offering fast and stable detection; RMCD also detected the shift but showed higher variability in Phase I, making it less stable; while RMVE lagged behind, signalling the shift more slowly than the others. These findings highlight that robust procedures, though designed for contaminated data, retain their efficiency in clean settings, with ROC and SGK emerging as the most reliable alternatives to the Classical chart.

8.6 Summary

This chapter proposed robust multivariate dispersion control charts based on the MEWMS and MEWMV frameworks with L_c norm formulations, incorporating the high breakdown ROC and SGK covariance estimators. Simulation studies conducted under diverse shift sizes, dimensions, and sample sizes showed that ROC and SGK based charts consistently surpassed the Classical, RMCD, and RMVE methods. They provided faster detection of dispersion shifts, greater robustness against contamination, and superior stability in higher dimensional contexts. In contrast, RMCD and RMVE offered only moderate improvements, while the Classical chart proved highly sensitive to outliers and largely ineffective under contamination.

A real data application using chemical process observations reinforced these findings. In uncontaminated conditions, ROC and SGK produced results comparable to the Classical chart, but their robustness ensured reliable performance when atypical data were present. Computational analysis further confirmed that these estimators achieve a favorable balance between accuracy and efficiency, being only marginally slower than the Classical chart while remaining substantially faster than RMCD and RMVE. Overall, the proposed ROC and SGK based charts offer a practical, robust, and computationally viable solution for multivariate dispersion monitoring, making them particularly relevant for industrial processes characterized by high dimensionality and potential data contamination.

CHAPTER 9

INDUSTRIAL APPLICATIONS OF PROPOSED CONTROL CHARTS

9.1 Introduction

In modern industrial practice, maintaining consistent product quality is essential for both customer satisfaction and sustainable competitiveness. Statistical process control (SPC) provides a scientific basis for monitoring and improving process stability, enabling early detection of variations that may affect product quality. However, conventional control charts often exhibit sensitivity to outliers or deviations from normality—conditions frequently encountered in real manufacturing settings. To address these limitations, this chapter demonstrates the application of the proposed robust bivariate and multivariate control charts for monitoring both mean and dispersion, using real-world industrial data.

9.2 Industry Contexts and Data Sources

To validate the practical relevance and performance of the proposed charts, data were collected from two industrial settings representing distinct production domains:

1. **Tholur Foods Pvt Ltd**, a leading food-processing company located in Kerala, and
2. **Star Pipes & Fittings**, a reputed manufacturer of PVC, UPVC, and CPVC piping system.

9.2.1. Case Study I – Tholur Foods Pvt Ltd (Bivariate Case)

Tholur Foods Pvt Ltd is a well-established brand known for producing traditional and authentic spice-based food products, ensuring consistency through rigorous quality control practices. As stated on their website, *"We are on a constant search to produce the best quality food products by using indigenous spices available all over India"*.

Figure 9.1: Tholur Tender Mango Pickle (250g)

For this study, data were obtained from the company's "*Tender Mango Pickle*" production process, focusing on two key quality characteristics: *Salinity*, and *pH*. These variables jointly influence the sensory and preservative quality of the product. Hence, the bivariate control charts proposed in this thesis were applied to simultaneously monitor these correlated attributes. In Phase I, 25 observations were considered, and in Phase II, 5 observations were used.

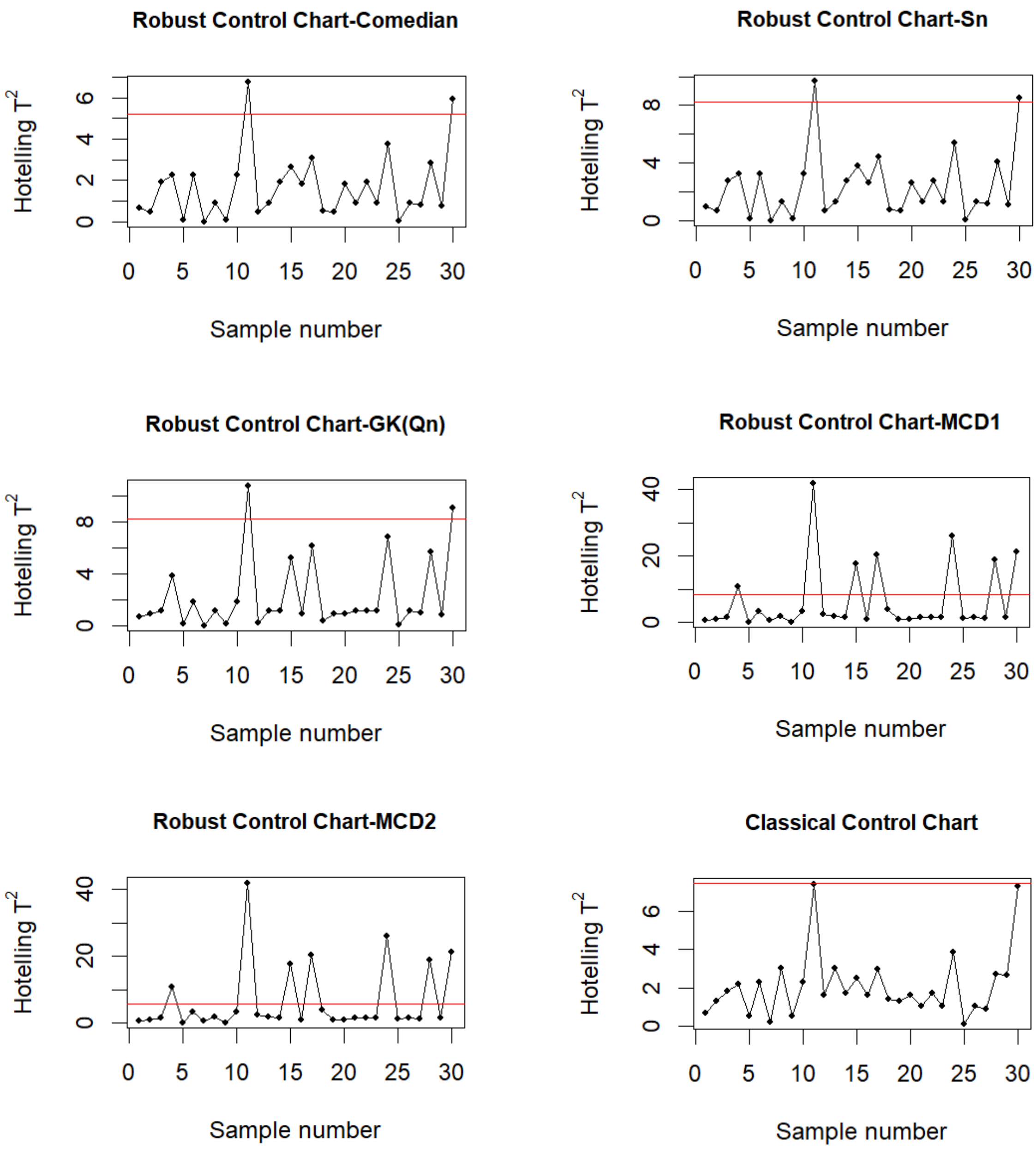


Figure 9.2: Bivariate Hotelling's T^2 Control charts for location for Tholur data

Since Tholur Foods maintains a robust quality assurance system, the data collected were found to be free from outliers. To examine the performance of the proposed control charts under data contamination, one deliberate outlier was introduced each in Phase I (observation 11) and Phase II (observation 30) data. The dataset for this case comprises 25 observations in Phase I and 5 observations in Phase II.

Figure 9.2 presents the Bivariate Hotelling's T^2 control charts for the Tholur data, constructed using both classical and robust methods. The charts include observations from both Phase I and Phase II, allowing a comprehensive assessment of process stability across stages. The plot clearly shows that the Comedian, Sn, and GK-based robust control charts successfully detect the out-of-control observations (11 and 30) in both Phase I and Phase II, whereas the classical chart fails to identify these shifts. Although the MCD-based chart is able to detect the contaminated observations in both phases, it also produces several false alarms, indicating reduced stability under clean conditions.

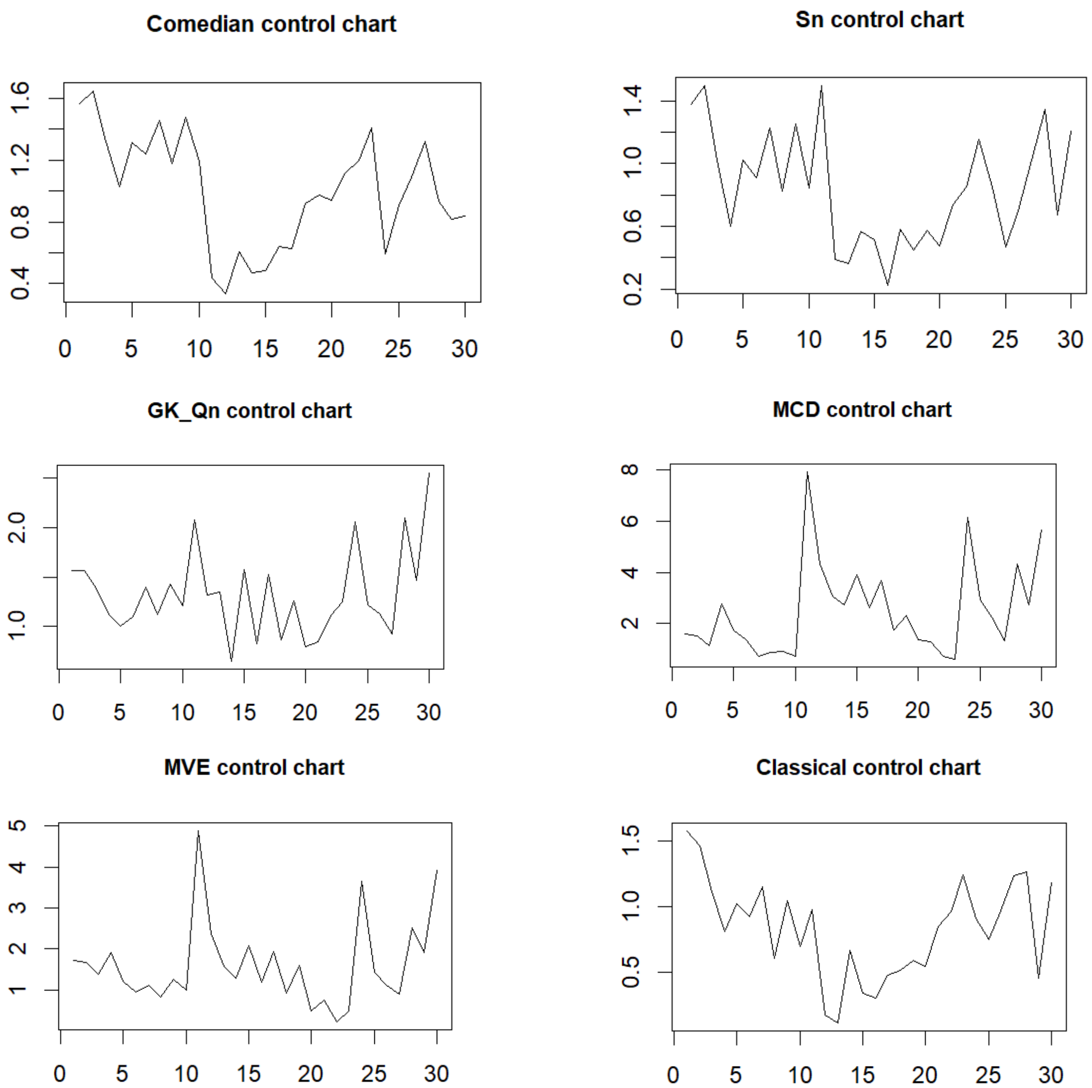


Figure 9.3: Bivariate control charts for dispersion for Tholur data using MEWMVL$_1$ statistic with $\omega = 0.4$, and $\lambda = 0.2$.

Figure 9.3 presents the Bivariate control charts for dispersion for the Tender Mango Pickle data constructed using the $MEWMVL_1$ statistic with $\omega = 0.4$ and $\lambda = 0.2$. The figure indicates that all observations remain within the control limits, signifying that the process variability is stable. This outcome is expected because the deliberately introduced mean shift does not alter the covariance structure of the data. Under such uncontaminated conditions, the classical control chart performs optimally. Consistent with this, both of the proposed robust control charts exhibit patterns similar to those of the classical chart, confirming that they retain efficiency when the data are free from contamination.

9.2.2 Case Study II – Star Pipes & Fittings (Multivariate Case)

Star Pipes & Fittings is a reputed Indian manufacturer of plumbing and piping systems, known for high-precision engineering and stringent quality standards. The company describes its products as *"extremely durable with a high strength-to-weight ratio, manufactured under state-of-the-art facilities."*

Figure 9.4: PY 63 (Single wye 45° without inspection door, solvent cement type 63mm)

For this study, the data correspond to the PVC fitting product **PY 63**, with five essential quality characteristics measured at different points in the production process:

1. Wall thickness at plain end
2. Wall thickness at socket
3. Mean inside diameter of socket at midpoint
4. Mean outside diameter of spigot portion
5. Socket depth

The proposed multivariate control charts for mean and dispersion were employed to monitor these interrelated dimensions simultaneously, providing a comprehensive picture of the manufacturing variability. In Phase I, 20 observations were considered, and in Phase II, 15 observations were used.

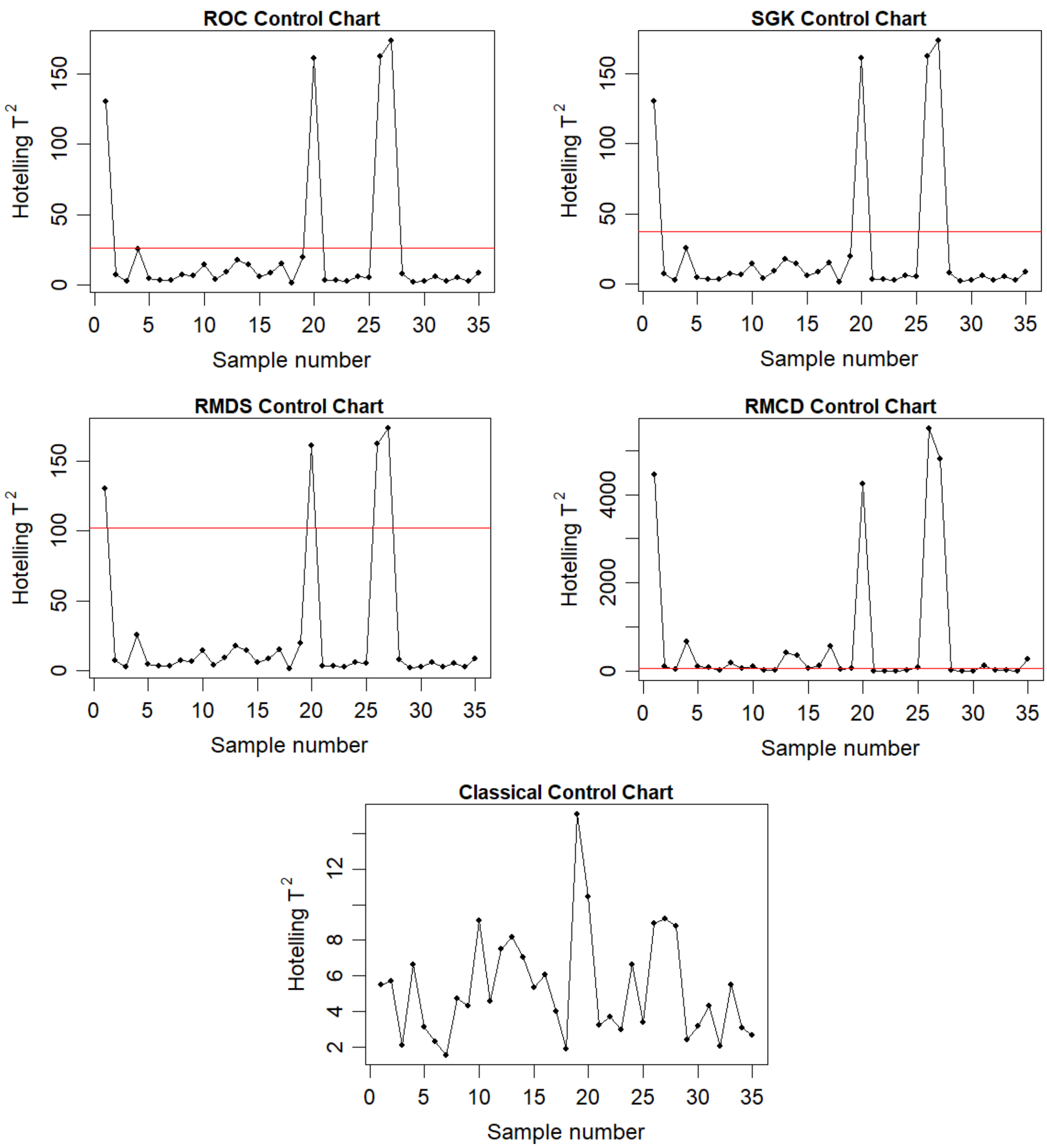


Figure 9.5: Multivariate Hotelling's T^2 Control charts for mean for Star data

Similar to the Tholur dataset, the data collected from Star Pipes & Fittings were originally free from outliers. To evaluate the robustness and detection capability of the proposed multivariate control charts under contamination, two observations were deliberately contaminated in each of Phase I (observations 1 and 20) and Phase II (observations 26 and 27) data. The dataset for this case includes 20 observations in Phase I and 15 observations in Phase II.

Figure 9.5 presents the Multivariate Hotelling's T^2 control charts for monitoring the mean vector of the Star data, comparing the performance of classical and robust methods (ROC, SGK, RMDS, and RMCD). The charts include both Phase I and Phase II observations, with deliberate contamination introduced to assess robustness. It is evident that the ROC, SGK, and RMDS control charts effectively identify the contaminated observations in both phases while maintaining stability for the uncontaminated samples, demonstrating their strong resistance to the influence of outliers and high sensitivity to mean shifts. In contrast, the classical Hotelling's T^2 chart fails to detect the true out-of-control points, indicating its vulnerability to contamination and masking effects. Although the RMCD-based chart also detects the contaminated observations, it exhibits inflated T^2 values and flags several normal observations as out of control, suggesting its relative instability and tendency to produce false alarms under minor variations.

Figure 9.6 presents the Multivariate Exponentially Weighted Moving Variance Likelihood (MEWMVL) control charts for monitoring the dispersion of the Star data, constructed with parameters $\omega = 0.4$ and $\lambda = 0.2$. The charts compare the performance of classical and robust estimators. The classical MEWMVL chart shows a gradual increase in the statistic without any clear signal beyond the control limit, indicating its inability to detect changes in process variability under contamination. In contrast, the robust control charts exhibit a distinct upward shift in the MEWMVL statistic between the 6th and 8th observations, accurately signaling an increase in process dispersion. Among them, the ROC and SGK charts display sharper peaks and faster detection, while RMCD and RMVE also capture the shift but with slightly smoother transitions, reflecting their stability.

9.3 Summary

The analysis presented in this chapter demonstrates the practical effectiveness of the proposed robust bivariate and multivariate control charts in real industrial environments. Using datasets from Tholur Foods Pvt. Ltd. and Star Pipes & Fittings, the results clearly highlight the superiority of robust estimators such as Comedian, Sn, GK, ROC, SGK, and RMDS over the classical and MCD-based methods. In both case studies, the robust charts accurately detected true out-of-control observations while minimizing false alarms, thereby achieving an optimal balance between sensitivity and stability. Conversely, the classical charts failed to identify process shifts, and the RMCD charts exhibited excessive false signaling under clean conditions. These findings confirm that the proposed robust framework offers consistent and reliable performance across different types of industrial processes, making it a valuable tool for real-world process monitoring and quality improvement.

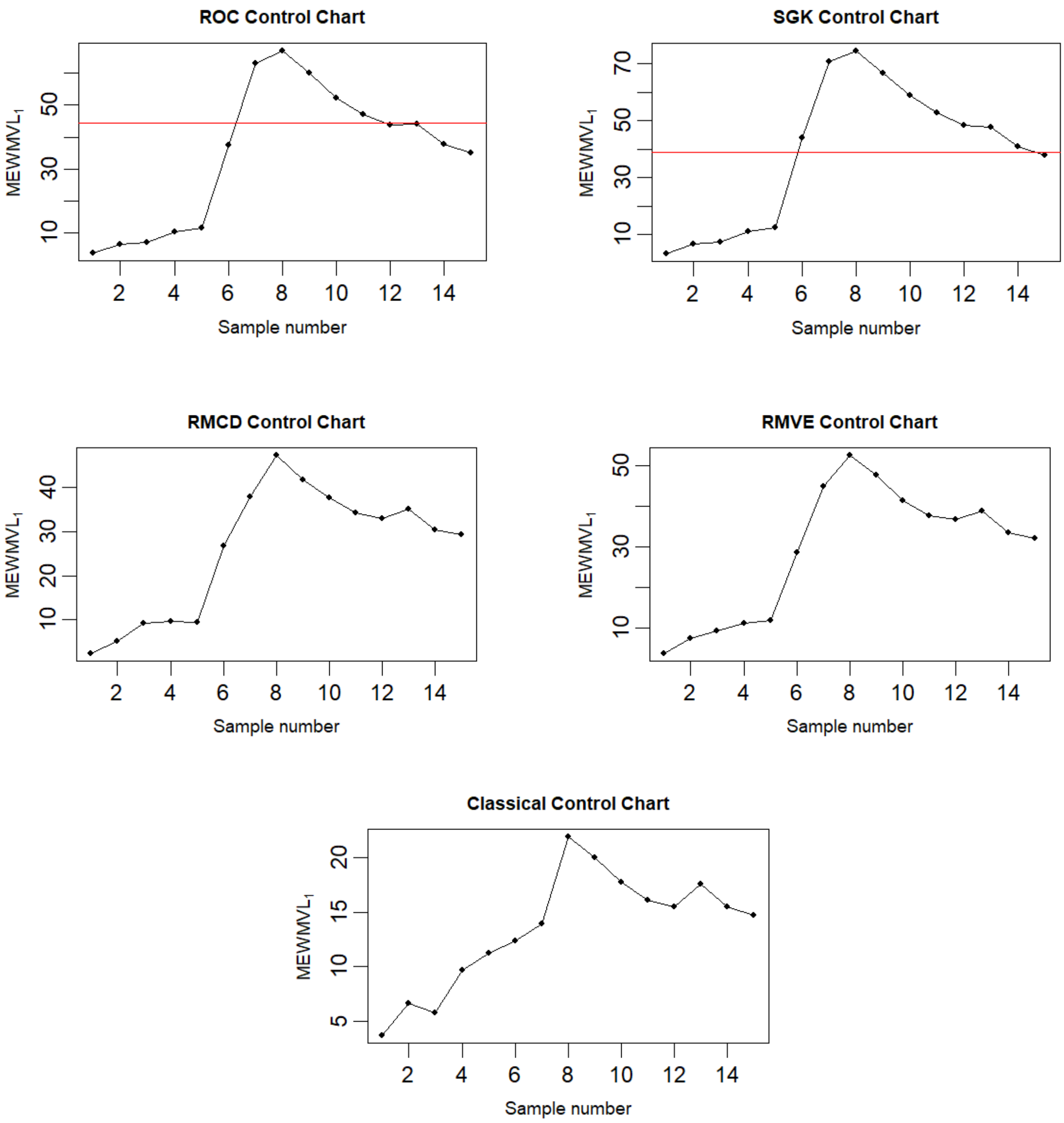


Figure 9.6: Multivariate control charts for dispersion for Star data using MEWMVL$_1$ statistic with $\omega = 0.4$, and $\lambda = 0.2$.

CHAPTER 10

CONCLUSIONS

This research addressed a critical challenge in statistical quality control: the reliable monitoring of bivariate and multivariate processes in the presence of outliers, contamination, and high-dimensional data. Traditional control charts, though widely applied, are highly sensitive to atypical observations, often resulting in inflated false alarm rates or failure to detect genuine process shifts. The primary objective of this thesis was to develop robust estimators and control chart methodologies that overcome these limitations, providing accurate, efficient, and practical tools for contemporary quality management across industrial and service-oriented processes.

A major contribution of this work is the development of the SGK covariance estimator, a robust multivariate estimation method based on the Gnanadesikan-Kettenring approach. Simulation studies demonstrated that SGK reliably identifies contaminated observations while maintaining very low false detection rates across a broad spectrum of scenarios, including high-dimensional, low-sample-size datasets, and data following multivariate normal, t, and exponential distributions. Comparative evaluations with established robust methods, such as MCD, RMD-S, ROC, S_n, and OGK, confirmed that SGK achieves a superior balance between robustness and computational efficiency. Furthermore, Q_n was established as the preferred robust scale estimator due to its consistently high success rate and accuracy across various conditions, while S_n may be preferred when computational efficiency is critical.

Building on robust estimation, bivariate control charts were developed for individual observations using high-breakdown estimators, including Comedian, S_n Covariance and GK. These charts effectively maintain in-control performance comparable to classical T^2 charts under clean data, while outperforming them in the presence of outliers during Phase I monitoring. Among the proposed methods, S_n Covariance and GK consistently demonstrated high detection rates with low false alarm rates, whereas Comedian exhibited the fastest computational performance. For rational subgroups, robust bivariate charts were constructed to incorporate group-wise information, improving parameter estimation and control limit stability. Simulation studies revealed that these subgroup-based charts provide enhanced detection of shifts while maintaining low false alarms, particularly under contamination or small sample conditions, highlighting their practical relevance for industrial applications where subgrouping is feasible.

Extending to multivariate processes, robust control charts for individual observations were developed using ROC and SGK estimators. Simulation studies and real-data applications confirmed that these charts outperform classical Hotelling's T^2, RMCD, and RMVE methods in detecting shifts in process location, particularly under contamination. The ROC-based charts exhibited slightly better performance in higher-dimensional and heavily contaminated settings, whereas SGK provided reliable detection for moderate dimensions and contamination levels. For multivariate rational subgroups, robust charts were formulated to leverage group-level information, which improved the estimation of in-control parameters, stabilized control limits, and facilitated rapid detection of shifts in higher-dimensional processes.

The thesis also addressed the monitoring of process variability, a critical aspect of quality control. Robust bivariate dispersion charts for individual observations were developed using S_n Covariance, GK_S_n, Comedian, and GK_Q_n within the MEWMS and MEWMV frameworks. These charts efficiently detect shifts in process spread while maintaining low false alarm rates, even in the presence of correlation and contamination. Applications to real datasets confirmed that S_n Covariance, GK_S_n, and Comedian achieve the optimal balance between robustness, sensitivity, and computational efficiency. For rational subgroups, bivariate variability charts demonstrated improved performance, particularly in stabilizing control limits and enhancing shift detection in Phase II monitoring. Multivariate dispersion charts were constructed using ROC and SGK estimators with L_c-norm formulations. Simulation studies revealed superior performance over classical, RMCD, and RMVE methods in detecting variability shifts, particularly in high-dimensional and contaminated settings. Real-world applications validated their reliability, robustness, and computational efficiency, demonstrating their practical utility for monitoring multivariate process variability under realistic industrial conditions.

The practical implications of this research are substantial. The proposed robust control charts enhance process monitoring by reliably detecting out of control conditions, reducing unnecessary interventions caused by false alarms, and improving overall decision making. They are particularly valuable for high dimensional processes where interdependencies among quality characteristics cannot be ignored. Their computational efficiency ensures feasibility for real-time applications, while robustness to outliers guarantees that parameter estimates remain accurate, supporting consistent quality control and operational reliability.

CHAPTER 11

RECOMMENDATIONS

- **Development of Robust Variability Control Charts for Rational Subgroups:** Extending robust dispersion monitoring from individual observations to rational subgroups will allow more accurate detection of shifts in process spread at the subgroup level, particularly under contamination.
- **Extension to Attribute and Mixed-Type Data:** Develop robust control charts capable of handling discrete, ordinal, or categorical quality characteristics, broadening applicability in service industries and healthcare.
- **Integration with Machine Learning:** Incorporate adaptive techniques, predictive monitoring, and anomaly detection using machine learning methods combined with robust estimators.
- **High-Dimensional Process Monitoring:** Explore dimensionality reduction techniques such as robust PCA, sparse covariance estimation, or factor analysis to handle very high-dimensional datasets efficiently.
- **Process Optimization Integration:** Combine robust monitoring with multi-objective optimization frameworks to improve decision-making considering cost, quality, and resource efficiency simultaneously.
- **Sector-Specific Industrial Applications:** Validate and adapt the proposed methods across diverse domains, including pharmaceuticals, food processing, chemical manufacturing, and service-oriented industries.
- **Hybrid and Adaptive Approaches:** Develop hybrid models combining continuous and categorical data, and adaptive control charts that update dynamically with evolving process conditions.

List of Publications

1. Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2023). Enhancing multivariate control charts for individual observations using ROC estimates. *Stochastics and Quality Control, 38*(2), 109–118.
2. Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2024). Bivariate robust control charts for individual observations. *Communications in Statistics - Simulation and Computation*, 1–11. https://doi.org/10.1080/03610918.2024.2391866
3. Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2025). Unveiling outliers with robust covariance matrix estimation: A shrinkage approach. *Statistical Papers,* 66(74). https://doi.org/ 10.1007/s00362-025-01694-x
4. Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2025). Robust bivariate control chart for process monitoring with rational subgroups. *Quality and Reliability Engineering International*, 41(7), 3232-3243. https://doi.org/10.1002/qre.70022

List of Presentations

1. Presented a paper entitled "Robust bivariate control charts for individual observations with S_n covariance" in IX International Symposium on Statistics and Optimization, held during October 28-29, 2023 at the Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh, India
2. Presented a paper entitled "RGK_{T_n} Chart: A Robust Approach for monitoring Individual Bivariate Observations" in the national seminar on Recent Developments and Applications in Statistics and Probability" organized by Department of Statistics, St. Thomas college (Autonomous), Thrissur, during 14-15 February 2024.
3. Presented a paper entitled "SGK Approach- A Robust Method for Detecting Multiple Outliers from Multivariate Data" in the Tenth International Conference on Statistics for Twenty-first Century-2024 (ICSTC-2024) organized by International Statistical Fraternity (ISF), Department of Statistics and School of Physical and Mathematical Sciences, University of Kerala, Trivandrum, held during 13 - 16 December, 2024.
4. Presented a paper entitled "Robust bivariate control charts for process monitoring" in the International conference on Advanced Data Analytics and Statistics (ICADS 2025) organized by the Department of Statistics, Vimala College (Autonomous), Thrissur in association with the Kerala Statistical Association (KSA) during 23-25 January 2025.

Bibliography

1. Abu-Shawiesh, M. O. A., Golam Kibria, B. M., & George, F. (2014). A robust bivariate control chart alternative to the Hotelling's T^2 control chart. *Quality and Reliability Engineering International, 30*(1), 25–35.
2. Alfaro, J. L., & Ortega, J. F. (2008). A robust alternative to Hotelling's T^2 control chart using trimmed estimators. *Quality and Reliability Engineering International, 24*(5), 601–611. https://doi.org/10.1002/qre.925
3. Alfaro, J. L., & Ortega, J. F. (2009). A comparison of robust alternatives to Hotelling's T^2 control chart. *Journal of Applied Statistics, 36*(12), 1385–1396. https://doi.org/10.1080/02664760802553055
4. Alt, F. B. (1985). Multivariate quality control. In S. Kotz & N. Johnson (Eds.), *The encyclopedia of statistical sciences* (Vol. 6, pp. 110–122). John Wiley & Sons.
5. Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). John Wiley & Sons.
6. Anggraini, D. (2021). Development of multivariate quality control and quality assurance models for antenatal care service in Indonesia. *Bulletin of the Australian Mathematical Society, 103*(2), 346–348. https://doi.org/10.1017/S0004972720000921
7. Atkinson, A. C. (1994). Fast very robust methods for the detection of multiple outliers. *Journal of the American Statistical Association, 89*, 1329–1330. https://doi.org/10.1080/01621459.1994.10476885
8. Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Wiley.
9. Bodecchi, L. M., Cocchi, M., Malagoli, M., Manfredini, M., & Marchetti, A. (2005). Application of infrared spectroscopy and multivariate quality-control methods in PVC manufacturing. *Analytica Chimica Acta, 554*(1–2), 207–217. https://doi.org/10.1016/j.aca.2005.07.051
10. Brownlee, K. A. (1965). *Statistical theory and methodology in science and engineering.* Wiley.
11. Cabana, E., Lillo, R. E., & Laniado, H. (2021). Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Statistical Papers, 62*, 1583–1609. https://doi.org/10.1007/s00362-020-01203-8

12. Cabana, E., & Lillo, R. E. (2022). Robust multivariate control chart based on shrinkage for individual observations. *Journal of Quality Technology, 54*(4), 415–440. https://doi.org/10.1080/00224065.2021.1931532
13. Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 29*(3), 231–237. https://doi.org/10.2307/2346068
14. Campbell, N. A. (1989). *Bushfire mapping using NOAA AVHRR data* (Technical report). CSIRO.
15. Chenouri, S., Variyath, A. M., & Steiner, S. H. (2009). A multivariate robust control chart for individual observations. *Journal of Quality Technology, 41*(3), 259–271. https://doi.org/10.1080/00224065.2009.11917770
16. Chiang, L. H., & Colegrove, L. F. (2007). Industrial implementation of on-line multivariate quality control. *Chemometrics and Intelligent Laboratory Systems, 88*(2), 143–153. https://doi.org/10.1016/j.chemolab.2007.01.001
17. Daudin, J. J., Duby, C., & Trecourt, P. (1988). Stability of principal component analysis studied by the bootstrap method. *Statistics: A Journal of Theoretical and Applied Statistics, 19*(2), 241–258. https://doi.org/10.1080/02331888808800340
18. DeMiguel, V., Martin-Utrera, A., & Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance, 37*(8), 3018–3034. https://doi.org/10.1016/j.jbankfin.2013.03.016
19. Deming, W. E. (1986). *Out of the crisis.* MIT Press.
20. Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association, 76*(374), 354–362. https://doi.org/10.1080/01621459.1981.10477611
21. Donoho, D. L. (1982). *Breakdown properties of multivariate location estimators* (Technical Report). Harvard University. http://www-stat.stanford.edu/~donoho/Reports/Oldies/BPMLE.pdf
22. Donoho, D. L., & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics, 20*(4), 1803–1827. https://doi.org/10.1214/aos/1176349020
23. Donoho, D. L., & Huber, P. J. (1983). The notion of breakdown point. In P. J. Bickel, K. A. Doksum, & J. L. Hodges (Eds.), *A festschrift for Erich L. Lehmann* (pp. 157–184). Belmont, CA.

24. Ennis, D. M., & Bi, J. (2000). Multivariate quality control with applications to sensory data. *Journal of Food Quality, 23*(6), 541–552. https://doSi.org/ 10.1111/j. 1745-4557.2000.tb00580.x
25. Falk, M. (1997). On MAD and comedians. *Annals of the Institute of Statistical Mathematics, 49*(4), 615–644. https://doi.org/10.1023/A:1003258024248
26. Genton, M. G., & Ma, Y. (1999). Robustness properties of dispersion estimators. *Statistics & Probability Letters, 44*(4), 343–350.
27. Gervini, D. (2003). A robust and efficient adaptive reweighted estimator of multivariate location and scatter. *Journal of Multivariate Analysis, 84*(1), 116–144. https://doi.org/10.1016/S0047-259X(02)00021-1
28. Gnanadesikan, R., & Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics, 28*(1), 81–124. https://doi.org/10.2307/2528961
29. Hall, P., & Welsh, A. H. (1985). Limit theorems for the median deviation. *Annals of the Institute of Statistical Mathematics, 37*(1), 27–36. https://doi.org/10.1007/BF02481078
30. Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*(346), 383–393.
31. Hampel, F. R. (1968). *Contributions to the theory of robust estimation* (Doctoral dissertation, University of California, Berkeley, CA).
32. Hampel, F. R. (1971). A generalized qualitative definition of robustness. *Annals of Mathematical Statistics, 42*(6), 1887–1896. https://doi.org/10.1214/aoms/1177699355
33. Hampel, F. R., Ronchetti, E., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions.* Wiley.
34. Healy, J. D. (1987). A note on multivariate CUSUM procedures. *Technometrics, 29*(4), 409–412. https://doi.org/10.1080/00401706.1987.10488238
35. Hodges, J. L. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In L. M. LeCam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 227–254). University of California Press.
36. Hubert, M., Rousseeuw, P. J., & Segaert, P. (2015). Multivariate functional outlier detection. *Statistical Methods & Applications, 24*(2), 177–202. https://doi.org/10.1007/s10260-014-0264-6

37. Hubert, M., Rousseeuw, P. J., & Van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science, 23*(1), 92–119. https://doi.org/10.1214/07-STS239
38. Huwang, L., Yeh, A. B., & Wu, C. W. (2007). Monitoring multivariate process variability for individual observations. *Journal of Quality Technology, 39*(3), 258–278. https://doi.org/10.1080/00224065.2007.11917768
39. Ialongo, C. (2025). Use of Hotelling's T^2 multivariate control chart for effective monitoring of a laboratory test with a 3-level quality control scheme. *Biochemia Medica, 35*(2), 020701. https://doi.org/10.11613/BM.2025.020701
40. Jackson, J. E. (1985). Multivariate quality control. *Communications in Statistics - Theory and Methods, 14*(11), 2657–2688. https://doi.org/10.1080/03610928508829069
41. Jackson, J. E., & Morris, R. H. (1957). An application of multivariate quality control to photographic processing. *Journal of the American Statistical Association, 52*(280), 186–199. https://doi.org/10.1080/01621459.1957.10501591
42. Jensen, W. A., Birch, J. B., & Woodall, W. H. (2007). High breakdown estimation methods for Phase I multivariate control charts. *Quality and Reliability Engineering International, 23*(5), 615–629. https://doi.org/10.1002/qre.796
43. Juran, J. M., Godfrey, A. B., Hoogstoel, R. E., & Schilling, E. G. (1999). *Juran's quality handbook* (5th ed.).
44. Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical data analysis based on the L1-norm and related methods* (pp. 217–230). Elsevier.
45. Ledoit, O., & Wolf, M. (2003a). Honey, I shrunk the sample covariance matrix. *UPF Economics and Business Working Paper, 691.*
46. Ledoit, O., & Wolf, M. (2003b). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10*(5), 603–621. https://doi.org/10.1016/S0927-5398(03)00007-0
47. Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

48. Lopuhaä, H. P., & Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics, 19*(1), 229–248. https://doi.org/10.1214/aos/1176347978
49. Lowry, C. A., & Montgomery, D. C. (1995). A review of multivariate control charts. *IIE Transactions, 27*(6), 800–810. https://doi.org/10.1080/07408179508936771
50. Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics, 34*(1), 46–53. https://doi.org/10.1080/00401706.1992.10485212
51. Mahpouya, F., Shahriari, H., & Roghanian, E. (2022). Design of a robust T^2 control chart: A re-sampling approach. *Quality and Reliability Engineering International, 38*(2), 924–940. https://doi.org/10.1002/qre.3064
52. Maleki, F., Mehri, S., Aghaie, A., & Shahriari, H. (2020). Robust T^2 control chart using median-based estimators. *Quality and Reliability Engineering International, 36*(6), 2187–2201. https://doi.org/10.1002/qre.2735
53. Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics, 4*(1), 51–67. https://doi.org/10.1214/aos/1176343602
54. Maronna, R. A., & Yohai, V. J. (1995). The behavior of the Stahel-Donoho robust multivariate estimator. *Journal of the American Statistical Association, 90*(429), 330–341. https://doi.org/10.1080/01621459.1995.10476439
55. Maronna, R. A., & Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics, 44*(4), 307–317. https://doi.org/10.1198/004017002188618660
56. Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust statistics: Theory and methods.* John Wiley & Sons.
57. Mason, R. L., Tracy, N. D., & Young, J. C. (1995). Decomposition of T^2 for multivariate control chart interpretation. *Journal of Quality Technology, 27*(2), 99–108. https://doi.org/10.1080/00224065.1995.11979734
58. Memar, A. O., & Niaki, S. T. A. (2009). New control charts for monitoring covariance matrix with individual observations. *Quality and Reliability Engineering International, 25*(7), 821–838. https://doi.org/10.1002/qre.1004
59. Montgomery, D. C. (2020). *Introduction to statistical quality control.* John Wiley & Sons.

60. Montgomery, D. C., & Mastrangelo, C. M. (1991). Some statistical process control methods for autocorrelated data. *Journal of Quality Technology, 23*(3), 179–193. https://doi.org/10.1080/00224065.1991.11978863
61. Oakland, J. S. (2014). *Total quality management and operational excellence: Text with cases.* Routledge.
62. Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters, 1*(6), 327–332. https://doi.org/10.1016/0167-7152(83)90041-1
63. Page, E. S. (1954). Continuous inspection schemes. *Biometrika, 41*(1/2), 100–115. https://doi.org/10.2307/2332728
64. Peña, D., & Prieto, F. J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics, 43*(3), 286–310. https://doi.org/10.1198/00401700152673865
65. Quesenberry, C. P. (2001). The multivariate short-run snapshot Q chart. *Quality Engineering, 13*(4), 679–683. https://doi.org/10.1081/QEN-100107667
66. Roberts, S. W. (2000). Control chart tests based on geometric moving averages. *Technometrics, 42*(1), 97–101. https://doi.org/10.1080/00401706.2000.10485967
67. Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association, 91*(435), 1047–1061. https://doi.org/10.1080/01621459.1996.10476918
68. Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79*(388), 871–880. https://doi.org/10.1080/01621459.1984.10478017
69. Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications, 8*, 283–297.
70. Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association, 88*(424), 1273–1283. https://doi.org/10.2307/2291267
71. Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics, 41*(3), 212–223. https://doi.org/10.1080/00401706.1999.10485670

72. Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(1), 73–79. https://doi.org/10.1002/widm.30
73. Rousseeuw, P. J., & Leroy, A. M. (2003). *Robust regression and outlier detection.* John Wiley & Sons.
74. Rousseeuw, P. J., & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association, 85*(411), 633–639. https://doi.org/10.1080/01621459.1990.10474920
75. Ryan, T. P. (2011). *Statistical methods for quality improvement.* John Wiley & Sons.
76. Sajana, O. K., & Sajesh, T. A. (2020). Sn covariance. *Communications in Statistics - Theory and Methods, 49*(24), 6133–6138. https://doi.org/10.1080/03610926.2019.1628275
77. Sajana, O. K., & Sajesh, T. A. (2022). Multidimensional outlier detection and robust estimation using Sn covariance. *Communications in Statistics - Simulation and Computation, 51*(7), 3912–3922. https://doi.org/10.1080/03610918.2020.1725820
78. Sajesh, T. A. (2023). Fast and robust bivariate control charts for individual observations. *Reliability: Theory & Applications, 18*(4(76)), 296–308.
79. Sajesh, T. A., & Srinivasan, M. R. (2012). Outlier detection for high dimensional data using the Comedian approach. *Journal of Statistical Computation and Simulation, 82*(5), 745–757. https://doi.org/10.1080/00949655.2010.545335
80. Santos-Fernández, E. (2012). *Multivariate statistical quality control using R.* Springer.
81. Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology, 4*(1), Article 32. https://doi.org/10.2202/1544-6115.1175
82. Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*. John Wiley & Sons.
83. Shewhart, W. A. (1931). *Economic control of quality of manufactured product.* D. Van Nostrand Company.
84. Stahel, W. A. (1981). Breakdown of covariance estimators. Fachgruppe für Statistik, Eidgenössische Techn. Hochsch.

85. Surtihadi, J., Raghavachari, M., & Runger, G. (2004). Multivariate control charts for process dispersion. *International Journal of Production Research, 42*(15), 2993–3009. https://doi.org/10.1080/00207540410001710682
86. Taguchi, G., Taguchi, G., Elsayed, E. A., & Hsiang, T. C. (1989). *Quality engineering in production systems.* McGraw-Hill Companies.
87. Tang, P. F., & Barnett, N. S. (1996). Dispersion control for multivariate processes. *Australian Journal of Statistics, 38*(3), 235–251. https://doi.org/10.1111/j.1467-842X.1996.tb01373.x
88. Tukey, J. W. (1975, August). Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians* (Vol. 2, pp. 523-531).
89. Tyler, D. E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *The Annals of Statistics, 22*(2), 1024–1044. https://doi.org/10.1214/aos/1176325624
90. Vargas, J. A. (2003). Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology, 35*(4), 367–376. https://doi.org/10.1080/00224065.2003.11980233
91. Variyath, A. M., & Vattathoor, J. (2014). Robust control charts for monitoring process variability in phase I multivariate individual observations. *Quality and Reliability Engineering International, 30*(6), 795–812. https://doi.org/10.1002/qre.1576
92. Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2024). Bivariate robust control charts for individual observations. *Communications in Statistics - Simulation and Computation.* https://doi.org/10.1080/03610918.2024.2391866
93. Vijayalakshmi, S., Sebastian, N., & Sajesh, T. A. (2025). Unveiling outliers with robust covariance matrix estimation: A shrinkage approach. *Statistical Papers, 66*(4), 1–23.
94. Wheeler, D. J., & Chambers, D. S. (1992). *Understanding statistical process control.* Knoxville.
95. Wilks, S. S. (1962). *Mathematical statistics.* John Wiley & Sons.
96. Willems, G., Pison, G., Rousseeuw, P. J., & Van Aelst, S. (2002). A robust Hotelling test. *Metrika, 55*(1), 125–138. https://doi.org/10.1007/s001840200191
97. Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology, 32*(4), 341–350.

https://doi.org/10.1080/00224065.2000.11979887

98. Woodall, W. H., & Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology, 31*(4), 376–386. https://doi.org/10.1080/00224065.1999.11979752
99. Woodall, W. H., & Montgomery, D. C. (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology, 46*(1), 78–94. https://doi.org/10.1080/00224065.2014.11917931
100. Woodall, W. H., & Ncube, M. M. (1985). Multivariate CUSUM quality-control procedures. *Technometrics, 27*(3), 285–292. https://doi.org/10.1080/00401706.1985.10487927
101. Yeh, A. B., Huwang, L., & Wu, C. W. (2005). A multivariate EWMA control chart for monitoring process variability with individual observations. *IIE Transactions, 37*(11), 1023–1035. https://doi.org/10.1080/07408170500232263
102. Yeh, A. B., Huwang, L., & Wu, Y. F. (2004). A likelihood-ratio-based EWMA control chart for monitoring variability of multivariate normal processes. *IIE Transactions, 36*(9), 865–879. https://doi.org/10.1080/07408170490509609
103. Yeh, A., Lin, D., Zhou, H., & Venkataramani, C. (2003). A multivariate exponentially weighted moving average control chart for monitoring process variability. *Journal of Applied Statistics, 30*(5), 507–536. https://doi.org/10.1080/0266476031000139536
104. Zuo, Y. (2003). Projection-based depth functions and associated medians. *The Annals of Statistics, 31*(5), 1460–1490. https://doi.org/10.1214/aos/1074290080

APPENDIX A

Table A.1: SRs of outlier detection methods for Normal distribution (θ_2 = 5 and ξ_2 = 0.1)

n	***p***	***ε***	C_{Sn}	C_{MAD}	C_{Qn}	S_n	**ROC**	**OGK**	**MCD**	**RMD-S**
100	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.960	1.000
		0.3	1.000	1.000	1.000	1.000	0.576	0.936	0.000	0.978
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	0.721	0.980	0.000	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.140	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	0.969	1.000	0.000	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	0.990	1.000	0.000	1.000
500	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	0.933	0.975	0.000	0.941
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	0.964	1.000	0.000	0.890
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.533	0.733	1.000	1.000	0.989	1.000	0.000	0.478
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.867	0.600	1.000	1.000	1.000	1.000	0.000	0.160
1000	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	0.901	0.988	0.000	0.901
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	0.967	1.000	0.000	0.736
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.600	0.800	1.000	1.000	0.999	1.000	0.000	0.177
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.800	0.933	1.000	1.000	1.000	1.000	0.000	0.037

Table A.2: FDRs of outlier detection methods for Normal distribution ($\theta_2 = 5$ and $\xi_2 = 0.1$)

n	*p*	*ε*	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	0.000	0.000	0.000	0.094	0.001	0.038	0.059	0.028
		0.2	0.000	0.000	0.000	0.113	0.006	0.080	0.055	0.005
		0.3	0.000	0.000	0.000	0.105	0.006	0.060	0.048	0.000
		0.1	0.000	0.000	0.000	0.092	0.002	0.035	0.413	0.000
	10	0.2	0.000	0.000	0.000	0.000	0.000	0.048	0.110	0.025
		0.3	0.000	0.000	0.000	0.000	0.005	0.085	0.094	0.002
		0.1	0.000	0.000	0.000	0.001	0.003	0.065	0.434	0.000
		0.2	0.000	0.000	0.000	0.001	0.001	0.036	0.615	0.000
	20	0.3	0.000	0.000	0.000	0.022	0.000	0.056	0.269	0.029
		0.1	0.000	0.000	0.000	0.050	0.003	0.098	0.346	0.002
		0.2	0.000	0.000	0.000	0.028	0.001	0.075	0.478	0.000
		0.3	0.000	0.000	0.000	0.017	0.000	0.037	0.592	0.000
	30	0.1	0.000	0.001	0.000	0.028	0.000	0.070	0.290	0.021
		0.2	0.000	0.000	0.000	0.048	0.001	0.100	0.363	0.001
		0.3	0.000	0.000	0.000	0.026	0.001	0.075	0.436	0.000
		0.1	0.000	0.000	0.000	0.019	0.000	0.047	0.521	0.000
500	5	0.2	0.000	0.000	0.000	0.093	0.000	0.030	0.028	0.023
		0.3	0.000	0.000	0.000	0.133	0.004	0.077	0.025	0.006
		0.1	0.000	0.000	0.000	0.120	0.002	0.056	0.022	0.000
		0.2	0.000	0.000	0.000	0.116	0.001	0.029	0.330	0.000
	10	0.3	0.000	0.000	0.000	0.000	0.000	0.030	0.030	0.026
		0.1	0.000	0.000	0.000	0.000	0.002	0.080	0.027	0.003
		0.2	0.000	0.000	0.000	0.000	0.001	0.059	0.229	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.031	0.653	0.000
	20	0.1	0.000	0.000	0.000	0.010	0.000	0.037	0.040	0.024
		0.2	0.000	0.000	0.000	0.036	0.001	0.083	0.129	0.002
		0.3	0.000	0.000	0.000	0.023	0.000	0.058	0.499	0.000
		0.1	0.000	0.000	0.000	0.012	0.000	0.031	0.897	0.000
	30	0.2	0.000	0.000	0.000	0.011	0.000	0.039	0.061	0.023
		0.3	0.000	0.000	0.000	0.039	0.000	0.083	0.238	0.001
		0.1	0.000	0.000	0.000	0.025	0.000	0.061	0.638	0.000
		0.2	0.000	0.000	0.000	0.013	0.000	0.034	0.975	0.000
1000	5	0.3	0.000	0.000	0.000	0.089	0.000	0.026	0.028	0.001
		0.1	0.000	0.000	0.000	0.139	0.004	0.077	0.025	0.005
		0.2	0.000	0.000	0.000	0.125	0.002	0.054	0.022	0.000
		0.3	0.000	0.000	0.000	0.114	0.001	0.030	0.322	0.000
	10	0.1	0.000	0.000	0.000	0.000	0.000	0.028	0.030	0.001
		0.2	0.000	0.000	0.000	0.000	0.002	0.079	0.025	0.003
		0.3	0.000	0.000	0.000	0.000	0.001	0.057	0.195	0.000
		0.1	0.000	0.000	0.000	0.000	0.000	0.028	0.635	0.000
	20	0.2	0.000	0.000	0.000	0.010	0.000	0.037	0.031	0.024
		0.3	0.000	0.000	0.000	0.036	0.001	0.078	0.090	0.002
		0.1	0.000	0.000	0.000	0.023	0.000	0.057	0.449	0.000
		0.2	0.000	0.000	0.000	0.012	0.000	0.030	0.933	0.000
	30	0.3	0.000	0.000	0.000	0.008	0.000	0.034	0.036	0.001
		0.1	0.000	0.000	0.000	0.038	0.000	0.082	0.160	0.001
		0.2	0.000	0.000	0.000	0.023	0.000	0.059	0.667	0.000
		0.3	0.000	0.000	0.000	0.012	0.000	0.030	0.998	0.000

Table A.3: SRs of outlier detection methods for multivariate *t* distribution (θ_2 = 5 and ξ_2 = 0.1)

n	*p*	*ε*	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.800	1.000
		0.3	1.000	1.000	1.000	1.000	0.945	0.990	0.032	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.960	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.054	1.000
		0.3	1.000	1.000	1.000	1.000	0.987	1.000	0.056	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.028	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.024	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.055	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.012	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.030	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.036	1.000
500	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.961	1.000
		0.3	1.000	1.000	1.000	1.000	0.984	1.000	0.031	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.032	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.055	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.048	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.053	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.079	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.039	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.062	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.093	1.000
1000	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.033	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.035	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.058	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.039	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.059	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.080	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.045	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.066	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.108	1.000

Table A.4: FDRs of outlier detection methods for multivariate t distribution (θ_2 = 5 and ξ_2 =0.1)

n	***p***	***ε***	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$	$\mathbf{S}_\mathbf{n}$	**ROC**	**OGK**	**MCD**	**RMD-S**
100	5	0.1	0.000	0.000	0.000	0.090	0.000	0.036	0.236	0.000
		0.2	0.051	0.051	0.047	0.224	0.108	0.216	0.228	0.116
		0.3	0.053	0.053	0.044	0.228	0.087	0.168	0.226	0.070
		0.1	0.038	0.038	0.035	0.291	0.057	0.116	0.450	0.029
	10	0.2	0.000	0.000	0.000	0.000	0.048	0.000	0.351	0.002
		0.3	0.104	0.104	0.086	0.099	0.146	0.262	0.316	0.162
		0.1	0.070	0.070	0.077	0.111	0.111	0.204	0.484	0.095
		0.2	0.058	0.058	0.064	0.155	0.077	0.150	0.590	0.040
	20	0.3	0.000	0.000	0.000	0.026	0.000	0.070	0.379	0.002
		0.1	0.161	0.161	0.158	0.267	0.170	0.296	0.431	0.198
		0.2	0.155	0.155	0.146	0.212	0.132	0.241	0.493	0.120
		0.3	0.110	0.110	0.120	0.154	0.095	0.177	0.571	0.055
	30	0.1	0.000	0.000	0.000	0.027	0.000	0.078	0.340	0.002
		0.2	0.242	0.242	0.250	0.292	0.191	0.317	0.391	0.221
		0.3	0.202	0.202	0.217	0.235	0.151	0.260	0.449	0.141
		0.1	0.150	0.150	0.156	0.170	0.102	0.191	0.543	0.063
500	5	0.2	0.000	0.000	0.000	0.092	0.000	0.028	0.235	0.001
		0.3	0.042	0.042	0.045	0.250	0.110	0.214	0.214	0.123
		0.1	0.037	0.037	0.039	0.303	0.086	0.170	0.200	0.075
		0.2	0.033	0.032	0.032	0.378	0.058	0.119	0.425	0.035
	10	0.3	0.000	0.000	0.000	0.000	0.000	0.031	0.326	0.001
		0.1	0.068	0.068	0.068	0.116	0.141	0.261	0.310	0.162
		0.2	0.058	0.058	0.059	0.151	0.111	0.211	0.504	0.101
		0.3	0.052	0.052	0.051	0.205	0.077	0.151	0.645	0.047
	20	0.1	0.000	0.000	0.000	0.009	0.000	0.036	0.432	0.001
		0.2	0.107	0.107	0.101	0.270	0.174	0.301	0.515	0.203
		0.3	0.084	0.084	0.086	0.217	0.132	0.244	0.653	0.124
		0.1	0.076	0.077	0.076	0.161	0.099	0.181	0.784	0.060
	30	0.2	0.000	0.000	0.000	0.009	0.000	0.035	0.482	0.001
		0.3	0.131	0.131	0.126	0.298	0.190	0.326	0.572	0.226
		0.1	0.114	0.114	0.110	0.241	0.149	0.264	0.694	0.141
		0.2	0.094	0.094	0.096	0.177	0.107	0.194	0.831	0.066
1000	5	0.3	0.000	0.000	0.000	0.091	0.000	0.027	0.226	0.001
		0.1	0.041	0.041	0.043	0.267	0.111	0.217	0.211	0.126
		0.2	0.040	0.040	0.041	0.319	0.086	0.171	0.195	0.080
		0.3	0.033	0.033	0.032	0.401	0.059	0.121	0.419	0.037
	10	0.1	0.000	0.000	0.000	0.000	0.000	0.028	0.326	0.001
		0.2	0.068	0.068	0.065	0.130	0.142	0.266	0.310	0.168
		0.3	0.059	0.059	0.058	0.167	0.109	0.208	0.480	0.100
		0.1	0.052	0.052	0.053	0.217	0.082	0.159	0.647	0.052
	20	0.2	0.000	0.000	0.000	0.007	0.000	0.031	0.440	0.001
		0.3	0.093	0.093	0.094	0.269	0.171	0.302	0.520	0.205
		0.1	0.086	0.086	0.086	0.217	0.138	0.242	0.671	0.129
		0.2	0.075	0.075	0.071	0.158	0.097	0.177	0.809	0.065
	30	0.3	0.000	0.000	0.000	0.009	0.000	0.034	0.500	0.002
		0.1	0.119	0.119	0.117	0.299	0.192	0.327	0.600	0.228
		0.2	0.100	0.099	0.102	0.243	0.149	0.263	0.729	0.141
		0.3	0.082	0.083	0.082	0.177	0.105	0.194	0.877	0.067

Table A.5: SRs of outlier detection methods for multivariate exponential distribution (θ_2 = 5 and ξ_2 = 1)

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.990	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.341	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.550	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.343	1.000
500	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.565	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.014	1.000
1000	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.801	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.004	1.000

Table A.6: FDRs of outlier detection methods for multivariate exponential distribution (θ_2 = 5 and ξ_2 = 1)

n	***p***	ε	**C$_{Sn}$**	**C$_{MAD}$**	**C$_{Qn}$**	**S$_n$**	**ROC**	**OGK**	**MCD**	**RMD-S**
100	5	0.1	0.000	0.000	0.000	0.103	0.001	0.042	0.030	0.001
		0.2	0.000	0.000	0.000	0.039	0.000	0.033	0.016	0.000
		0.3	0.000	0.000	0.000	0.028	0.002	0.022	0.007	0.000
		0.1	0.000	0.000	0.000	0.024	0.000	0.008	0.005	0.000
	10	0.2	0.000	0.000	0.000	0.000	0.000	0.044	0.070	0.002
		0.3	0.000	0.000	0.000	0.000	0.001	0.037	0.054	0.000
		0.1	0.000	0.000	0.000	0.000	0.000	0.025	0.026	0.000
		0.2	0.000	0.000	0.000	0.000	0.000	0.022	0.017	0.000
	20	0.3	0.000	0.000	0.000	0.022	0.000	0.057	0.232	0.001
		0.1	0.000	0.000	0.000	0.019	0.000	0.044	0.176	0.000
		0.2	0.000	0.000	0.000	0.012	0.001	0.043	0.102	0.000
		0.3	0.000	0.000	0.000	0.006	0.000	0.020	0.234	0.000
	30	0.1	0.000	0.006	0.000	0.033	0.000	0.073	0.282	0.002
		0.2	0.000	0.000	0.000	0.021	0.000	0.052	0.196	0.000
		0.3	0.000	0.000	0.000	0.009	0.000	0.038	0.221	0.000
		0.1	0.000	0.000	0.000	0.005	0.000	0.022	0.260	0.000
500	5	0.2	0.000	0.000	0.000	0.095	0.000	0.026	0.001	0.000
		0.3	0.000	0.000	0.000	0.042	0.000	0.016	0.001	0.000
		0.1	0.000	0.000	0.000	0.042	0.000	0.018	0.001	0.000
		0.2	0.000	0.000	0.000	0.041	0.000	0.005	0.001	0.000
	10	0.3	0.000	0.000	0.000	0.000	0.000	0.030	0.002	0.001
		0.1	0.000	0.000	0.000	0.000	0.000	0.026	0.001	0.000
		0.2	0.000	0.000	0.000	0.000	0.000	0.017	0.001	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.012	0.001	0.000
	20	0.1	0.000	0.000	0.000	0.009	0.000	0.037	0.004	0.001
		0.2	0.000	0.000	0.000	0.006	0.000	0.029	0.004	0.000
		0.3	0.000	0.000	0.000	0.004	0.000	0.023	0.003	0.000
		0.1	0.000	0.000	0.000	0.001	0.000	0.015	0.003	0.000
	30	0.2	0.000	0.000	0.000	0.013	0.000	0.041	0.008	0.001
		0.3	0.000	0.000	0.000	0.008	0.000	0.031	0.008	0.000
		0.1	0.000	0.000	0.000	0.005	0.000	0.028	0.008	0.000
		0.2	0.000	0.000	0.000	0.002	0.000	0.018	0.005	0.000
1000	5	0.3	0.000	0.000	0.000	0.093	0.000	0.027	0.001	0.000
		0.1	0.000	0.000	0.000	0.047	0.000	0.019	0.001	0.000
		0.2	0.000	0.000	0.000	0.047	0.000	0.013	0.001	0.000
		0.3	0.000	0.000	0.000	0.051	0.000	0.007	0.001	0.000
	10	0.1	0.000	0.000	0.000	0.000	0.000	0.029	0.002	0.001
		0.2	0.000	0.000	0.000	0.000	0.000	0.023	0.001	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.017	0.001	0.000
		0.1	0.000	0.000	0.000	0.000	0.000	0.010	0.001	0.000
	20	0.2	0.000	0.000	0.000	0.009	0.000	0.032	0.002	0.001
		0.3	0.000	0.000	0.000	0.004	0.000	0.024	0.002	0.000
		0.1	0.000	0.000	0.000	0.004	0.000	0.022	0.001	0.000
		0.2	0.000	0.000	0.000	0.001	0.000	0.012	0.001	0.000
	30	0.3	0.000	0.000	0.000	0.007	0.000	0.033	0.003	0.001
		0.1	0.000	0.000	0.000	0.005	0.000	0.026	0.003	0.000
		0.2	0.000	0.000	0.000	0.003	0.000	0.025	0.002	0.000
		0.3	0.000	0.000	0.000	0.002	0.000	0.015	0.002	0.000

Table A.7: SRs of outlier detection methods for multivariate exponential distribution (θ_2 = 5 and ξ_2 = 0.1)

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.960	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.960	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.720	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
500	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
1000	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000

Table A.8: FDRs of outlier detection methods for multivariate exponential distribution (θ_2 = 5 and ξ_2 = 0.1)

n	*p*	*ε*	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	0.000	0.000	0.000	0.103	0.001	0.042	0.030	0.001
		0.2	0.000	0.000	0.000	0.033	0.000	0.027	0.023	0.001
		0.3	0.000	0.000	0.000	0.050	0.001	0.022	0.010	0.000
		0.1	0.000	0.000	0.000	0.145	0.001	0.020	0.033	0.000
	10	0.2	0.000	0.000	0.000	0.000	0.000	0.044	0.070	0.002
		0.3	0.000	0.000	0.000	0.000	0.000	0.039	0.059	0.000
		0.1	0.000	0.000	0.000	0.000	0.002	0.045	0.048	0.000
		0.2	0.000	0.000	0.000	0.000	0.001	0.022	0.594	0.000
	20	0.3	0.000	0.000	0.000	0.022	0.000	0.057	0.232	0.001
		0.1	0.000	0.000	0.000	0.018	0.000	0.054	0.157	0.000
		0.2	0.000	0.000	0.000	0.017	0.001	0.044	0.467	0.000
		0.3	0.000	0.000	0.000	0.020	0.000	0.028	0.582	0.000
	30	0.1	0.000	0.006	0.000	0.033	0.000	0.073	0.291	0.002
		0.2	0.000	0.000	0.000	0.017	0.000	0.063	0.243	0.000
		0.3	0.000	0.000	0.000	0.014	0.000	0.061	0.437	0.000
		0.1	0.000	0.000	0.000	0.018	0.000	0.033	0.525	0.000
500	5	0.2	0.000	0.000	0.000	0.095	0.000	0.026	0.002	0.000
		0.3	0.000	0.000	0.000	0.054	0.000	0.019	0.002	0.000
		0.1	0.000	0.000	0.000	0.135	0.000	0.017	0.001	0.000
		0.2	0.000	0.000	0.000	0.315	0.000	0.012	0.001	0.000
	10	0.3	0.000	0.000	0.000	0.000	0.000	0.030	0.002	0.001
		0.1	0.000	0.000	0.000	0.000	0.000	0.027	0.002	0.000
		0.2	0.000	0.000	0.000	0.000	0.000	0.026	0.002	0.000
		0.3	0.000	0.000	0.000	0.002	0.000	0.019	0.616	0.000
	20	0.1	0.000	0.000	0.000	0.009	0.000	0.037	0.005	0.001
		0.2	0.000	0.000	0.000	0.008	0.000	0.032	0.003	0.000
		0.3	0.000	0.000	0.000	0.007	0.000	0.037	0.408	0.000
		0.1	0.000	0.000	0.000	0.004	0.000	0.017	0.909	0.000
	30	0.2	0.000	0.000	0.000	0.013	0.000	0.041	0.009	0.001
		0.3	0.000	0.000	0.000	0.008	0.000	0.032	0.010	0.000
		0.1	0.000	0.000	0.000	0.007	0.000	0.029	0.591	0.000
		0.2	0.000	0.000	0.000	0.004	0.000	0.014	0.987	0.000
1000	5	0.3	0.000	0.000	0.000	0.093	0.000	0.027	0.001	0.000
		0.1	0.000	0.000	0.000	0.080	0.000	0.020	0.001	0.000
		0.2	0.000	0.000	0.000	0.171	0.000	0.015	0.001	0.000
		0.3	0.000	0.000	0.000	0.347	0.000	0.009	0.001	0.000
	10	0.1	0.000	0.000	0.000	0.000	0.000	0.029	0.001	0.001
		0.2	0.000	0.000	0.000	0.000	0.000	0.028	0.001	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.026	0.001	0.000
		0.1	0.000	0.000	0.000	0.004	0.000	0.013	0.574	0.000
	20	0.2	0.000	0.000	0.000	0.009	0.000	0.032	0.002	0.001
		0.3	0.000	0.000	0.000	0.005	0.000	0.025	0.002	0.000
		0.1	0.000	0.000	0.000	0.005	0.000	0.027	0.280	0.000
		0.2	0.000	0.000	0.000	0.001	0.000	0.014	0.946	0.000
	30	0.3	0.000	0.000	0.000	0.007	0.000	0.033	0.003	0.001
		0.1	0.000	0.000	0.000	0.007	0.000	0.028	0.003	0.000
		0.2	0.000	0.000	0.000	0.005	0.000	0.031	0.574	0.000
		0.3	0.000	0.000	0.000	0.002	0.000	0.018	0.999	0.000

Table A.9: SRs of outlier detection methods for affinely transformed data (θ_2 = 5 and ξ_2 = 0.1)

n	***p***	***ε***	$\mathbf{C}_{Sn}$	$\mathbf{C}_{MAD}$	$\mathbf{C}_{Qn}$	$\mathbf{S}_\mathbf{n}$	**ROC**	**OGK**	**MCD**	**RMD-S**
100	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	0.970	1.000	1.000	1.000	1.000	0.983
		0.3	0.900	0.900	0.958	0.999	0.626	0.967	0.000	0.842
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	0.900	0.900	0.990	1.000	1.000	1.000	0.000	1.000
		0.3	0.900	0.900	0.990	0.787	0.725	0.998	0.000	0.906
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.160	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.600	0.600	1.000	1.000	0.819	1.000	0.000	0.906
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.900	1.000	1.000	0.997	0.914	1.000	0.000	0.935
500	5	0.1	1.000	1.000	0.999	1.000	1.000	1.000	1.000	0.996
		0.2	0.900	0.900	0.990	1.000	1.000	1.000	1.000	0.990
		0.3	0.800	0.817	0.990	1.000	0.539	0.980	0.000	0.800
	10	0.1	1.000	0.900	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	0.900	0.900	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.800	0.800	0.970	0.860	0.693	1.000	0.000	0.598
	20	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	0.800	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.3	0.700	0.900	0.960	1.000	0.870	1.000	0.000	0.364
	30	0.1	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000
		0.2	0.900	1.000	0.980	1.000	1.000	1.000	0.000	1.000
		0.3	0.700	0.800	0.980	1.000	0.964	1.000	0.000	0.225
1000	5	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	0.900	1.000	0.980	1.000	1.000	1.000	1.000	1.000
		0.3	0.800	0.700	0.970	1.000	0.572	0.989	0.000	0.809
	10	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	0.800	1.000	0.970	1.000	1.000	1.000	0.000	1.000
		0.3	0.600	0.600	0.950	0.795	0.653	1.000	0.000	0.462
	20	0.1	1.000	1.000	0.970	1.000	1.000	1.000	0.000	1.000
		0.2	0.900	1.000	0.980	1.000	1.000	1.000	0.000	1.000
		0.3	0.800	0.900	0.920	1.000	0.785	1.000	0.000	0.478
	30	0.1	1.000	1.000	0.990	1.000	1.000	1.000	0.000	1.000
		0.2	1.000	1.000	0.950	1.000	1.000	1.000	0.000	1.000
		0.3	0.400	0.500	0.950	1.000	0.968	1.000	0.000	0.006

Table A.10: FDRs of outlier detection methods for affinely transformed data (θ_2 = 5 and ξ_2 = 0.1)

n	*p*	*ε*	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	0.000	0.000	0.000	0.169	0.012	0.107	0.054	0.020
		0.2	0.000	0.000	0.000	0.092	0.007	0.086	0.052	0.022
		0.3	0.000	0.000	0.000	0.075	0.005	0.059	0.037	0.009
		0.1	0.000	0.007	0.000	0.152	0.007	0.035	0.430	0.002
	10	0.2	0.000	0.000	0.000	0.001	0.005	0.110	0.113	0.013
		0.3	0.000	0.000	0.000	0.000	0.004	0.091	0.088	0.022
		0.1	0.000	0.000	0.001	0.001	0.005	0.072	0.429	0.007
		0.2	0.000	0.000	0.000	0.010	0.006	0.042	0.605	0.001
	20	0.3	0.000	0.000	0.000	0.066	0.003	0.115	0.271	0.009
		0.1	0.000	0.000	0.000	0.053	0.003	0.101	0.351	0.019
		0.2	0.000	0.000	0.000	0.040	0.002	0.077	0.468	0.001
		0.3	0.064	0.111	0.000	0.033	0.004	0.046	0.597	0.000
	30	0.1	0.000	0.013	0.000	0.067	0.001	0.118	0.286	0.006
		0.2	0.001	0.001	0.000	0.050	0.002	0.102	0.365	0.014
		0.3	0.000	0.000	0.000	0.057	0.001	0.078	0.431	0.001
		0.1	0.000	0.000	0.001	0.034	0.003	0.046	0.520	0.000
500	5	0.2	0.000	0.000	0.000	0.168	0.006	0.102	0.028	0.019
		0.3	0.000	0.000	0.000	0.096	0.004	0.078	0.022	0.012
		0.1	0.000	0.000	0.000	0.095	0.005	0.060	0.019	0.004
		0.2	0.006	0.000	0.000	0.144	0.004	0.033	0.338	0.008
	10	0.3	0.000	0.000	0.000	0.001	0.004	0.104	0.031	0.014
		0.1	0.000	0.000	0.000	0.000	0.002	0.084	0.028	0.012
		0.2	0.000	0.000	0.000	0.000	0.003	0.064	0.237	0.002
		0.3	0.000	0.000	0.002	0.007	0.002	0.034	0.645	0.001
	20	0.1	0.000	0.000	0.000	0.053	0.001	0.102	0.038	0.008
		0.2	0.000	0.000	0.000	0.042	0.001	0.083	0.128	0.015
		0.3	0.000	0.000	0.000	0.037	0.001	0.067	0.503	0.002
		0.1	0.000	0.000	0.001	0.026	0.002	0.034	0.895	0.000
	30	0.2	0.000	0.000	0.000	0.054	0.000	0.104	0.059	0.006
		0.3	0.000	0.000	0.000	0.044	0.000	0.086	0.241	0.015
		0.1	0.000	0.000	0.000	0.040	0.001	0.067	0.627	0.001
		0.2	0.003	0.000	0.000	0.027	0.001	0.037	0.979	0.000
1000	5	0.3	0.000	0.000	0.000	0.169	0.006	0.100	0.025	0.021
		0.1	0.000	0.000	0.000	0.096	0.004	0.076	0.023	0.006
		0.2	0.000	0.000	0.000	0.090	0.004	0.058	0.023	0.002
		0.3	0.002	0.000	0.000	0.172	0.005	0.031	0.322	0.005
	10	0.1	0.000	0.000	0.000	0.000	0.003	0.101	0.028	0.017
		0.2	0.000	0.000	0.000	0.000	0.002	0.081	0.023	0.011
		0.3	0.000	0.000	0.000	0.001	0.003	0.062	0.205	0.000
		0.1	0.001	0.000	0.000	0.006	0.003	0.034	0.645	0.001
	20	0.2	0.000	0.000	0.000	0.053	0.001	0.103	0.031	0.011
		0.3	0.000	0.000	0.000	0.041	0.001	0.085	0.090	0.009
		0.1	0.000	0.000	0.000	0.038	0.001	0.066	0.443	0.001
		0.2	0.000	0.000	0.000	0.023	0.001	0.036	0.933	0.000
	30	0.3	0.000	0.000	0.000	0.053	0.000	0.102	0.036	0.007
		0.1	0.000	0.000	0.000	0.040	0.000	0.086	0.152	0.010
		0.2	0.000	0.000	0.000	0.039	0.001	0.064	0.663	0.000
		0.3	0.000	0.000	0.000	0.026	0.000	0.035	0.997	0.000

Table A.11: SRs of outlier detection methods for Normal distribution (θ_2 = 5 and ξ_2 = 1)

n	*p*	*ε*	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	RMD-S
50	100	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	300	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	100	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	300	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.12: FDRs of outlier detection methods for Normal distribution (θ_2 = 5 and ξ_2 = 1)

n	*p*	*ε*	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	RMD-S
50	100	0.1	0.003	0.007	0.004	0.068	0.000	0.118	0.008
		0.2	0.000	0.000	0.000	0.041	0.000	0.086	0.000
		0.3	0.000	0.000	0.000	0.029	0.000	0.056	0.000
	200	0.1	0.000	0.003	0.000	0.078	0.000	0.233	0.003
		0.2	0.000	0.000	0.000	0.055	0.000	0.146	0.000
		0.3	0.000	0.000	0.000	0.026	0.000	0.071	0.000
	300	0.1	0.000	0.000	0.000	0.111	0.000	0.257	0.000
		0.2	0.000	0.000	0.000	0.063	0.000	0.199	0.000
		0.3	0.000	0.000	0.000	0.040	0.000	0.116	0.000
100	100	0.1	0.000	0.004	0.001	0.001	0.000	0.001	0.000
		0.2	0.000	0.000	0.000	0.000	0.000	0.001	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.001	0.000
	200	0.1	0.000	0.000	0.000	0.001	0.000	0.002	0.000
		0.2	0.000	0.000	0.000	0.001	0.000	0.001	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.001	0.000
	300	0.1	0.000	0.000	0.000	0.001	0.000	0.003	0.000
		0.2	0.000	0.000	0.000	0.001	0.000	0.002	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.001	0.000

Table A.13: SRs of outlier detection methods for Normal distribution (θ_2 = 5 and ξ_2 = 0.1)

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	RMD-S
50	100	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	300	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	100	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	200	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	300	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.14: FDRs of outlier detection methods for Normal distribution (θ_2 = 5 and ξ_2 = 0.1)

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	RMD-S
50	100	0.1	0.000	0.000	0.000	0.084	0.000	0.224	0.000
		0.2	0.000	0.000	0.000	0.088	0.000	0.138	0.000
		0.3	0.000	0.000	0.000	0.040	0.000	0.097	0.000
	200	0.1	0.000	0.002	0.000	0.142	0.000	0.284	0.000
		0.2	0.000	0.000	0.000	0.103	0.003	0.198	0.000
		0.3	0.000	0.000	0.000	0.060	0.000	0.117	0.000
	300	0.1	0.000	0.000	0.000	0.151	0.000	0.309	0.000
		0.2	0.000	0.000	0.000	0.130	0.000	0.230	0.000
		0.3	0.000	0.000	0.000	0.063	0.000	0.166	0.000
100	100	0.1	0.000	0.003	0.000	0.058	0.000	0.138	0.003
		0.2	0.000	0.000	0.000	0.055	0.000	0.094	0.000
		0.3	0.000	0.000	0.000	0.033	0.000	0.070	0.000
	200	0.1	0.000	0.000	0.000	0.072	0.000	0.218	0.000
		0.2	0.000	0.000	0.000	0.074	0.000	0.141	0.000
		0.3	0.000	0.000	0.000	0.056	0.000	0.094	0.000
	300	0.1	0.000	0.000	0.000	0.087	0.000	0.247	0.000
		0.2	0.000	0.000	0.000	0.101	0.000	0.168	0.000
		0.3	0.000	0.000	0.000	0.053	0.000	0.104	0.000

Table A.15: F Score for Normal distribution (θ_2 = 5 and ξ_2 = 1)

n	*p*	*ε*	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	1.000	1.000	1.000	0.679	0.980	0.754	0.876	0.979
		0.2	0.999	0.999	0.999	0.825	0.987	0.893	0.932	0.998
		0.3	1.000	1.000	1.000	0.898	0.995	0.961	0.976	0.956
	10	0.1	1.000	1.000	1.000	1.000	0.982	0.722	0.732	0.989
		0.2	1.000	1.000	1.000	0.998	0.997	0.898	0.898	1.000
		0.3	1.000	1.000	1.000	0.999	0.999	0.963	0.951	1.000
	20	0.1	1.000	1.000	1.000	0.813	0.994	0.717	0.523	0.987
		0.2	1.000	1.000	1.000	0.959	0.998	0.884	0.757	1.000
		0.3	1.000	1.000	1.000	0.983	1.000	0.960	0.362	1.000
	30	0.1	1.000	1.000	1.000	0.849	1.000	0.727	0.528	1.000
		0.2	1.000	1.000	1.000	0.947	0.999	0.875	0.390	1.000
		0.3	1.000	0.999	0.999	0.985	1.000	0.964	0.313	1.000
500	5	0.1	1.000	1.000	1.000	0.634	0.981	0.759	0.904	0.973
		0.2	1.000	1.000	1.000	0.810	0.995	0.903	0.954	0.999
		0.3	1.000	1.000	1.000	0.883	0.997	0.960	0.974	0.975
	10	0.1	1.000	1.000	1.000	0.999	0.991	0.741	0.881	0.986
		0.2	1.000	1.000	1.000	1.000	0.998	0.907	0.958	1.000
		0.3	1.000	1.000	1.000	0.999	0.999	0.968	0.979	0.935
	20	0.1	1.000	1.000	1.000	0.844	0.997	0.719	0.860	0.992
		0.2	1.000	1.000	1.000	0.959	1.000	0.905	0.950	1.000
		0.3	1.000	1.000	1.000	0.990	1.000	0.969	0.589	0.617
	30	0.1	1.000	1.000	1.000	0.852	0.998	0.741	0.818	0.995
		0.2	1.000	1.000	1.000	0.952	1.000	0.900	0.886	1.000
		0.3	1.000	1.000	1.000	0.986	1.000	0.967	0.119	0.242
1000	5	0.1	1.000	1.000	1.000	0.620	0.987	0.748	0.904	0.975
		0.2	1.000	1.000	1.000	0.806	0.997	0.913	0.959	0.999
		0.3	1.000	1.000	1.000	0.883	0.998	0.966	0.977	0.954
	10	0.1	1.000	1.000	1.000	0.999	0.993	0.744	0.898	0.986
		0.2	1.000	1.000	1.000	1.000	0.999	0.904	0.959	1.000
		0.3	1.000	1.000	1.000	1.000	1.000	0.968	0.976	0.858
	20	0.1	1.000	1.000	1.000	0.857	0.999	0.739	0.886	0.994
		0.2	1.000	1.000	1.000	0.953	0.999	0.898	0.950	1.000
		0.3	1.000	1.000	1.000	0.988	1.000	0.968	0.367	0.304
	30	0.1	1.000	1.000	1.000	0.856	0.999	0.737	0.868	0.996
		0.2	1.000	1.000	1.000	0.959	1.000	0.902	0.951	1.000
		0.3	0.893	1.000	1.000	0.986	1.000	0.968	0.079	0.060

Table A.16: F Score for multivariate t distribution (θ_2 = 5 and ξ_2 = 1)

n	p	ε	C_{Sn}	C_{MAD}	C_{Qn}	S_n	ROC	OGK	MCD	RMD-S
100	5	0.1	0.828	0.828	0.828	0.492	0.656	0.517	0.520	0.655
		0.2	0.921	0.920	0.921	0.703	0.863	0.752	0.730	0.884
		0.3	0.913	0.916	0.938	0.802	0.698	0.756	0.848	0.913
	10	0.1	0.717	0.717	0.717	0.690	0.613	0.465	0.416	0.574
		0.2	0.854	0.854	0.854	0.821	0.809	0.704	0.655	0.823
		0.3	0.852	0.909	0.931	0.881	0.837	0.814	0.627	0.927
	20	0.1	0.603	0.603	0.603	0.463	0.568	0.438	0.408	0.527
		0.2	0.786	0.786	0.786	0.701	0.784	0.676	0.351	0.815
		0.3	0.755	0.823	0.853	0.842	0.867	0.797	0.363	0.919
	30	0.1	0.476	0.476	0.476	0.421	0.545	0.408	0.282	0.497
		0.2	0.708	0.708	0.708	0.687	0.775	0.664	0.268	0.793
		0.3	0.517	0.637	0.833	0.833	0.884	0.795	0.338	0.822
500	5	0.1	0.826	0.826	0.826	0.480	0.667	0.510	0.515	0.642
		0.2	0.931	0.930	0.930	0.683	0.855	0.745	0.722	0.869
		0.3	0.435	0.635	0.782	0.799	0.604	0.746	0.835	0.733
	10	0.1	0.770	0.770	0.770	0.673	0.614	0.461	0.423	0.577
		0.2	0.892	0.894	0.892	0.828	0.822	0.702	0.642	0.832
		0.3	0.286	0.400	0.438	0.889	0.845	0.807	0.754	0.633
	20	0.1	0.688	0.688	0.688	0.454	0.559	0.424	0.355	0.520
		0.2	0.852	0.853	0.852	0.690	0.786	0.671	0.521	0.797
		0.3	0.140	0.170	0.213	0.840	0.892	0.821	0.366	0.402
	30	0.1	0.641	0.641	0.641	0.426	0.544	0.405	0.331	0.498
		0.2	0.821	0.824	0.823	0.675	0.770	0.654	0.284	0.776
		0.3	0.201	0.200	0.196	0.828	0.889	0.811	0.388	0.351
1000	5	0.1	0.830	0.830	0.830	0.475	0.666	0.506	0.511	0.637
		0.2	0.933	0.934	0.932	0.680	0.856	0.747	0.720	0.868
		0.3	0.235	0.405	0.598	0.793	0.614	0.742	0.832	0.674
	10	0.1	0.773	0.773	0.773	0.673	0.612	0.458	0.420	0.576
		0.2	0.894	0.894	0.894	0.822	0.816	0.701	0.635	0.827
		0.3	0.258	0.289	0.336	0.888	0.853	0.793	0.747	0.542
	20	0.1	0.705	0.705	0.706	0.450	0.567	0.423	0.349	0.523
		0.2	0.854	0.854	0.854	0.691	0.787	0.667	0.541	0.794
		0.3	0.124	0.129	0.153	0.842	0.895	0.823	0.368	0.337
	30	0.1	0.661	0.661	0.660	0.428	0.544	0.406	0.318	0.495
		0.2	0.790	0.827	0.715	0.671	0.765	0.650	0.290	0.771
		0.3	0.252	0.250	0.248	0.826	0.885	0.811	0.383	0.388

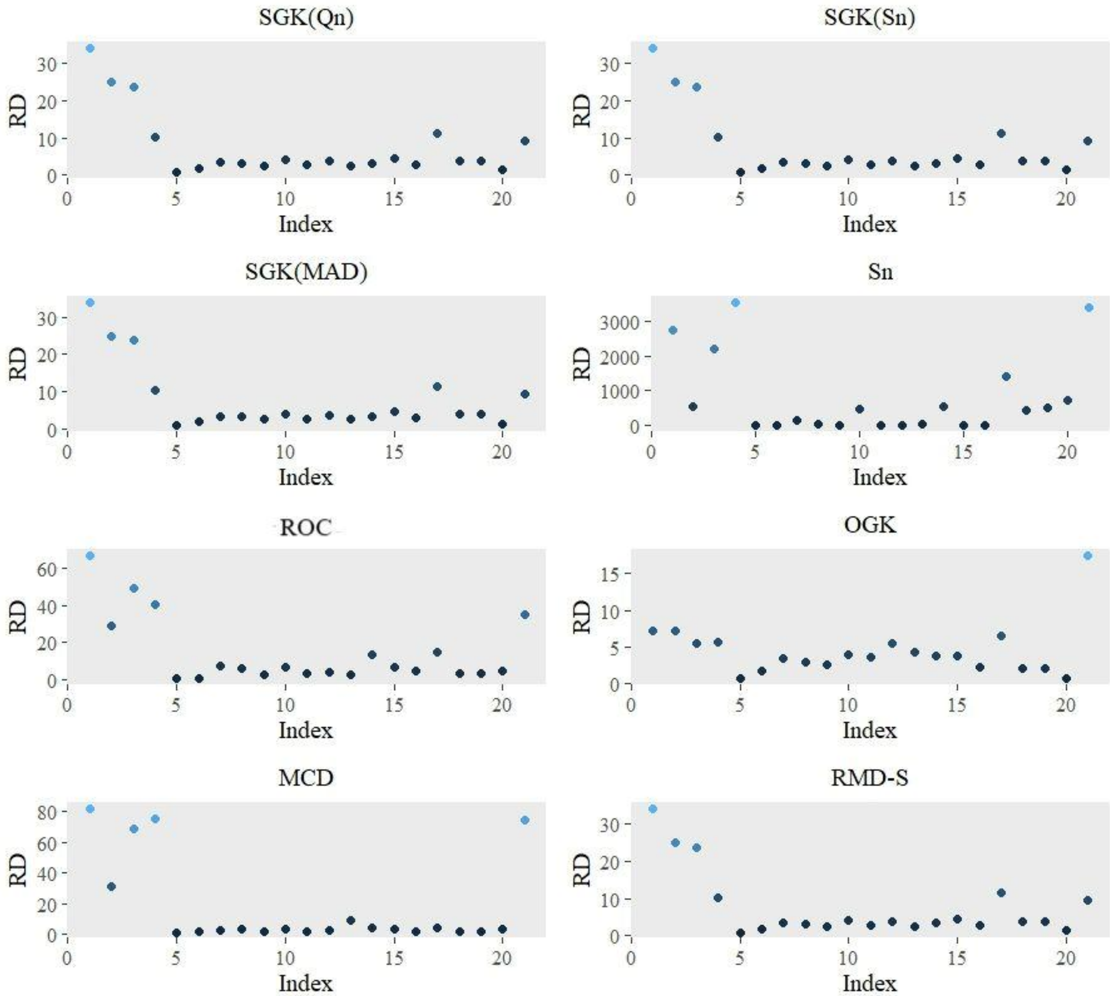


Figure A.1: Robust Mahalanobis distance plots for *Stack Loss* data

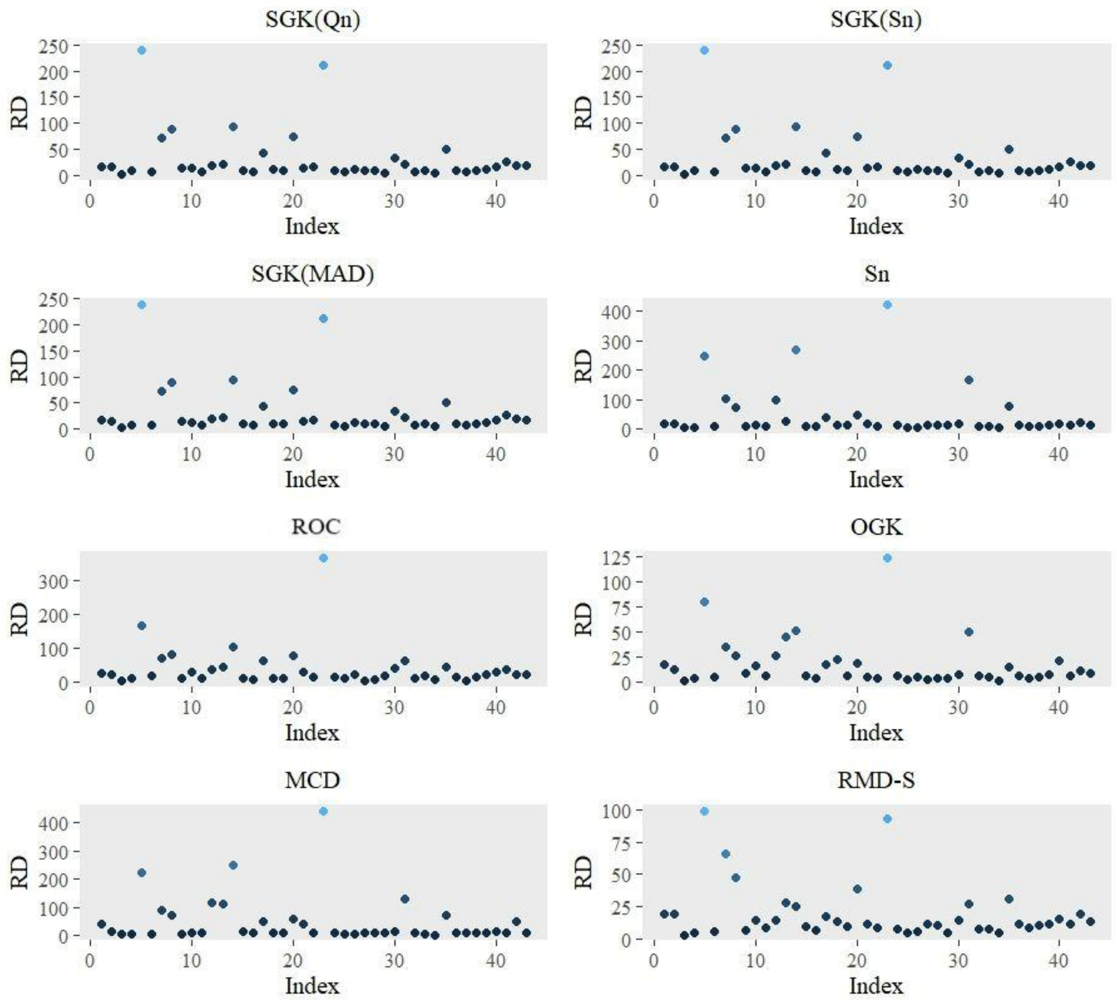


Figure A.2: Robust Mahalanobis distance plots for *USJudgeRatings* data

APPENDIX B

Table B.1: SR (FAR) in Phase II when there are case A outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, 3)^T$

n	***ε***	**S_n Covariance**	**Comedian**	**GK**
5	0	1 (0.024)	1 (0.022)	1 (0.025)
	0.1	1 (0.011)	1 (0.014)	1 (0.009)
	0.2	1 (0.003)	1 (0.004)	1 (0.002)
	0.3	1 (0.002)	1 (0.002)	1 (0.001)
10	0	1 (0.03)	1 (0.031)	1 (0.032)
	0.1	1 (0.002)	1 (0.01)	1 (0.003)
	0.2	1 (0.001)	1 (0.002)	1 (0.002)
	0.3	1 (0.003)	1 (0.004)	1 (0.006)
20	0	1 (0.032)	1 (0.033)	1 (0.031)
	0.1	1 (0.009)	1 (0.014)	1 (0.01)
	0.2	1 (0.002)	1 (0.004)	1 (0.005)
	0.3	1 (0.001)	1 (0.001)	1 (0.001)
50	0	1 (0.037)	1 (0.039)	1 (0.038)
	0.1	1 (0.007)	1 (0.009)	1 (0.008)
	0.2	1 (0.002)	1 (0.003)	1 (0.002)
	0.3	1 (0)	1 (0)	1 (0)
100	0	1 (0.027)	1 (0.026)	1 (0.027)
	0.1	1 (0.007)	1 (0.009)	1 (0.009)
	0.2	1 (0)	1 (0.002)	1 (0.001)
	0.3	1 (0.001)	1 (0.001)	1 (0.001)

Table B.2: SR (FAR) in Phase II when there are case A outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, 0)^T$

n	***ε***	**S_n Covariance**	**Comedian**	**GK**
5	0	1 (0.032)	1 (0.034)	1 (0.036)
	0.1	1 (0.026)	1 (0.028)	1 (0.027)
	0.2	1 (0.01)	1 (0.013)	1 (0.012)
	0.3	1 (0.013)	1 (0.017)	1 (0.014)
10	0	1 (0.029)	1 (0.032)	1 (0.031)
	0.1	1 (0.015)	1 (0.017)	1 (0.016)
	0.2	1 (0.014)	1 (0.019)	1 (0.014)
	0.3	1 (0.019)	1 (0.022)	1 (0.021)
20	0	1 (0.025)	1 (0.035)	1 (0.029)
	0.1	1 (0.019)	1 (0.021)	1 (0.018)
	0.2	1 (0.014)	1 (0.013)	1 (0.014)
	0.3	1 (0.016)	1 (0.015)	1 (0.017)
50	0	1 (0.03)	1 (0.032)	1 (0.03)
	0.1	1 (0.021)	1 (0.027)	1 (0.019)
	0.2	1 (0.014)	1 (0.015)	1 (0.016)
	0.3	1 (0.008)	1 (0.008)	1 (0.009)
100	0	1 (0.034)	1 (0.034)	1 (0.034)
	0.1	1 (0.02)	1 (0.021)	1 (0.018)
	0.2	1 (0.023)	1 (0.021)	1 (0.02)
	0.3	1 (0.017)	1 (0.017)	1 (0.018)

Table B.3: SR (FAR) in Phase II when there are case A outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, -3)^T$

n	***ε***	***S_n* Covariance**	**Comedian**	**GK**
5	0	1 (0.024)	1 (0.031)	1 (0.03)
	0.1	1 (0.011)	1 (0.018)	1 (0.012)
	0.2	1 (0.004)	1 (0.009)	1 (0.004)
	0.3	1 (0)	1 (0.003)	1 (0.002)
10	0	1 (0.032)	1 (0.037)	1 (0.042)
	0.1	1 (0.012)	1 (0.017)	1 (0.012)
	0.2	1 (0.003)	1 (0.003)	1 (0.003)
	0.3	1 (0.001)	1 (0.002)	1 (0.001)
20	0	1 (0.032)	1 (0.028)	1 (0.031)
	0.1	1 (0.008)	1 (0.015)	1 (0.011)
	0.2	1 (0.004)	1 (0.006)	1 (0.004)
	0.3	1 (0.001)	1 (0.004)	1 (0.002)
50	0	1 (0.034)	1 (0.037)	1 (0.033)
	0.1	1 (0.01)	1 (0.012)	1 (0.008)
	0.2	1 (0.003)	1 (0.005)	1 (0.003)
	0.3	1 (0)	1 (0.002)	1 (0.002)
100	0	1 (0.036)	1 (0.035)	1 (0.037)
	0.1	1 (0.009)	1 (0.011)	1 (0.008)
	0.2	1 (0.002)	1 (0.004)	1 (0.003)
	0.3	1 (0.001)	1 (0.003)	1 (0.003)

Table B.4: SR (FAR) in Phase II when there are case A outliers in Phase I data with $\boldsymbol{\mu}_1 = (10, 10)^T$

n	***ε***	***S_n* Covariance**	**Comedian**	**GK**
5	0	1 (0.036)	1 (0.04)	1 (0.034)
	0.1	1 (0.005)	1 (0.008)	1 (0.005)
	0.2	1 (0)	1 (0.005)	1 (0)
	0.3	1 (0)	1 (0)	1 (0)
10	0	1 (0.025)	1 (0.026)	1 (0.026)
	0.1	1 (0.005)	1 (0.012)	1 (0.004)
	0.2	1 (0)	1 (0.001)	1 (0)
	0.3	1 (0)	1 (0)	1 (0)
20	0	1 (0.035)	1 (0.037)	1 (0.035)
	0.1	1 (0.007)	1 (0.011)	1 (0.007)
	0.2	1 (0.002)	1 (0.005)	1 (0.002)
	0.3	1 (0)	1 (0)	1 (0)
50	0	1 (0.031)	1 (0.028)	1 (0.029)
	0.1	1 (0.005)	1 (0.011)	1 (0.006)
	0.2	1 (0)	1 (0.001)	1 (0.001)
	0.3	1 (0)	1 (0)	1 (0)
100	0	1 (0.025)	1 (0.024)	1 (0.024)
	0.1	1 (0.011)	1 (0.013)	1 (0.008)
	0.2	1 (0.002)	1 (0.003)	1 (0.003)
	0.3	1 (0)	1 (0)	1 (0)

Table B.5: Success rate (false alarm rate) in Phase II when there are case B outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, 3)^T$

n	***ε***	***S_n* Covariance**	**Comedian**	**GK**
5	0	0.997 (0.037)	0.95 (0.056)	0.999 (0.047)
	0.1	0.981 (0.015)	0.89 (0.031)	0.993 (0.032)
	0.2	0.951 (0.021)	0.853 (0.027)	0.978 (0.032)
	0.3	0.903 (0.014)	0.781 (0.02)	0.964 (0.037)
10	0	1 (0.018)	0.982 (0.036)	1 (0.032)
	0.1	0.994 (0.008)	0.955 (0.026)	1 (0.013)
	0.2	0.982 (0.011)	0.919 (0.035)	1 (0.016)
	0.3	0.957 (0.017)	0.839 (0.022)	1 (0.01)
20	0	1 (0.028)	1 (0.039)	1 (0.043)
	0.1	1 (0.006)	0.991 (0.018)	1 (0.013)
	0.2	0.999 (0.007)	0.972 (0.019)	1 (0.013)
	0.3	0.993 (0.006)	0.923 (0.036)	1 (0.009)
50	0	1 (0.02)	1 (0.027)	1 (0.038)
	0.1	1 (0.005)	1 (0.008)	1 (0.011)
	0.2	1 (0.005)	0.999 (0.007)	1 (0.006)
	0.3	1 (0.003)	0.988 (0.018)	1 (0.006)
100	0	1 (0.024)	1 (0.023)	1 (0.031)
	0.1	1 (0.003)	1 (0.01)	1 (0.015)
	0.2	1 (0.001)	1 (0.003)	1 (0.008)
	0.3	1 (0.003)	0.996 (0.013)	1 (0.011)

Table B.6: Success rate (false alarm rate) in Phase II when there are case B outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, 0)^T$

n	***ε***	***S_n* Covariance**	**Comedian**	**GK**
5	0	0.995 (0.033)	0.935 (0.044)	0.998 (0.054)
	0	1 (0.016)	0.988 (0.025)	0.974 (0.044)
	0.1	1 (0.013)	0.998 (0.015)	0.981 (0.048)
	0.2	1 (0.012)	1 (0.012)	0.992 (0.036)
10	0.3	0.999 (0.022)	0.978 (0.045)	1 (0.036)
	0	1 (0.007)	1 (0.006)	0.998 (0.028)
	0.1	1 (0.013)	1 (0.017)	0.999 (0.03)
	0.2	1 (0.01)	1 (0.01)	1 (0.024)
20	0.3	1 (0.016)	0.996 (0.025)	1 (0.035)
	0	1 (0.013)	1 (0.015)	1 (0.024)
	0.1	1 (0.01)	1 (0.008)	1 (0.03)
	0.2	1 (0.013)	1 (0.013)	1 (0.026)
50	0.3	1 (0.029)	1 (0.033)	1 (0.048)
	0	1 (0.014)	1 (0.015)	1 (0.026)
	0.1	1 (0.01)	1 (0.011)	1 (0.027)
	0.2	1 (0.014)	1 (0.013)	1 (0.022)
100	0.3	1 (0.012)	1 (0.014)	1 (0.025)
	0	1 (0.006)	1 (0.007)	1 (0.015)
	0.1	1 (0.013)	1 (0.012)	1 (0.023)
	0.2	1 (0.021)	1 (0.019)	1 (0.026)

Table B.7: Success rate (false alarm rate) in Phase II when there are case B outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, -3)^T$

n	***ε***	**S_n Covariance**	**Comedian**	**GK**
5	0	0.993 (0.038)	0.921 (0.066)	0.995 (0.054)
	0.1	1 (0.007)	1 (0.009)	1 (0.005)
	0.2	1 (0.007)	1 (0.014)	1 (0.002)
	0.3	1 (0)	1 (0.001)	1 (0)
10	0	1 (0.024)	0.978 (0.034)	1 (0.034)
	0.1	1 (0.007)	1 (0.008)	1 (0.001)
	0.2	1 (0.003)	1 (0.007)	1 (0.001)
	0.3	1 (0)	1 (0.001)	1 (0)
20	0	1 (0.023)	0.996 (0.031)	1 (0.044)
	0.1	1 (0.008)	1 (0.011)	1 (0.004)
	0.2	1 (0.001)	1 (0.006)	1 (0.001)
	0.3	1 (0.001)	1 (0.003)	1 (0.001)
50	0	1 (0.019)	1 (0.023)	1 (0.036)
	0.1	1 (0.006)	1 (0.008)	1 (0.003)
	0.2	1 (0.002)	1 (0.006)	1 (0)
	0.3	1 (0)	1 (0)	1 (0)
100	0	1 (0.022)	1 (0.022)	1 (0.03)
	0.1	1 (0.009)	1 (0.008)	1 (0.003)
	0.2	1 (0.002)	1 (0.003)	1 (0.001)
	0.3	1 (0.002)	1 (0.002)	1 (0.002)

Table B.8: Success rate (false alarm rate) in Phase II when there are case B outliers in Phase I data with $\boldsymbol{\mu}_1 = (10, 10)^T$

n	***ε***	**S_n Covariance**	**Comedian**	**GK**
5	0	0.994 (0.026)	0.933 (0.05)	0.997 (0.05)
	0.1	0.965 (0.025)	0.906 (0.043)	0.993 (0.023)
	0.2	0.944 (0.017)	0.871 (0.025)	0.994 (0.016)
	0.3	0.883 (0.009)	0.791 (0.014)	0.994 (0.008)
10	0	1 (0.023)	0.973 (0.04)	1 (0.044)
	0.1	0.991 (0.022)	0.964 (0.031)	1 (0.012)
	0.2	0.984 (0.011)	0.929 (0.018)	1 (0.001)
	0.3	0.944 (0.007)	0.876 (0.027)	1 (0.002)
20	0	1 (0.027)	0.997 (0.035)	1 (0.042)
	0.1	1 (0.005)	0.994 (0.01)	1 (0.011)
	0.2	0.998 (0.001)	0.976 (0.017)	1 (0.005)
	0.3	0.99 (0.005)	0.934 (0.022)	1 (0.004)
50	0	1 (0.019)	1 (0.022)	1 (0.029)
	0.1	1 (0.008)	1 (0.009)	1 (0.01)
	0.2	1 (0.001)	0.999 (0.007)	1 (0.001)
	0.3	1 (0)	0.992 (0.014)	1 (0)
100	0	1 (0.017)	1 (0.02)	1 (0.028)
	0.1	1 (0.004)	1 (0.008)	1 (0.007)
	0.2	1 (0)	1 (0.001)	1 (0.001)
	0.3	1 (0)	1 (0.004)	1 (0)

Table B.9: Success rate (false alarm rate) in Phase II when there are case C outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, 3)^T$

n	ε	S_n Covariance	Comedian	GK
5	0	0.991 (0.035)	0.931 (0.049)	0.999 (0.063)
	0.1	1 (0.016)	1 (0.017)	1 (0.01)
	0.2	1 (0.004)	1 (0.01)	1 (0.003)
	0.3	1 (0)	1 (0.004)	1 (0.005)
10	0	0.999 (0.026)	0.98 (0.045)	1 (0.039)
	0.1	1 (0.01)	1 (0.015)	1 (0.006)
	0.2	1 (0.004)	1 (0.006)	1 (0.004)
	0.3	1 (0)	1 (0.002)	1 (0.002)
20	0	1 (0.022)	0.998 (0.035)	1 (0.033)
	0.1	1 (0.012)	1 (0.013)	1 (0.009)
	0.2	1 (0.004)	1 (0.01)	1 (0.004)
	0.3	1 (0)	1 (0)	1 (0.003)
50	0	1 (0.021)	1 (0.022)	1 (0.032)
	0.1	1 (0.01)	1 (0.011)	1 (0.005)
	0.2	1 (0.003)	1 (0.008)	1 (0.004)
	0.3	1 (0)	1 (0.002)	1 (0.004)
100	0	1 (0.007)	1 (0.012)	1 (0.025)
	0.1	1 (0.008)	1 (0.008)	1 (0.006)
	0.2	1 (0.002)	1 (0.004)	1 (0.002)
	0.3	1 (0.002)	1 (0.002)	1 (0.006)

Table B.10: Success rate (false alarm rate) in Phase II when there are case C outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, 3)^T$

n	ε	S_n Covariance	Comedian	GK
5	0	0.996 (0.026)	0.943 (0.04)	0.997 (0.046)
	0.1	1 (0.024)	0.965 (0.045)	0.863 (0.14)
	0.2	1 (0.029)	0.995 (0.06)	0.878 (0.118)
	0.3	1 (0.066)	1 (0.091)	0.978 (0.131)
10	0	1 (0.02)	0.98 (0.043)	1 (0.043)
	0.1	1 (0.031)	0.997 (0.044)	0.947 (0.135)
	0.2	1 (0.045)	1 (0.063)	0.957 (0.15)
	0.3	1 (0.072)	1 (0.094)	0.99 (0.108)
20	0	1 (0.02)	0.995 (0.03)	1 (0.029)
	0.1	1 (0.013)	1 (0.026)	0.996 (0.138)
	0.2	1 (0.037)	1 (0.053)	0.983 (0.132)
	0.3	1 (0.064)	1 (0.084)	1 (0.092)
50	0	1 (0.017)	1 (0.015)	1 (0.027)
	0.1	1 (0.02)	1 (0.028)	1 (0.101)
	0.2	1 (0.04)	1 (0.058)	1 (0.134)
	0.3	1 (0.055)	1 (0.078)	1 (0.079)
100	0	1 (0.025)	1 (0.024)	1 (0.038)
	0.1	1 (0.026)	1 (0.032)	1 (0.113)
	0.2	1 (0.035)	1 (0.05)	1 (0.104)
	0.3	1 (0.084)	1 (0.109)	1 (0.108)

Table B.11: Success rate (false alarm rate) in Phase II when there are case C outliers in Phase I data with $\boldsymbol{\mu}_1 = (3, 3)^T$

n	***ε***	**S_n Covariance**	**Comedian**	**GK**
5	0	0.991 (0.035)	0.931 (0.049)	0.999 (0.063)
	0.1	1 (0.016)	1 (0.017)	1 (0.01)
	0.2	1 (0.004)	1 (0.01)	1 (0.003)
	0.3	1 (0)	1 (0.004)	1 (0.005)
10	0	0.999 (0.026)	0.98 (0.045)	1 (0.039)
	0.1	1 (0.01)	1 (0.015)	1 (0.006)
	0.2	1 (0.004)	1 (0.006)	1 (0.004)
	0.3	1 (0)	1 (0.002)	1 (0.002)
20	0	1 (0.022)	0.998 (0.035)	1 (0.033)
	0.1	1 (0.012)	1 (0.013)	1 (0.009)
	0.2	1 (0.004)	1 (0.01)	1 (0.004)
	0.3	1 (0)	1 (0)	1 (0.003)
50	0	1 (0.021)	1 (0.022)	1 (0.032)
	0.1	1 (0.01)	1 (0.011)	1 (0.005)
	0.2	1 (0.003)	1 (0.008)	1 (0.004)
	0.3	1 (0)	1 (0.002)	1 (0.004)
100	0	1 (0.007)	1 (0.012)	1 (0.025)
	0.1	1 (0.008)	1 (0.008)	1 (0.006)
	0.2	1 (0.002)	1 (0.004)	1 (0.002)
	0.3	1 (0.002)	1 (0.002)	1 (0.006)

Table B.12: Success rate (false alarm rate) in Phase II when there are case C outliers in Phase I data with $\boldsymbol{\mu}_1 = (10, 10)^T$

n	***ε***	**S_n Covariance**	**Comedian**	**GK**
5	0	0.993 (0.034)	0.932 (0.046)	1 (0.051)
	0.1	0.967 (0.023)	0.921 (0.035)	0.998 (0.023)
	0.2	0.932 (0.021)	0.877 (0.027)	0.999 (0.002)
	0.3	0.898 (0.015)	0.807 (0.021)	1 (0.001)
10	0	1 (0.019)	0.977 (0.037)	1 (0.031)
	0.1	0.998 (0.013)	0.962 (0.036)	1 (0.01)
	0.2	0.979 (0.007)	0.936 (0.012)	1 (0.002)
	0.3	0.95 (0.006)	0.861 (0.018)	1 (0.001)
20	0	1 (0.016)	0.996 (0.028)	1 (0.029)
	0.1	1 (0.009)	0.997 (0.029)	1 (0.02)
	0.2	0.997 (0.002)	0.988 (0.018)	1 (0)
	0.3	0.992 (0.005)	0.928 (0.013)	1 (0.002)
50	0	1 (0.014)	1 (0.017)	1 (0.029)
	0.1	1 (0.008)	1 (0.011)	1 (0.011)
	0.2	1 (0.001)	1 (0.002)	1 (0.001)
	0.3	1 (0.003)	0.993 (0.012)	1 (0.001)
100	0	1 (0.017)	1 (0.022)	1 (0.036)
	0.1	1 (0.004)	1 (0.007)	1 (0.009)
	0.2	1 (0)	1 (0.004)	1 (0.002)
	0.3	1 (0.001)	1 (0.003)	1 (0.001)

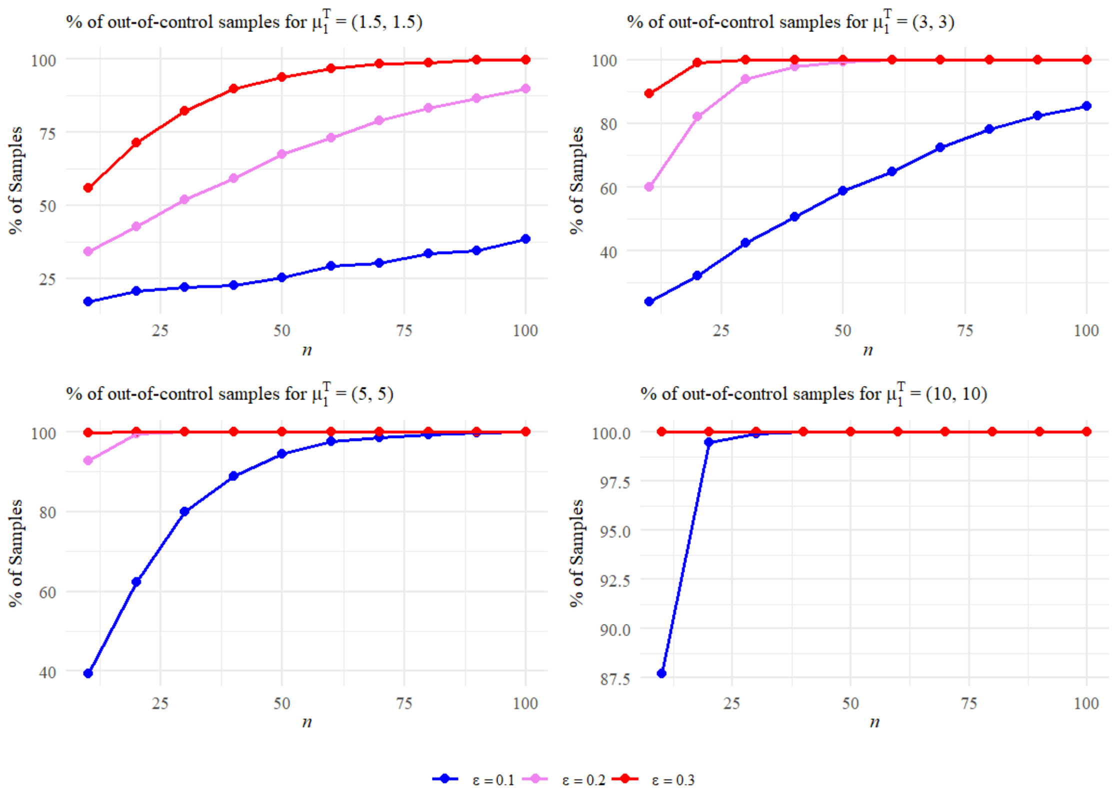


Figure B.1: Average percentage of out-of-control samples under Case C contamination

APPENDIX C

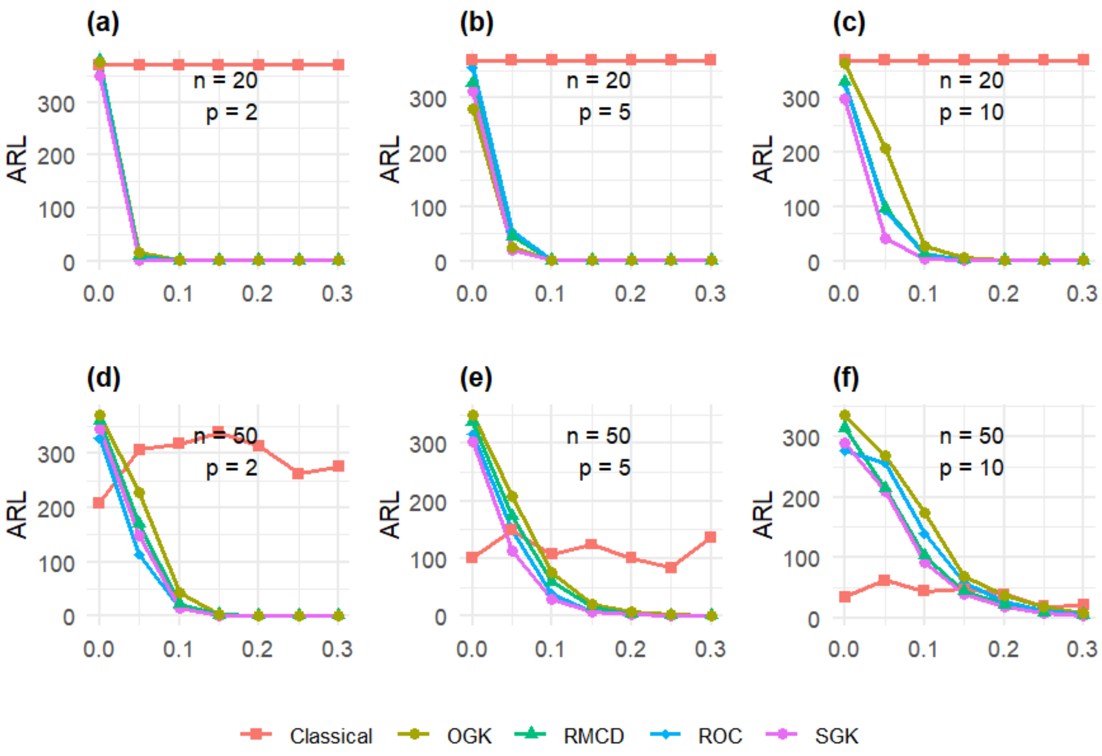


Figure C.1: ARL plots of multivariate control charts based on ROC, SGK, RMCD, OGK and Classical estimators under Case C contamination for $\varepsilon = 0.1$ and $\delta^2 = 150$.

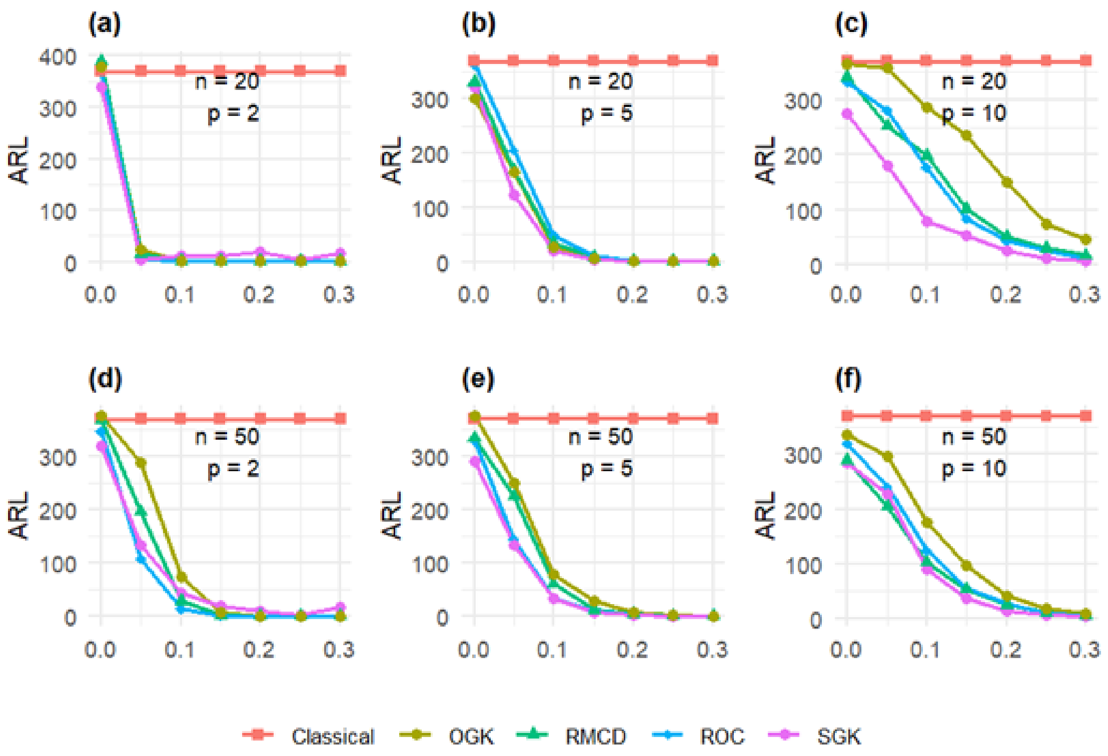


Figure C.2: ARL plots of multivariate control charts based on ROC, SGK, RMCD, OGK and Classical estimators under Case C contamination for $\varepsilon = 0.2$ and $\delta^2 = 150$.

APPENDIX D

Table D.1: Upper control limits for MEWMS statistics (Comedian control chart)

n	*ω*	$MEWMSL_1$			$MEWMSL_2$			MEWMS		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.1	2.04	3.99	6.36	2.83	11.80	31.48	3.67	5.44	7.82
	0.2	2.25	3.85	5.72	3.50	11.00	25.34	3.90	5.46	7.25
	0.3	2.60	4.31	6.11	4.75	14.08	29.03	4.20	5.93	7.69
	0.4	2.97	4.97	6.91	6.26	19.08	38.12	4.51	6.57	8.53
	0.5	3.38	5.71	8.02	8.24	25.29	51.99	4.87	7.25	9.54
	0.6	3.78	6.52	9.10	10.39	33.30	68.64	5.20	8.02	10.65
	0.7	4.17	7.33	10.31	12.78	42.56	88.41	5.52	8.75	11.76
	0.8	4.67	8.20	11.54	16.14	54.14	109.93	5.94	9.57	12.96
	0.9	5.10	9.06	12.81	19.37	66.04	136.58	6.30	10.36	14.16
100	0.1	1.55	2.77	4.76	1.51	5.41	16.85	3.20	4.31	6.14
	0.2	1.87	2.97	4.20	2.34	6.40	13.29	3.55	4.66	5.77
	0.3	2.25	3.56	4.81	3.47	9.33	17.85	3.87	5.22	6.44
	0.4	2.64	4.23	5.73	4.82	13.49	25.57	4.18	5.85	7.35
	0.5	3.02	4.92	6.63	6.43	18.53	34.64	4.51	6.50	8.21
	0.6	3.41	5.65	7.65	8.31	24.73	47.09	4.83	7.17	9.20
	0.7	3.84	6.43	8.74	10.64	32.32	61.81	5.20	7.88	10.22
	0.8	4.27	7.19	9.78	13.27	40.80	78.32	5.56	8.59	11.22
	0.9	4.75	8.04	11.02	16.48	51.44	98.60	5.95	9.40	12.37

Table D2: Upper control limits for MEWMS statistics (S_n Covariance)

n	*ω*	MEWMSL1			MEWMSL2			MEWMS		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.1	1.73	3.52	5.72	1.81	8.90	25.33	3.23	4.90	7.06
	0.2	1.87	3.19	4.93	2.27	7.35	18.45	3.46	4.73	6.34
	0.3	2.18	3.63	5.13	3.23	9.72	20.51	3.76	5.22	6.64
	0.4	2.56	4.20	5.77	4.51	13.25	26.37	4.07	5.74	7.34
	0.5	2.93	4.89	6.74	6.01	18.43	36.14	4.38	6.41	8.25
	0.6	3.30	5.59	7.73	7.73	24.24	48.03	4.69	7.04	9.19
	0.7	3.69	6.31	8.67	9.71	31.05	60.51	5.01	7.72	10.14
	0.8	4.15	7.15	9.97	12.52	40.62	81.82	5.40	8.51	11.33
	0.9	4.57	7.92	10.94	15.23	49.96	97.56	5.75	9.18	12.23
100	0.1	1.39	2.52	4.18	1.19	4.28	12.69	2.99	3.98	5.55
	0.2	1.69	2.62	3.72	1.81	4.86	10.21	3.32	4.29	5.30
	0.3	2.06	3.22	4.38	2.82	7.54	14.42	3.66	4.88	6.02
	0.4	2.43	3.86	5.16	4.04	11.02	20.47	3.97	5.46	6.75
	0.5	2.82	4.53	6.04	5.50	15.43	28.61	4.29	6.08	7.61
	0.6	3.20	5.24	7.01	7.20	21.07	39.33	4.61	6.72	8.54
	0.7	3.61	5.96	8.00	9.26	27.53	51.86	4.95	7.39	9.51
	0.8	4.02	6.73	9.10	11.60	35.41	67.30	5.29	8.10	10.53
	0.9	4.47	7.48	10.11	14.49	44.07	83.20	5.67	8.79	11.48

Table D3: Upper control limits for MEWMS statistics (Classical)

n	ω	$MEWMSL_1$			$MEWMSL_2$			MEWMS		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.1	1.60	3.15	4.80	1.45	6.93	17.18	2.94	4.48	6.15
	0.2	1.58	2.58	4.16	1.52	4.45	12.62	3.14	4.07	5.48
	0.3	1.88	2.85	3.95	2.25	5.68	11.19	3.45	4.44	5.43
	0.4	2.22	3.39	4.44	3.29	8.27	14.72	3.75	4.96	5.95
	0.5	2.59	3.95	5.11	4.54	11.49	19.99	4.04	5.48	6.65
	0.6	2.96	4.59	5.93	6.03	15.77	27.41	4.36	6.07	7.40
	0.7	3.34	5.22	6.71	7.88	20.63	35.34	4.68	6.64	8.17
	0.8	3.76	5.88	7.54	10.00	26.60	45.32	5.02	7.24	8.95
	0.9	4.20	6.60	8.44	12.62	33.84	57.32	5.39	7.92	9.84
100	0.1	1.26	2.36	4.04	0.95	3.66	11.82	2.83	3.75	5.35
	0.2	1.55	2.33	3.38	1.50	3.70	8.11	3.18	3.99	4.88
	0.3	1.91	2.86	3.74	2.39	5.79	10.22	3.51	4.51	5.38
	0.4	2.28	3.47	4.48	3.50	8.77	15.19	3.83	5.09	6.09
	0.5	2.66	4.10	5.28	4.85	12.48	21.51	4.14	5.66	6.86
	0.6	3.04	4.79	6.17	6.44	17.26	29.92	4.45	6.28	7.71
	0.7	3.44	5.47	7.10	8.33	22.89	39.84	4.78	6.90	8.60
	0.8	3.85	6.15	7.97	10.56	29.11	50.48	5.12	7.53	9.40
	0.9	4.31	6.88	8.97	13.29	36.69	64.69	5.50	8.20	10.34

Table D.4: Upper control limits for MEWMS statistics (MCD control chart)

n	ω	$MEWMSL_1$			$MEWMSL_2$			MEWMS		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.1	2.93	5.88	9.50	6.21	26.02	70.20	4.67	7.59	11.16
	0.2	3.10	5.57	8.61	6.90	23.20	55.97	4.87	7.34	10.35
	0.3	3.55	6.40	9.61	9.14	31.00	71.63	5.28	8.17	11.36
	0.4	3.97	7.17	10.62	11.46	39.13	86.29	5.65	8.93	12.37
	0.5	4.42	8.13	11.96	14.33	50.90	111.38	6.04	9.84	13.75
	0.6	4.90	9.19	13.63	17.72	65.26	147.77	6.49	10.86	15.35
	0.7	5.33	10.16	15.09	21.06	79.92	178.07	6.84	11.79	16.78
	0.8	5.92	11.46	16.88	26.17	101.92	225.69	7.37	13.03	18.63
	0.9	6.45	12.47	18.59	31.07	121.12	272.82	7.84	14.00	20.13
100	0.1	1.80	3.22	5.29	2.17	7.44	20.88	3.55	4.87	6.76
	0.2	2.15	3.46	4.89	3.20	8.81	18.29	3.90	5.23	6.60
	0.3	2.59	4.22	5.80	4.69	13.13	25.37	4.28	5.95	7.54
	0.4	3.00	4.97	6.83	6.38	18.54	36.17	4.63	6.66	8.54
	0.5	3.41	5.75	7.92	8.32	25.29	48.72	4.99	7.42	9.60
	0.6	3.81	6.53	9.02	10.51	33.04	63.66	5.33	8.13	10.68
	0.7	4.29	7.50	10.35	13.42	43.84	84.97	5.73	9.06	12.01
	0.8	4.75	8.36	11.62	16.59	54.49	107.49	6.11	9.87	13.20
	0.9	5.25	9.29	12.95	20.36	67.41	133.18	6.55	10.75	14.43

Table D.5: Upper control limits for MEWMS statistics (MVE)

n	ω	$MEWMSL_1$			$MEWMSL_2$			MEWMS		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.1	2.01	4.15	6.88	2.73	12.72	36.26	3.51	5.58	8.35
	0.2	2.20	3.89	6.04	3.30	11.20	27.68	3.76	5.44	7.57
	0.3	2.60	4.42	6.38	4.69	14.64	31.87	4.14	5.99	7.93
	0.4	2.96	5.08	7.21	6.16	19.51	40.22	4.47	6.67	8.78
	0.5	3.35	5.81	8.23	8.03	25.94	53.24	4.83	7.37	9.85
	0.6	3.75	6.64	9.44	10.06	34.04	69.83	5.18	8.15	11.00
	0.7	4.18	7.50	10.69	12.62	43.86	91.34	5.55	9.01	12.26
	0.8	4.68	8.48	12.04	16.02	56.14	115.89	6.01	9.92	13.57
	0.9	5.13	9.32	13.29	19.29	67.53	139.64	6.40	10.77	14.84
100	0.1	1.47	2.64	4.45	1.32	4.79	14.74	3.00	4.06	5.81
	0.2	1.78	2.80	4.04	2.04	5.62	12.09	3.37	4.42	5.55
	0.3	2.14	3.40	4.57	3.08	8.44	16.00	3.71	5.02	6.21
	0.4	2.52	4.07	5.45	4.36	12.33	22.85	4.04	5.65	7.06
	0.5	2.92	4.82	6.52	5.90	17.52	33.17	4.38	6.37	8.11
	0.6	3.29	5.52	7.47	7.58	23.28	43.97	4.70	7.03	8.99
	0.7	3.69	6.31	8.61	9.71	30.69	59.40	5.05	7.78	10.11
	0.8	4.12	7.10	9.71	12.18	39.15	74.78	5.41	8.53	11.20
	0.9	4.57	7.94	10.80	15.11	48.95	93.09	5.79	9.33	12.26

Table D.6: Upper control limits for MEWMS statistics (GK_Qn)

n	ω	$MEWMSL_1$			$MEWMSL_2$			MEWMS		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.1	1.71	3.49	5.67	1.74	8.75	24.72	3.17	4.86	6.96
	0.2	1.77	2.99	4.61	1.97	6.32	15.76	3.37	4.57	6.04
	0.3	2.11	3.45	4.89	2.99	8.66	18.04	3.71	5.07	6.47
	0.4	2.49	4.06	5.57	4.23	12.34	24.18	4.02	5.67	7.18
	0.5	2.85	4.70	6.47	5.65	16.76	33.22	4.32	6.25	8.03
	0.6	3.23	5.45	7.37	7.37	22.67	43.13	4.64	6.92	8.89
	0.7	3.65	6.13	8.35	9.48	29.20	55.92	4.98	7.57	9.88
	0.8	4.06	6.89	9.51	11.88	37.21	73.50	5.32	8.28	10.93
	0.9	4.51	7.70	10.68	14.71	46.55	93.42	5.69	9.03	12.04
100	0.1	1.34	2.46	4.12	1.09	4.05	12.36	2.93	3.89	5.44
	0.2	1.64	2.55	3.71	1.71	4.57	9.97	3.29	4.22	5.25
	0.3	2.01	3.13	4.18	2.68	7.03	13.11	3.62	4.79	5.83
	0.4	2.40	3.78	5.02	3.90	10.57	19.47	3.94	5.40	6.67
	0.5	2.78	4.44	5.91	5.35	14.92	27.13	4.26	6.01	7.51
	0.6	3.16	5.12	6.84	6.99	20.11	37.08	4.57	6.62	8.42
	0.7	3.57	5.89	7.90	9.06	26.80	50.09	4.91	7.36	9.39
	0.8	3.99	6.62	8.95	11.42	33.98	64.19	5.26	8.01	10.39
	0.9	4.44	7.43	9.95	14.24	43.23	79.32	5.64	8.75	11.32

Table D.7: Upper control limits for MEWMS statistics (GK__**S**$_n$)

n	ω	$MEWMSL_1$			$MEWMSL_2$			MEWMS		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.1	1.79	3.60	5.83	1.95	9.36	26.07	3.34	5.02	7.25
	0.2	1.93	3.35	5.25	2.47	8.14	21.14	3.56	4.96	6.75
	0.3	2.28	3.81	5.46	3.56	10.69	22.49	3.88	5.45	7.08
	0.4	2.63	4.40	6.14	4.80	14.57	29.45	4.18	6.02	7.75
	0.5	3.02	5.11	7.09	6.39	19.91	39.80	4.50	6.68	8.73
	0.6	3.42	5.85	8.20	8.32	26.63	53.19	4.85	7.39	9.75
	0.7	3.82	6.63	9.24	10.48	34.30	69.12	5.18	8.12	10.76
	0.8	4.24	7.43	10.33	13.09	43.25	86.72	5.52	8.85	11.84
	0.9	4.74	8.32	11.67	16.52	54.95	110.88	5.94	9.72	13.09
100	0.1	1.42	2.64	4.51	1.23	4.78	15.19	3.03	4.14	5.90
	0.2	1.71	2.70	3.91	1.90	5.14	11.22	3.37	4.39	5.51
	0.3	2.09	3.29	4.44	2.94	7.85	14.90	3.71	4.96	6.09
	0.4	2.47	3.92	5.24	4.17	11.38	21.00	4.02	5.55	6.90
	0.5	2.85	4.62	6.16	5.64	16.10	29.66	4.34	6.19	7.77
	0.6	3.24	5.31	7.14	7.39	21.43	40.39	4.66	6.82	8.71
	0.7	3.64	6.07	8.12	9.45	28.36	53.02	5.00	7.52	9.67
	0.8	4.08	6.85	9.29	12.01	36.69	69.56	5.36	8.28	10.77
	0.9	4.54	7.66	10.33	14.94	46.10	86.68	5.74	8.99	11.76

Table D.8: Upper control limits for MEWMV statistics with $\lambda = 0.1$ (Comedian)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.2	7.00	9.28	11.31	28.63	53.57	84.79	3.29	4.56	6.03
	0.3	4.14	6.12	7.90	11.67	26.72	47.31	3.57	5.05	6.48
	0.4	3.69	5.82	7.81	9.82	25.66	48.39	3.83	5.57	7.24
	0.5	3.79	6.23	8.48	10.49	30.17	57.82	4.08	6.11	7.96
	0.6	3.98	6.69	9.19	11.64	35.07	69.42	4.42	6.73	8.88
	0.7	4.22	7.30	10.14	13.16	42.45	85.42	4.69	7.43	9.96
	0.8	4.61	8.06	11.29	15.88	52.09	106.00	5.03	8.07	10.90
	0.9	5.08	8.99	12.75	19.25	65.22	135.69	5.33	8.66	11.86
100	0.2	6.86	8.44	9.70	26.21	40.93	56.26	3.01	3.94	4.87
	0.3	3.69	5.19	6.45	9.02	18.35	29.50	3.29	4.43	5.45
	0.4	3.31	4.97	6.42	7.76	18.33	31.49	3.57	4.97	6.17
	0.5	3.39	5.32	7.01	8.29	21.51	38.60	3.85	5.51	6.92
	0.6	3.61	5.85	7.80	9.45	26.37	48.73	4.11	6.08	7.79
	0.7	3.92	6.49	8.75	11.19	32.89	61.63	4.41	6.67	8.62
	0.8	4.28	7.21	9.73	13.35	40.87	77.53	4.71	7.28	9.51
	0.9	4.72	8.01	10.89	16.32	50.62	97.33	5.04	7.91	10.38

Table D.9: Upper control limits for MEWMV statistics with λ = 0.1 (S_n Covariance)

n	ω	MEWMVL$_1$			MEWMVL$_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.2	5.83	7.27	8.53	19.14	31.43	46.50	2.92	3.97	5.18
	0.3	3.41	4.91	6.25	7.72	16.84	28.73	3.17	4.35	5.44
	0.4	3.11	4.81	6.32	6.86	17.19	30.81	3.45	4.87	6.22
	0.5	3.22	5.13	6.84	7.45	20.09	37.23	3.69	5.34	6.84
	0.6	3.43	5.64	7.62	8.49	24.46	46.89	3.96	5.92	7.69
	0.7	3.77	6.32	8.55	10.24	31.37	60.56	4.26	6.48	8.53
	0.8	4.09	6.97	9.49	12.15	38.23	73.86	4.54	7.10	9.38
	0.9	4.51	7.76	10.65	14.80	47.47	94.79	4.85	7.69	10.24
100	0.2	6.22	7.44	8.42	21.08	31.11	41.09	2.83	3.63	4.41
	0.3	3.34	4.63	5.70	7.28	14.39	22.57	3.10	4.10	5.01
	0.4	3.04	4.54	5.78	6.52	14.96	24.96	3.39	4.65	5.72
	0.5	3.14	4.88	6.39	7.06	18.06	31.73	3.64	5.14	6.43
	0.6	3.36	5.36	7.16	8.08	22.09	40.51	3.91	5.67	7.15
	0.7	3.69	6.01	7.96	9.78	27.94	51.12	4.19	6.27	8.06
	0.8	4.04	6.71	8.98	11.80	35.19	65.86	4.49	6.85	8.85
	0.9	4.45	7.41	9.96	14.39	42.99	80.39	4.80	7.45	9.67

Table D.10: Upper control limits for MEWMV statistics with λ = 0.1 (Classical)

n	ω	MEWMVL$_1$			MEWMVL$_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.2	5.15	5.90	6.53	14.08	19.15	26.76	2.71	3.46	4.47
	0.3	3.00	4.00	4.83	5.73	10.38	16.01	2.97	3.82	4.59
	0.4	2.78	3.95	4.90	5.33	11.09	17.37	3.24	4.28	5.12
	0.5	2.92	4.30	5.37	5.98	13.64	22.02	3.50	4.76	5.73
	0.6	3.18	4.81	6.07	7.12	17.37	28.67	3.78	5.24	6.42
	0.7	3.48	5.39	6.88	8.62	22.24	37.93	4.05	5.76	7.08
	0.8	3.85	6.03	7.73	10.60	27.98	47.64	4.34	6.28	7.78
	0.9	4.27	6.74	8.67	13.11	34.98	60.49	4.65	6.83	8.51
100	0.2	5.88	6.75	7.39	18.30	24.39	29.84	2.73	3.43	4.17
	0.3	3.16	4.21	5.03	6.38	11.50	16.62	3.02	3.88	4.61
	0.4	2.88	4.15	5.15	5.77	12.26	19.14	3.29	4.37	5.23
	0.5	3.01	4.49	5.68	6.39	14.93	24.51	3.56	4.88	5.92
	0.6	3.26	5.01	6.43	7.55	19.00	32.37	3.83	5.38	6.61
	0.7	3.57	5.60	7.20	9.12	24.10	41.01	4.11	5.92	7.37
	0.8	3.92	6.26	8.13	11.08	30.30	53.14	4.41	6.47	8.15
	0.9	4.34	6.92	9.03	13.57	37.15	65.10	4.72	7.06	8.90

Table D.11: Upper control limits for MEWMV statistics with λ = 0.1 (MCD)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.2	9.25	13.60	17.83	52.75	121.16	216.15	4.06	6.12	8.57
	0.3	5.51	8.74	12.06	20.81	54.90	106.26	4.37	6.66	9.16
	0.4	4.90	8.27	11.82	17.22	50.81	105.97	4.69	7.35	10.09
	0.5	4.81	8.45	12.36	16.85	53.99	116.02	4.99	8.05	11.17
	0.6	5.05	9.15	13.53	18.85	64.10	137.80	5.34	8.98	12.64
	0.7	5.33	10.00	14.66	21.20	77.05	169.27	5.71	9.86	14.01
	0.8	5.66	10.77	15.92	23.96	90.19	200.16	6.00	10.57	15.07
	0.9	6.17	11.86	17.23	28.55	109.48	235.82	6.40	11.45	16.21
100	0.2	7.81	9.93	11.89	34.72	59.38	90.16	3.30	4.41	5.60
	0.3	4.24	6.05	7.65	11.80	24.96	40.86	3.62	4.98	6.26
	0.4	3.78	5.83	7.72	10.09	24.79	44.77	3.92	5.62	7.16
	0.5	3.84	6.17	8.25	10.61	28.53	52.55	4.21	6.22	8.01
	0.6	4.06	6.78	9.26	12.01	35.17	67.38	4.51	6.88	9.02
	0.7	4.34	7.44	10.17	13.83	42.75	82.32	4.81	7.53	10.00
	0.8	4.76	8.26	11.41	16.68	53.10	103.97	5.15	8.22	10.95
	0.9	5.20	9.10	12.60	20.02	64.99	127.41	5.51	8.96	12.08

`

Table D.12: Upper control limits for MEWMV statistics with λ = 0.1 (MVE)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.2	6.26	8.35	10.41	24.53	47.65	77.45	3.13	4.46	6.07
	0.3	3.80	5.71	7.57	10.24	24.25	43.74	3.43	4.94	6.49
	0.4	3.52	5.68	7.76	8.96	24.54	47.26	3.71	5.47	7.18
	0.5	3.60	5.99	8.28	9.38	27.46	54.23	4.00	6.09	8.10
	0.6	3.82	6.60	9.25	10.61	33.43	68.51	4.27	6.69	8.91
	0.7	4.16	7.35	10.41	12.56	41.82	84.80	4.58	7.44	9.98
	0.8	4.50	8.13	11.50	14.87	51.76	105.31	4.89	8.05	11.12
	0.9	4.90	8.91	12.60	17.53	61.84	126.07	5.26	8.82	12.11
100	0.2	6.16	7.39	8.43	21.54	32.55	44.87	2.85	3.70	4.62
	0.3	3.39	4.75	5.93	7.76	15.86	25.55	3.13	4.20	5.15
	0.4	3.10	4.69	6.05	6.81	16.28	27.87	3.42	4.73	5.90
	0.5	3.22	5.08	6.70	7.40	19.46	35.06	3.69	5.32	6.73
	0.6	3.45	5.62	7.53	8.52	24.09	44.74	3.96	5.89	7.53
	0.7	3.74	6.27	8.42	10.07	30.24	56.53	4.24	6.48	8.44
	0.8	4.10	6.93	9.45	12.13	37.47	71.44	4.54	7.10	9.25
	0.9	4.49	7.71	10.54	14.65	46.35	88.60	4.87	7.75	10.21

Table D.13: Upper control limits for MEWMV statistics with $\lambda = 0.1$ (GK_Q_n)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.2	5.66	7.01	8.32	17.76	28.96	44.74	2.85	3.82	5.00
	0.3	3.31	4.73	6.00	7.19	15.24	25.75	3.13	4.26	5.39
	0.4	3.02	4.61	6.02	6.38	15.53	27.36	3.38	4.71	5.91
	0.5	3.15	4.97	6.63	7.07	18.58	34.45	3.66	5.28	6.76
	0.6	3.36	5.47	7.36	8.06	22.96	43.17	3.92	5.80	7.52
	0.7	3.68	6.12	8.40	9.72	29.07	56.39	4.20	6.39	8.35
	0.8	4.03	6.76	9.22	11.74	35.61	68.81	4.50	6.96	9.14
	0.9	4.41	7.54	10.33	14.13	44.38	86.22	4.82	7.55	10.03
100	0.2	6.10	7.23	8.14	20.04	28.79	37.91	2.79	3.56	4.37
	0.3	3.29	4.52	5.56	6.99	13.60	21.04	3.08	4.06	4.92
	0.4	2.98	4.41	5.63	6.20	14.01	23.50	3.35	4.57	5.60
	0.5	3.09	4.77	6.20	6.77	16.98	29.51	3.61	5.08	6.33
	0.6	3.33	5.27	6.95	7.93	21.18	38.01	3.89	5.62	7.09
	0.7	3.63	5.90	7.87	9.46	26.85	49.60	4.18	6.21	7.96
	0.8	4.00	6.57	8.80	11.54	33.54	62.54	4.47	6.79	8.71
	0.9	4.41	7.32	9.80	14.13	41.83	77.06	4.77	7.37	9.55

Table D.14: Upper control limits for MEWMV statistics with $\lambda = 0.1$ (GK_S_n)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.2	6.15	7.91	9.59	21.33	37.31	58.59	6.95	8.44	9.86
	0.3	3.60	5.23	6.76	8.53	18.98	32.69	4.73	6.12	7.42
	0.4	3.27	5.11	6.82	7.56	19.31	35.58	4.35	5.95	7.47
	0.5	3.34	5.39	7.30	8.01	21.87	41.13	4.29	6.11	7.77
	0.6	3.57	5.98	8.17	9.20	27.37	53.28	4.38	6.52	8.49
	0.7	3.86	6.58	9.14	10.83	33.94	66.80	4.54	6.97	9.22
	0.8	4.20	7.24	10.02	12.83	41.06	80.89	4.75	7.47	9.84
	0.9	4.62	8.04	11.26	15.58	50.94	103.33	5.00	8.07	10.85
100	0.2	6.33	7.60	8.64	21.80	32.19	42.59	7.10	8.18	9.07
	0.3	3.42	4.76	5.91	7.63	15.19	24.11	4.59	5.74	6.71
	0.4	3.08	4.60	5.90	6.66	15.37	25.86	4.19	5.53	6.65
	0.5	3.17	4.96	6.49	7.18	18.41	32.49	4.15	5.73	7.07
	0.6	3.43	5.49	7.35	8.38	23.04	42.06	4.25	6.09	7.74
	0.7	3.73	6.14	8.29	10.05	29.20	54.33	4.43	6.58	8.42
	0.8	4.08	6.83	9.17	12.10	36.39	67.16	4.64	7.08	9.18
	0.9	4.48	7.55	10.22	14.62	44.46	83.89	4.88	7.62	9.94

Table D.15: Upper control limits for MEWMV statistics with λ = 0.2 (Comedian)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.3	23.87	29.77	34.84	306.54	490.84	681.23	2.96	4.17	5.35
	0.4	6.83	9.46	11.75	29.34	59.61	96.27	3.20	4.65	6.01
	0.5	4.97	7.61	10.05	17.45	43.13	78.47	3.40	5.09	6.62
	0.6	4.47	7.27	9.86	14.75	40.66	77.76	3.64	5.57	7.38
	0.7	4.48	7.58	10.38	14.97	45.39	88.33	3.88	6.08	8.10
	0.8	4.72	8.22	11.42	16.57	53.72	107.74	4.16	6.66	8.99
	0.9	5.07	8.91	12.59	19.23	64.20	131.95	4.45	7.26	9.80
100	0.3	32.70	37.86	41.61	552.02	750.95	907.99	2.74	3.69	4.52
	0.4	6.33	8.41	10.08	24.49	44.82	67.50	2.98	4.15	5.16
	0.5	4.50	6.63	8.40	14.11	31.61	51.98	3.20	4.59	5.83
	0.6	4.12	6.43	8.42	12.32	31.38	55.41	3.42	5.05	6.42
	0.7	4.16	6.77	9.05	12.70	35.63	66.02	3.67	5.56	7.20
	0.8	4.40	7.32	9.90	14.24	42.42	80.06	3.92	6.07	7.93
	0.9	4.72	8.00	10.91	16.46	50.81	96.99	4.19	6.59	8.65

Table D.16: Upper control limits for MEWMV statistics with λ = 0.2 (S_n Covariance)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.3	20.92	24.59	27.60	227.60	322.57	414.69	2.65	3.62	4.55
	0.4	5.77	7.71	9.29	20.33	38.08	57.55	2.87	4.05	5.15
	0.5	4.20	6.24	7.96	12.31	28.51	48.08	3.08	4.44	5.66
	0.6	3.87	6.10	8.13	10.88	28.53	53.03	3.30	4.89	6.34
	0.7	3.96	6.51	8.87	11.47	33.28	63.79	3.54	5.39	7.04
	0.8	4.16	7.02	9.62	12.69	39.04	75.35	3.79	5.90	7.73
	0.9	4.52	7.75	10.65	14.92	47.77	93.25	4.03	6.35	8.41
100	0.3	30.17	33.61	36.08	463.19	578.72	671.41	2.59	3.43	4.16
	0.4	5.83	7.59	9.00	20.46	35.74	51.62	2.82	3.86	4.74
	0.5	4.14	5.99	7.51	11.84	25.66	41.53	3.04	4.29	5.36
	0.6	3.82	5.87	7.62	10.52	26.12	45.26	3.27	4.73	5.98
	0.7	3.89	6.21	8.22	11.03	29.96	53.81	3.49	5.20	6.65
	0.8	4.11	6.75	9.00	12.38	35.71	66.29	3.75	5.70	7.34
	0.9	4.47	7.43	9.93	14.54	43.61	80.21	4.00	6.18	8.03

Table D.17: Upper control limits for MEWMV statistics with λ = 0.2 (Classical)

n	ω	MEWMVL$_1$			MEWMVL$_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.3	19.26	20.47	21.13	186.25	210.16	224.12	2.48	3.19	3.80
	0.4	5.24	6.50	7.46	16.01	24.92	33.40	2.71	3.58	4.29
	0.5	3.81	5.26	6.38	9.88	19.23	28.66	2.92	3.96	4.76
	0.6	3.56	5.22	6.52	9.06	20.19	32.42	3.15	4.38	5.31
	0.7	3.67	5.59	7.03	9.73	23.88	39.11	3.38	4.78	5.86
	0.8	3.94	6.12	7.85	11.20	29.06	49.13	3.62	5.26	6.45
	0.9	4.29	6.74	8.68	13.29	35.47	60.33	3.89	5.68	7.06
100	0.3	28.77	30.44	31.40	415.47	464.97	494.95	2.52	3.25	3.83
	0.4	5.56	7.00	8.08	18.18	29.20	39.14	2.75	3.65	4.37
	0.5	3.97	5.54	6.78	10.76	21.45	32.68	2.97	4.07	4.95
	0.6	3.69	5.50	6.93	9.78	22.55	36.69	3.20	4.51	5.57
	0.7	3.77	5.83	7.50	10.30	26.08	44.03	3.43	4.95	6.16
	0.8	4.01	6.34	8.24	11.67	31.35	54.37	3.67	5.41	6.80
	0.9	4.36	6.97	9.05	13.80	37.61	65.55	3.94	5.89	7.45

Table D.18: Upper control limits for MEWMV statistics with λ = 0.2 (RMCD)

n	ω	MEWMVL$_1$			MEWMVL$_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.3	29.82	41.32	52.34	502.41	1016.46	1736.11	3.64	5.56	7.67
	0.4	8.92	13.41	18.04	50.89	121.89	225.89	3.92	6.15	8.44
	0.5	6.44	10.50	14.55	28.92	79.57	155.66	4.16	6.68	9.41
	0.6	5.80	10.04	14.29	24.57	75.20	155.45	4.46	7.36	10.36
	0.7	5.65	10.19	14.70	23.77	78.94	167.00	4.70	7.97	11.39
	0.8	5.79	10.77	15.89	25.15	90.02	196.34	5.01	8.74	12.70
	0.9	6.18	11.93	17.80	28.89	111.23	247.79	5.34	9.46	13.57
100	0.3	35.50	42.16	47.75	662.07	964.30	1284.59	3.04	4.19	5.26
	0.4	7.16	9.69	11.90	31.28	60.35	94.27	3.28	4.69	5.94
	0.5	5.10	7.63	9.84	17.84	41.15	70.21	3.52	5.16	6.63
	0.6	4.66	7.40	9.90	15.74	41.15	74.37	3.76	5.69	7.46
	0.7	4.64	7.70	10.52	15.79	45.69	86.07	4.02	6.27	8.29
	0.8	4.87	8.36	11.51	17.66	54.35	106.05	4.30	6.85	9.16
	0.9	5.20	9.04	12.43	20.15	64.07	124.73	4.58	7.41	9.92

Table D.19: Upper control limits for MEWMV statistics with λ = 0.2 (MVE)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.3	21.62	26.37	30.72	260.85	419.54	603.72	2.84	4.07	5.36
	0.4	6.28	8.76	11.07	26.06	54.19	89.56	3.10	4.57	6.03
	0.5	4.66	7.20	9.67	15.65	39.14	71.05	3.32	5.06	6.69
	0.6	4.33	7.15	9.75	13.73	39.15	74.89	3.56	5.59	7.48
	0.7	4.39	7.61	10.66	14.16	44.67	89.89	3.82	6.16	8.25
	0.8	4.56	8.03	11.35	15.23	50.64	100.73	4.07	6.77	9.25
	0.9	4.91	8.84	12.58	17.71	60.96	125.44	4.33	7.25	9.92
100	0.3	29.57	32.56	35.07	455.53	570.40	689.53	2.63	3.53	4.34
	0.4	5.88	7.72	9.22	21.51	38.77	57.31	2.86	3.97	4.97
	0.5	4.21	6.16	7.84	12.42	27.62	45.73	3.08	4.42	5.60
	0.6	3.91	6.12	8.08	11.04	28.38	51.08	3.30	4.89	6.29
	0.7	3.98	6.49	8.70	11.57	32.41	59.62	3.55	5.40	7.02
	0.8	4.20	7.05	9.52	12.83	38.47	72.21	3.78	5.89	7.66
	0.9	4.54	7.75	10.59	15.02	46.74	89.10	4.06	6.43	8.48

Table D.20: Upper control limits for MEWMV statistics with λ = 0.2 (GK_Q_n)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.3	20.33	23.40	25.79	211.75	284.29	351.41	2.62	3.55	4.46
	0.4	5.63	7.42	8.91	19.00	34.30	50.73	2.83	3.95	4.96
	0.5	4.07	5.98	7.61	11.42	25.53	42.83	3.05	4.37	5.58
	0.6	3.80	5.94	7.82	10.42	26.61	47.36	3.27	4.82	6.22
	0.7	3.87	6.30	8.49	10.90	30.90	57.70	3.50	5.28	6.81
	0.8	4.09	6.86	9.29	12.21	36.56	69.85	3.73	5.74	7.53
	0.9	4.43	7.53	10.34	14.26	44.46	86.23	4.01	6.27	8.36
100	0.3	29.45	32.15	34.13	437.95	523.59	593.21	2.57	3.39	4.11
	0.4	5.73	7.41	8.72	19.55	33.51	47.86	2.80	3.81	4.69
	0.5	4.08	5.87	7.34	11.42	24.41	39.12	3.03	4.26	5.31
	0.6	3.78	5.80	7.53	10.28	25.25	43.51	3.24	4.70	5.90
	0.7	3.85	6.10	8.03	10.75	28.56	50.99	3.49	5.16	6.58
	0.8	4.08	6.65	8.84	12.08	34.48	62.85	3.73	5.64	7.22
	0.9	4.45	7.34	9.84	14.40	42.08	77.98	3.97	6.12	7.95

Table D.21: Upper control limits for MEWMV statistics with $\lambda = 0.2$ (GK_S_n)

n	ω	$MEWMVL_1$			$MEWMVL_2$			MEWMV		
		95%	99%	99.73%	95%	99%	99.73%	95%	99%	99.73%
50	0.3	21.65	26.12	29.76	243.89	363.06	480.56	16.82	20.00	22.59
	0.4	6.05	8.24	10.15	22.29	43.23	67.03	5.72	7.29	8.65
	0.5	4.38	6.57	8.54	13.28	31.04	52.99	4.48	6.06	7.50
	0.6	4.03	6.48	8.71	11.76	31.61	58.36	4.12	5.91	7.56
	0.7	4.08	6.80	9.26	12.18	35.79	68.46	4.04	6.04	7.88
	0.8	4.29	7.35	10.23	13.52	42.33	84.07	4.08	6.36	8.43
	0.9	4.62	8.06	11.16	15.66	51.05	101.10	4.21	6.75	9.03
100	0.3	30.45	34.04	36.70	471.71	592.80	694.19	2.62	3.50	4.26
	0.4	5.95	7.77	9.25	21.25	37.37	54.06	2.86	3.94	4.87
	0.5	4.21	6.12	7.71	12.20	26.57	43.02	3.08	4.37	5.48
	0.6	3.88	6.03	7.86	10.86	27.28	47.67	3.30	4.84	6.16
	0.7	3.94	6.35	8.43	11.29	30.97	55.89	3.53	5.30	6.81
	0.8	4.17	6.89	9.25	12.73	37.02	68.70	3.79	5.80	7.50
	0.9	4.51	7.57	10.20	14.87	44.93	84.13	4.04	6.32	8.23

ARL Plots

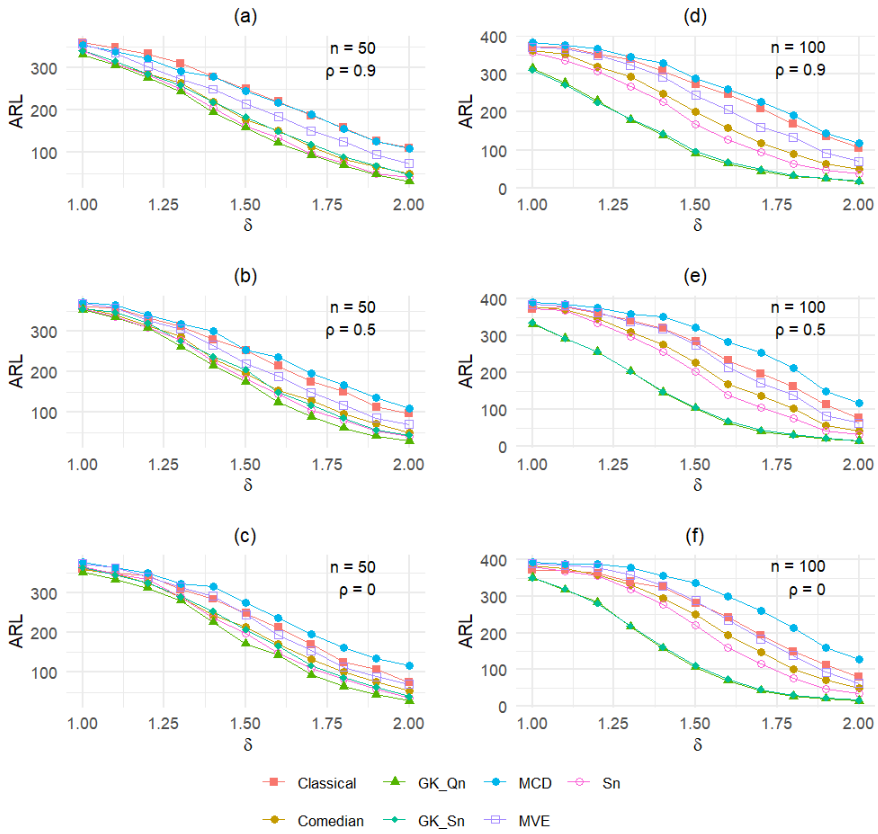


Figure D.1: ARL plots for MEWMSL$_1$ statistic with $\varepsilon = 0.1$ and $\omega = 0.2$.

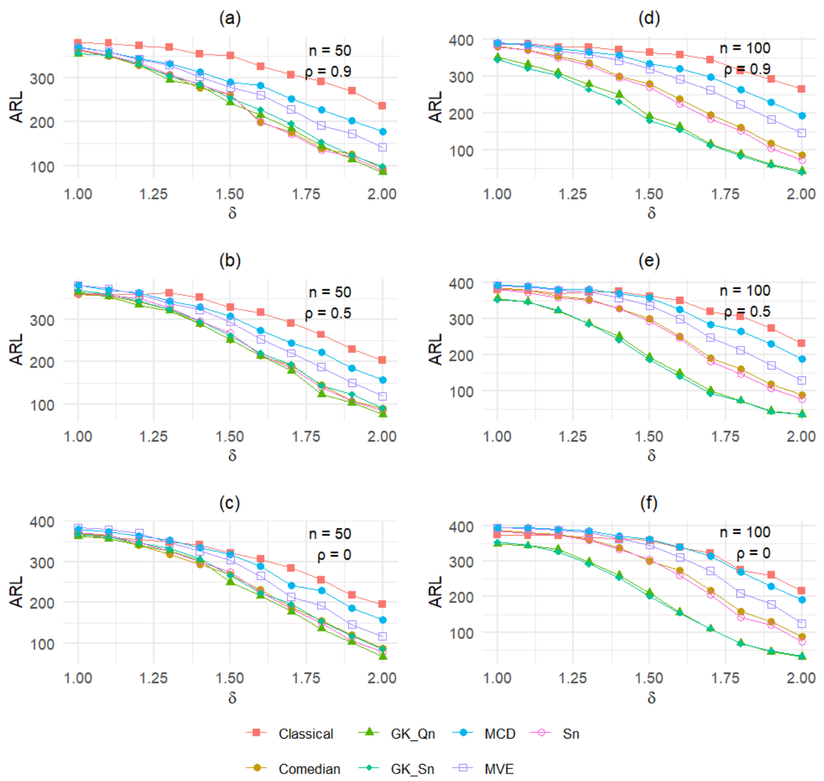


Figure D.2: ARL plots for MEWMSL$_1$ statistic with $\varepsilon = 0.2$ and $\omega = 0.2$.

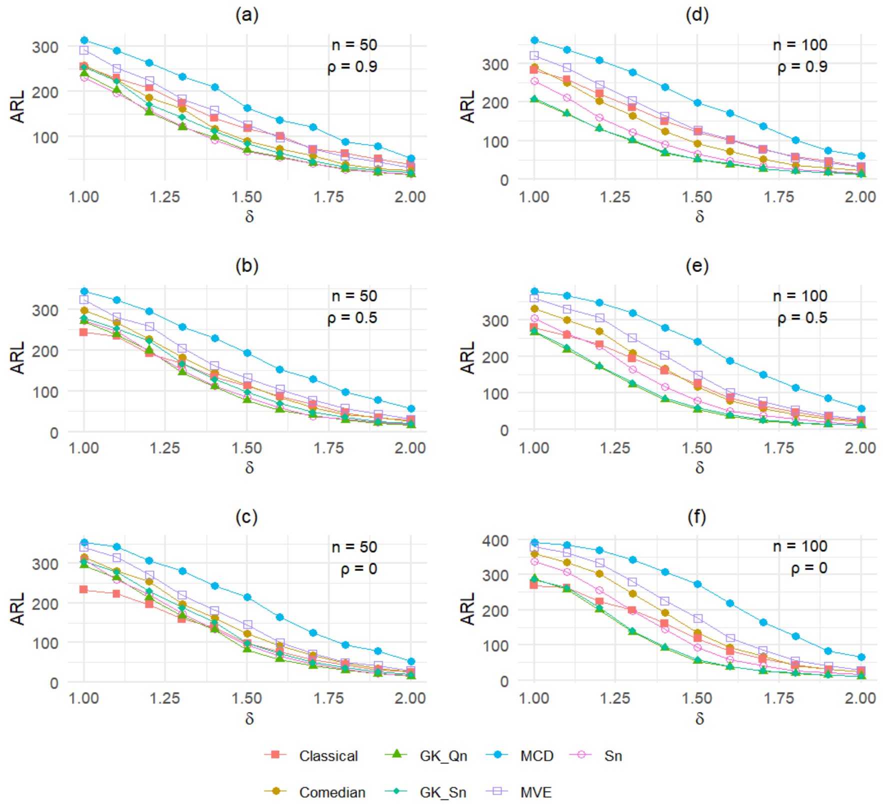


Figure D.3: ARL plots for MEWMSL$_1$ statistic with $\varepsilon = 0.1$ and $\omega = 0.4$.

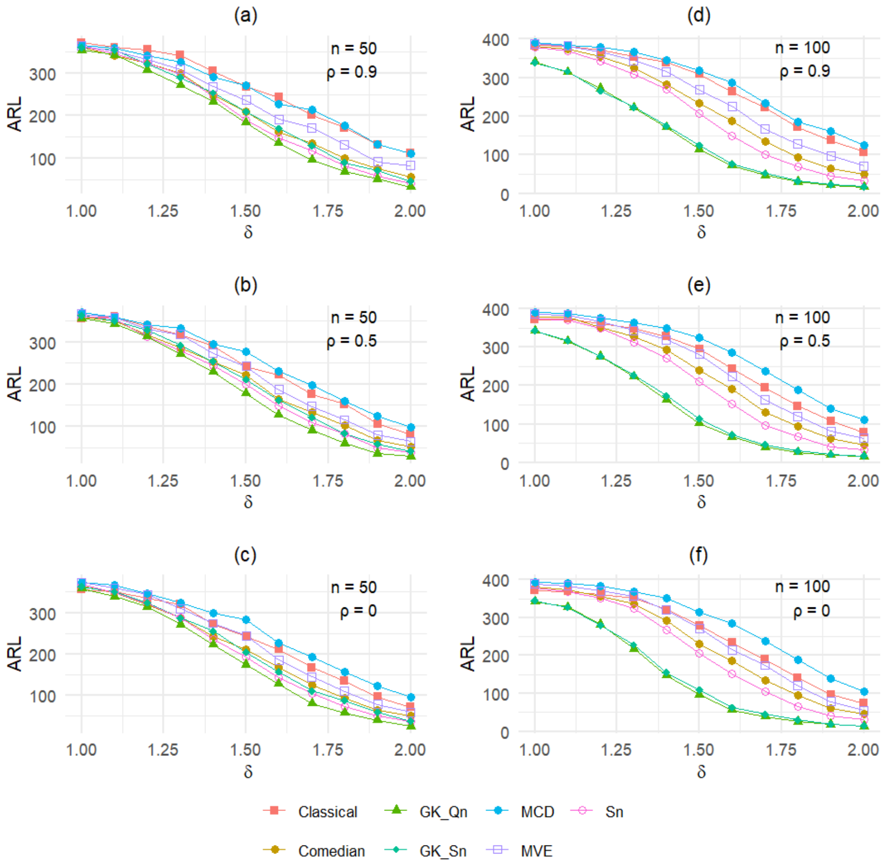


Figure D.4: ARL plots for MEWMSL$_2$ statistic with $\varepsilon = 0.1$ and $\omega = 0.2$.

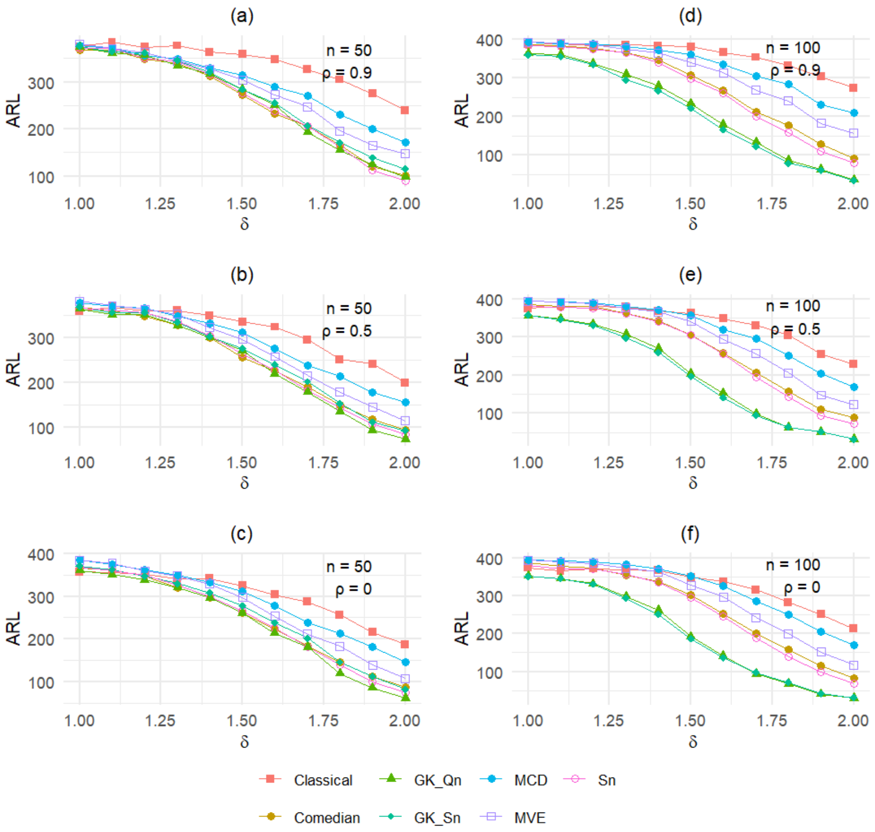


Figure D.5: ARL plots for MEWMSL$_2$ statistic with $\varepsilon = 0.2$ and $\omega = 0.2$.

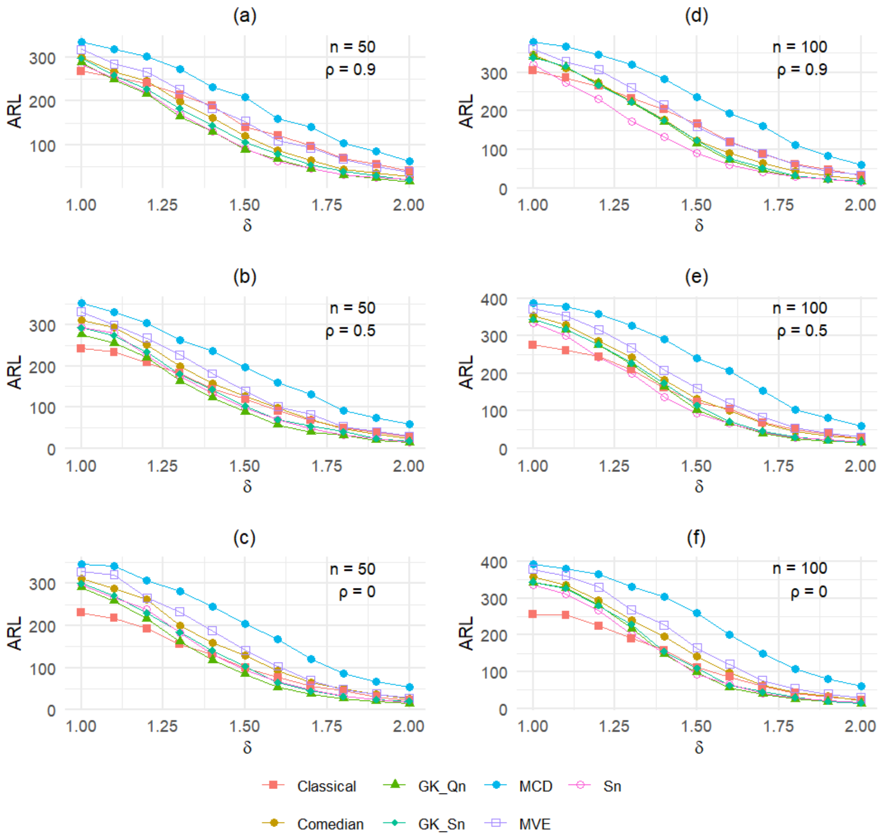


Figure D.6: ARL plots for MEWMSL$_2$ statistic with $\varepsilon = 0.1$ and $\omega = 0.4$.

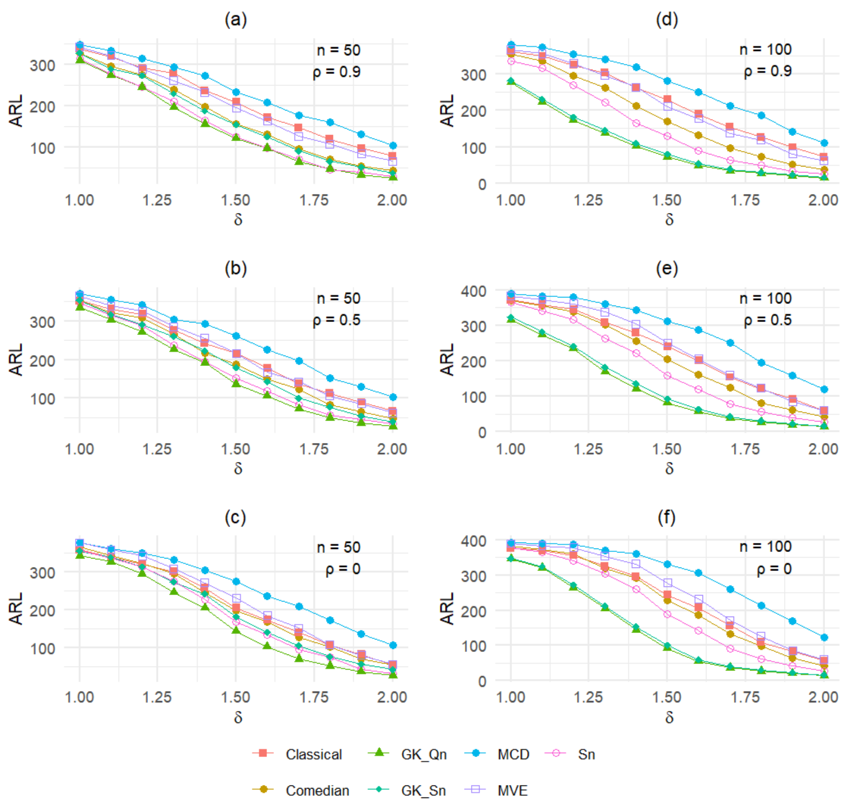


Figure D.7: ARL plots for MEWMS statistic with $\varepsilon = 0.1$ and $\omega = 0.2$.

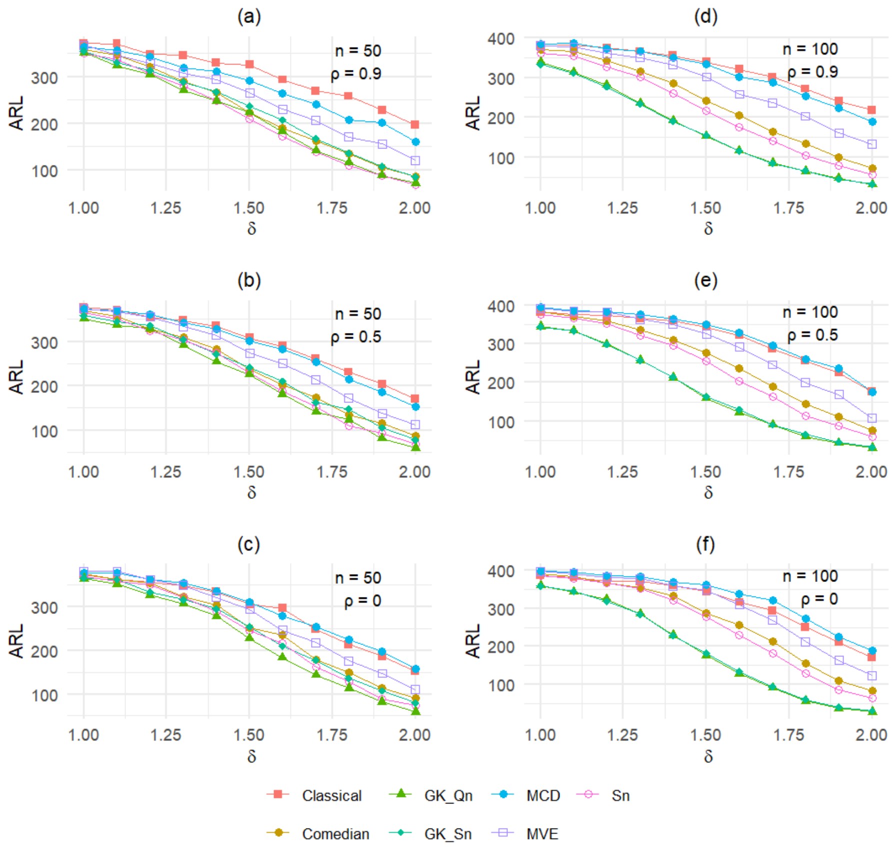


Figure D.8: ARL plots for MEWMS statistic with $\varepsilon = 0.2$ and $\omega = 0.2$.

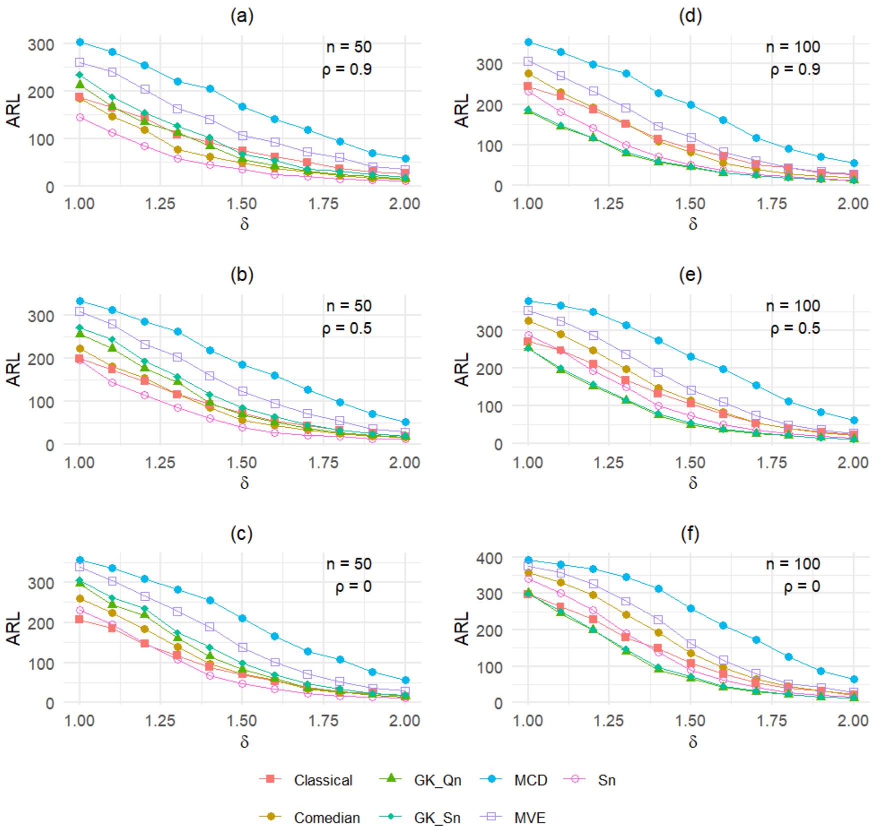


Figure D.9: ARL plots for MEWMS statistic with $\varepsilon = 0.1$ and $\omega = 0.4$.

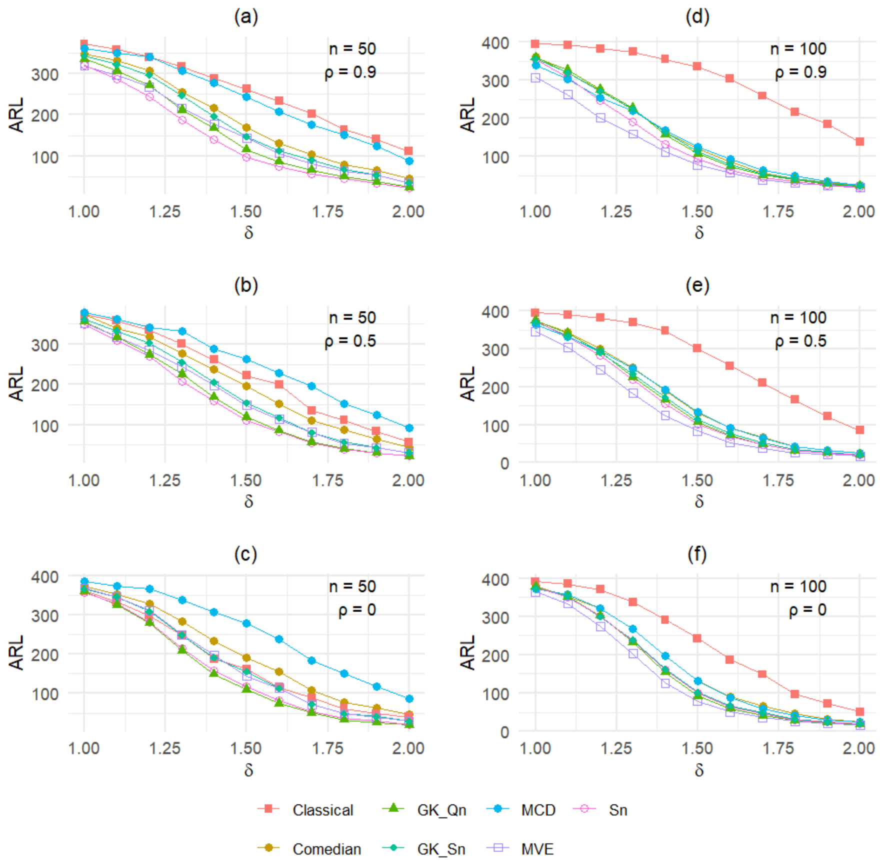


Figure D.10: ARL plots for MEWMVL$_1$ statistic with $\varepsilon = 0.1$, $\omega = 0.2$ and $\lambda = 0.1$.

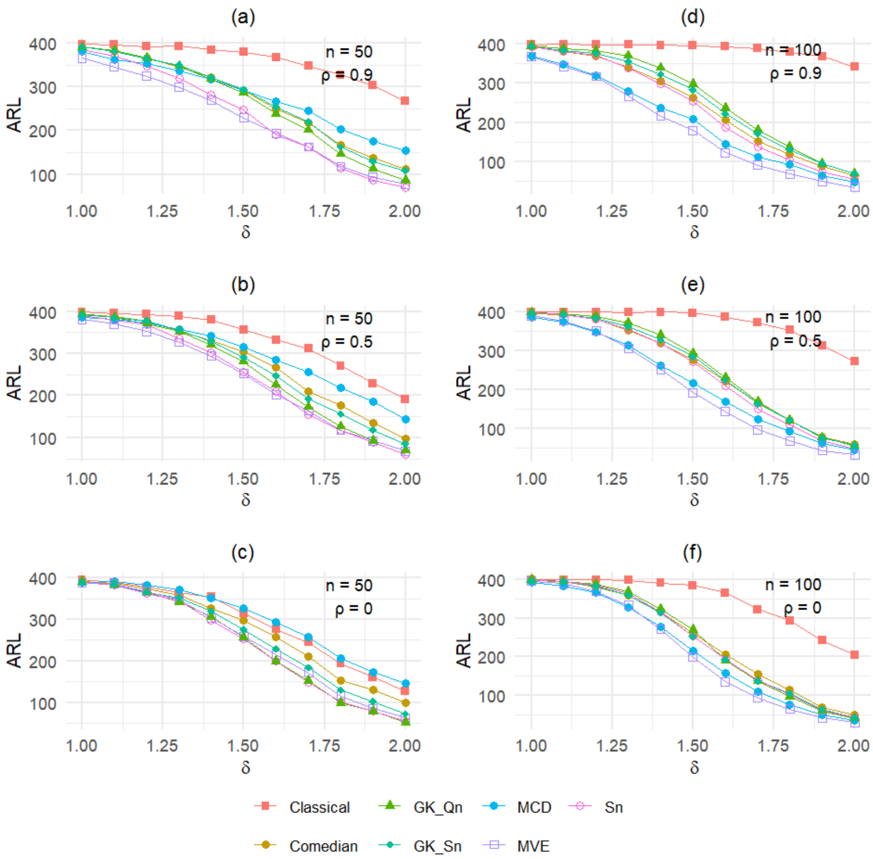


Figure D.11: ARL plots for MEWMVL$_1$ statistic with $\varepsilon = 0.2$, $\omega = 0.2$ and $\lambda = 0.1$.

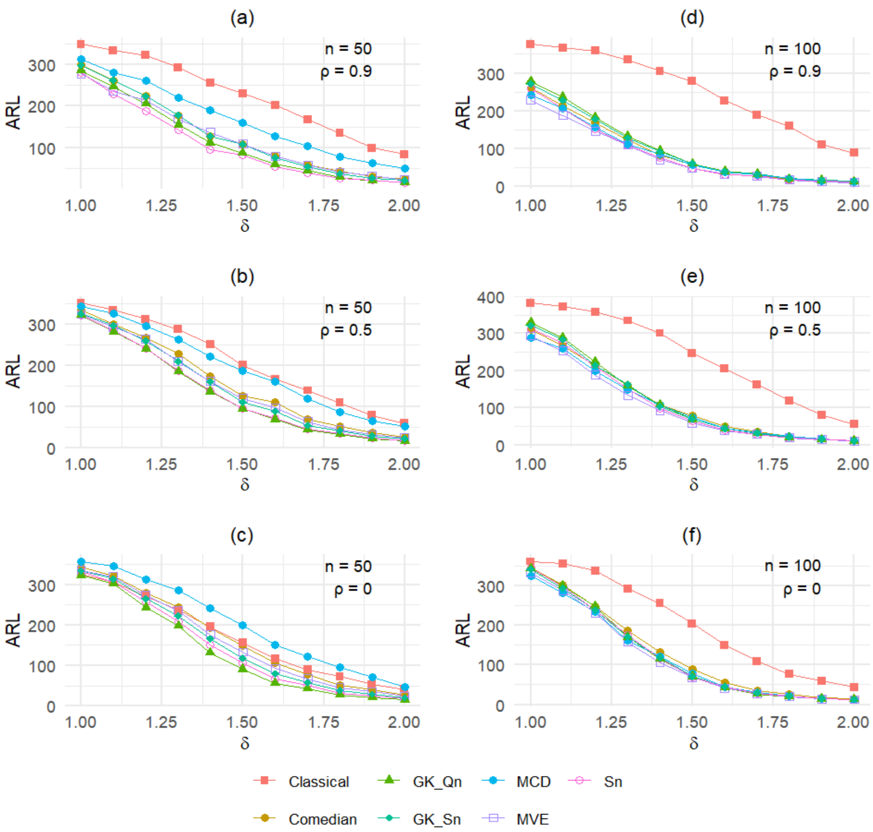


Figure D.12: ARL plots for MEWMVL$_1$ statistic with $\varepsilon = 0.1$, $\omega = 0.4$ and $\lambda = 0.2$

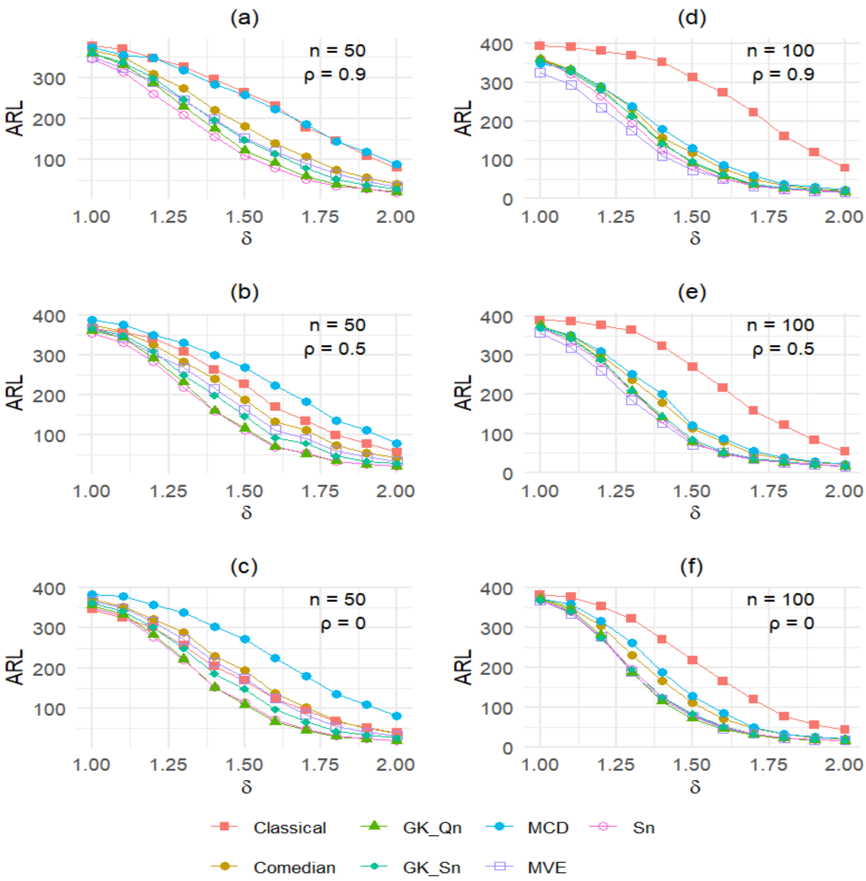


Figure D.13: ARL plots for MEWMVL$_2$ statistic with $\varepsilon = 0.1$, $\omega = 0.2$ and $\lambda = 0.1$.

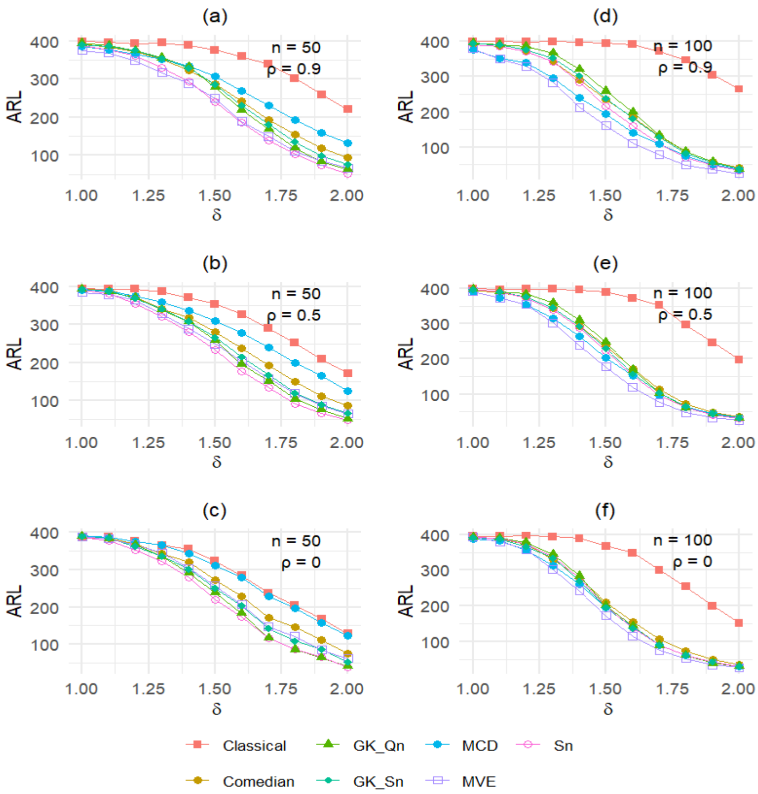


Figure D.14: ARL plots for MEWMVL$_2$ statistic with $\varepsilon = 0.2$, $\omega = 0.2$ and $\lambda = 0.1$.

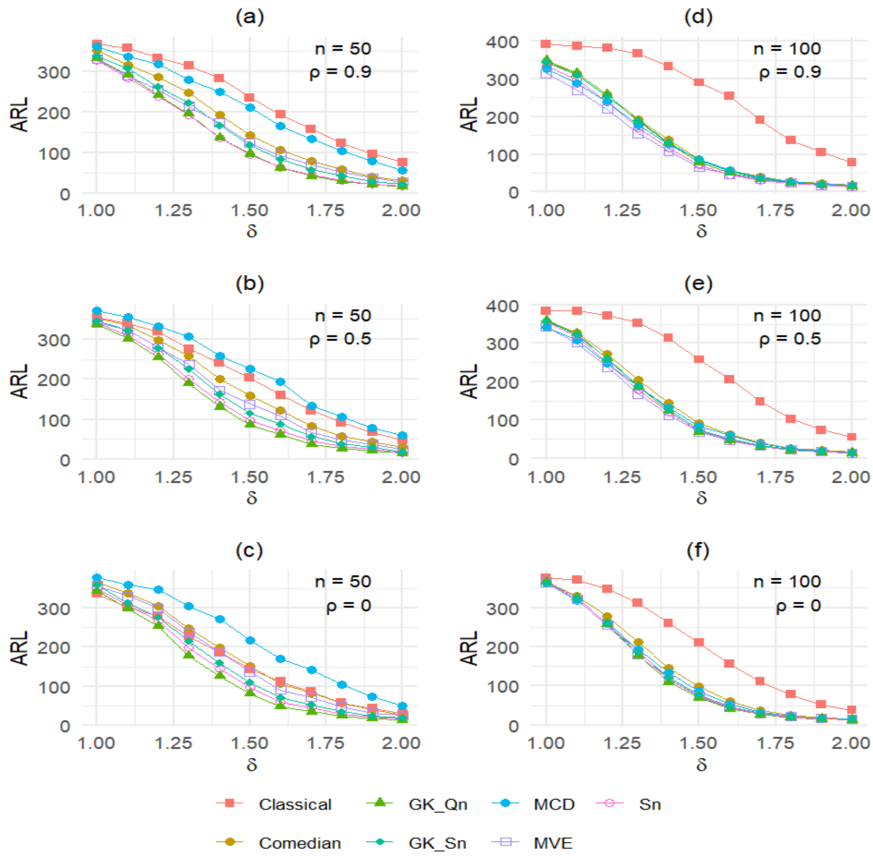


Figure D.15: ARL plots for MEWMVL$_2$ statistic with $\varepsilon = 0.1$, $\omega = 0.4$ and $\lambda = 0.2$.

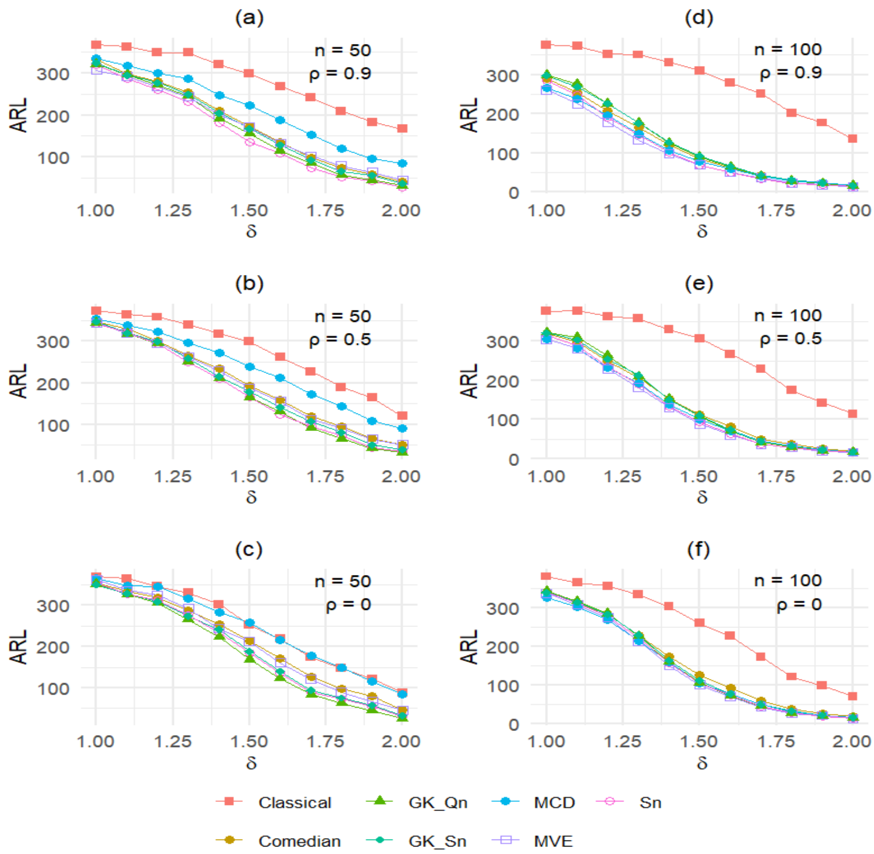


Figure D.16: ARL plots for MEWMV statistic with $\varepsilon = 0.1$, $\omega = 0.2$ and $\lambda = 0.1$.

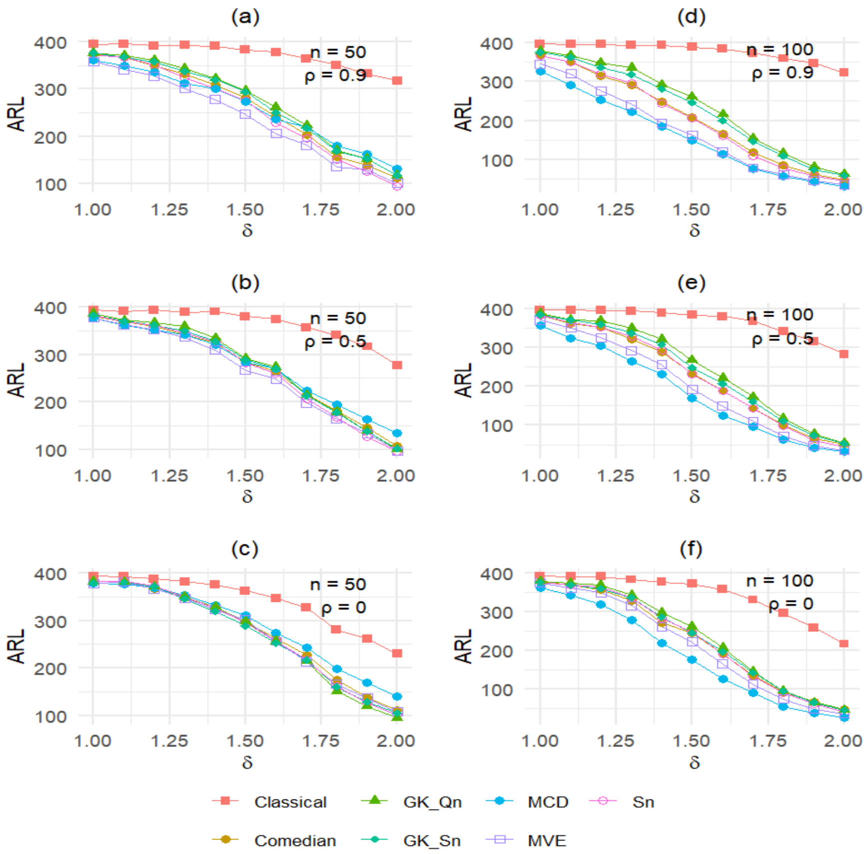


Figure D.17: ARL plots for MEWMV statistic with $\varepsilon = 0.2$, $\omega = 0.2$ and $\lambda = 0.1$.

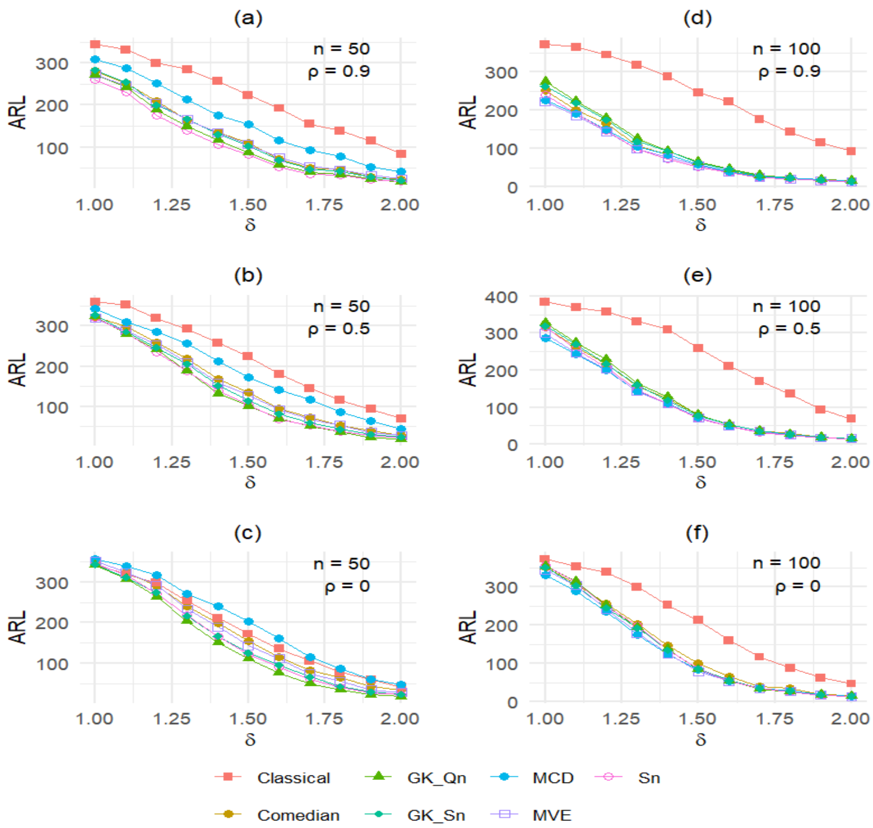


Figure D.18: ARL plots for MEWMV statistic with $\varepsilon = 0.1$, $\omega = 0.4$ and $\lambda = 0.2$.