A MULTIVARIATE APPROACH FOR OUTLIER DETECTION IN TIME SERIES DATA

By

Sidra Bibi

PIDE2017FMPHILETS06

Supervised by

Dr. Amena Urooj

Department of Econometrics and Statistics

Pakistan Institute of Development Economics
Islamabad
2020

Pakistan Institute of Development Economics

CERTIFICATE

This is to certify that this thesis entitled: **"A Multivariate Approach for Outliner Detection in Time Series Data"** submitted by Ms. Sidra Bibi is accepted in its present form by the Department of Economics & Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree of **Master of Philosophy in Econometrics.**

External Examiner:

Dr. Saqlain Raza
Assistant Professor,
Department of Economics, COMSATS,
Islamabad

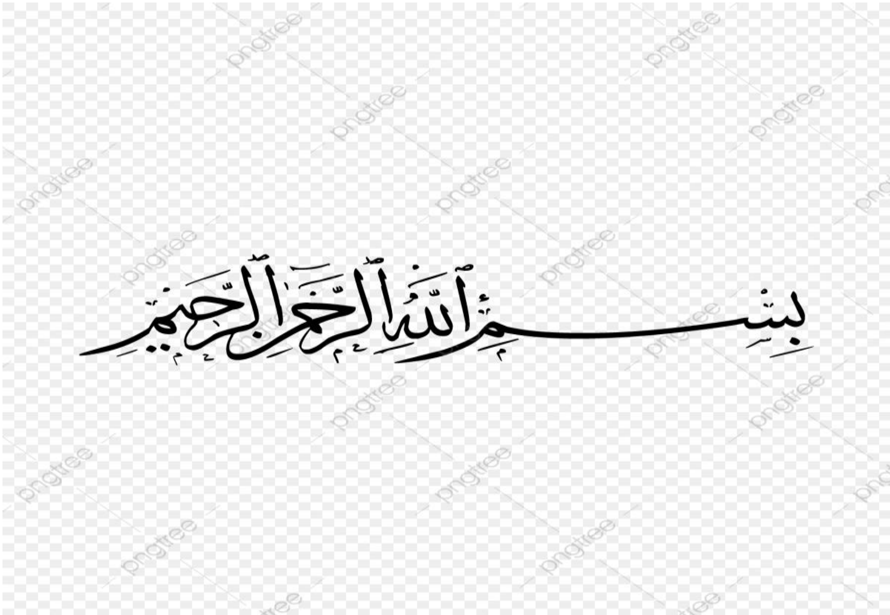
Supervisor:

Dr. Amena Urooj
Assistant Professor
PIDE, Islamabad

Head, Department of Economics & Econometrics:

Dr. Karim Khan
Associate Professor/Head
Department of Economics &Econometrics
PIDE, Islamabad

In the name of Allah, the Greatest Compassionate and most Benevolent



And Allah made it not except as a sign of good tidings for you and to reassure your hearts thereby, and victory is not except from Allah, the Exalted in Might, the Wise-(Surah Ali Imran, Ayah 126)

Author's Declaration

I sidra bibi hereby state that my MPhil thesis titled "A Multivariate Approach for Outlier Detection in Time Series" is my own work and has not been submitted previously by me for taking any degree from this University "Pakistan Institute of development Economics" or anywhere else in the country/world.

At any time is my statement is found to be incorrect even after my Graduation the university has the right to withdraw my MPhil degree.

Date: Signature of Student

Sidra Bibi

Dedications

This work is modestly devoted to all my respected treasures in life:

To my beloved family

Attended as my motivations and strength during thundery days

Especially to my supervisor

Who always beside me even for better for worse

And

Most of all to the one who gave me a life,

strength a faith to overcome all difficulties

my

Allah

ACKNOWLEDGMENT

First of all I would like to thank **Allah S.W.T**, the most sympathetic and generous that no well secrecies and mysteries of the universe who guide us in dimness and help us in problems and all respect to **His** messenger **Hazrat Muhammad (S.A.W)**, **Who** has guide **His** ummah to pursue knowledge from cradle to grave and allow us to win integrity of life and **Whose** educations has helped as an inspiration of light for mortality and the hours of misery and dimness. It is because of **His** blessing; it is possible to us to finish this thesis.

Therefore, I would like to acknowledge with thanks, the individuals who had guided me. Firstly, I would like to thanks my kinds sincerely and laudable research Supervisor **Dr. Amena Urooj**, she had guided me on how to complete this work, for their very affectionate and kind attitude guidance encouraging discussion and keen interest throughout my research work that enabled me to carry out my research work more confidently.

I am grateful to **Dr. Hafsa Hina** for providing their constant help and guidance in my studies and research. I express my special thanks to all my friends, who had helped and shared ideas with me. Lastly, I want to also thanks my beloved and respectable parents and my helping sisters, who always prayed well for me and giving their time to hear my problems.

Thanks also all staff in PIDE.

Sidra Bibi

Table of Contents

Contents **Page**

LIST OF TABLES

CONTENTS **PAGE**

LIST OF FIGURES

List Page

LIST OF ABBREVIATION

VAR	Vector Auto Regressive
SLS	Seasonal Level Shift
AO	Additive Outlier
IO	Innovative outlier
LS	Level shift
TC	Transient Change
MTS	Multivariate Time Series
J-max	Joint Maximum
MAO	Multivariate Additive Outlier
MIO	Multivariate Innovative Outlier
MLS	Multivariate Level shift
MTC	Multivariate Transient Change
MSLS	Multivariate Seasonal Level Shift
MARIMA	Multivariate Auto Regressive Integrated Moving Average
ARIMA	Auto Regressive Integrated Moving Average
AIK	Akaike Information Criteria
MVOD	Multivariate Voronoi Diagram
ESACF	Extended Sample Autocorrelation Function
PMD	Pakistan Metrological Department

ABSTRACT

This study has provided a method for detection of multivariate seasonal level shift in multivariate time series based on joint test statistics which follows a chi-square distribution and also provides a modified joint maximum test statistic of Tsay *et al*. (2000) method based on one test statistics for multivariate analysis. The J-maximum test statistic is modified by including multivariate seasonal level shift (MSLS). We have detected five types of outliers MSLS, MIO, MAO, MLS and MTC[1] and obtained power, size and empirical critical values using simulation and application on monthly time series data of Pakistan, and also checked their impact on the model parameter estimates, covariance matrix and standard error of the residuals. We have observed that multivariate SLS give good performance with large sample size in terms of power and size. Multivariate SLS is not much confusing with other types of outliers in $\mathrm{VAR}(0)(1)_{12}$ and $\mathrm{VAR}(1)(1)_{12}$ processes. we have observed that multivariate SLS along with other types of outliers drastically affect all the estimates and J and J-maximum test statistics for MSLS along with other types of outlier depends on dimension, sample size, order and structure of the model.

We have used real data example to detect outliers by using monthly time series data of temperature, rainfall and humidity for three stations of Pakistan and concluded that MSLS including other types of outliers in one series cause outlier in another series. We have also observed that estimates and standard error of the residuals have clear changes after adjusting the outliers in the series. At the end we have concluded that MSLS along with other types of outlier badly affect the estimates, analysis, results and decision taken on the

[1] MIO: Multivariate Innovative Outlier, MAO: multivariate Additive Outlier, MLS: multivariate level shift and MTC: multivariate transient shift and MSLS: multivariate seasonal level shift.

basis of these results. There is need to detect and adjust the MSLS along with other types of outliers in the data series to make the results reliable.

Key Word: MSLS (multivariate seasonal level shift), MAO (multivariate additive outlier), MIO (multivariate innovative outlier), MLS (multivariate level shift), MTC (multivariate transient change), VAR (vector auto regressive), Pakistan, simulation, $\mathrm{VAR}(0)(1)_{12}$ and $\mathrm{VAR}(1)(1)_{12}$.

CHAPTER 1

INTRODUCTION

Identification of multivariate outlier detection in time series has become important section of analysis, ignorance of which leads to misspecification of model and unreliable results of analysis. In seasonal modification and automatic time series modeling outlier detection also has an importance. For example, oscillating from banking deception to robotics as it allows anomaly detection in the system. Outliers do cause misleading conclusions, time series prediction and for this motive, numerous outlier detection techniques and robust estimation procedures have been anticipated previously for single level time series analysis, however, very limited are noted for vector time series.

Various studies related to outlier identification in the time series analysis attentive on solo series but we work on more than one data series in multivariate time series analysis. Several reasons are the cause of multivariate outlier detection in time series, for example one series outlier may cause to the presence of outlier in another series and also an outlier of sensible size affecting all the series may be unheeded in univariate time series analysis because univariate methods fail to associate information about the outliers among the component series. These outliers can be more easily noticed in multivariate time series analysis. Model choice is then more intricate and longer and more vulnerable to errors, which then disturb estimate.

By definition outliers are points that are distant from remaining observations. As a result, they can potentially skew or bias any analysis performed on the dataset. It is therefore very important to detect and effectively deal with outliers. They may indicate a variability in a measurement, experimental errors or an innovation. In other words, an outlier is an observation that diverges from an overall pattern on a sample. Outliers do not need to be extreme values. In the process of producing, collecting, processing, and analyzing data, outliers can come from many sources and hide in many dimensions. Those which are not a product of an error are called novelties. Most common causes of outliers in a data set are data entry errors, measurement errors, experimental errors, intentional, data processing errors, sampling errors and natural.

Multivariate outliers can be found in an n-dimensional space (of n-features). Looking at distributions in n-dimensional spaces can be very difficult for the human brain that is why we need to train a model to do it for us.

The outliers are classified in few categories; the multivariate additive outliers (MAO) affect only individual observation of the sequence and not the upcoming values, the multivariate innovational outliers (MIO) have a passing influence on the series like an novelty, the multivariate level shift (MLS) rise or diminution of all the observations at a specified point of the series by a constant, the multivariate temporary change (MTC) rise or decline harshly the level of the series which speedily returns to its original level exponentially, the last type is multivariate seasonal level shift (MSLS). An unexpected change which disturb only specific season of a year are called the seasonal outliers. These can be divided into AO and SLS. SLS is considered as a specific kind of level shift. In general, the MAO and the MIO are considered as non-typical observations whereas the MTC and the MLS as structural changes. The problem of outlier detection in multivariate time series is a complex problem because the different components can be affected by different types of outliers. In this thesis we will study outlier detection directly in multivariate time series framework with respect to different types of outliers and their impact on model selection and performance of other statistics in the presence of outliers.

Via explanation outliers are points that are distant from remaining observations. As a result, they can possibly skew or bias any analysis done on the dataset. It is therefore very important to detect and effectively deal with outliers. They may indicate a variability in a measurement, experimental errors or an innovation. In other words, an outlier is an observation that diverges from an overall pattern on a sample. Outliers do not need to be extreme values.

Multivariate outliers can be found in an n-dimensional space (of n-features). Looking at distributions in n-dimensional spaces can be very difficult for the human brain that is why we need to train a model to do it for us.

Outliers can have a dramatic impact on the results of common multivariate statistical analysis. For example, they can distort correlation coefficients (Marascuilo and Serlin, 1988; Osborne and Overbay, 2004), and create problems in regression analysis, even leading to the presence of collinearity among the set of predictor variables in multiple regression (Pedhazur, 1997). Distortions to the correlation may in turn lead to biased sample estimates, as outliers artificially impact the degree of linearity present between a pair of variables (Osborne and Overbay, 2004). In addition, methods based on the correlation coefficient such as factor analysis and structural equation modeling are also negatively impacted by the presence of outliers in data (Brown, 2006). Cluster analysis is particularly sensitive to outliers with a distortion of cluster results when outliers are the center or starting point of the analysis (Kaufman and Rousseeuw, 2005). Outliers can also themselves form a cluster, which is not truly representative of the broader array of values in the population. Outliers have also been shown to detrimentally impact testing for mean differences using ANOVA through biasing group means where they are present (Osborne and Overbay, 2004).

While outliers can be problematic from a statistical perspective, it is not always advisable to remove them from the data. When these observations are members of the target population, their presence in the dataset can be quite informative regarding the nature of the population e.g. Mourao-Miranda *et al*. (2011).

To remove outliers from the sample in this case would lead to loss of information about the population at large. In such situations, outlier detection would be helpful in terms of identifying members of the target population who are unusual when compared to the rest, but these individuals should not be removed from the sample Zijlstra *et al.* (2011).

Outlier detection has become a field of interest for many researchers and practitioners and is now one of the main tasks of time series data mining. Outlier detection has been studied in a variety of application domains such as credit card fraud detection, intrusion detection in cybersecurity, or fault diagnosis in industry. In particular, the analysis of outliers in time series data examines anomalous behaviors across time [Gupta et al. 2014a].

Detecting outliers is of major importance for almost any quantitative discipline. In machine learning and in any quantitative discipline the quality of data is as important as the quality of a prediction or classification model. The univariate method does not always work well. The multivariate method tries to solve that by building a model using all the data available, and then cleaning those instances with errors above a given value. Much research work is found in multivariate case for example, four types of outlier was explored by Tsay *et al*. (2000) in multilevel case Additive outlier, Innovative outlier, Level shift and Transient change by using multivariate ARIMA model and suggested two test statistics; Joint maximum test and component maximum test for detecting outliers. They have concluded that a multivariate outlier depends not only on its size and underlying model, but also on the interactions between the size and the dynamic structure of the model.

In practical life outlier detection has a very important role. Such as abnormalities in climate change (rainfall, temperature, humidity, fog etc.), air pollution, water pollution, water level in sea seriously affect to ecosystem which may cause different diseases and agriculture sectors i.e. fluctuation in agriculture product prices. If outliers are not detected and adjusted in the data then the analysis becomes wrong and we cannot predict future in reliable manner. Hence, we cannot think about proper solution to deal with the problem of pollution, irregular rainfall, and extreme temperature etc. When outliers are detected and adjusted in the data and analysis become reliable resulting in effective policy implication and predictions work well.

We have noted that seasonal patterns do exist in multivariate time series so there is needs of handling outliers in seasonal patterns in multivariate time series as well. One of the special kind of MLS is the MSLS which occurs in seasonal VAR (p)(P^s) at specific time point t such that for $1 \leq t \leq n$, and reoccur in each year in the same season called S after occurence of the MSLS and the affect of MSLS remains up to consequential seasons.

Much studies are found for detecting seasonal pattern in univariate time series in Pakistan and other countries including Kaiser and Maravall (2001) who have detected seasonal outliers in univariate time series. Sidra *et al*. (2015) have detected Seasonal variation of fine particulate matter in residential microenvironments of Lahore, Pakistan. Asghar and Urooj (2017) have explored the correct identification of seasonal outliers using most commonly applied test statistics and evaluated the performance of seasonal level shift (SLS) by means of empirical level of significance, power of the test for sensitivity in detecting changes, and the vulnerability to masking of outliers by misspecification frequencies. Their empirical study based on Pakistan.

It is noted that seasonal patterns may face seasonal outliers in multivariate time series too and therefore, needs handling of such outliers in seasonal patterns. Upto our knowledge, there is no such study guiding the insight about the multivariate seasonal outlier. In this study we attempt to look at the impact of SLS in multivariate time series framework directly.

1.1 **Literature Gap**

In past single level outlier detection have Frequent focus. Previously exploration about existence of outliers in the multilevel time series and their detection was ignored may be due to its computational difficulties. However, Tsay *et al*. (2000) and some other have explored outliers and their detection in multivariate time series. Up to our knowledge, no one study is found for outlier detection in multivariate time in case of Pakistan, however, in case of other countries various studies are available for outlier detection in multivariate time series and some exploring their theoretical aspects, however all these studies detect various types of outliers in multivariate time series but no one study is done for recognition of MSLS in multilevel time series around the world.

However, we have noted from existing literature the presence of outlier in one series may cause outlier in another series so we can think that along other outliers seasonal outlier in one series may also cause seasonal or other type of outlier in another series and there may exist the possibility of seasonal level shift and it impacts series, model, estimates, inference and predictions and up to our knowledge no one study was found in the world for the detection of SLS in multivariate time series. So, it's needed to be explored. Ignorance of all these may affect data analysis results and forecast. So, it is necessary to detect SLS along with other outliers in the vector time series. Therefore, we plan on the way to fill this gap by suggesting a method for detection of seasonal level shift in multivariate time series. We will also search for existence of multivariate outlier including SLS in case of Pakistan.

1.2 **Significance of the study**

Importance of our study is that, it attempts to provide a modified Tsay's (2000) multivariate outlier's detection method by including seasonal level shift (SLS). The seasonal level shift is not explored in multivariate structure up till now, we introduce seasonal level shift in multivariate time series using seasonal VAR model and apply the procedure as suggested by Tsay *et al.* (2000) based on one test statistics for detection of multivariate outliers i.e. joint maximum test statistics.

We explore five types of outliers in multivariate time series framework named as multivariate additive outlier (MAO), multivariate innovative outlier (MIO), multivariate level shift (MLS), and multivariate transient change (MTC), multivariate seasonal level shift (MSLS). We plan to identify effect of the existence of MSLS on multivariate time series, suggest its detection procedure and examine the performance of the suggested method.

In many cases multivariable observations cannot be detected as outliers when each variable is considered independently. Outlier detection is possible only when multivariate analysis is performed, and the interactions among different variables are compared within the class of data. Here it is needed to detect and adjust MSLS along with other types of outliers in time series data to make results reliable.

1.3 **Objectives**

i. To study the Impact of SLS on multivariate time series in terms of effect on estimates and standard errors.

ii. To suggest multivariate outlier's detection method by including seasonal outliers (SLS) among other type of outliers.

iii. To evaluate performance of suggested procedure in terms of power and size.

iv. To examine the performance of suggested procedure for outlier detection in multivariate time series by using time series data of Pakistan.

1.4 **Framework of the Study**

Mostly techniques for outlier detection focus on each variable autonomously, meanwhile outliers can affect the estimated mean and standard deviation, However, using vigorous methods for outlier detection can decrease or eliminate the effect of outliers on estimates of location and spread. However, up to our knowledge there is no one method is present to detect SLS directly in multivariate time series. We provide a modified Tsay's methods of multivariate outlier detection based on one test statistics joint test statistics by including SLS. We also used this modified method in simulation study to obtain empirical size, power and quantiles of the test statistics and to identify the impact on estimates and standard errors and covariances matrix of the analysis. We have employed this method in real data example to obtain the significance results of SLS in real world data.

1.5 **Organization of the Study**

This study is organized in six chapters. In chapter 1 we included introduction, significance of the study, objectives of the study, framework of the study. in chapter 2 we introduce the concept of SLS in multivariate time series along with its impact. We would discuss the detection of multivariate outliers by adding AO, IO, LS, TC. in chapter 3 we modify and suggest an iterative procedure for estimating multivariate outliers in seasonal VAR model by using one test statistics, joint maximum test statistics. In chapter 4 we use simulation to obtain finite sample critical values and empirical power and empirical level of significance of the test statistics for all five type of outlier in our study and for overall test statistics named as joint maximum test statistics. In chapter 5 we included empirical analysis, we shall demonstrate a real-world example by using few monthly time series data series. In chapter 6 we included conclusion and policy recommendations.

CHAPTER 2

LITERATUTRE REVIEW

Most of the time series data is affected by outliers and structural breaks both in univariate and multivariate structure. These occurred due to unexpected shock which may be interna or external. The outlier effects in multivariate time series were ignored earlier due to complexities in detection and handling procedures, however, in recent times some studies have focused in designing methods for outlier detection in multivariate framework. We can use nowadays numerous outlier detection method to testing each type behavior. Yet, up to our knowledge Seasonal level shift is ignored in multivariate time series. Existing literature can be divided into three parts, 1st part of the literature covers the studies about detection and handling of outliers in univariate time series, while 2nd part of literature discusses the studies related to handling of outliers in multivariate time series and 3rd part includes seasonal time series.

2.1 Case of Univariate time series

Fox (1972) was adopted two models with two outliers types the AO and the IO and identified outlier and checked their impact on time series. Likelihood and approximate likelihood ratios criteria for the adopted models obtained and to compare the power function with the earlier approaches.

Hillmer (1984) and Ledolter (1989) has explored effect of the AO on forecast. Both the forecast error for the period in which it occurs and the forecast error of subsequent periods affected by an additive outlier. Suggestion of this study is the method for the observation adjustment, which remove the biasness of forecast due to the AO. Ledolter (1989) study was concluded that AO near the forecast origin affect badly to forecast as compare to other condition.

Tsay (1986) in his research explored the problem of time series model specification in the existence of outliers. He proposed an iterative procedure for identifying the outliers, for removing their effects, and for specifying a tentative model for the underlying process. The procedure is basically grounded on the iterative estimation procedure of Chang and Tiao (1983) and the extended sample autocorrelation function (ESACF) model identification method of Tsay and Tiao (1984). An example is given. Properties of the proposed procedure are discussed. He used U.K spirits data to check the performance of method. While the projected method cannot reveal what caused the series to behave incoherently, it can often pinpoint those observations that deserve the special treatments. For spirits data, it not only recognized the two additive outliers but also led to the discovery of an intervention. Consequently, a substantial (75%) reduction in residual mean square was obtained. In conclusion, besides the parametric approach implemented in this article, there are many other methods for the treatment of outliers in time series.

Tsay (1988) was identified a method for identifying and handling four types of outliers by univariate approach in time series, he has planned an efficient iterative method. For identifying known and unknown outliers and their location, iterative procedure with two steps recommended by this study which contains estimation, specification, detection and removing cycle. This is the most common used method for outlier identification.

Chang *et al*. (1988) in their study was used LR criteria for testing and existence of AO and IO and for distinguishing between them he was explored the criteria. An iterative procedure was adopted by this study for identifying and estimating the time series parameters using ARIMA models in the existence of these outliers. Conclusion of this study is that the proposed method earlier in the literature for AR (1) coefficient is favorable.

Thury (1992) has imposed a seasonal adjustment study which is model grounded for identifying the impact of outliers, this study results was confirmed that outliers badly affect the seasonal modification which is model grounded and proposed that by exclusion of the outliers in the raw data, model grounded seasonal adjustment can be enhanced.

Chen and Liu (1993) have directed a study, which explore in the existence of four outlier forecasting in time series, four outlier types are included. Conclusion of this study is that if outliers exist near the forecast origin the time series is most affected as compare to other condition this also recommend that removal of outlier can give better forecast.

Balke (1993) was identified that, the procedure of Tsay (1988) does not work pleasingly, when there is LS exist and established it with examples. He identified that estimation and specification of the initial ARMA process affected by LS, also, this problem affected the detection and exclusion step. Furthermore, the Tsay method for distinguishing between LS and IO, misidentify level shifts. To deal with this problem given a simple extension of Tsay's procedure was given by Balke (1993), who has specified and estimate an initial ARMA (0, 0), before the initial ARMA specification. One of the CR procedure is suggested based on two phases, first for disturbance search and the reduction phase. This extension was claimed to minimize many problems of Tsay's procedure and easy to handle the possibility of level shift at the unknown period comparatively.

Chen and Liu (1993) was proposed a method for estimating of model parameters, outliers' effect by the joint estimation technique in time series. For identification and adjustment of outliers he has proposed an Iterative procedure for all outlier types that is four. This Study was confirmed that the power and misidentification inversely related to critical value. The study also concluded that the AO, TC, and LS can cause substantial bias in the estimation of the model parameter and, while effect of IO is not as much of serious on the model parameters.

Fomby and Balke (1994) was used a modified procedure for outliers as suggested by the Balke (1993) for identifying frequency, timing, and persistence of large shocks detection performance. They concluded that, outliers when related to business cycles then outliers are gathered together, therefore, there may exist incongruity between outlier's behaviors in real versus nominal series. They found out the significant evidence that infrequent large shocks have an important contribution to the variability of the macroeconomic time series. Series with outliers indicate excess kurtosis and the skewness along with non-linearity in the data.

Trivez (1995) has steered a method to scrutinize the forecasts with the Level Shifts and the Temporary Changes in ARIMA models. Here results identified that prediction interval affected by the LS and TC and the inaccuracy of point forecast becomes meaningfully enlarged. It is also noted that this effect depends upon the distance of occurrence from the forecast origin and as well as on auto regressive integrated moving average process.

Vaage (2000) has assessed two different procedures. Of the Tsay (1988) and the Balke (1993) respectively and has determined that the existence of level shift might lead to misidentification and loss of the test power. Study investigated that, how the model of Balke’s perform in both cases. Study confirmed that the Balke's procedure outperforms when the restrictions of Balke are removed.

Kaiser and Maravall (2001) premeditated the automatic outlier recognition and adjustment by considering distinct types of outliers. Simulation and real example was imposed in this study for examining SLS in detailed. outlier types included are the additive, innovative, level shift, and transitory change outliers. Study results confirmed that SLS have important properties in the time series and declared it as important type of outlier and suggested to replace it the IO. They concluded that when we considered the automatic outlier detections therefore, the innovative outlier (IO.) must not be castoff from outliers list.

Furusjo, *et al*. (2005) have employed QSAR models for four different environmental endpoints to demonstrate the importance of appropriate training set selection and how the reliability of QSAR predictions can be enlarged by outlier diagnostics. All models displayed reliable results; test set prediction errors were very analogous in magnitude to training set estimation errors when prediction outlier diagnostics were used to perceive and remove outliers in the prediction data. Test set prediction errors for elements classified as outliers were much larger. The difference in the number of outliers amongst models with a randomly and analytically selected training demonstrates well the need of illustrative training data.

Smith (2005) has steered a study, in which an extended STOPBREAK model of Engle and Smith (1999) was used named stochastic permanent breaks model by allowing for richer dynamics for forecasting in the existence of the level shift and displays that its forecasts is better than numerous alternatives. And said that the model STOPBREAK outpaces numerous alternative models.

Charles (2004) and the Charles (2006) employed GARCH models with outliers in financial data. He was applied on real data an outlier recognition methodology which is based on the Franses (1999) and Chen and Liu (1993). The study approves that all the excess kurtosis is removed after the adjustment of outliers from two series. Therefore normality is not rejected.

Nair *et al* (2006) have inspected the damage detection and localization algorithm based on time series and applied ARMA model. Study confirms that damage recognition and localization algorithms obtained from linear systems are effective for the stationary signals.

Galeano *et al*. (2006) and Baragona and Battaglia (2004) have used projection persuit method for detecting outlier in multivariate time series. Baragona and Battaglia (2007), have detected outliers in multivariate time series by Independent Component Analysis.

Helbling and Cleroux (2009) in their study they have reviewed numerous methods for sleuthing outliers in vector time series and planned new method of detecting outliers in the multivariate time series model, an experimental method based on graph of the influence function and another consisting of testing for the presence of outliers. They have measured all the methods from a hypothetical point of view.

Wang (2011) has projected an well-organized distance-based algorithm for noticing outlying samples in vector time series datasets.

Emerson and Emerson (2011) have discovered the performance of the outlier-sum statistic (Tibshirani and Hastie; Biostatistics, 2007), a projected method for identifying genes for which only a subset of a group of samples or patients displays differential expression levels. This study engrossed on this method as an example of how inattentiveness to standard statistical theory can lead to methods that exhibit some thoughtful problems. Results showed that, the proposed method offers little benefit even in the most idealized scenarios, and grieves from various limitations including difficulty of calibration, high false positive rates owing to its asymmetric treatment of groups, poor power or discriminatory ability under many alternatives, and poorly defined application to one-sample settings. They were incapable to reproduce their discoveries and conversed numerous unwanted and improbable features of their results.

Karanjit and Upadhyaya (2012) have attempted to structure and present a extensive overview of the detailed research on outlier detection techniques in varied research areas and applications also trying to the highlight richness and difficulty connected with each application domain. They have distinguished simple outliers from the complex outliers and defined two types of complex outliers, contextual and the collective outliers.

Zwilling and Wang (2014) have presented a general method for identifying outliers in vector time series grounded on a Voronoi diagram, which we call Multivariate Voronoi Outlier Detection (MVOD).

Micenkov´a, *et al*. (2015) have planned BORE (a Bagged Outlier Representation Ensemble) which uses unconfirmed outlier scoring functions (OSFs) as features in a supervised learning framework. BORE is able to adapt to arbitrary OSF feature representations, to the imbalance in labeled data as well as to prediction-time constraints on computational cost. They have established the good performance of BORE compared to the variety of competing methods in the non-budgeted and the budgeted outlier recognition problem on 12 real-world datasets. Results exhibited that BORE is the only method capable of handling budget restraints at test time. They have showed that it effectively selects a subset of the features that provide upright overall outlier detection performance.

Urooj and Asghar (2017) was used Chen and Liu (1993) method to discovered the performance of the outlier detection test statistic. The results show that the sampling distribution of TC is the less concentrated as compare to additive innovative and level shift. Empirical critical values for 1% 5% 10% with AR (1) process are higher as compare to other. The study also determined that TC confused with IO and AO and at the extremes it become equal to IO and AO.

Asghar and Urooj (2017 b) was used Kaiser and Maravall (2001) method and scrutinized the identification of seasonal level shift in SARIMA model. Study results confirms that for the identification of SLS detection in the case of SAR (1) and SMA (1) models the size of SLS affects the sampling distribution of test statistic. The empirical quantile is inversely related to sample size n and the coefficients of model. The small sample size and large coefficients values is not good for empirical power of the test. The study not recorded the impact of SLS on forecasts.

From above literature we concluded that all five outlier's types names AO, IO, LS, TC, SLS are explored for univariate time series in different studies and are applied on real world data in case of Pakistan and in case of other countries. These analyses have shown that there is significant impact of outliers on series, models, estimate, inferences and prediction. It is also shown that when these outliers are detected and adjusted in the data then analysis and future predictions become reliable. We now moved towards the literature covering the existence of outliers and their handling in multivariate time series.

2.2 Case of Multilevel Time series

Tsay *et al.*(2000*)* have derived a method to detect the four outliers' types directly in multivariate time series via simulation and empirical study by using multivariate ARIMA model. For detection of outliers they generated two statistics as joint maximum and component maximum test statistics. They highlighted difference amongst ingle level and multievel outliers and investigated dynamic effect of vector outliers on individual components. In empirical study they have used two real data examples. In simulation study, they have used multidimensional AR (6) and AR (1) models to obtains empirical critical values of the test statistics for bivariate and trivariate and sample size is 50, 100 and 200 and generate 10,000 realizations. This study concluded that the multivariate outlier's effect be subject to not only on its magnitude and the primary models but also on the collaboration between the magnitude and the dynamic structure of model. The later feature does not seem in univariate case. A multivariate outlier can bring various type of outliers for the single variable models. By associating result of univariate and multivariate outlier recognition one can gain awareness into the characteristic of an outlier.

Pan, *et al*. (2000) have proposed a procedure to detect several outliers in multi-level data. Simulation based artificial data was generated for analysis. Gaussian approach was used for the analysis of high dimensional data Results showed that for multi-dimensional data, extra points scattered consistently on the sphere must be produced. Similarly, the number of these points should rise progressively up until a comparatively stable result was made. Both real-world data analysis and simulation study display that the projected procedure is suitable in exercise.

Pena and Prieto (2001) have projected a method to recognize outliers in multi-level samples, grounded on the analysis of single-level projections onto directions that resemble to extremes for the kurtosis coefficient. A thorough analysis has been directed on the possessions of the kurtosis coefficient in polluted univariate samples and on the affiliation between directions to outliers and extremes for the kurtosis in the multi-level case. They determined that a method that figures a huge set of random directions will be extra influential as compare to another one that calculates a small number of definite directions in case if we have a large set of random uniformly scattered outliers in multi dimension. In another way, when the outliers appear along definite directions, a method that we search for these directions is anticipated to be very beneficial. These results emphasize the returns of combining random and specific directions in the search for multi-level outliers.

Tsay *et al.*(2006) developed a procedure for a multi-level approach to detect outliers in time series by using projection pursuit methods. They showed that through projection directions identification of outliers can be more influential than outlier identification directly in multilevel series. The optimal directions for the detecting outliers are found by numerical optimization of the kurtosis coefficient of projected series. They have planned an iterative procedure for detection and handling of multilevel outliers grounded on a single component search in these optimal directions. In contrast with the existing methods, the proposed procedure can identify outliers without prespecifying a vector ARMA model for the data. The Monte Carlo study and real data analysis results identified The good performance of the projected method.

Helbling and Cleroux (2009) have detected the outliers in multi-level time series models based on the multilevel autocorrelation and VARMA process on a simulation based generated data. They measured numerous methods in multilevel time series for detection of outliers. Some methods are based on the tests of hypotheses and others are based on the projection pursuit and ICA. They familiarized the coefficient of multilevel autocorrelation, gained its influence function together with its distribution. They also planned new methods of detecting outliers in the multilevel time series model, a heuristic method based on the graph of the influence function and another comprising of testing for the occurrence of outliers. All the methods measured in this paper have been seen from a theoretical point of view. Numerical comparisons would be interesting and remain to be done.

Cheng, *et al***. (**2009) have presented an anomaly sleuthing algorithm in noisy multilevel time series data. To capture the dependence relationships among variables, they employed kernel matrix alignment method based on simulation data. Time series data was used from 2000 to 2005. They concluded that algorithm used in our study is effective which is explained from both real and artificial data sets.

Baragona and Battagalia (2007) have detected the outliers in multilevel time series. For recognizing the positions of multiple outliers in vector time series, they have employed ICA as a implement. Model applied on both real and artificial data. For artificial data simulation-based method was used and for real data time series quarterly data was used from 1990 Q1 to 2000Q4. ARIMA and algorithmic technique was used. Findings showed that the projected method did best for single outlier recognition, as long as the outlier pattern is examined, even at the end of time series, and good performances were detected also for patches and LS.

Galeano, *et al.* (2011) have tracked multilevel outlier detection method in time series. Both time series and simulation-based data was used. For time series data was used from 1940 to 2010. VARMA, ARMA was used for multivariate and univariate analysis. Results showed that the projected procedure can recognize outliers without prespecifying a vector ARMA model for the data. Both from the Monte Carlo study and real data analysis, the good performance of the proposed method was observed.

Furwa (2017) work on outlier detection for skewed distribution: bivariate case and extended a technique for outlier detection SSSBB (Split Sample Skewness Adjusted) for the bivariate case and compared the result with the robust Mahalanobis distance technique considering various types of distributions. For comparing SSSBB and Mahalanobis distance, this study was used Monte Carlo Simulations. Four distributions are considered in this study named normal distribution, chi-square, gamma and beta distributions and different sample sizes are taken, to evaluate the performance of SSSBB for bivariate data and the study initiate that SSSBB does well as compared to Mahalanobis distance, in all the cases considered in the study. On the basis of the area of hurdle and outlier perceived ratio, the results show that SSSBB is a better method for normal as well as skewed data sets because SSSBB (Split Sample Skewness Adjusted) technique detects the possible outliers in the specified area of fence.

2.3 **Case of Seasonal Time Series**

Pal *et al*. (2013) have investigated the everchanging seasonality and increase in the frequency of precipitation by using wet and dry seasons data of the U.S. Mazzoni and Rezende, (2003) have investigated a tetragonopterinae from the ubatiba river, Brazil and identified a seasonal diet shift. Noonari *et al*. (2015) have used main vegetables prices of

Sindh Pakistan for examining the flexibility and seasonal variations. Imran *et al*. (2014) have investigated an analytical study of variations in the monsoon patterns over Pakistan.

2.4 Summary

From above literature we have noted that outliers in multivariate time series have significant impact on modeling, series, estimates, inferences and future prediction. When outliers are detected and adjusted in the data then analysis and future predictions become reliable. However, we also noted that detection of Seasonal level shift is totally ignored in multivariate time series, however numerous method can be used for outlier detection according to their type. So, we will fulfil this gap by detecting SLS in multivariate time series using monthly time series data of Pakistan.

CHAPTER 3

DATA AND METHODOLOGY

In this study we consider multivariate time series models with outliers using seasonal VAR (p)(P^s) model and study the effect of these outliers and their detection using Tsay *et al.* (2000) procedure by modifying the one test statistics namely joint maximum test for detecting five types of outliers. We explore all five types of outliers by using Simulation and empirical study.

3.1 **Seasonal Vector Auto Regressive Model**

For the multilevel time series analysis, the multilevel AR model is only the utmost effective, stretchy, and easy to use models. Particularly to describe the dynamic behavior of economic and financial time series and to predict, the vector AR model has confirmed to be valuable. The multilevel AR is a natural extension of the single level AR model to dynamic multivariate time series. It elaborates theory-based simultaneous equations models however often provides superior forecasts to those from univariate time series models. The VAR models can be made conditional on the potential future paths of specified variables in the model therefore Predictions from VAR models are quite flexible. furthermore, the vector AR model is also used for structural inference and policy analysis. In structural analysis, certain assumptions about the causal structure of the data under investigation are imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables in the model are summarized. These causal impacts are frequently summarized with IRF and forecast error variance decompositions, ARIMAX requires MLE which is commonly slow therefore vector AR model can be estimated using OLS or GLS which are normally fast.

Seasonality in macroeconomic data refers to systematic and recurrent variation in the data within the year. For example, monthly retail sales spike every year in December, as Christmas approaches, and airfares and motel rates increase during tourist season. Unmodeled seasonality tends to violate the constant parameter assumption of standard linear VAR models. There are a several remedies depending on the type of seasonal variation in the data. There are two types of seasonality i.e. deterministic seasonality and stochastic seasonality.

It is necessary to distinguish between deterministic seasonal variation and stochastic seasonal variation. The most common form of deterministic seasonal variation involves adding seasonal dummies to the VAR model. For example, consider the stable auto regression

$$A(L)y_t = \nu_i + u_t,$$

Where ν_i is the intercept associated with the i^{th} season and u_t is white noise error term. Quarterly models would typically include four seasonal dummies, and monthly models would include twelve seasonal dummies, where each dummy variable takes on a value of 1 for the quarter (or month) in question and zero otherwise. While it is common to include one dummy for each quarter (or month), one dummy would obviously be enough if we knew that there is no seasonality in the other quarters (or months). Models with seasonal dummies impart a deterministic pattern. Seasonality is perfectly predictable and, every year, the seasonal effect is the same. Although seasonal dummies are often used when modeling I (0) processes, it is worth stressing that it is equally possible for an I (1) process in levels to have deterministic seasonal.

In Stochastic seasonality in VAR with stationary seasonal process, one can model seasonal effects as random, which allows seasonality to be less than perfectly predictable and permits seasonal effects to evolve over time. Such randomness is appealing because there is no reason to expect seasonality to be time-invariant. For example, one would expect seasonal patterns in airfares to evolve with changes in market structure and seasonal energy consumption patterns to evolve with new technologies. Some forms of stochastic seasonality are easy to model. For example, for a VAR process

$A_s (L_s) Y_t = \nu + u_t$,

Where $A_s(L_s)$ involves only seasonal lags(s). The autocorrelation function of Y_t spikes at lags s, 2s, 3s, . . ., rs. The usually observed seasonal factor is s = 4 for quarterly data and s = 12 for monthly data. To allow for more sophisticated no seasonal auto covariance, a multiplicative model of the form

$A(L) A_s(L_s) Y_t = \nu + u_t$

may be considered, where A(L) is a standard VAR operator of order p, say. Multiplying the operator show that a standard VAR process with sufficiently large lag order nests the purely seasonal VAR or multiplicative model, if all roots of the VAR operator are out of the unit circle, however, some roots are the complicated pairs with periodic cyclicity. This means that even a standard VAR model may generate seasonality following a stationary stochastic process, provided, we include enough lags. One drawback of such model is that stationary seasonal processes tend not to exhibit the regularity commonly associated with seasonal effects unless the seasonal roots approach the unit circle. This observation suggests that this seasonal model without the addition of seasonal dummies will be of limited relevance for applied work. In this study we are considering VAR model with stochastic seasonality.

Mercy & Kihoro (2015) have estimated a vector AR model for a reformed seasonal univariate time series for prediction by using seasonal univariate time series data. Blazsek, *et al.* (2018) have introduced the Seasonal-QVAR model for the global real economic activity and world crude oil production, that classifies unseen seasonality in linear VAR and VARMA models. Lof and Frances (2000) have conducted a study for an empirical forecasting y using periodic and seasonal cointegration models for bivariate quarterly observed time series and contain both the single and multiple equation methods. A multilevel AR model in first differences with and the without cointegration limitations is as well comprised in the analysis where it assists as benchmark.

Both the seasonal and non-seasonal factors are included in the seasonal VAR model. The generally used notation for seasonal VAR model is $VAR(p)(P)_s$, With p : non-seasonal vector AR order, P = seasonal vector AR order, S = seasonal frequency of VAR model.

For stochastic seasonality, in formal the VAR model can be written as

Seasonal VAR model: $A_s(L_s)\, X_t = \alpha + u_t$ with $t = 1,\cdots ,n$,

$Y_t = X_t + \alpha(B)\omega_i I_t^{(h)}$ i= AO, IO, LS, TC, SLS

3.1.1 Multilevel outliers in timeseries

Let $X_t = (X_{1t},\cdots,X_{rt})'$ be a r-dimensional vector representing a multivariate VAR time series (vector autoregressive)

A multivariate seasonal VAR model is given by

$A(L)\, A_s(L_s)\, X_t = \alpha + u_t$ with $t = 1,\cdots,n$, …. 3.1

Where $X_t = (X_{1t}, X_{2t}, X_{3t}, \ldots\ldots X_{rt})'$,

$\alpha = (\alpha_{11}, \alpha_{21}, \alpha_{31}, \ldots\ldots\ldots, \alpha_{r1})'$,

$u_t = (u_{1t}, u_{2t}, u_{3t}, \ldots\ldots\ldots, u_{rt})'$

and L is the r × r and matrix backshift operator such that $LX_t = X_{t-1}$ and

Non seasonal lags of VAR: $A(L) = I - A_1L - \cdots - A_pL^p$

Seasonal lags of VAR: $A_s(L_s) = 1 - A_1L^S - \ldots - A_PL^{PS}$

are matrix polynomials of orders p, α is a r-dimensional constant vector and $u_t = (u_{1t},\cdots,u_{rt})'$ is a chain of autonomous white noise vectors of zero means and covariance matrices Σ.

Given an observed multivariate time series in which we add outlier term A_t where $A_t = \alpha(B)\omega_i I_t^{(h)}$ and actual series X_t then it becomes like observed series Y_t, $Y = \acute{Y}_1, \ldots, \acute{Y}_n$ with $Y_t = (Y_{1t},\cdots,Y_{rt})'$, the existence of outliers can be modelized as:

$Y_t = X_t + \alpha(B)\omega_i I_t^{(h)}$ i= AO, IO, LS, TC, SLS …. 3.2

Where $I_t^{(h)}$ is an indicator variable which characterizing the outlier at time h that is $I_t^{(h)}$ = 1 and $I_t^{(h)} = 0$ if $t \neq h$ and $\omega = (\omega_1, \cdots, \omega_r)'$ is the initial impact on the data series and X_t follows. α(B) is the dynamic effect of outlier on the series and changed with the type of outlier.

In this study we attempt to detect outliers in multivariate time series. We use seasonal VAR model and the method for detection of outliers as suggested by Tsay *et al.*(2000) by modifying it for four types of outliers.

For stochastic seasonality, we can write seasonal vector AR model as

(1) $A(L)\, A_s(L_s)\, X_t = \alpha + u_t$

$X_t = (X_{1t}, X_{2t}, X_{3t}, \ldots, X_{rt})'$

$\alpha = (\alpha_{11}, \alpha_{21}, \alpha_{31}, \ldots, \alpha_{r1})'$

$u_t = (u_{1t}, u_{2t}, u_{3t}, \ldots, u_{rt})'$

The seasonal component is not supported by the basic VAR. Direct modeling of the seasonal component of the series supported by an extended VAR model which is known as Seasonal VAR. The non-seasonal terms simply multiplied by the seasonal terms.

A(L) and $A_s(L^s)$ are the matrix polynomials of finite degree p with order $r \times r$, α is r-dimensional constant vector, L is backshift and $u_t = (u_{1t}, u_{2t}, \ldots, u_{rt})'$ is a classification of autonomously and identically distributed the Gaussian random vectors with zero mean and positive fixed covariance matrix $\sum$.

Let assume that the observed series is Y_t and the parameters are known but Y_t series is affected by outliers at time h such that observed series is $Y_t = X_t + \alpha(B)\,\omega_i \varepsilon^h_t$ and $Y_t = (Y_{1t}, Y_{2t}, \ldots\ldots Y_{rt})'$ and $\omega_i = (\omega_1, \omega_2, \ldots\ldots\ldots \omega_r)'$ is the magnitude of outlier in series Xt, I^h_t is the indicator variable for time point h and equal 1 if t = h and equal 0 if t not = h. $\alpha(L)$ is changed with the type of outlier. In the below we will identify theocratically and mathematically five types of outliers directly in multivariate case, given as.

VAR model with outliers can be written as

$$Yt = X_t + \alpha(L)\omega_i I^h_t + u_t$$

Here X_t is actual series without outlier and Yt is series with outlier which involves actual series X_t and error terms u_t. Where $u_t = A(L)A_s(L^s)\ X_t$, $X_t = (\pi(L))^{-1}\ u_t$ and $\pi(L) = A(L)A_s(L^s)$ which are the seasonal and non-seasonal roots of the VAR model.

$\alpha_{IO}(L) = \pi(L)I$ for multivariate innovational outlier and $\alpha_{AO}(L) = I$ for multivariate Additive outlier and $\alpha_{LS}(L) = (1 - L)^{-1}I$ for multilevel level shift and $\alpha_{TC}(L) = \{D(\delta)\}^{-1}$ for multilevel transient change. $D(\delta)$ is $r \times r$ diagonal matrix with diagonal elements $[(1-\delta_1 L), (1-\delta_2 L), \ldots\ldots\ldots\ldots, (1-\delta_r L)]$ And $0 < \delta < 1$ and for easiness we will assume that $\delta_1 = \ldots = \delta_r = \delta$. For multivariate seasonal level shift $\alpha_{sls}(L) = \frac{1}{\nabla s} - \frac{1}{s\nabla}$ where $\nabla = 1-L$ and $\nabla s = 1-L^s$ and s=4, s=12 for quarterly and monthly data respectively.

3.1.2 Lag Selection in VAR (p)(P)s Model

In case of time series data with seasonal frequency 4 and 12, we have to estimate VAR model with 4^{th} and 12^{th} lags respectively. In case of annual time series data, we faced the problem of enough observation for VAR model, we need to take a lag length, which is neither too small nor too large, however to eradicate residual autocorrelation in VAR models' lags are imperative. For optimal lag length, lag length criteria specify a positive way of choosing the optimal lag after estimating the initial VAR model. Mostly used lag selection criteria established by econometricians are HQ, SIC, AIC, LR and BIC, etc. Usually, we preferred that lag length for which the values of most of these lag length criteria are diminished. But this is not enough for lag selection, there is need to checked the residual autocorrelation of the estimated VAR model, finally for lag length selection. Additionally, VAR models are over parameterized and lags corrode the degrees of freedom and declines the strength of diagnostic tests because of multiple lags, therefore, to select the minimum lags that eradicates VAR residual autocorrelation we need to perform the lag exclusion tests.

Generally, the preferred lag selection criteria are AIC because it is favorable for small sample forecasting structures. However, for large sample size the BIC and HQ works well and preferred the true order of the VAR model by comparing to the order. $r+pr^2$ is used to calculate the number of coefficients in a VAR (p) (P)s model for r time series is, r is the number of intercepts, each lag p adds an $r \times r$ matrix of coefficients.

3.2 **Detection Procedure of Outliers**

Distant data may be frequently observed that, it does not fit the common pattern in multilevel time series. Manifestations of outliers and shifts are impulsive measures that may harshly distort the analysis of the multilevel time series. For occasion, the undetected outliers and shifts badly affects the model building, seasonal valuation, and forecasting. Structure dependence of the multilevel time series gives increase to the well-known swamping and masking effect that misinform using most outliers credentials techniques.

3.2.1 Steps for Detection of Outliers

i. We will Assume no outlier at the start of analysis and build a Seasonal VAR model.

ii. Calculate $\hat{a}_t$ as residuals and $\hat{\pi}_i$ is the vector of estimated coefficients of VAR (Vector Autoregressive) model.

iii. Calculate the effect of each type of outlier at each time point. Estimate the vector $\hat{\omega}_{i,\,h.}$ where i= IO, AO, LS, TC, SLS.

iv. Then we will test the implication of multivariate outlier at time index h, we consider the null hypothesis ω=0 versus the alternative ω ≠ 0. One test statistic used is J-maximum which treats components of ω jointly. Joint maximum test statistic follow noncentral ${\chi_k}^2(\eta_i)$ distribution with noncentrality parameters $\eta_i = \omega' \Sigma_{i,h}^{-1} \omega$. by the null hypothesis $H_o : \omega = 0$, the $J_{i,h}$ follow the chi-squared distribution with the degrees of freedom r, following Tsay *et al.*(2000).

$$J_{i,h} = \hat{\omega}_{i,h}' \, \Sigma_{i,h}^{-1} \, \hat{\omega}_{i,h} \quad \text{where i= IO,AO, LS, TC, SLS} \qquad \ldots \qquad 3.3$$

Here ω is magnitude of outlier on the series $\Sigma_{i,h}^{-1}$ is the covariance matrix. We define general test statistics as

$$J_{max}(i, h_i) = \max_{h} j_{i,h}$$

Wherever h_i represents the time index where the extreme of test statistics happens, h_i shows the maximum of test statistics among all the five test for AO, IO, LS, TC, SLS, for example if the joint test statistics value of SLS is greater than the other four then here we write it as $j_{max}(SLS, h_{SLS}) = \max_{h} j_{SLS,h}$.

We detect outlier in multilevel time series by using joint test statistics, if we detect IO in MVTS then by the null hypothesis that there is no outlier in the sample and supposition that model of Xt is acknowledged the value $j_{max}(IO, h_{IO})$ is the extreme value of a random sample, sample size=n with a chi-square distribution with r degree of freedom IO= innovative outlier. Consequently, the asymptotic distribution of $j_{max}(IO, h_{IO})$ can be gained by using the extreme value distribution, however, for other all of the four joint test statistics named MAO, MLS, MTC and MSLS is the maximum of a dependent sample with chi square distribution with r degree of freedom, therefore their asymptotic distributions are more complex, because it is depending on the serial dependence of $\{ J_{i,h}\}$. Tsay's *et al*. (2000) identified that "when we estimate the outlier size ω, we can see that the serial correlation is of $\{ j_{LS,h}\}_{h=1}^{n}$ are tougher than those of $\{ j_{i,h}\}$ for i= MIO, MAO, MTC this is because of nondecaying factor persuaded by the operator $(1\text{-}B)^{-1}$ therefore $\widehat{\omega}_{LS,h}$ comprises all the filtered values of $\hat{e}_t$ for t ≥h. Accordingly, the asymptotic distribution of $J_{max}(LS, h_{LS})$ is more focused as compare to other joint test statistics. The degree of attention depends on cumulative π values therefore, the critical value of $J_{max}(LS, h_{LS})$ generally lesser as compare to other joint test statistics."

In case of univariate time series analysis if a single $j_{max}(i, h_i)$ is substantial at time poin h_o, we categorize a multilevel outlier of the type i at h_o, where i= MAO, MLS, MTC, MSLS. For multiple significant $j_{max}(i, h_i)$, we recognize the type of outlier grounded on the test statistics which has the minimum detected p-value at time index h_o and probability of that test statistics is lesser than 0.05 level of significance then we declared that type of the outlier at time index h_o at the 5% level of significance. When outlier is declared then its impact on the real time series is removed by using the results for detected series and adjusted series is preserved as a new data set and the detection procedure is repeated, we dismiss the detection practice when there is no significant outlier is noticed.

3.2.2 Derivation of Equations with and without Outliers

VAR model equation without outliers (Actual series)

$A(L^s) A(L) X_t = \alpha + u_t,$

This equation shows the main structure of VAR model with seasonal and non-seasonal lags

$X_t = (X_{1t}, X_{2t}, X_{3t} \ldots X_{rt})'$

X_i is i × j matrix where each column is k× 1 vector.

$\alpha = (\alpha_{11}, \alpha_{21}, \alpha_{31} \ldots \alpha_{r1})'$

$u_t = (u_{1t}, u_{2t}, u_{3t} \ldots u_{rt})'$

$X_t = (\pi(L))^{-1}u_t$ and $\pi(L)= A(L)A_s(L^s)$ which are the seasonal and non-seasonal roots of the VAR model.

3.2.2.1 Observed series for Additive outlier

An Additive outlier denotes a surprising variation in one of the observations. It can appear because of a recording or measurement error or other single effect. The observed series for AO and its impact on residuals is given below:

Outlier term $At= \omega_{i,h}\ \alpha_i(L)\ I^h_t$ and i = AO, IO, LS, TC, SLS

$$Y_t = X_t + \omega_{AO,h}\ \alpha_{AO}(L)\ I^h_t \qquad \ldots\ldots. \qquad 3.5$$

$X_t = X_t = (\pi(L))^{-1}u_t$, $\alpha_{AO}(L) = 1$, $X_t = (X_{1t}. X_{2t}, X_{3t} \ldots. X_{rt})'$

$Y_t = (\pi(L))^{-1}u_t + \omega_{AO,h}\ \alpha_{AO}(L)\ I^h_t$

$Y_t = (\pi(L))^{-1}\ u_t + \omega_{AO,}\ h\ I^h_t$

For in terms of residuals Multiply $\pi(L)$ on both side

$\pi(L)\ Y_t = u_t + \omega_{AO,h}\ \pi(L)\ I^h_t$, $\pi(L)\ Y_t = a_t$

$a_t = u_t + \omega_{AO,h}\ \pi(L)\ I^h_t$, $\pi(L) = (1 - \pi(L))$

For multivariate at $= (1- \sum_{i=1}^{\infty} \pi_i\ (L^i)\omega_{AO,}\ h\ I^h_t + u_t$

$$a_t = (I^h_t - \sum_{i=1}^{\infty} \pi_i\ I^h_{t-i})\omega_{AO,}\ h + u_t \qquad \ldots \qquad 3.6$$

$u_t \sim N(0, \sum)$ from GLS (Generalized least square estimators) $\omega_{AO,h}$

Therefore, when $\pi(L)\neq 1$, an AO at time index h for a vector AR(P) model, it will affect a_t for t = h, h + 1,⋯ ,h + p.

$$\hat{\omega}_{AO,h} = -\left(\sum_{i=0}^{n-h} \pi_i' \, \Sigma^{-1} \, \pi_i\right)^{-1} \sum_{i=0}^{n-h} \pi_i' \, \Sigma^{-1} a_{h+i}, \qquad \pi_o = 1. \qquad \ldots. \qquad 3.7$$

Where ω is the magnitude of MAO on the data series and this impact finished after the time period of outlier, π_i is the coefficient matrix for all lags of VAR model. Σ is the covariance matrix and L is the backshift lag operator $LX = X_{t-1}$.

3.2.2.2 Observed series for Level shift

When a surprising move occurs in all the observations of the multilevel detected time series after some time point, we call it as MLS. Multivariate level shifts distressing all the components of a multilevel time series, sometimes we called it structural breaks, because it yield a permanent effect on the multilevel series. The observed series in case of LS and its impact on residuals is given below:

$$Y_t = X_t + \omega_{LS,}\, h\, \alpha_{LS}(L)\, I^h_t \qquad \ldots\ldots. \qquad 3.8$$

$$X_t = (\pi(L))^{-1} u_t, \quad \alpha_{LS}(L) = (1 - L)^{-1}, \qquad Y_t = (Y_{1t}.\ Y_{2t},\ Y_{3t} \ldots\ldots\ Y_{rt})'$$

$$Y_t = (\pi(L))^{-1} u_t + \omega_{LS,h}\, \alpha_{LS}(L)\, I^h_t$$

$$Y_t = (\pi(L))^{-1} u_t + \omega_{LS,}\, (1 - L)^{-1}\, h\, I^h_t$$

For in terms of residuals Multiply π(L) on both side

$$\pi(L)\, Y_t = u_t + \omega_{LS,h}\, \pi(L)\, (1 - L)^{-1} I^h_t, \qquad \pi(L)\, Y_t = a_t$$

$$a_t = u_t + \omega_{LS,h}\, \pi(L)(1 - L)^{-1} I^h_t, \qquad \pi(L) = (1 - \pi(L))$$

For multivariate $a_t = (1- \pi_i (L^i))\omega_{LS,}\, h\, (1 – L)^{-1} I^h_t + u_t$

$a_t = (1 – L)^{-1} (I^h_t - \sum_{i=1}^{\infty} \pi_i I^h_{t-i})\omega_{LS,h} + u_t$ …. 3.9

$u_t \sim N(0, \sum)$ from GLS (Generalized least square estimators) $\omega_{LS,h}$

Here it is clear that a LS at time index h affects all filtered values of a_t for $t \geq h$.

$\hat{\omega}_{LS,h} = - (1 - L)^{-1} (\sum_{i=0}^{n-h} \pi_i^{'} \Sigma^{-1} \pi_i)^{-1} \sum_{i=0}^{n-h} \pi_i^{'} \Sigma^{-1} a_{h+i}$, $\pi_o = 1$. 3.10

Where ω is the size of LS on the data series and this impact is same for all the observations after the time period of its occurrence, π_i is the coefficient matrix for all lags of VAR model. Σ is the covariance matrix and L is the backshift lag operator $LX = X_{t-1}$.

3.2.2.3 Observed series for Transient change outlier

When an unexpected change occurs on the specific values of a time series that vanishes after a short period of time, we call it as MTC. A MTC can become an MAO or an MLS depending on the value of δ, if δ equals zero, the TC can be considered as an AO. When δ is one, the TC takes the property of a LS. In our study we took δ =0.6, which allows around ine periods of decreasing effects because O. 6^9=0.010. The observed series with TC and its impact on residuals is given below:

$Y_t = X_t + \omega_{TC,h} \alpha_{TC}(B) I^h_t$ … 3.11

$X_t=(\pi(L))^{-1}u_t$, $\alpha_{TC}(L) = (1 – \delta L)^{-1}$, $Y_t= (Y_{1t}. Y_{2t}, Y_{3t}….Y_{rt})^{'}$

$Y_t = (\pi(L))^{-1}u_t + \omega_{TC,h}\, \alpha_{TC}(L)\, I^h{}_t$

$Y_t = (\pi(L))^{-1}u_t + \omega_{TC,h}\, (1 - \delta L)^{-1}\, I^h{}_t$

For in terms of residuals Multiply $\pi(L)$ on both side

$\pi(L)\, Y_t = u_t + \omega_{TC,h}\, \pi(L)\, (1 - \delta L)^{-1}\, I^h{}_t$, $\pi(L)\, Y_t = a_t$

$a_t = u_t + \omega_{TC,h}\, \pi(L)(1 - \delta L)^{-1}\, I^h{}_t$, $\pi(L) = (1 - \pi(L))$

For multivariate $a_t = (1- \sum_{i=1}^{\infty} \pi_i\, (L^i))\omega_{TC,h}\, (1 - \delta L)^{-1}\, I^h{}_t + u_t$

$a_t = ((1 - \delta L)^{-1} I^h{}_t - (1 - \delta L)^{-1} \sum_{i=1}^{\infty} \pi_i\, I^h{}_{t-i})\omega_{TC,h} + u_t$

$a_t = (1 - \delta L)^{-1}\, (I^h{}_t - \sum_{i=1}^{\infty} \pi_i\, I^h{}_{t-i})\omega_{TC,h} + u_t$ ….. 3.12

$u_t \sim N(0, \sum)$ from GLS (Generalized least square estimators) $\omega_{TC,h}$

Because we took δ=0.6 which is less than 1, it makes cleared that a MTC at time index h affects all error term values for $t \geq h$, but the effects falloff gradually as t - h rises.

$\hat{\omega}_{TC,h} = - (1 - \delta L)^{-1} (\sum_{i=0}^{n-h} \pi_i' \ \Sigma^{-1} \ \pi_i)^{-1} \ \sum_{i=0}^{n-h} \pi_i' \ \Sigma^{-1} \ a_{h+i}$, $\pi_o = 1$. 3.13

Where ω is the initial impact of TC o the series and this impact falloffs gradually with time at the rate δ, where $0 < \delta < 1$. π_i is the coefficient matrix for all lags of VAR model. Σ is the covariance matrix and L is the backshift lag operator $LX= X_{t-1.}$

3.2.2.4 Multivariate Seasonal level shift with VAR model

MSLS affect only specific months or the quarters of each year in data after its occurrence. One of the special kinds of multilevel level shift is known as MSLS which occurs in seasonal VAR (0)(1)s at specific time point and repeated its pattern in the every year at same time period, its effect caries up to the subsequent seasons. Observed series with the MSLS and its impact on residuals is given below:

$X_t = (\pi(L))^{-1} u_t$, and $\pi(L)= A(L)A_s(L^s)$ $X_t = (X_{1t}. X_{2t}, X_{3t}\ldots. X_{rt})'$

$A_s(L^s) = 1 - A_1L^s - A_2L^{s2} - \ldots\ldots - A_pL^{sp}$ are the seasonal roots of VAR model.

$A(L) = 1 - A_1L - A_2L^2 - \ldots\ldots - A_pL^p$ are the non-seasonal roots of the VAR model

Now observed series for SLS is

Yt = Xt + At

At = $\omega_{SLS,h}\ \alpha_{SLS}(L)\ I^h{}_t$, $\alpha_{SLS}(L) = (\frac{1}{1-L^s} - \frac{1}{s(1-L)})$, $Y_t =(Y_{1t}, Y_{2t}, Y_{3t}\ldots.Y_{rt})'$

$I^h{}_t$ is a variable we can say it indicator variable, that is , $I^h{}_t$ =1 for t=h and 0 elsewhere. For j= SLS, the magnitude of jth. outlier is ω_j, dynamics of outliers is determent by $\alpha_j(L)$. Y_t is the detected series with MSLS which is given below

$Y_t = X_t$ $t < h$, $Y_t = X_t + A_t$ $t \geq h$, h is the time point when SLS occurs and t is time.

$Y_t =X_t + \omega_{SLS,h}\ \alpha_{SLS}(L)\ I^h{}_t$

$Y_t = (\pi(L))^{-1} u_t + \omega_{SLS,h}\ \alpha_{SLS}(L)\ I^h{}_t$, $\alpha_{SLS}(L) = S(L)$

S(L) computes the dynamic effect of SLS on the series by using a lag operator on the $(\pi(L))^{-1}$ roots. Kaiser and Maravall (2001) was suggested the basic SLS, that is charted by taking $S(L)=\frac{1}{1-L^s}$ and this make an impact on trend for removing this a dynamic weights are defined as $S(L) = (\frac{1}{1-L^s} - \frac{1}{s(1-L)})$ Therefore purely a seasonal effect was produces by the SLS defined, though, this dynamic factor $S(L) = [(1 + L^s + L^{2s} + L^{3s} + \cdots) - 1/S(1 + L + L^2 + L^3 + \cdots)]$

$$S(L) = (1-\frac{1}{S}) - \frac{L}{S} - \frac{L^2}{S} - \ldots + (1-\frac{1}{S}) L^S - \frac{L^S}{S} - \frac{L^{S+1}}{S} - \ldots + (1-\frac{1}{S}) L^{2S} - \frac{L^{2S+1}}{S} - \ldots$$

does not yield $S_o(L) = 1$, therefore there is need to stabilize to dynamic impact which is completed as recommended by Palate (2006), that is, we have to multiply S(L) by the feature $\frac{S}{S-1}$ as $S(L) = \frac{S}{S-1}(\frac{1}{1-L^s} - \frac{1}{s(1-L)})$ simplifying gives $S(L) = \frac{1-\frac{S}{S-1}L+\frac{1}{S-1}L^S}{(1-L^S)((1-L)}$ where S = number of observations per year. Therefore, Equation in univariate time series becomes

$$Y_t = X_t + \omega_{SLS,h} \frac{1-\frac{S}{S-1}L+\frac{1}{S-1}L^S}{(1-L^S)((1-L)} I^h_t$$

The observed series which specifies the existence of SLS effects in univariate time series for numerous seasons at same months or quarter in each year with outlier magnitude ω_{sls}. Is shown in the above equation.

Now in multivariate case this equation for observed series is written as

$$Y_t = (\pi(L))^{-1} u_t + \omega_{SLS,}h[\{(1 - L^S)^{-1} (1-L)^{-1}\}\{1 - S(S-1)^{-1}L + (S-1)^{-1}L^S\}] I^h_t \ldots 3.14$$

$X_t = (\pi(L))^{-1}u_t$, $X_t = (X_{1t}, X_{2t}, X_{3t} \ldots X_{rt})'$

$Y_t = (Y_{1t}, Y_{2t}, Y_{3t} \ldots Y_{rt})'$

The outlier is introduced in the model by generating a variable following Urooj and Asghar(2017)

$\nu_t^T = \alpha SLS(L)\, \varepsilon^h_t$, = 0 for t < T , 1 for for t = T + Sj, $-(S-1)^{-1}$ for t= (T + Sj + 1),…., (T+Sj+S-1) ; j= 0, 1 , 2, 3,….. …… 3.15

Effect of outliers on residuals:-

For in terms of residuals Multiply π(L) on both side

$\pi(L)\, Y_t = u_t + \omega_{SLS,h}\, \pi(L)\, \nu_t^T$

$\pi(L)\, Y_t = a_t$

$a_t = u_t + \omega_{SLS,h}\, \pi(L)\, (-(S-1)^{-1})\, \varepsilon^h_t$ for t= (T + Sj + 1),…., (T+Sj+S-1); j= 0, 1 , 2, 3,…..

$a_t = u_t$ for $t < T$

$a_t = u_t + \omega_{SLS,h}\, \pi(L)$ for t = T + Sj

$\pi(B) = (1 - \pi(L))$

For multivariate at $= (1 - \sum_{i=1}^{\infty} \pi_i (L^i))\, \omega_{TC,h}\, (-(S-1)^{-1}) + u_t$

$a_t = \{(-(S-1)^{-1}) - (-(S-1)^{-1}) \sum_{i=1}^{\infty} \pi_i (L^i)]\, \omega_{SLS,h} + u_t$

$$a_t = -(S-1)^{-1}\{1 - \sum_{i=1}^{\infty} \pi_i (L^i)]\omega_{SLS,h} + u_t \quad \ldots.. 3.16$$

$u_t \sim N(0, \sum)$ from GLS (Generalized least square estimators) $\omega_{SLS,h}$

Here it is clear that a SLS at time index h affects all filtered values at some specific season in each year for $t \geq h$.

$$\widehat{\omega}_{SLS,h} = -(S-1)^{-1} [(\sum_{i=0}^{n-h} \pi_i' \Sigma^{-1} \pi_i)^{-1} \sum_{i=0}^{n-h} \pi_i' \Sigma^{-1} a_{h+i}], \quad \pi_o = 1. \quad \ldots\ldots 3.17$$

Where ω is magnitude of MSLS on the data series and this impact is same for all the observations occurs on the same season in every year after the time period of its occurrence, π_i is the coefficient matrix for all lags of VAR model. Σ is the covariance matrix and L is the backshift lag operator $LX = X_{t-1}$.

3.2.2.5 Observed series for Innovative outlier

For a MIO at time point of outlier h, all the impact of the outliers is confined in $\hat{a}_h$ therefore we evaluate the MIO outlier magnitude by using $\widehat{\omega}_{ih} = \hat{a}_h$ where i shows multivariate innovational outliers.

$$Y_t = X_t + \omega_{IO,h} \alpha_{IO}(L) I^h_t, \quad \alpha_{IO}(L) = \pi(L) \quad \ldots\ldots. 3.18$$

For in terms of residuals

$$a_t = I^h_t \omega_{IO,h} + u_t$$

Therefore, an IO only affects a one filtered value at time of the existence, and we calculate the initial impact of innovative outlier by using $\hat{\omega}_{ih} = \hat{a}_h$

The covariance matrix of estimator is

$\sum_{i,h} = \sum_{i=0}^{n-h} \hat{\pi}_i^{'} \sum^{-1} \hat{\pi}_i$, i=AO, IO, LS, TC, SLS ….. 3.19

To check the worth of outlier at time point h we took the null hypothesis H_o: ω_{IO}=0 versus the substitute H_A: ω_{IO}≠0. If IO is in the current error or shock therefore Innovation outliers can affect future values of the series. On the other hand, AO can affect only the present observation and may result from dictation errors Fox (1972). An IO can affect subsequent observations, and therefore it has lesser impact on parameter estimators and the model selection s compare to additive outlier, the resultant estimators, parameters and the particular model can be quite changed from the true ones with AO.

We shall demonstrate detection of these five types of outliers by an example from real world on the climatic data of Pakistan.

3.2.3 Joint test

Joint tests are often the most appropriate test to use in multivariate hypothesis testing, it considers the all the coefficient of independent component in regression equal zero, if the p-value becomes less than 0.05 at 5% critical value then we reject this hypothesis and concluded that these coefficients are not equal to zero there is some relationship exist among dependent and independent variables.

For example, in our study for detecting SLS in multivariate time series data, we run a multivariate regression through generalized least square estimators for calculating ω_{SLS} initial impact of SLS on the filtered series a_t, given below

$$a_t = -(S-1)^{-1}\{1 - \sum_{i=1}^{\infty} \pi_i(L^i)]\omega_{SLS,h} + u_t$$

$u_t \sim N(0, \sum)$ from GLS (Generalized least square estimators) $\omega_{SLS,h}$

$a_t = a_{1t}, a_{2t}, a_{3t}$, residuals obtained from a multivariate VAR model for trivariate case of model 3.1.

S=12 and S=4, for monthly and quarterly data respectively

π_i are the coefficient matrix for all lags in VAR model, ω_{SLS} are the size of SLS for all three variables in VAR model at time period h.

We will obtain ω_{SLS} of seasonal level shift from this regression then we test the significance of ω_{SLS} through joint test statistics for checking the presence of SLS in our data series. $\omega_{SLS} = \omega_{SLS1}, \omega_{SLS2}, \omega_{SLS,3}$. Null and alternative hypothesis is given below:

H_o: $\omega_{SLS1} = \omega_{SLS2} = \omega_{SLS3} = 0$

H_A: $\omega_{SLS1} \neq \omega_{SLS2} \neq \omega_{SLS3} \neq 0$

Now the joint test statistic which we use to test this hypothesis is given below

$$J_{SLS,h} = \hat{\omega}'_{SLS,h} \Sigma^{-1}_{SLS,h} \hat{\omega}_{SLS,h}$$

Here ω is initial impact of SLS on the series $\Sigma_{\mathrm{SLS,h}}^{-1}$ is its covariance matrix, J_{SLS} follow chi-square distribution with "r" degree of freedom, r= number of variables used in VAR model. J_{SLS} takes all ω jointly, If the p-value of J_{SLS} becomes less than significance level at 5%, then we reject null hypothesis and confirms the presence of SLS in time series data. $J_{SLS,h}$ is the modified test statistics of Tsay's *et al.* (2000) test statistics which we have provide in our thesis. some example for joint test in multivariate analysis are given below:

Uriel, E. (2013) impose hypothesis testing in the multiple regression model by using joint test statistics, Smyth, G. K. (2004) used the Joint null criterion for multiple hypothesis tests, Kodde & Palm (1986) was tested jointly to the equality and inequality restrictions by using wald criteria, Hossain & Majumder (2018) proposed a joint test to test hypothesis, Srivastava & Khatri (2009) discussed joint tests statistics in an introduction to multivariate statistics, Wang & Weiss (2018) was used joint test in the methods of multilevel hypothesis testing and evaluated the significant individual change, Xia *et al.* (2018) have detected false discovery rate control in high-dimensional multivariate regression by using Joint test, Pratikno, B. (2012) was used joint test statistics to test the hypothesis in analysis.

For the case of IO at time point of outlier call h, all information related to outlier is confined in a_h. $\omega_{IO,h}$= a_h, thus, joint test statistics becomes like:

$$J_{IO,h} = \widehat{\omega}_{\mathrm{IO,h}}^{'} \Sigma_{\mathrm{IO,h}}^{-1} \widehat{\omega}_{\mathrm{IO,h}}$$

Null and alternative hypothesis is given below:

H_o: $\omega_{IO1} = \omega_{IO2} = \omega_{IO3} = 0$

H_A: $\omega_{IO1} \neq \omega_{IO2} \neq \omega_{IO3} \neq 0$

IO indicates innovative outlier, $\omega_{IO,h}$ is the size of outlier, $\Sigma_{IO,h}$ is the covariance matrix. Joint test statistics treats all the components of $\omega_{IO,h}$ as a multivariate measure. The $J_{IO,h}$ follow the chi-square distribution with r degrees of freedom for a given time h.

For the case of AO at time index h, for calculating $\omega_{AO,h}$, we used filtered series a_t by using GLS estimators which is explained in section 3.2.2.1, joint test statistics for AO becomes like:

$$J_{AO,h} = \hat{\omega}'_{\mathrm{AO,h}} \Sigma^{-1}_{\mathrm{AO,h}} \hat{\omega}_{\mathrm{AO,h}}$$

Null and alternative hypothesis is given below:

H_o: $\omega_{AO1} = \omega_{AO2} = \omega_{AO3} = 0$

H_A: $\omega_{AO1} \neq \omega_{AO2} \neq \omega_{AO3} \neq 0$

AO indicates additive outlier, $\omega_{AO,h}$ is the size of AO, $\Sigma_{AO,h}$ is the covariance matrix. Joint test statistics treats all the components of $\omega_{AO,h}$ as a multivariate measure. The $J_{AO,h}$ follow the chi-square distribution with r degrees of freedom for a given time h.

For the case of LS at time index h, for calculating $\omega_{LS,h}$, we used filtered series a_t by using GLS estimators which is explained in section 3.2.2.2, joint test statistics for LS becomes like:

$$J_{LS,h} = \hat{\omega}'_{\mathrm{LS,h}} \Sigma^{-1}_{\mathrm{LS,h}} \hat{\omega}_{\mathrm{LS,h}}$$

Null and alternative hypothesis is given below:

H_o: $\omega_{LS1} = \omega_{LS2} = \omega_{LS3} = 0$

H_A: $\omega_{LS1} \neq \omega_{LS2} \neq \omega_{LS3} \neq 0$

LS indicates level shift, $\omega_{LS,h}$ is the size of LS, $\Sigma_{LS,h}$ is the covariance matrix. Joint test statistics treats all the components of $\omega_{LS,h}$ as a multivariate measure. The $J_{LS,h}$ follow the chi-square distribution with r degrees of freedom for a given time h.

For the case of TC at time index h, for calculating $\omega_{TC,h}$, we used filtered series a_t by using GLS estimators which is explained in section 3.2.2.3, joint test statistics for TC becomes like:

$$J_{TC,h} = \widehat{\omega}'_{TC,h}\, \Sigma^{-1}_{TC,h}\, \widehat{\omega}_{TC,h}$$

Null and alternative hypothesis is given below:

H_o: $\omega_{TC1} = \omega_{TC2} = \omega_{TC3} = 0$

H_A: $\omega_{TC1} \neq \omega_{TC2} \neq \omega_{TC3} \neq 0$

TC indicates transient change, $\omega_{TC,h}$ is the size of TC, $\Sigma_{TC,h}$ is the covariance matrix. Joint test statistics treats all the components of $\omega_{TC,h}$ as a multivariate measure. The $J_{TC,h}$ follow the chi-square distribution with r degrees of freedom for a given time h.

Framework of the study

The study will comprise of two main parts simulation experiment and empirical study.

3.2.4 Simulation experiment

We intend to study the existence of various types and outliers including the SLS in multivariate time series its location detection and impact are our main focus. We will use simulation to find out the power of the proposed joint test statistics in detection of MSLS. In this chapter we investigate finite sample empirical critical values and empirical power of the test statistics using simulation and use seasonal $VAR(0)(1)_{12}$ models for obtaining observed quintiles of the joint test statistics for trivariate case and sample size is n= 150 and 200. We generate 1000 realizations. Here number of parameters in VAR model = $r + r^2 p$ here, r is number of components and p is lag numbers, in this study we use 3 number of variables and 12th lag only to compute a seasonal VAR model. We will estimate a VAR model for each realization with proper order by OLS, then will obtain the residuals and covariance matrix and then we will calculate the joint test statistics by using the estimated parameters, their results, then used it to identify the location, magnitude and effect of five types of outliers. Although we will calculate the α and empirical power of the joint test statistics for the case of multivariate SLS.

We employ the $VAR\ (0)(1)_{12}$ model to obtain the empirical quantiles for sample n=150, 200 and $VAR\ (1)(1)_{12}$ model to obtain the empirical quantiles for sample=200 for r=3 "r" is number of variables. We took δ=0.6 generally in most studies δ value taken is 0.6. Two vectors $VAR\ (0)(1)_{12}$ and $VAR\ (1)(1)_{12}$ models for trivariate case are given below:

For VAR $(0)(1)_{12}$

$X_{1t}= A'_{1,1,12}.\ X_{1,t-12} + A'_{1,2,12}.\ X_{2,t-12} + A'_{1,3,12}.\ X_{3,t-12} + u_{1t}$

$X_{2t}= A'_{2,1,12}.\ X_{1,t-12} + A'_{2,2,12}.\ X_{2,t-12} + + A'_{2,3,12}.\ X_{3,t-12} + u_{2t}$

$X_{3t}= A'_{3,1,12}.\ X_{1,t-12} + A'_{3,2,12}.\ X_{2,t-12} + + A'_{3,3,12}.\ X_{3,t-12} +u_{3t}$

$$\begin{bmatrix} x1t \\ x2t \\ x3t \end{bmatrix} = \begin{bmatrix} \text{A } 1,1,12 & \text{A } 1,2,12 & \text{A } 1,3,12 \\ \text{A } 2,1,12 & \text{A } 2,2,12 & \text{A } 2,3,12 \\ \text{A } 3,1,12 & \text{A } 3,2,12 & \text{A } 3,3,1\ 2 \end{bmatrix} \begin{bmatrix} \mathrm{X}_{1,\mathrm{t}-12} \\ \mathrm{X}_{2,\mathrm{t}-12} \\ \mathrm{X}_{3,\mathrm{t}-12} \end{bmatrix} + \begin{bmatrix} \mathrm{u1t} \\ \mathrm{u2t} \\ \mathrm{u3t} \end{bmatrix}$$

For VAR(1)(1)12

$$X_{1t} = A'_{1,1,1}X_{1,t-1} + A'_{1,2,1}X_{2,t-1} + A'_{1,3,1}X_{3,t-1} + A'_{1,1,12}X_{1,t-12} + A'_{1,2,12}X_{2,t-12} + A'_{1,3,12}X_{1,t-12} + u_{1t}$$

$$X_{2t} = A'_{2,1,1}X_{1,t-1} + A'_{2,2,1}X_{2,t-1} + A'_{2,3,1}X_{3,t-1} + A'_{2,1,12}X_{1,t-12} + A'_{2,2,12}X_{2,t-12} + A'_{2,3,12}X_{2,t-12} + u_{2t}$$

$$X_{3t} = A'_{3,1,1}X_{1,t-1} + A'_{3,2,1}X_{2,t-1} + A'_{3,3,1}X_{3,t-1} + A'_{3,1,12}X_{1,t-12} + A'_{3,2,12}X_{2,t-12} + A'_{3,3,12}X_{1,t-12} + u_{1t}$$

$$\begin{bmatrix} x1t \\ x2t \\ x3t \end{bmatrix} = \begin{bmatrix} \text{A}.\ 1,1,1 & \text{A}.\ 1,2,1 & \text{A}.\ 1,3,1 \\ \text{A}.\ 2,1,1 & \text{A}.\ 2,2,1 & \text{A}.\ 2,3,1 \\ \text{A}.\ 3,1,1 & \text{A}.\ 3,2,1 & \text{A}.\ 3,3,1 \end{bmatrix} \begin{bmatrix} \mathrm{X}_{1,\mathrm{t}-1} \\ \mathrm{X}_{2,\mathrm{t}-1} \\ \mathrm{X}_{3,\mathrm{t}-1} \end{bmatrix} + \begin{bmatrix} \text{A}.\ 1,1,12 & \text{A}.\ 1,2,12 & \text{A}.\ 1,3,12 \\ \text{A}.\ 2,1,12 & \text{A}.\ 2,2,12 & \text{A}.\ 2,3,12 \\ \text{A}.\ 3,1,12 & \text{A}.\ 3,2,12 & \text{A}.\ 3,3,1\ 2 \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{X}_{1,\mathrm{t}-12} \\ \mathrm{X}_{2,\mathrm{t}-12} \\ \mathrm{X}_{3,\mathrm{t}-12} \end{bmatrix} + \begin{bmatrix} \mathrm{u1t} \\ \mathrm{u2t} \\ \mathrm{u3t} \end{bmatrix}$$

3.3 **Empirical analysis**

In Empirical analysis we use one real data example on monthly time series data with three variables of Pakistan. We use rainfall, temperature and "Precipitation and Humidity Altitude" data from 2008 to 2020 for Lahore, Faisalabad, and Karachi Pakistan and here we have 151 observations in the data. It is observed that seasonality in temperature of these cities is associated with the seasonality in rainfall and humidity which intern affects badly to the ecosystem of these cities.

Pakistan is situated in a region which is prone to climate change where both summer and winter rainfalls occurred. Seasonality exist in climatic data of Pakistan there may exist the possibility of SLS and detection of SLS in this data is very useful if predicted well in time. Issuing of accurate forecast of seasonal rainfall might be very useful. The abnormalities in rainfall, temperature and humidity at any area seriously affect the ecosystem of that area. The duration of summer season in Pakistan is about six months comprising of two major sub seasons pre-monsoon (April-June) and Monsoon (July-September) and the duration of winter season in Pakistan is about five months (November to march). Various studies are found whose work on ecosystem, climate change like SR khan (2001), Kiran and Ain (2006), Akram and Hamid (2015), Iqbal *et al*. (2014), Hussain and Mustafa (2016), Saifullah (2017), worked on climate change in Pakistan, Inalizadeh *et al*. (2011) worked on importance of outlier detection in spatial analysis of wind erosion in Iran, Asghar and Urooj (2012), have detected structural breaks by using automatic model selection and forecasted the wheat and rice prices for Pakistan, Irshad and Hussain (2017) worked on analysis of ecological efficiency and its influencing factors in developing countries, Shah *et al*. (2017), worked on minimum temperature analysis and trends in Pakistan, Rahman and Hasan (2017) worked on modeling and forecasting of carbon dioxide emissions in Bangladesh but no one worked for detecting seasonal level shift in multivariate time series. In this study we will detect seasonal level shift in multivariate time series by using rainfall, temperature and humidity data of Lahore, Faisalabad and Karachi. Then we will explore the location, magnitude, size and impact of seasonal level shift in rainfall, temperature and humidity and its impact on ecosystem of Lahore by linking all these with each other. We will apply the modified Tsay *et al.* (2000) technique for detection of outliers. The technique is modified for the inclusion of SLS in multivariate time series and we will use seasonal VAR model.

Detection of SLS may play an important role in seasonal data analysis, as like in pas we can see that changing seasons affected the air quality as a result of the increased or decreased diffusion of Impurities Li and Lin (2003); Ramachandran *et al*. (2003); Hanninen *et al*. (2011). Variation in the concentration of particulates over the seasons also caused by the disparity in the wind swiftness, comparative humidity and temperature.

Climate change significantly affects the water possessions, agriculture, forests, biodiversity, environment, and health sectors which eventually affect the socio-economic of a country. Climate has great effect on the construction for its electricity ingesting and building performance. During summer and winter season the overall heating and cooling necessities in the buildings is measured by climate.

climate change has the most significant impact on Pakistan out of the all countries. A UK-based worldwide risk accessing firm named Verisk Maplecroft, has graded Pakistan on 22nd number in the word, who is most affected by the Climate Change, by Vulnerability Index 2016 (CCVI); and three cities of Pakistan are amongst the 69 well thought-out highly affected by climate change named Lahore on 7^{th} place, Faisalabad on 22^{nd} and Karachi on 25^{th}.

these weather conditions harmfully affected the agriculture system of the Pakistan. Climate change impacts includes high temperatures, seasonal droughts, hefty rains, torrents, infrequent rains, lacking seasonal rainfall, viruses, scarcity of fresh drinking water and allergy, he said, water sector, agriculture sector, tourism sector and health sectors etc. affected by these climate changes.

Lahore faced huge level of the seasonal discrepancy in monthly rainfall. Also, Lahore experiences dangerous seasonal discrepancy in the supposed humidity. changes in climate badly affected the Pakistan. This condition of Pakistan is due to two reasons. 1^{st} is geological location of Pakistan is on the world map and 2^{nd} is due to changing climate. Several sectors and aspects of Pakistan are under serious danger. Ali (2013) was explored that climate change affected the agricultural sector in terms of fluctuation in temperature and rainfall, population sensitivity depends on water, human migration crisis, food and shelter, coastal belts occurs due to rise in level of sea, glaciers are melted because of rise in temperature etc.

Data Description

Here we attempt to measure SLS in rainfall due to SLS in humidity and temperature which causes disturbances in ecosystem of Lahore, Faisalabad and Karachi. In light of the above deliberations, this study aims to detect seasonal level shift and non-seasonal outliers in climate change (temperature, rainfall, humidity) and to examine the cause of these outlier and give suggestion that how to deal with these outliers.

We took data from a website world weather. Online

1TABLE 3.1 DATA DESCRIPTION

Variables	Frequency	Source
Rainfall	mm/ month	CDPC of PMD/ world weather. Online
Temperature	$°_C$ / month	CDPC of PMD/ worldweather. online
Humidity	Percentage / month	CDPC of PMD/ worldweather. online

CHAPTER 4

SIMULATION EXPERIMENT

In this study our aim is to provide a modified Tsay's multivariate outlier's detection method by including seasonal level shift s (SLS). Different researchers have built various models and techniques to detect multivariate outlier in time series but up to our knowledge the seasonal level shift is not explored in multivariate structure up till now, we familiarized seasonal level shift in multivariate time series using seasonal VAR model and applied the procedure as suggested by Tsay *et al.* (2000) grounded on one test statistics for detection of multivariate outliers i.e. joint maximum test statistics.

In simulation analysis we use statistical outlier detection method which rely on the statistical approaches that assume a distribution or probability model to fit the given dataset. Under the distribution assumed to fit the dataset, the outliers are those points that do not agree with or conform to the underlying model of the data. By using simulation, we will test the power and size of test statistics for different order and sample size, we will check that how much work good used test statistics for outlier detection.

We explored five types of outliers in multivariate time series framework named as multivariate additive outlier (MAO), multivariate innovative outlier (MIO), multivariate level shift (MLS), multivariate transient change (MTC) and multivariate seasonal level shift (MSLS). We have identified the effect of existence of multivariate SLS on multivariate time series, suggested its detection procedure and examined the performance of the suggested method in terms of power and size impact on estimates using simulation and calculated the critical values for all five types of outlier for sample size 150, 200 using seasonal and non-seasonal VAR. Power is a theoretical concept and we see it in repeated sample and used for checking the performance of test. The power of a statistical test measures the test's ability to detect a specific alternate hypothesis. In general, the power of a test *is* the probability that the test will reject the null hypothesis when a specified alternative is true. In case of real world when we use test statistic to detect the outliers and this test detected outlier in the real-world data then we compare the detected outlier in the data and history of real world. If in real world this unusual observation exists, it means that our test performance is good if outliers not exist in that time period in real world scenario then it means that this is the weakness of test statistics and here test statistic detected outlier erroneously. We explored all five types of outlier including seasonal level shift (SLS) by using one real data example on climatic data of Pakistan.

While by modifying the Tsay's *et al.* (2000) technique for multivariate outlier detection, we will explore five types of outlier named AO, IO, LS, TC, SLS directly in multivariate time series, also examined the presentation of suggested procedure for outlier recognition in multivariate time series by using simulation and time series data of Pakistan. We used data of rainfall, temperature and humidity for three stations Lahore, Faisalabad and Karachi of Pakistan, applied seasonal model and Tsay's *et al*. (2000) technique to detect outlier in multivariate time series. Data will be taken from January 2008 to July 2020, we have 151 total number of observations.

Seasonal VAR model structure following "time series analysis and its applications" third edition by robert H. shumway & david S. stoffer

$$(1 - \pi L^{12})\, x_t = u_t \qquad \ldots \qquad 4.1$$

x_t- πx_{t-12} = u_t

x_t = πx_{t-12} + u_t

For multivariate

(1- $\sum_{i=1}^{\infty} \pi_i$ L^i) $x_{t,}$= u_t

x_t= $\sum_{i=1}^{\infty} \pi_i$ L^i) x_t + u_t ……. i=s, s=12 for monthly data

x_t= ($\sum \pi_{12}$ L^{12}) x_t + u_t ….x_t=.x_{1t} , x_{2t}, x_{3t} … 4.2

4. Simulation Results

In simulation we estimated a seasonal and no seasonal VAR model to obtain the empirical significance level, empirical power, test statistics empirical critical values and we checked the impact of outliers on estimates and residuals with different sample size, and also provide a modified a Tsay's multivariate outlier detection method based on one test statistics join test statistics by including SLS in multivariate structure. The detailed analysis is given below:

4.1 For Sample size 150

In the below, we have generated 1000 realization for trivariate case with n=150 by normal distribution.

4.1.1 Simulation when no outlier in the series

In start, assuming there is no outlier we have estimated a multivariate seasonal VAR$(0)(1)_{12}$ with 12[th] lag only model then obtained residuals u_{it} and covariance matrix $\sum$ and coefficient matrix π then we have estimated outlier size ω and covariance matrix of estimators then we have calculated joint test statistics and then we have calculated joint maximum test statistics for all five types of outliers. These series are free of outliers because we have generated them from normal distribution, we have done all these process on these series by using seasonal $\mathrm{VAR}(0)(1)_{12}$ for calculating the empirical significance level/size. First of all, we have calculated joint test statistics for five types of outliers AO, TC, LS, SLS and then compare their results by joint maximum test and then finalized the results.

Type I error: rejecting true null = α

Empirical level of significance=size=number of detected outlier when there is no chance of outlier/total number of iterations

2Table 4.1 Empirical level of significance

Empirical level of significance Sample size: 150 realization: 1000						
		AO	IO	LS	TC	SLS
No of outlier detected		19	55	14	555	147
Empirical level of significance	joint test statistic	0.019	0.055	0.014	0.555	0.147
	For overall joint maximum test statistics	0.547				

Note: (detailed tables of these results will be provided on demand)

Here this table show the erroneously detection of outlier when there is no outie in the series.

4.1.2 Simulation with outlier

We have generated 3 data series of sample n=150 by normal distribution in RStudio and make 1000 realization of trivariate then we have added five types of outlier on same time point at t=85. We have calculated seasonal VAR(0)(1)12 model for each realization and obtained residuals, covariance matrix, coefficient matrix for each equation then we calculated initial impact of outlier named as ω_{AO} , ω_{LS} , ω_{TC} , ω_{SLS} by using GLS(generalized least square) estimators and calculated $\omega_{IO,h} = a_{t,h}$ as suggested in procedure above in chapter 3, and their covariance matrix and then we have calculated joint test statistics following chi-square distribution for five types of outliers names j_{AO}, j_{IO}, j_{LS}, j_{TC}, j_{SLS} for each realization, at which time point p-value of test statistic value is minimum and probability of that statistic value is less than 0.05 we considered it as outlier in the series, then we have compare these four type outliers statistics with each other by the rule of joint maximum test statistic, in each realization which one outlier statistic is greater than other statistic value we declared that type of outlier in this realization.

Joint test statistic results

Null hypothesis H_o: there is no outlier in the series

Alternative hypothesis H_A: there is outlier in the series

This is false null because there is outlier in the series which we have added on the specific time point for checking the test statistic for detection of outlier. Therefore, in this test there is a possibility of committing

Type II error: rejection of false null.

The Significance level used is α=0.05.

This Joint test is following the chi-square distribution, when calculated value is larger than critical value or probability is less than 0.05, we have reject the null and declared that there is outlier in the series, at which time point p-value of test statistic is minimum and less than 0.05 that time point is considered as outlier as suggested by Tsay *et al.* (2000).

1TABLE 4.2 POWER OF THE FIVE JOINT TEST STATISTICS

Empirical power of the test statistics sample size: 150 significance level: 0.05 realization: 1000						
		AO	IO	LS	TC	SLS
No of outlier detected		605	998	730	683	623
Empirical power	joint test statistic	60.5%	99.8%	73%	68.3%	62.3%
	For overall joint maximum test statistics	99.8%				
β = no outlier detected when there is outlier in the series	joint test statistic	0.395	0.002	0.27	0.317	0.377
	For overall joint maximum test statistics	0.002				

2TABLE 4.3 EMPIRICAL QUANTILES OF THE FIVE J_{MAX} (I, H_I) STATISTICS

Sample Size	Test	Probabilities 50%	75%	90%	95%	97.5%	99%
Trivariate case		$VAR(0)(1)_{12}$					
150	J_{max} (AO, h_{AO})	17.8499	26.469	42.335	50.287	60.118	74.769
	J_{max} (IO, h_{IO})	38.1253	44.989	50.753	54.954	57.424	60.565
	J_{max} (LS, h_{LS})	13.566	17.495	21.403	28.1	26.483	30.939
	J_{max} (TC, h_{TC})	18.42	33.49	53.59	63.79	75.12	90.64
	J_{max} (SLS, h_{SLS})	16.5737	27.564	46.443	64.514	76.444	99.776

Here i= AO, IO, LS, TC, SLS defined above in chapter 3, based on 1000 realizations the model used are seasonal VAR (0)(1)12 given above in chapter 3.

From this table we make this explanation. First as usual, empirical critical values of J_{max}(LS, hLS) are lesser as compare to other four joint test. The J_{max} (i, h_i) for i= MAO, MIO are nearer to each other retaining that a common critical value we can use for these two test. J_{max}(i, h_i) for i= TC, SLS are closer retaining that a common critical value we can use for these test for AO and IO, for TC and SLS. Our simulations suggest that 52.62 for AO and IO, 64.152 for TC and SLS with 5% level of significance can be used as estimated critical value for sample n=150. In short, the empirical quantile of joint test statistics, particularly for J_{max} (LS, h_{LS}) depends on the sample size, dimension and the model structure. From these results we obtained that both the Theory and practice for this test statistics needed more work. Here we can see clearly that for sample size 150 with seasonal VAR model, the multivariate IO critical values are not much confused with multivariate SLS critical values these are have a significance difference. The same concept was observed for univariate SLS by Asghar and Urooj (2017) that, in SAR (1) SLS is generally not confused with other types of outliers.

Note: (detailed results will be provided on demand)

4.1.2.1 Coefficient matrix for VAR $(0)(1)_{12}$ with and without outliers

These results are based on 1000 times estimated parameters of model, by sampling distribution, the matrix of expected coefficient, expected (π) matrix, expected covariance matrix, eigen values vector, with and without outliers in trivariate case. VAR is used for multilevel time series. The structure of VAR is that each component of VAR model is a linear function of its past lags and the past lags of other variables. In the covariance matrix the diagonal entries are the variances and the other entries are the covariances. Hence, the covariance matrix is sometimes called the variance-covariance matrix. The covariance matrix can be also expressed as

$$\sum= \begin{bmatrix} \sigma\ (x1, x1) & \sigma\ (x1, x2) & \sigma\ (x1, x3) \\ \sigma\ (x2, x1) & \sigma\ (x2, x2) & \sigma\ (x2, x3) \\ \sigma\ (x3, x1) & \sigma\ (x3, x2) & \sigma\ (x3, x3) \end{bmatrix}$$

When diagonal elements are equal 1and off diagonal elements are equal 0, this means that our data series are white noise and variables are independent of each other in case of residuals we can say that the error term are independent of each other.

4.1.2.1.1 Coefficient and covariance matrix for VAR (0)(1)12 model without outliers

Covariance matrix of residuals

$$\sum=\begin{bmatrix} 0.975 & 0.003 & 0.0003 \\ 0.003 & 0.976 & -0.002 \\ 0.0003 & -0.002 & 0.977 \end{bmatrix}$$

Π_{12} matrix (with 12th lag)

$$\Pi_{12}=\begin{bmatrix} -0.00165 & -0.00104 & -0.00176 \\ -0.00358 & 0.000367 & -0.00288 \\ -0.00799 & -0.00291 & -0.0000644 \end{bmatrix}$$

Covariance matrix of estimators for Transient change outlier

$$\sum=\begin{bmatrix} 0.949 & -0.00078 & -0.000307 \\ -0.000786 & 0.998 & 0.0233 \\ -0.000307 & 0.0233 & 1.0087 \end{bmatrix}$$

Covariance matrix of estimators for Level Shift

$$\sum=\begin{bmatrix} 0.948 & 0.0019 & -0.00148 \\ 0.00193 & 1.00131 & 0.0241 \\ -0.00148 & 0.0241 & 1.0032 \end{bmatrix}$$

Covariance matrix of estimators for Additive outlier

$$\sum=\begin{bmatrix} 0.951 & 0.0025 & -0.00051 \\ 0.0025 & 0.999 & 0.0215 \\ -0.000510 & 0.0215 & 1.00048 \end{bmatrix}$$

Covariance matrix of estimators for Seasonal Level Shift

$$\sum=\begin{bmatrix} 0.951 & -0.000017 & -0.00167 \\ -0.000017 & 1.00023 & 0.0232 \\ -0.00167 & 0.0232 & 1.00383 \end{bmatrix}$$

The coefficients matrix has eigen values (-0.00636, 0.00353, 0.00148,)

Here these results show that all diagonal entries of covariance matrix entries are almost equal 1 and off diagonal entries are almost equal to zero this show the white noise of data, in coefficient matrix for 12th lag of the model are closer to zero this means that in this multivariate data all the variables not depending on its own lag value and lag value of other variable in data this also show the normality of data.

In the below we show the results of seasonal VAR model with outliers.

4.1.2.1.2 Coefficient and covariance matrix for VAR (0)(1)12 model with outliers

For additive outlier

Coefficient matrix

$$\Pi_{12}=\begin{bmatrix} -0.0027 & 0.00556 & -0.00391 \\ -0.0021 & 0.00646 & -0.0037 \\ -0.00164 & 0.0054 & -0.00544 \end{bmatrix}$$

Covariance matrix of residuals for Additive outlier

$$\sum=\begin{bmatrix} 9.402 & 8.423 & 8.434 \\ 8.423 & 9.415 & 8.439 \\ 8.434 & 8.434 & 9.429 \end{bmatrix}$$

Covariance matrix of estimators for Additive outlier

$$\sum=\begin{bmatrix}9.946 & 8.986 & 8.997\\ 8.986 & 9.994 & 9.01\\ 8.997 & 9.01 & 10.009\end{bmatrix}$$

The coefficient matrix Π_{12} have eigen values (-0.00398, 0.003357, -0.00105).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12th lag of the model are closer to zero but have a little difference than the results with normal data this means that additive outlier have a little bit impact on coefficient of VAR model but badly affects the covariance matrix.

For Innovative outlier

Coefficient matrix

$$\Pi_{12}=\begin{bmatrix}-0.367 & 0.037 & -0.0409\\ -0.0404 & 0.367 & -0.0403\\ -0.0411 & 0.037 & -0.366\end{bmatrix}$$

<u>Covariance matrix of residuals for Innovative outlier</u>

$$\sum=\begin{bmatrix}1.094 & 0.117 & 0.121\\ 0.117 & 1.096 & 0.1207\\ 0.121 & 0.1207 & 0.1207\end{bmatrix}$$

<u>Covariance matrix of estimators for Innovative outlier</u>

$$\sum=\begin{bmatrix}1.094 & 0.117 & 0.121\\ 0.117 & 1.096 & 0.1207\\ 0.121 & 0.1207 & 0.1207\end{bmatrix}$$

The coefficient matrix Π_{12} have eigen values (0.446, 0.329, 0.325).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a small difference as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise but it show that innovative outlier have little bit impact on covariance matrix than AO, coefficient matrix for 12th lag of the model are closer to zero but have a little difference than the results with normal data this means that IO have a little bit impact on coefficient of VAR model.

For Level Shift

Coefficient matrix

$$\Pi 12=\begin{bmatrix} 0.327 & 0.333 & 0.329 \\ 0.327 & 0.329 & 0.331 \\ 0.331 & 0.333 & 0.324 \end{bmatrix}$$

<u>Covariance matrix of residuals for Level Shift</u>

$$\sum=\begin{bmatrix} 7.65 & 6.67 & 6.65 \\ 6.67 & 7.66 & 6.66 \\ 6.65 & 6.66 & 7.61 \end{bmatrix}$$

<u>Covariance matrix of estimators for Level Shift</u>

$$\sum=\begin{bmatrix} 8.585 & 7.639 & 7.613 \\ 7.639 & 8.659 & 7.651 \\ 7.613 & 7.651 & 8.603 \end{bmatrix}$$

The coefficient matrix Π_{12} have eigen values (0.989, -0.0059, -0.0025).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12th lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that level shift has a huge impact on coefficient of VAR model, also badly affects the covariance matrix.

For Transient Change

Coefficient matrix

$$\Pi 12=\begin{bmatrix} 0.226 & 0.229 & 0.254 \\ 0.230 & 0.222 & 0.257 \\ 0.229 & 0.228 & 0.254 \end{bmatrix}$$

<u>Covariance matrix of residuals for Transient Change outlier</u>

$$\sum=\begin{bmatrix} 25.098 & 24.128 & 24.103 \\ 24.128 & 25.102 & 24.1006 \\ 24.103 & 24.1006 & 25.059 \end{bmatrix}$$

<u>Covariance matrix of estimators for Transient Change outlier</u>

$$\sum=\begin{bmatrix} 29.054 & 28.113 & 28.078 \\ 28.113 & 29.123 & 28.106 \\ 28.078 & 28.106 & 29.049 \end{bmatrix}$$

The coefficient matrix Π12 have eigen values (0.711, -0.006, -0.0013).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that TC have a huge impact on coefficient of VAR model, also badly affects the covariance matrix.

For Seasonal level shift

Coefficient matrix

$$\Pi 12=\begin{bmatrix} 0.3003 & 0.310 & 0.3061 \\ 0.3036 & 0.3031 & 0.3097 \\ 0.3039 & 0.310 & 0.302 \end{bmatrix}$$

Covariance matrix of residuals for Seasonal level shift Level Shift

$$\sum=\begin{bmatrix} 1.981 & 1.0003 & 1.0127 \\ 1.0003 & 1.971 & 1.0009 \\ 1.0127 & 1.0009 & 1.993 \end{bmatrix}$$

Covariance matrix of estimators for Seasonal Level Shift

$$\sum=\begin{bmatrix} 1.966 & 1.0062 & 1.0151 \\ 1.0062 & 2.001 & 1.028 \\ 1.0151 & 1.028 & 2.016 \end{bmatrix}$$

The coefficient matrix Π_{12} have eigen values (0.916, -0.00712, -0.00373).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a small difference as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that SLS have a huge impact on coefficient of VAR model, small effect on covariance matrix.

From these results we make these observations, firstly, Π weights of $VAR(0)(1)_{12}$ model with n=150 for data series with outliers becomes much larger than the Π weights of $VAR(0)(1)_{12}$ for data series without outlier and have changed sign, this difference clearly identified that all five types of outliers badly and significantly affects the estimation and estimated parameters of VAR model and also affects results which can make our results doubtful. Π weights of VAR $(0)(1)_{12}$ with seasonal level shift also have clearly much larger weights than the Π weights with normal data series, this means that with other types of outliers there is necessary to detect and adjust the SLS in the data series in another case this can distort the whole analysis and can make results doubtful.

Secondly, for sample size Π weights with additive outlier have very small difference than the Π weights with normal data this is less sensitive for our analysis results than those of other four types of outliers, Π weights with LS outlier are much larger than the Π weights with normal data and also much larger than the Π weights with other four types of outliers. Hence LS outlier is much sensitive for our analysis, results, estimates and make results doubtful than other types of outlier. SLS, IO and TC have much larger Π weights than the Π weights with normal data but have less than the LS Π weights, SLS, IO and TC also much sensitive for our analysis, estimates and results but less sensitive than the LS outlier with sample size n=150. However, there is clearly seen that estimated coefficient and covariance matrix of innovational outlier are not much closer to seasonal level shift we can say that for sample size n=150 with seasonal VAR model we are not confused in innovational outlier and SLS. For making results clearer further, we will use large sample size for these outliers in multivariate structure. As mentioned by Asghar and Urooj (2017) univariate SLS detection give more better results for large sample size.

Thirdly, the eigen vectors of Π weights with AO have less difference than with the normal data, this also identified that AO is less sensitive for analysis than other type of outliers. The eigen vectors for Π weights with LS have large difference than with the normal data and also larger than the eigen vectors for Π weights with other type of outlier, this also clarify that LS is highly sensitive for our analysis and results. SLS, IO and TC outlier eigen vectors for Π weights also have significant difference than with normal data these are also much sensitive for our results of analysis but less sensitive than the level shift.

Fourthly, covariance matrix of residuals and covariance matrix of estimators for all five types of outliers have much larger difference than for the normal data. Outliers also badly affects the covariance structure and make results doubtful.

In the below we have simulated the data with seasonal $VAR(0)(1)_{12}$ model and sample size n=200.

4.2 For Sample size 200 with VAR $(0)(1)_{12}$

In the below, we have generated 1000 realization for trivariate case with n=200 by normal distribution.

4.2.1 Simulation when no outlier in the series

In start assuming there is no outlier we have estimated a multivariate seasonal VAR$(0)(1)_{12}$ with 12^{th} lag only model then obtained residuals u_{it} and covariance matrix of residuals $\sum$ and coefficient matrix π then I have estimated outlier size ω_i and covariance matrix of estimators then we have calculated joint test statistics and then we have calculated joint maximum test statistics for all five types of outliers. These series are free of outliers because we have generated them from normal distribution, we have done all these process on these series by using seasonal VAR (0)(1)12 for calculating the empirical significance level/size. First of all, we have calculated joint test statistics for five types of outliers AO, IO, TC, LS, SLS and then compare their results by joint maximum test and then

Type I error: rejecting true null

Empirical level of significance=size=no of detected outlier when there is no chance of outlier/total number of iterations.

3TABLE 4.4 EMPIRICAL LEVEL OF SIGNIFICANCE

Sample size: 200 Level of significance: 0.05 realization: 1000						
		AO	IO	LS	TC	SLS
no of outlier detected		65	45	3	445	50
Empirical level of significance	joint test statistic	0.065	0.045	0.003	0.445	0.05
	For overall joint maximum test statistics	0.493				

By comparing these results with the results of 4.2.1 we have observed that empirical level of significance has decline for all types of outliers except AO, by increasing sample size for seasonal VAR $(0)(1)_{12}$ model. We have concluded that empirical level of significance for SLS and all other four types of outliers heavily depends upon the sample size.

(detailed tables of these results will be provided on demand)

4.2.2 Simulation with outliers

We have generated 3 data series of sample n=200 by normal distribution and make 1000 realization of trivariate then we have added five types of outlier on same time point at t=85. We have calculated seasonal $VAR(0)(1)_{12}$ model for each realization and obtained residuals, covariance matrix, coefficient matrix for each equation then we calculated

ω_{AO}, ω_{LS}, ω_{TC}, ω_{SLS} by using GLS(generalized least square) estimators and calculated $\omega_{IO,h}= a_{t,h}$ as suggested in procedure above in chapter 3 and their covariance matrix and then we have calculated joint test statistics following chi-square distribution for four types of outliers names $j_{AO}, j_{IO}, j_{LS}, j_{TC}, j_{SLS}$ for each realization, at which time point p-value of test statistic value is minimum and probability of that statistic value is less than 0.05 we considered it as outlier in the series at that time point, then we have compare these four type outliers statistics results with each other by the rule of joint maximum test statistic, in each realization which one outlier statistic is greater than other statistic value we declared that type of outlier in this realization.

Joint test statistic results

Null hypothesis. H_o : there is no outlier in the series

Alternative hypothesis. H_A : there is outlier in the series

This is fall null because there is outlier in the series which we have added on the specific time point for checking the test statistic for detection of outlier.

Type II error: rejection of false null.

Significance level α=0.05

Joint test statistics follow chi-square distribution, when calculated value is larger by comparing the critical value or probability value is less than 0.05 then we will reject the null and there is outlier in the series, at which time point p-value of test statistic is minimum that time point is considered as outlier as suggested by Tsay *et al.* (2000).

4TABLE 4.5 POWER OF THE FIVE JOINT TEST STATISTICS

sample size: 200 significance level: 0.05 realization: 1000						
		AO	IO	LS	TC	SLS
no of outlier detected		745	1000	894	983	899
Empirical level of significance of test	joint test statistic	74.5%	100%	89.4%	98.3%	89.9%
	for over joint maximum test statistics	100%				
β = no outlier detected when there is outlier in the series	joint test statistic	0.255	0	0.106	0.017	0.101
	for over joint maximum test statistics	0				

From this table comparing with table 4.2 we observed that β, power of the J test depends on sample size with increasing sample size therefore power of the test statistics increased and β value declined. We have also clearly observed that empirical power of test statistics of multivariate joint test statistics for SLS have much difference for large sample size and have 89.9% empirical power, therefore we can say that joint test statistics for multivariate SLS is a good test with large sample size.

5TABLE 4.6 EMPIRICAL QUANTILES OF THE FIVE J_{MAX} (I, H_I) STATISTICS

Sample	Test	Probabilities					
Size		50%	75%	90%	95%	97.5%	99%
Trivariate case		$VAR(0)(1)_{12}$					
200	J_{max} (AO, h_{AO})	28.11	35.909	50.02	58.43	59.67	89.87
	J_{max} (IO, h_{IO})	39.013	45.425	51.805	55.294	58.424	61.177
	J_{max} (LS, h_{LS})	20.14	23.49	28.44	30.84	32.16	35.54
	J_{max} (TC, h_{TC})	22.15	36.99	53.65	69.24	76.58	91.23
	J_{max} (SLS, h_{SLS})	24.776	42.923	70.469	85.687	109.26	120.67

Here i= AO, IO, LS, TC, SLS defined above based on 1000 realizations the model used are seasonal VAR (0)(1)12 given above in chapter 3.

From this table we make this explanation. 1st as usual, empirical critical values of J_{max}(.LS, h_{LS}) are lesser as compare to other four joint test. $J_{max.}$(i, h_i) for i= AO, IO are closer, retaining that a common critical value we can use for these two test. $J_{max.}$(i, $h_{i.}$) for i= TC, SLS are closer, retaining that a common critical value we can use for these two test, our simulations suggest that 56.862 for AO and IO, 77.46 for TC and SLS can be used as estimated critical value for sample n=200. This approximate critical value increases with increasing the sample size and also mostly the empirical critical values are increased with increase in sample size, in short the empirical critical values of the J test statistics, especially for $J_{max.}$(.LS, $h_{LS.}$) depends on the sample size, dimension and the model structure. From these results we have observed that both the Theory and practice of this test needed attention. Here we can see clearly that for sample size 200 with seasonal VAR model, the multivariate IO critical values are not much confused with multivariate SLS critical values these are have a significance difference. The same concept was observed for univariate SLS by Asghar and Urooj (2017) that, in SAR (1) SLS is generally not confused with other types of outliers.

Therefore, we concluded that SLS with large sample size and with seasonal VAR model for monthly seasonal frequency is not much confused with other types of outlier and have significant impact on all estimates.

4.2.2.1 Coefficient matrix for VAR $(0)(1)_{12}$ with and without outliers

These results are based on 1000 times estimated parameters of model, by average sampling distribution, we have taken expected matrix of 1000 matrix for coefficient matrix/π matrix, covariance matrix, eigen values vector, before and after adding outliers in trivariate case.

4.2.2.1.1 Coefficient and covariance matrix for VAR $(0)(1)_{12}$ model without outliers

Covariance matrix of residuals

$$\sum=\begin{bmatrix} 0.984 & -0.0027 & -0.0021 \\ -0.0027 & 0.985 & -0.0013 \\ -0.0021 & 10.0013 & 0.987 \end{bmatrix}$$

Π12 matrix (with 12^{th} lag)

$$\Pi 12=\begin{bmatrix} 0.00101 & 0.0011 & 0.0015 \\ 0.00023 & 0.00105 & -0.00058 \\ 0.00205 & -0.0023 & -0.00108 \end{bmatrix}$$

Covariance matrix of estimators for Transient change outlier

$$\sum=\begin{bmatrix} 1.07 & 0.04 & 0.09 \\ 0.04 & 1.07 & 0.09 \\ 0.09 & 0.09 & 1.08 \end{bmatrix}$$

Covariance matrix of estimators for Level Shift

$$\sum=\begin{bmatrix} 0.967 & -0.00204 & -0.00027 \\ -0.00204 & 1.00322 & 0.0164 \\ -0.00027 & 0.0164 & 1.0051 \end{bmatrix}$$

Covariance matrix of estimators for Additive outlier

$$\sum=\begin{bmatrix} 0.968 & -0.00251 & -0.00022 \\ -0.00251 & 1.0009 & 0.01509 \\ -0.000228 & 0.01509 & 1.0041 \end{bmatrix}$$

Covariance matrix of estimators for Seasonal Level Shift

$$\sum=\begin{bmatrix} 0.967 & -0.0030 & 0.00062 \\ -0.0030 & 1.0004 & 0.0151 \\ 0.00062 & 0.0151 & 1.0054 \end{bmatrix}$$

Here these results with large sample size also show that all diagonal entries of covariance matrix entries are almost equal 1 and off diagonal entries are almost equal to zero this show the white noise of data, in coefficient matrix for 12^{th} lag of the model are closer to zero this means that in this multivariate data all the variables not depending on its own lag value and lag value of other variable in data this also show the normality of data.

In the below we show the results of seasonal VAR model with outliers.

4.2.2.1.2 Coefficient and covariance matrix for VAR (0)(1)$_{12}$ model with outliers

For additive outlier

Coefficient matrix

$$\Pi_{12}=\begin{bmatrix} -0.00379 & 0.000242 & 0.00272 \\ -0.00908 & 0.00065 & 0.00763 \\ -0.00899 & -0.00138 & 0.0101 \end{bmatrix}$$

Covariance matrix of residuals for Additive outlier

$$\sum=\begin{bmatrix} 5.72 & 4.73 & 4.73 \\ 4.737 & 5.70 & 4.73 \\ 4.73 & 4.73 & 5.72 \end{bmatrix}$$

Covariance matrix of estimators for Additive outlier

$$\sum = \begin{bmatrix} 5.99 & 5.01 & 5.02 \\ 5.01 & 6.01 & 5.03 \\ 5.02 & 5.03 & 6.02 \end{bmatrix}$$

The coefficient matrix Π_{12} have eigen values (0.0062, 0.0016, -0.00091).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag of the model are closer to zero but have a little difference than the results with normal data this means that additive outlier have a little bit impact on coefficient of VAR model but badly affects the covariance matrix.

For Innovative outlier

Coefficient matrix

$$\Pi 12 = \begin{bmatrix} -0.0009 & 0.0057 & -0.000203 \\ 0.000767 & 0.000644 & 0.00501 \\ -0.00708 & 0.0054 & -0.00186 \end{bmatrix}$$

Covariance matrix of residuals for Innovative outlier

$$\sum = \begin{bmatrix} 1.069 & 0.082 & 0.083 \\ 0.083 & 1.061 & 0.082 \\ 0.083 & 0.0826 & 1.061 \end{bmatrix}$$

Covariance matrix of estimators for Innovative outlier

$$\sum = \begin{bmatrix} 1.069 & 0.082 & 0.083 \\ 0.083 & 1.061 & 0.082 \\ 0.083 & 0.0826 & 1.061 \end{bmatrix}$$

The coefficient matrix Π_{12} have eigen values (-0.00845, 0.00639, 0.00639).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a small difference as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise but it show that innovative outlier have little bit impact on covariance matrix than AO, coefficient matrix for 12th lag of the model are closer to zero but have a little difference than the results with normal data this means that IO have a little bit impact on coefficient of VAR model.

For Level Shift

Coefficient matrix

$$\Pi_{12}=\begin{bmatrix} 0.328 & 0.330 & 0.334 \\ 0.329 & 0.328 & 0.334 \\ -0.333 & 0.329 & 0.328 \end{bmatrix}$$

<u>Covariance matrix of residuals for Level Shift</u>

$$\sum=\begin{bmatrix} 6.09 & 5.11 & 5.10 \\ 5.11 & 6.10 & 5.11 \\ 5.10 & 5.11 & 6.09 \end{bmatrix}$$

<u>Covariance matrix of estimators for Level Shift</u>

$$\sum=\begin{bmatrix} 6.48 & 5.52 & 5.52 \\ 5.52 & 6.53 & 5.54 \\ 5.52 & 5.54 & 6.52 \end{bmatrix}$$

The coefficient matrix Π12 have eigen values (0.992, -0.00576, -0.00188).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12th lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that level shift has a huge impact on coefficient of VAR model, also badly affects the covariance matrix.

For Transient Change

Coefficient matrix

$$\Pi 12=\begin{bmatrix} 0.268 & 0.273 & 0.283 \\ 0.269 & 0.266 & 0.289 \\ 0.272 & 0.271 & 0.281 \end{bmatrix}$$

Covariance matrix of residuals for Transient Change outlier

$$\sum=\begin{bmatrix} 20.19 & 19.19 & 19.20 \\ 19.19 & 20.18 & 19.19 \\ 19.20 & 19.19 & 20.17 \end{bmatrix}$$

Covariance matrix of estimators for Transient Change outlier

$$\sum=\begin{bmatrix} 22.34 & 21.36 & 21.36 \\ 21.36 & 22.36 & 21.38 \\ 21.36 & 21.38 & 22.36 \end{bmatrix}$$

The coefficient matrix Π12 have eigen values (0.825332, -0.0017, -0.0017).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that level shift has a huge impact on coefficient of VAR model, also badly affects the covariance matrix.

For Seasonal level shift

Coefficient matrix

$$\Pi12=\begin{bmatrix}0.3064 & 0.3136 & 0.3131\\ 0.3089 & 0.3104 & 0.317\\ 0.3116 & 0.3147 & 0.3074\end{bmatrix}$$

Covariance matrix of residuals for Seasonal level shift Level Shift

$$\sum=\begin{bmatrix}1.715 & 0.754 & 0.744\\ 0.754 & 1.576 & 0.762\\ 0.744 & 0.762 & 1.744\end{bmatrix}$$

Covariance matrix of estimators for Seasonal Level Shift

$$\sum=\begin{bmatrix}1.815 & 0.841 & 0.830\\ 0.841 & 1.829 & 0.835\\ 0.830 & 0.835 & 1.817\end{bmatrix}$$

The coefficient matrix Π12 have eigen values (0.934, -0.0069, -0.0032).

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a small difference as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that SLS have a huge impact on coefficient of VAR model, small effect on covariance matrix.

From these results we make these remarks, firstly, Π weights of $VAR(0)(1)_{12}$ model for data series with outliers becomes much larger than the Π weights of $VAR(0)(1)_{12}$ for data series without outlier and have changed sign, this difference clearly identified that all five types of outliers badly and significantly affects the estimation and estimated parameters of VAR model and also affects results which can make our results doubtful. Π weights of VAR $(0)(1)_{12}$ with seasonal level shift also have clearly much larger weights than the Π weights with normal data series, this means that with other types of outliers there is necessary to detect and adjust the SLS in the data series in another case this can distort the whole analysis and can make results defective.

Secondly, Π weights with multivariate innovative and additive outlier have very small difference than the Π weights with normal data this is less sensitive for our analysis results than those of other three types of outliers, Π weights with LS are much larger than the Π weights with normal data and also much larger than the Π weights with other four types of outliers. Hence LS is much sensitive for our analysis, results, estimates and make results doubtful than other types of outlier. SLS and TC have much larger Π weights than the Π weights with normal data but have less than the LS Π weights, SLS and TC also much sensitive for our analysis, estimates and results but less sensitive than the LS outlier.

However, there is clearly seen that estimated coefficient and covariance matrix of innovational outlier have huge difference from seasonal level shift we can say that for sample size n=200 with seasonal VAR model we are not confused in innovational outlier and SLS. By comparing the results with 2.2.1 we have observed that by increasing sample size we have more clarity between MIO (multivariate innovative outlier) and MSLS (multivariate seasonal level shift) results and empirical power and empirical level of significance of the test statistics also improved, as mentioned by Asghar and Urooj (2017) univariate SLS detection give more better results for large sample size.

Thirdly, the eigen vectors of Π weights with IO, AO have less difference than with the normal data, this also identified that IO and AO are less sensitive for analysis than other type of outliers with seasonal VAR model for sample size n=200. The eigen vectors for Π weights with LS have large difference than with the normal data and also larger than the eigen vectors for Π weights with other type of outlier, this also clarify that LS is highly sensitive for our analysis and results. SLS and TC outlier eigen vectors for Π weights also have significant difference than with normal data these are also much sensitive for our results of analysis but less sensitive than the level shift.

Fourthly, covariance matrix of residuals and covariance matrix of estimators for all five types of outliers have much larger difference than for the normal data. Outliers are also badly affecting the covariance matrix and make results doubtful.

Fifthly by comparing results of 4.2.1. and 4.2.2 we have detected that Π weights, ∑ weights, eigen values, empirical quantiles for J_{max} (i, h_i), power of the test statistics, β value, empirical level of significance all depends on sample size also with increasing sample size all these values are departs.

In the below we have simulated the data with $VAR(1)(1)_{12}$ model and sample size n=200.

4.3 For Sample size 200 with VAR $(1)(1)_{12}$

In the below, we have generated 1000 realization for trivariate case with n=200 by normal distribution.

4.3.1 Simulation when no outlier in the series

In start assuming there is no outlier we have estimated a multivariate $VAR(1)(1)_{12}$ with 1^{st} and 12^{th} lag only model then obtained residuals u_{it} and covariance matrix of residuals ∑

and coefficient matrix π then I have estimated outlier size ω_i and covariance matrix of estimators then we have calculated joint test statistics and then we have calculated joint maximum test statistics for all five types of outliers. These series are free of outliers because we have generated them from normal distribution, we have done all these process on these series by using VAR $(1)(1)_{12}$ for calculating the empirical significance level/size. First of all, we have calculated joint test statistics for five types of outliers AO, IO, TC, LS, SLS and then compare their results by joint maximum test and then finalized the results.

Type I error: rejecting true null

Empirical level of significance=size=no of detected outlier when there is no chance of outlier/total number of iterations

6TABLE 4.7 EMPIRICAL LEVEL OF SIGNIFICANCE

Sample size: 200 Level of significance: 0.05 realization: 1000						
		AO	IO	LS	TC	SLS
No of outlier detected		178	50	130	177	14
Empirical level of significance	joint test statistic	0.178	0.05	0.13	0.177	0.014
	For overall joint maximum test statistics	0.294				

From these results by comparing 4.2.1, 4.2.2 and 4.2.3 we have observed that empirical level of significance of SLS along with all other four types of outliers also depend on sample size, dimension and model structure, has oscillate by increasing sample size and changed model structure and dimension for VAR model.

(detailed tables of these results will be provided on demand)

In the below we provided a simulation results when there is outlier in the series.

4.3.2 Simulation with outlier

We have generated 3 data series of sample n=200 by normal distribution in RStudio and make 1000 realization of trivariate then we have added five types of outlier on same time point at t=85. We have calculated VAR(1)(1)$_{12}$ model for each realization and obtained residuals, covariance matrix, coefficient matrix for each equation then we calculated ω_{AO}, $\omega_{LS}, \omega_{TC}, \omega_{SLS}$ by using GLS(generalized least square) estimators and $\omega_{IO,h} = a_{t,h}$ as explained in chapter 3, their covariance matrix and then we have calculated joint test statistics following chi-square distribution for five types of outliers names j_{AO}, j_{IO}, j_{LS}, j_{TC}, j_{SLS} for each realization, at which time point p-value of test statistic value is minimum and probability of that statistic value is less than 0.05 we considered it as outlier in the series at that time point, then we have compare these four type outliers statistics results with each other by the rule of joint maximum test statistic, in each realization which one outlier statistic is greater than other statistic value we declared that type of outlier in this realization.

Joint test statistic results

Null hypothesis. H_o : there is no outlier in the series

Alternative hypothesis. H_A : there is outlier in the series

This is false null because there is outlier in the series which we have added on the specific time point for checking the test statistic for detection of outlier.

Type II error: rejection of false null.

Significance level α=0.05

Joint test statistics follow chi-square distribution, when calculated value is larger by comparing with critical value or probability is less than 0.05 then we reject the null and there is outlier in the series, at which time point p-value of test statistic is minimum that time point is measured as outlier as suggested by Tsay *et al.* (2000).

7TABLE 4.8 POWER OF THE FIVE JOINT TEST STATISTICS

Empirical power of the test statistics sample size: 200 significance level: 0.05 realization: 1000						
		AO	IO	LS	TC	SLS
No of outlier detected		582	999	500	401	784
Empirical level of significance	joint test statistic	58.2%	99.9%	50%	40.1%	78.4%
	For overall joint maximum test statistics	100%				
β = no outlier detected when there is outlier in the series	joint test statistic	0.418	0.001	0.5	0.599	0.216
	For overall joint maximum test statistics	0				

From this table comparing with table 4.2, 4.5 we observed that empirical β and empirical power of the $J(SLS, h_{SLS})$ test statistics along with other all four types of outlier not only depends on sample size but also on the dimension and model structure, with increasing sample size and changing model structure empirical power of the test statistics and β value deviates.

8TABLE 4.9 EMPIRICAL QUANTILES OF THE FIVE J_{MAX} (I, H_I) STATISTICS

Sample	Test	Probabilities					
Size		50%	75%	90%	95%	97.5%	99%
Trivariate case		$VAR(1)(1)_{12}$					
200	J_{max} (AO, h_{AO})	27.62	41.025	54.93	67.01	80.36	101.60
	J_{max} (IO, h_{IO})	39.19	44.895	51.312	55.434	58.258	62.523
	J_{max} (LS, h_{LS})	20.37	30.7	37.452	44.184	51.093	57.3664
	J_{max} (TC, h_{TC})	16.7	20.8	26.38	35.34	41.74	46.812
	J_{max} (SLS, h_{SLS})	20.97	33.195	54.352	70.256	83.945	99.6376

Here i= AO, IO, LS, TC, SLS defined above based on 1000 realizations the model used are seasonal VAR (1)(1)12 given above in chapter 3.

From this table we make this observation. 1^{st} against the expectation, empirical quantiles of $J_{max.}$(TC, h_{TC}) are lesser as compare to other four joint test. $J_{max.}$(i, h_i) for i= AO, IO, SLS are closer, retaining that a common critical value we can use for these three test statistics. Our simulations suggested that 64.23 can be used as estimated critical value for additive outlier, innovative outlier and seasonal level shift for sample n=200 with VAR $(1)(1)_{12}$. This approximate critical value changed with increasing the sample size and also changed with changing the model structure and also all the critical values are increased with increase in sample size and changed with change in model structure, In short, the empirical critical values of the joint test statistics, especially for J_{max} (LS, h_{LS}) depends on the sample size, dimension and the model structure. From these results we have concluded that both the theory and practice of this test statistics needed attention.

4.3.2.1 Coefficient matrix for VAR $(1)(1)_{12}$ with and without outliers in the series

These results are based on 1000 times estimated parameters of model, by average sampling distribution, we have taken expected matrix of 1000 matrix for coefficient matrix/π matrix, covariance matrix, eigen values vector, before and after adding outliers in trivariate case.

4.3.2.1.1 Coefficient and covariance matrix for VAR $(1)(1)_{12}$ model without outliers

Covariance matrix of residuals

$$\sum=\begin{bmatrix}0.969 & 0.003 & 0.003\\0.003 & 0.969 & 0.003\\0.003 & 0.003 & 0.969\end{bmatrix}$$

Π1 matrix (with 1^{st} lag)

$$\Pi1=\begin{bmatrix}0.0012 & 0.0025 & 0.0055\\0.0005 & -0.0039 & -0.0035\\0.0044 & -0.0022 & 0.0074\end{bmatrix}$$

Π12 matrix (with 12^{th} lag)

$$\Pi12=\begin{bmatrix}0.001 & 0.002 & 0.001\\0.0069 & -0.0037 & -0.0001\\0.0019 & -0.0015 & 0.0012\end{bmatrix}$$

Covariance matrix of estimators for Transient change outlier

$$\sum=\begin{bmatrix}0.949 & -0.0001 & 0.001\\0.0001 & 0.986 & 0.02\\0.001 & 0.02 & 0.987\end{bmatrix}$$

Covariance matrix of estimators for Level shift

$$\sum=\begin{bmatrix} 0.949 & -0.0001 & 0.001 \\ 0.0001 & 0.986 & 0.02 \\ 0.001 & 0.02 & 0.987 \end{bmatrix}$$

<u>Covariance matrix of estimators for Additive outlier</u>

$$\sum=\begin{bmatrix} 0.951 & -0.0007 & 0.001 \\ 0.0007 & 0.984 & 0.018 \\ 0.001 & 0.018 & 0.986 \end{bmatrix}$$

<u>Covariance matrix of estimators for Seasonal level shift</u>

$$\sum=\begin{bmatrix} 0.952 & -0.002 & 0.001 \\ 0.002 & 0.985 & 0.018 \\ 0.001 & 0.018 & 0.984 \end{bmatrix}$$

Here these results with large sample size also show that all diagonal entries of covariance matrix entries are almost equal 1 and off diagonal entries are almost equal to zero this show the white noise of data, in coefficient matrix for 12th lag of the model are closer to zero this means that in this multivariate data all the variables not depending on its own lag value and lag value of other variable in data this also show the normality of data.

In the below we show the results of seasonal VAR model with outliers.

4.3.2.1.2 Coefficient and covariance matrix for VAR (1)(1)$_{12}$ model with outliers

With additive outlier

Coefficient matrix

$$\Pi1=\begin{bmatrix} 0.0143 & -0.0082 & -0.0077 \\ 0.0112 & -0.00321 & -0.0079 \\ 0.0144 & -0.00739 & -0.0079 \end{bmatrix}$$

$$\Pi12=\begin{bmatrix} 0.00145 & -0.00077 & 0.0001 \\ -0.00178 & -0.00158 & 0.00422 \\ 0.00079 & -0.00216 & 0.00056 \end{bmatrix}$$

Covariance matrix of residuals for Additive outlier

$$\sum=\begin{bmatrix} 9.336 & 8.327 & 8.319 \\ 8.327 & 9.314 & 8.337 \\ 8.319 & 8.337 & 9.292 \end{bmatrix}$$

Covariance matrix of estimators for Additive outlier

$$\sum=\begin{bmatrix} 9.828 & 8.836 & 8.824 \\ 8.836 & 9.839 & 8.859 \\ 8.824 & 8.859 & 9.807 \end{bmatrix}$$

The two-coefficient matrix Π1 and Π12 have eigen values (0.00345, -0.00222, 0.001958) and (0.002077, -0.00296, 0.001314) respectively.

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag and 1^{st} lag of the model are closer to zero but have a little difference than the results with normal data this means that additive outlier have a little bit impact on coefficient of VAR model but badly affects the covariance matrix.

With Innovative outlier

Coefficient matrix

$$\Pi1=\begin{bmatrix} 0.000203 & -0.00285 & 0.00226 \\ 0.00128 & -0.00128 & -0.00225 \\ 0.00164 & 0.000851 & -0.000778 \end{bmatrix}$$

$$\Pi 12=\begin{bmatrix} -0.00302 & -0.0016 & 0.00029 \\ -0.00102 & -0.00042 & 0.000046 \\ -0.00464 & -0.00182 & -0.0003 \end{bmatrix}$$

<u>Covariance matrix of residuals for Innovative outlier</u>

$$\sum=\begin{bmatrix} 1.054 & 0.0832 & 0.0802 \\ 0.0832 & 1.055 & 0.085 \\ 0.0802 & 0.0852 & 1.046 \end{bmatrix}$$

<u>Covariance matrix of estimators for Innovative outlier</u>

$$\sum=\begin{bmatrix} 1.054 & 0.0832 & 0.0802 \\ 0.0832 & 1.055 & 0.085 \\ 0.0802 & 0.0852 & 1.046 \end{bmatrix}$$

The two-coefficient matrix Π1 and Π12 have eigen values (0.0041,-0.00404, 0.00176) and (-0.00302, -0.00079, 0.000068) respectively.

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a small difference as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise but it show that innovative outlier have little bit impact on covariance matrix than AO, coefficient matrix for 12^{th} lag of the model are closer to zero but have a little difference than the results with normal data this means that IO have a little bit impact on coefficient of VAR model.

For Level Shift

Coefficient matrix

$$\Pi 1=\begin{bmatrix} 0.333 & 0.323 & 0.340 \\ 0.341 & 0.316 & 0.339 \\ 0.342 & 0.323 & 0.332 \end{bmatrix}$$

$$\Pi12=\begin{bmatrix} -0.004 & 0.003 & 0.002 \\ -0.003 & -0.002 & 0.008 \\ -0.002 & -0.0006 & 0.005 \end{bmatrix}$$

<u>Covariance matrix of residuals for Level shift</u>

$$\sum=\begin{bmatrix} 14.328 & 13.317 & 13.308 \\ 13.317 & 14.328 & 13.346 \\ 13.308 & 13.346 & 14.304 \end{bmatrix}$$

<u>Covariance matrix of estimators for Level shift</u>

$$\sum=\begin{bmatrix} 16.177 & 15.181 & 15.171 \\ 15.181 & 16.205 & 15.223 \\ 15.171 & 15.223 & 16.181 \end{bmatrix}$$

The two-coefficient matrix Π1 and Π12 have eigen values (0.997766740, -0.008489779, -0.007211160) and (0.003249, -0.00094,0.00428) respectively.

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that level shift has a huge impact on coefficient of VAR model, also badly affects the covariance matrix.

For Transient Change

Coefficient matrix

$$\Pi1=\begin{bmatrix} 0.325 & 0.331 & 0.333 \\ 0.338 & 0.321 & 0.331 \\ 0.336 & 0.331 & 0.322 \end{bmatrix}$$

$$\Pi12=\begin{bmatrix} -0.0062 & -0.015 & 0.001 \\ -0.0026 & -0.015 & 0.0007 \\ -0.0015 & -0.015 & -0.00043 \end{bmatrix}$$

<u>Covariance matrix of residuals for Transient Change outlier</u>

$$\sum=\begin{bmatrix} 10.55 & 9.52 & 9.51 \\ 9.52 & 10.50 & 9.52 \\ 9.51 & 9.52 & 10.48 \end{bmatrix}$$

<u>Covariance matrix of estimators for Transient Change outlier</u>

$$\sum=\begin{bmatrix} 11.85 & 10.84 & 10.82 \\ 10.84 & 11.83 & 10.84 \\ 10.82 & 10.84 & 11.80 \end{bmatrix}$$

The two-coefficient matrix Π1 and Π12 have eigen values (0.990739, -0.01029, -0.01145) and (-0.01821, -0.00159, -0.00253) respectively.

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a large value as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12th lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that TC have a huge impact on coefficient of VAR model, also badly affects the covariance matrix.

with Seasonal level shift

Coefficient matrix

$$\Pi1=\begin{bmatrix} 0.00143 & 0.00201 & -2.402 \\ -0.0027 & 0.0013 & -3.123 \\ 0.00077 & -0.0017 & 5.0205 \end{bmatrix}$$

$$\Pi 12=\begin{bmatrix}0.402 & 0.422 & 0.342\\ 0.294 & 0.287 & 0.294\\ 0.295 & 0.295 & 0.282\end{bmatrix}$$

Covariance matrix of residuals for Seasonal level shift Level shift

$$\sum=\begin{bmatrix}1.63 & 0.54 & 0.53\\ 0.54 & 1.50 & 0.54\\ 0.53 & 0.54 & 1.50\end{bmatrix}$$

Covariance matrix of estimators for Seasonal level shift

$$\sum=\begin{bmatrix}1.525 & 0.462 & 0.453\\ 0.462 & 1.455 & 0.489\\ 0.453 & 0.489 & 1.456\end{bmatrix}$$

The two-coefficient matrix Π1 and Π12 have eigen values (0.004846, -0.00113, -0.00048) and (0.855719063,0.123519265, -0.006321527) respectively.

Here these results show that all diagonal entries of covariance matrix entries are not equal 1, have a small difference as compare to results of normal data and off diagonal entries are not equal to zero this show that our data series are not white noise, coefficient matrix for 12^{th} lag of the model have a large difference and show the significantly dependence of variables on its own lag value and the lagged value of other variables. This means that SLS have a huge impact on coefficient of VAR model, small effect on covariance matrix.

From these results we make these remarks, firstly, Π weights of $VAR(1)(1)_{12}$ model for data series with outliers becomes much larger than the Π weights of $VAR(1)(1)_{12}$ for data series without outlier and have changed sign, this difference clearly identified that all five types of outliers badly and significantly affects the estimation and estimated parameters of VAR model and also affects results which can make our results doubtful. Π weights of

VAR $(1)(1)_{12}$ with seasonal level shift also have clearly much larger weights than the Π weights with normal data series, this means that with other types of outliers there is necessary to detect and adjust the SLS in the data series in another case this can distort the whole analysis and can make results defective.

Secondly, Π weights with additive and innovative outlier have very small difference than the Π weights with normal data this is less sensitive for our analysis results than those of other three types of outliers on the other hand we can say that that in large sample size MIO(multivariate innovative outlier) are not confusing with other types of outliers, in addition we observed that MIO in large sample size with VAR model less affect the estimates than MAO(multivariate additive outlier), Π1 weights with LS outlier are much larger than the Π1 weights with normal data and also much larger than the Π1 weights with other four types of outliers however Π12 weights with SLS are much larger than the Π12 weights with normal data and also much larger than the Π12 weights with other four types of outliers. Hence SLS is much sensitive for our analysis, results and estimates, make results doubtful than other types of outlier with 12^{th} lag and LS is much sensitive with 1^{st} lag than other for VAR $(1)(1)_{12}$. LS and TC have much larger Π12 weights than the Π12 weights with normal data but have less than the SLS Π12 weights, LS and TC also much sensitive for our analysis, estimates and results but less sensitive than the SLS for VAR $(1)(1)_{12}$.

Thirdly, the eigen vectors of Π weights with MAO and MIO have less difference than with the normal data, this also identified that MAO and MIO are less sensitive for analysis than other type of outliers. The eigen vectors for Π1 weights with LS have large difference than with the normal data and also larger than the eigen vectors for Π1 weights with other type of outlier, this also clarify that LS is highly sensitive for our analysis and results with 1st lag. The eigen vectors for Π12 weights with SLS have large difference than with the normal data and also larger than the eigen vectors for Π12 weights with other type of outlier, this also clarify that SLS is highly sensitive for our analysis and results with 12th lag, TC outlier eigen vectors for Π weights also have significant difference than with normal data these are also much sensitive for our results of analysis but less sensitive than the seasonal level shift and Level shift.

Fourthly, covariance matrix of residuals and covariance matrix of estimators for all five types of outliers have much larger difference than for the normal data. Outliers are also badly affecting the covariance and make results doubtful.

Fifthly by comparing results of 4.2.1 and 4.2.2 and 4.2.3 we have detected that Π weights, ∑ weights, eigen values, empirical quantiles for J_{max} (i, h_i), power of the test statistics, β value, empirical level of significance all depends on sample size and dimension and model structure also with increasing sample size and changing dimension and structure of model all these values are departs.

4.4 On Average Residuals Standard Error Based on 1000 Realization

4.4.1 150 sample, seasonal VAR (0)(1)12

For sample size 150 with seasonal VAR (0)(1)12 model residual standard error for 1st equation is equal to 0.926, for 2nd equation is 0.901, for 3rd equation 0.9901, when series is generated from normal distribution, free of outlier.

For sample size 150 and data series with outlier residual standard error values becomes larger, when there is additive outlier in series, on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 3.145, for 2^{nd} equation 3.396, for 3^{rd} equation 3.254. when there is innovative outlier in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 0.975, for 2^{nd} equation 1.205, for 3^{rd} equation 0.987. when there is level shift in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 3.005, for 2^{nd} equation 2.856, for 3^{rd} equation 2.904. when there is transient change outlier in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 5.345, for 2^{nd} equation 5.106, for 3^{rd} equation 5.804.

when there is seasonal level shift in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 1.435, for 2^{nd} equation 1.345, for 3^{rd} equation 1.283.

4.4.2 200 sample, seasonal VAR (0)(1)12

For sample size 200 with seasonal VAR (0)(1)12 model residual standard error for 1^{st} equation is equal to 0.996, for 2^{nd} equation is 0.991, for 3^{rd} equation 1.001, when series is generated from normal distribution, free of outlier.

For sample size 200 when there is an outlier in the data series, residual standard error values become larger, when there is additive outlier in series, on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 2.556, for 2^{nd} equation 2.396, for 3^{rd} equation 2.754. when there is innovational outlier in series, on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 1.023, for 2^{nd} equation 0.956, for 3^{rd} equation 1.011. when there is level shift in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1^{st} equation 2.665, for 2^{nd} equation 2.356, for 3^{rd}

equation 2.204. when there is transient change outlier in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1st equation 4.545, for 2nd equation 4.606, for 3rd equation 4.704.

When there is seasonal level shift in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1st equation 1.335, for 2nd equation 1.245, for 3rd equation 1.323.

4.4.3 200 sample, seasonal VAR (1)(1)12

For sample size 200 with seasonal VAR (1)(1)12 model residual standard error for 1st equation is equal to 1.016, for 2nd equation is 1.001, for 3rd equation 0.991, when series is generated from normal distribution, free of outlier.

For sample size 200 when there is outlier in the data series, residual standard error values become larger, when there is additive outlier in series, on average residual standard error values with seasonal VAR (1)(1)12 for 1st equation 3.556, for 2nd equation 3.396, for 3rd equation 3.254. when there is innovative outlier in series, on average residual standard error values with seasonal VAR(1)(1)12 for 1st equation 0.982, for 2nd equation 1.021, for 3rd equation 0.991.when there is level shift in the series on average residual standard error values with seasonal VAR(1)(1)12 for 1st equation 3.965, for 2nd equation 3.756, for 3rd equation 3.804. when there is transient change outlier in the series on average residual standard error values with seasonal VAR (1)(1)12 for 1st equation 3.545, for 2nd equation 3.606, for 3rd equation 3.204.

When there is seasonal level shift in the series on average residual standard error values with seasonal VAR (0)(1)12 for 1st equation 1.335, for 2nd equation 1.245, for 3rd equation 1.223.

From these results of residuals standard error of VAR model, we have observed that all outliers affect badly to the standard error of residuals, there is no universally acceptable threshold for the residual standard error. This should be decided based on experience in the domain. In general, the smaller the residual standard error, the better the model fits the data. Firstly, series with outlier have larger standard error than the normal data series, standard error are less affected by the seasonal level shift and highly affected by the transient change outlier. Secondly, by changing the order LS standard error seriously affected here we can say that LS outlier is more sensitive to order of the VAR model than other outliers.

CHAPTER 5

EMPIRICAL ANALYSIS

In Empirical analysis we use one real data example on monthly time series data with three variables of Pakistan. We use rainfall, temperature and "Precipitation and Humidity Altitude" data from 1991 to 2018 for Lahore, Faisalabad and Karachi Pakistan and here we have 204 observations in the data. It is observed that seasonality in temperature of these cities is associated with the seasonality in rainfall and humidity which intern affects badly to the overall weather conditions of these cities.

Outlier recognition in multilevel time series held by employing a trivariate seasonal VAR $(0)(1)_{12}$ model for average monthly climatic data of three stations of Pakistan Faisalabad, Islamabad and Karachi for three variables temperature in degree centigrade, rainfall in milli meter and humidity in percentage from time period January 2008 to July 2020. We have 151 total number of observations. We modified joint maximum test statistics for outlier detection of Tsay *et al.* (2000) technique by adding seasonal level shift in multivariate time series and apply on real data series example. Figure 5.1, 5.2, 5.3 show the average of monthly time series climatic data for Faisalabad, Lahore and Karachi respectively.

1FIGURE 5.1.

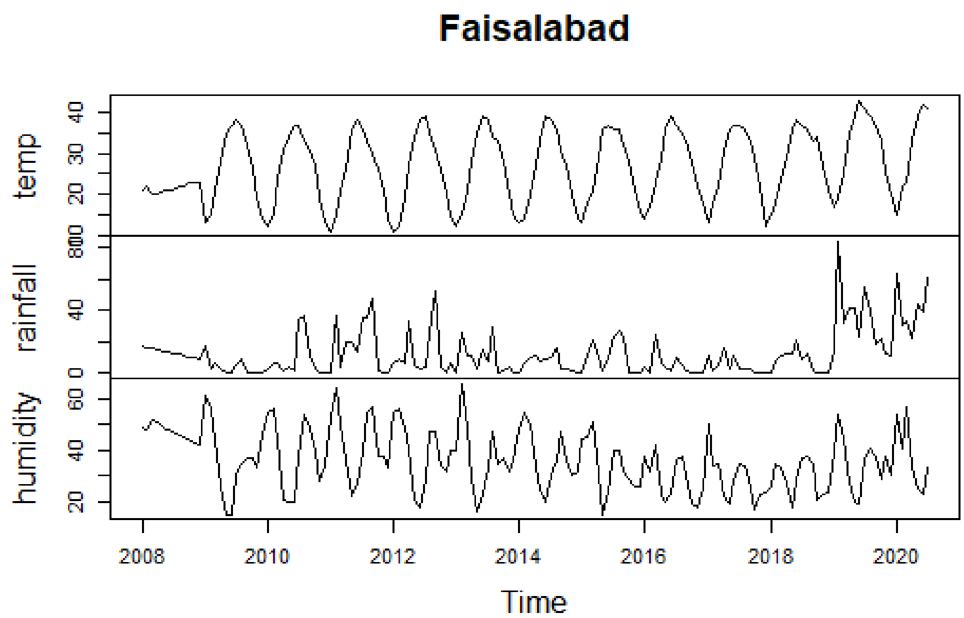


Figure 5.1: for monthly time series climatic data of Faisalabad

2FIGURE 5.2.

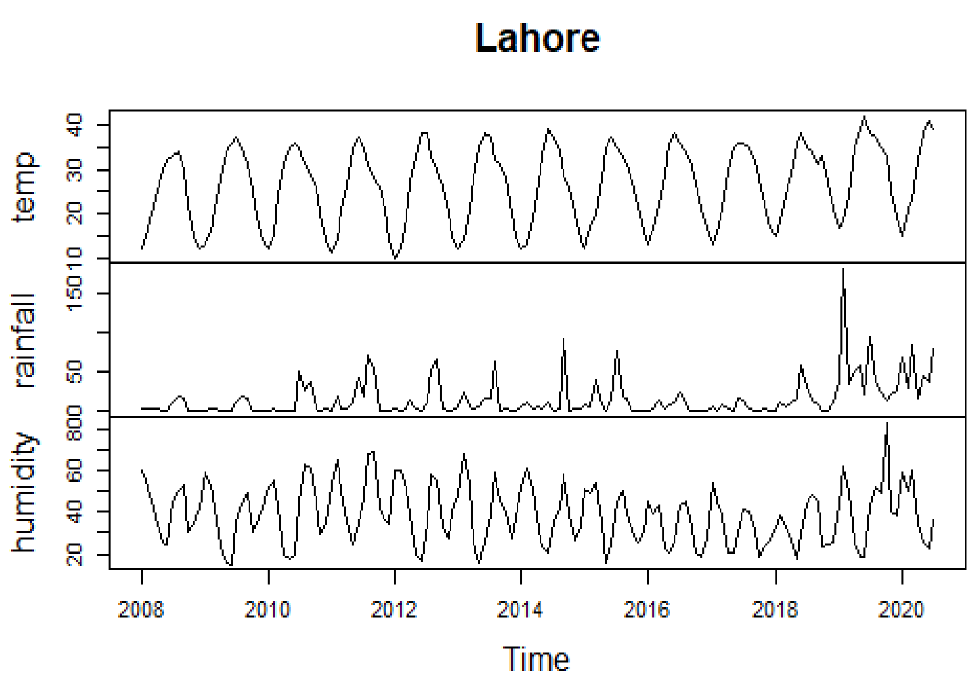


Figure 5.2: for monthly time series climatic data of Lahore

3FIGURE 5.3.

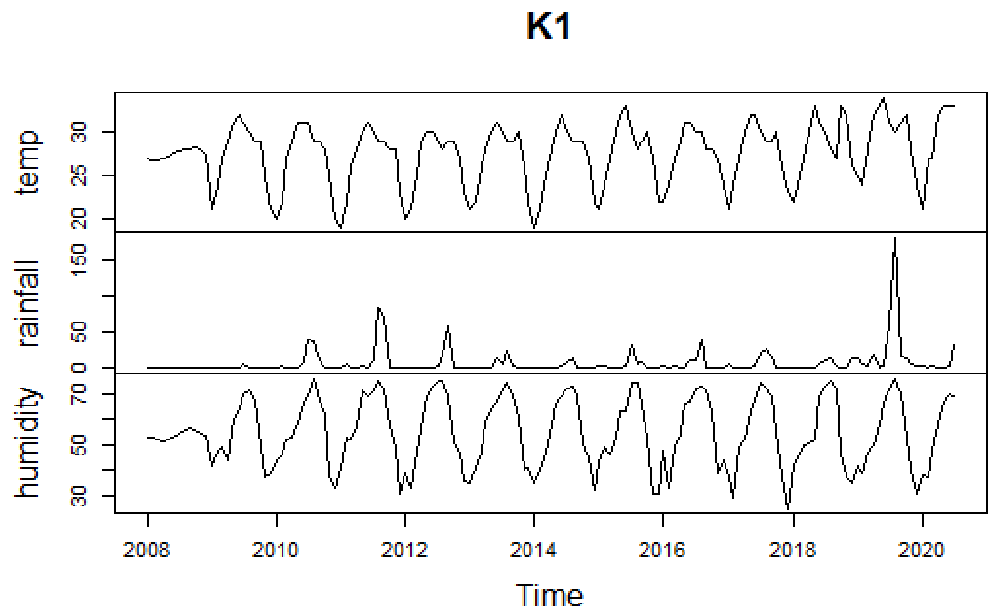


Figure 5.3: for monthly time series climatic data of Karachi

The seasonal VAR $(0)(1)_{12}$ model for climatic data is formed as: -

$temp_t$ = $A'_{1,1,12}$ $temp_{t-12}$ + $A'_{1,2,12}$ $rainfall_{t-12}$ + $A'_{1,3,12}$ $humidity_{t-12}$ $+u_{1t}$ … 5.1

$rainfall_t$ = $A'_{2,1,12}$ $temp_{t-12}$ + $A'_{2,2,12}$ $rainfall_{t-12}$ + $A'_{2,3,12}$ $humidity_{t-12}$ $+u_{2t}$ … 5.2

$humidity_t$=$A'_{3,1,12}$ $temp_{t-12}$ + $A'_{3,2,12}$ $rainfall_{t-12}$ + $A'_{3,3,12}$ $humidity_{t-12}$ $+u_{3t}$ … 5.3

Seasonal VAR $(0)(1)_{12}$ model in matrix form

$$\begin{bmatrix} temp_t \\ rainfall_t \\ humidity_t \end{bmatrix} = \begin{bmatrix} A._{1,1,12} & A._{1,2,12} & A._{1,3,12} \\ A._{2,1,12} & A._{2,2,12} & A._{2,3,12} \\ A._{3,1,12} & A._{3,2,12} & A._{3,3,12} \end{bmatrix} \begin{bmatrix} temp_{t-12} \\ rainfall_{t-12} \\ humidity_{t-12} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \quad . \ . \ .. \qquad 5.4$$

The detection results are conceded in table 5.3, 5.7, 5.11 based on 5% empirical critical value from table 4.3 for a seasonal VAR $(0)(1)_{12}$ model and sample size 150. There is only one outlier is noticed for Faisalabad at t=134, two outliers are perceived for Lahore at t= 134 and only one outlier is noticed at t=140, on one occasion, an outlier identified, we adjusted the its impact on the data and re-estimated the seasonal VAR $(0)(1)_{12}$ model. The estimated outlier size $\hat{\omega} = (\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3)'$ of all outliers are given in table 5.4, 5.8, 5.12 for Faisalabad, Lahore and Karachi respectively along with t ratios of estimates. The parameter estimates of the model 5.4 with and without outlier adjustment are given away in table 5.2, 5.6, 5.10 for Faisalabad, Lahore and Karachi respectively. The perceived outlier has noticeable effect on the seasonal parameters of the VAR model and residuals covariance matrix.

"An outlier in the single element may be persuaded by another element outlier and when we detect that outlier separately for each individual series, by using a marginal model it can result in overspecification of number of outliers. The multivariate joint detection of outlier could be more powerful than the univariate method.'' Tsay *et al.* (2000).

Firstly, the detected outlier SLS at t=134 for Faisalabad affects all the three-variable temperature, rainfall and humidity, this time point was the 11^{th} season with monthly data according to February and that is February 2019. In feburary 2019 there is highly maximum rainfall(83.8mm), temperature(19C^{o}) and humidity (54%) was recorded than the past and in feburary 2020 again this pattern is recorded this make a seasonal level shift in this data.

Rainfall reaches this extreme point due to high level of humidity and temperature at t=134 than the t<134, this make clear to this point that outlier in one series may be cause by outlier in another series this may ignored in univariate outlier detection. Secondly, the positive and significant estimates $\widehat{\omega}$ of SLS for Faisalabad at time point 134 shows that the high temperature and humidity caused extremeness in rainfall or seasonal level shift in rainfall. This information is not reliable if one wants to detect univariate outliers, finally the significance of $\widehat{\omega}$ at t=134 also specifies a probable combined structural break of the series at that time point. The negative estimated outlier affect shows the negative relation of input and output series with a delay of 11^{th} time periods

Firstly, the detected outlier TC and SLS at t=134 in iteration 1 and iteration 2 respectively for Lahore clearly affects all the three-variable temperature, rainfall and humidity, this time point was the 11^{th} season with monthly data according to feburary and that is feburary 2019. In feburary 2018 there is highly extreme rainfall (182 mm), temperature(18C^{o}) and humidity (60%) was recorded than the past and in feburary 2020 again this pattern is noted this highly affect the time series data analysis and future prediction may go to unreliable if we don't realize this outlier during the data analysis. Rainfall reaches this extreme point due to high level of humidity and temperature at t=134 as compare to the t<134, this make clear to this point that outlier in one series may cause by outlier in another series this may ignored in univariate outlier detection. Secondly, the positive and significant estimates $\widehat{\omega}$ of TC & SLS for Lahore at time point 134 shows that the high temperature and humidity caused extremeness in rainfall in rainfall. This information is not reliable if one wants to detect univariate outliers, finally the significance of $\widehat{\omega}$ at t=134 also specifies a probable combined structural break of the series at that time point.

The detected outlier AO at t=140 in iteration 1 for Karachi clearly affects two of the three-variable rainfall and humidity, this time point was the 11^{th} season with monthly data according to august and that is august 2019. In august 2019 there is highly extreme rainfall (181.9 mm), and humidity (76%) was recorded than all the past values, this highly affected the time series data analysis and future prediction may go to unreliable if we don't realize this outlier during the data analysis. Rainfall reaches this extreme point due to high level of humidity at t=140 as compare to the t<140, this make clear to this point that outlier in one series may cause by outlier in another series this may ignored in univariate outlier detection. Secondly, the positive and significant estimates $\hat{\omega}$ of AO for karachi at time point 140 shows that the high level of humidity caused extremeness in rainfall. Such information is not reliable if one wants to univariate outlier detection, finally the significance of $\hat{\omega}$ at t=140 also specifies a probable combined structural break of the series at that time point.

5.1 Multivariate Outlier Detection for Faisalabad

This real data example based on trivariate case on the monthly climatic data of Faisalabad, Pakistan from january 2008 up to July 2020. In very initial stage we selected lags of VAR model through VARselect command in R.

9TABLE 5.1. LAG SELECTION FOR VAR

AIC.(n)	HQ.(n)	SC.(n)	FPE.(n)
12	5	5	12

Here AIC and FPE select the lag 12 for VAR model hence we preferred the results of AIC and took the seasonal VAR model with 12^{th} lag.

10TABLE 5.2. REGRESSION RESULTS OF SEASONAL VAR MODEL

Before outlier adjustment					
Equation # 1 of model 5.4 for temperature		Equation # 2 of model 5.4 for rainfall		Equation # 3 of model 5.4 for humidity	
Coefficients	Estimates	Coefficients	Estimates	Coefficients	Estimates
$A_{.1,1,12}$	0.98481 (0.02144)	$A_{.2,1,12}$	0.311893 (0.0819)	$A_{.3,1,12}$	0.1907 (0.048)
$A_{.1,2,12}$	0.00223 (0.02515)	$A_{.2,2,12}$	0.4253 (0.096)	$A_{.3,2,12}$	-0.142 (0.056)
$A_{.1,3,12}$	0.03149 (0.01641)	$A_{.2,3,12}$	-0.0073 (0.062)	$A_{.3,3,12}$	0.873 (0.037)
After outlier adjustment					
Equation # 1 of model 5.4 for temperature		Equation # 2 of model 5.4 for rainfall		Equation # 3 of model 5.4 for humidity	
Coefficients	Estimates	Coefficients	Estimates	Coefficients	Estimates
$A_{1,1,12}$	-1.0540 (0.655)	$A_{2,1,12}$	-1.16 (0.412)	$A_{3,1,12}$	-1.782 (0.65)
$A_{1,2,12}$	3.936 (1.09)	$A_{2,2,12}$	3.28 (0.68)	$A_{3,2,12}$	3.66 (1.08)
$A_{1,3,12}$	0.155 (0.151)	$A_{2,3,12}$	0.074 (0.095)	$A_{3,3,12}$	0.992 (0.150)

Parameters estimate of model 5.4 before and after multivariate outlier detection, the value in parenthesis are standard error.

11TABLE 5.3. RESULTS FOR MULTIVARIATE OUTLIER DETECTION FOR FAISALABAD

	Joint maximum test results					Outlier	
Iterations	J_{max}(AO, h_{AO})	J_{max}(IO, h_{IO})	J_{max}(LS, h_{LS})	J_{max}(TC, h_{TC})	J_{max}(SLS, h_{SLS})	Time	type
1	<50.287	<54.954	<28.10	144.05(134)	185.462(134)	134	SLS
2	<50.287	79.43(134)	<28.10	<63.79	72.307(134)	134	IO
3	<50.287	112.23(146)	<28.10	168.64(146)	162.358(146)	146	TC
4	<50.287	65.34(146)	<28.10	<63.79	<64.154	146	IO
5	<50.287	<54.954	<28.10	<63.79	<64.154		
Criticeal.value	50.287	54.954	28.10	63.79	64.154		

Multilevel additive outlier, MAO, Multilevel innovative outlier, MIO, Multilevel level shift , MLS, Multilevel transient change, TC, multilevel seasonal level shift, SLS.

These are the results for multivariate outlier detection for monthly climatic data of Faisalabad at 5% critical value from January 2008 to July 2020, the number in brackets for joint tests is the corresponding time index.

In the above results we have adopted an iterative procedure for outlier detection, once we have identified an outlier in the series then we adjusted that type of outlier in the series and repeated the process until when there is no outlier is declared.

We iterated the outlier detection procedure five times in 1st iteration we identified only two outliers TC and SLS on time point t=134, however by the rule of joint maximum test statistics, J (SLS, h_{SLS}) outlier value is greater than the TC therefore in iteration one we have declared SLS. Then we adjusted the data series with SLS then repeated the same process in iteration 2, in 2nd iteration we declared an IO at time point t=134. Then we adjusted the data series with IO and repeat the same process on this adjusted series in iteration 3, in 3rd iteration we declared TC at time point t=146, Then we adjusted the data series with TC and repeat the same process on this adjusted series in iteration 4, in 4th iteration we declared IO at time point t=146, Then we adjusted the data series with IO and repeat the same process on this adjusted series in iteration 5, in 5th iteration we did not identified any outlier. Now this series is considered as outlier free data series.

12TABLE 5.4. ω_{SLS} , ω_{AO}, ω_{LS}, ω_{TC} RESULTS

Iterations	ω_{SLS1}(t)	ω_{SLS2}(t)	ω_{SLS3}(t)
1	30.84(2.22)	16.13(1.99)	29.33(2.31)
2	19.87(0.031)	7.655(0.022)	40.97(0.077)
3	61.064(0.0079)	29.079(0.0073)	78.97(0.012)
4	23.58(0.11)	12.088(0.115)	0.853(0.0047)
5	17.004(0.047)	8.648(0.049)	10.817(0.035)
Iterations	ω_{AO1}(t)	ω_{AO2}(t)	ω_{AO3}(t)
1	0.732(0.359)	-1.11(-0.44)	-1.53(-1.9)
2	-1.889(1.13)	-3.294(-1.41)	-1.115(-1.28)
3	-1.685(-0.75)	-2.719(-0.65)	1.182(-0.63)
4	2.129(1.16)	2.235(0.802)	-4.88(-1.92)
5	-1.406(-0.55)	-2.15(-0.45)	-1.101(-0.53)
Iterations	ω_{LS1}(t)	ω_{LS2}(t)	ω_{LS3}(t)
1	-1.205(0.702)	1.549(0.62)	-2.038(-2.48)
2	-1.037(-0.78)	-2.005(-0.63)	-0.402(-0.62)
3	-0.873(-0.38)	-1.969(-0.35)	-0.475(-0.22)
4	1.702(0.59)	0.993(0.15)	-3.386(-1.18)
5	4.055(0.73)	3.209(0.26)	8.194(-1.49)
Iterations	ω_{TC1}(t)	ω_{TC2}(t)	ω_{TC3}(t)
1	22.290(2.97)	-24.851 (-2.94)	-1.094(-0.68)
2	17.036(1.51)	-17.16 (-1.47)	-1.15 (-0.41)
3	17.502(0.896)	-17.235(-0.86))	-1.329(-0.27)
4	-2.647(-0.72)	7.049(1.56)	-4.427(-0.87)
5	-4.822(-1.204)	4.303(-0.94)	0.319(-0.08)
Iterations	ω_{IO1}(t)	ω_{IO2}(t)	ω_{IO3}(t)
1	0.148	75.346	21.03
2	31.724	48.044	36.867
3	63.22	45.98	60.74
4	34.35	23.96	36.045
5	59.851	38.891	58.987

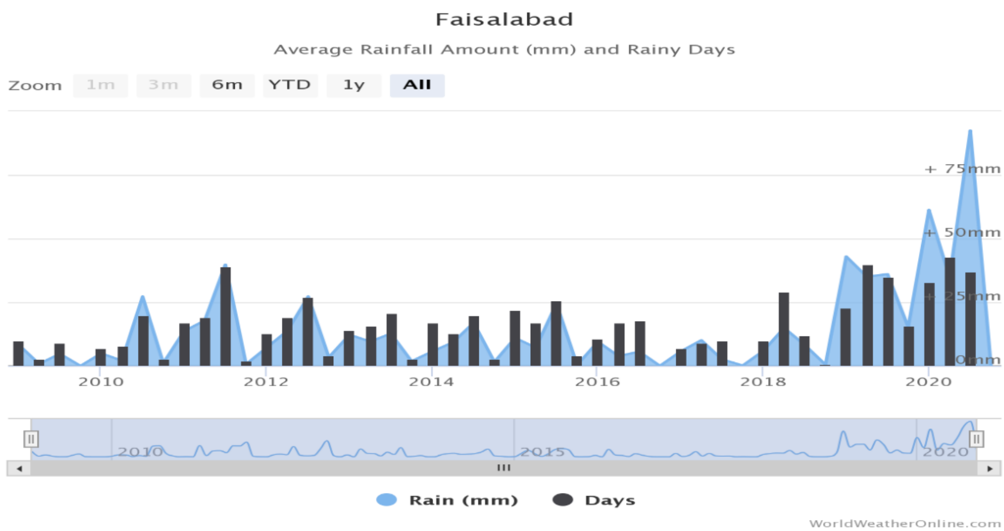
These results are for monthly climatic data of Faisalabad from January 2008 to July 2020, there is t-value of matching ω_{SLS} , ω_{AO}, ω_{LS}, ω_{TC} value in parenthesis respectively.

5.1.1 Residuals standard error before and after outlier adjustment

We have sample size 151 and use seasonal VAR (0)(1)12 model, residual standard error for 1st equation is equal to 3.647, for 2nd equation is 13.94, for 3rd equation 8.234.

By multivariate outlier detection in the data using joint test statistics we identified here SLS after adjusting this outlier in the series residual standard error become smaller, residual standard error for 1st equation is equal to 15.28, for 2nd equation is 9.165, for 3rd equation 15.15.

Here an image taken from world weather website is shown below:



In this image we can see that in feburary 2019 very extreme level of rainfall was recorded and in 2020 again this pattern was recorded, which show the shift of seasonal level for low level of rainfall to extreme level of rainfall for Faisalabad, this confirmed that SLS detected by out=r test statistics in time series data of Faisalabad is existed in real world data, which show the correct detection of SLS.

5.2 Multivariate outlier detection for Lahore

These results are for trivariate case on the average monthly climatic data of Lahore, Pakistan from January 2008 up to July 2020. In very initial stage we selected lags of VAR model through VARselect in R

13TABLE 5.5. LAG SELECTION

AIC.(n)	HQ.(n)	SC.(n)	FPE.(n)
12	12	10	12

Here AIC and FPE select the lag 12 for VAR model hence we preferred the results of AIC and took the seasonal VAR model with 12th lag.

14TABLE 5.6. REGRESSION RESULTS OF SEASONAL VAR MODEL

Before outlier adjustment					
Equation # 1 of model 5.4 for temperature		Equation # 2 of model 5.4 for rainfall		Equation # 3 of model 5.4 for humidity	
Coefficients	Estimates	Coefficients	Estimates	Coefficients	Estimates
$A_{.1,1,12}$	1.00307 (0.0115)	$A_{.2,1,12}$	0.3772 (0.1446)	$A_{\cdot 3,1,12}$	0.17541 (0.0544)
$A_{.1,2,12}$	0.00655 (0.00727)	$A_{.2,2,12}$	0.2799 (0.096)	$A_{\cdot 3,2,12}$	-0.10084 (0.0344)
$A_{\cdot 1,3,12}$	0.008674 (0.008027)	$A_{\cdot 2,3,12}$	0.09424 (0.10079)	$A_{\cdot 3,3,12}$	0.922 (0.0379)
After outlier adjustment					
Equation # 1 of model 5.4 for temperature		Equation # 2 of model 5.4 for rainfall		Equation # 3 of model 5.4 for humidity	
Coefficients	Estimates	Coefficients	Estimates	Coefficients	Estimates
$A_{\cdot 1,1,12}$	0.491 (0.024)	$A_{\cdot 2,1,12}$	0.636 (0.034)	$A_{\cdot 3,1,12}$	0.914 (0.131)
$A_{\cdot 1,2,12}$	0.990 (0.046)	$A_{\cdot 2,2,12}$	-0.363 (0.063)	$A_{\cdot 3,2,12}$	-0.785 (0.244)
$A_{\cdot 1,3,12}$	-0.103 (0.014)	$A_{\cdot 2,3,12}$	0.218 (0.0202)	$A_{\cdot 3,3,12}$	0.704 (0.078)

In above these are Parameters estimate of model 5.4 model, these are the results for before and after multivariate outlier detection, the value in parenthesis are standard error.

15TABLE 5.7. RESULTS FOR MULTIVARIATE OUTLIER FOR LAHORE

	Joint maximum test results					outlier	
Iterations	J_{max}(AO, h_{AO})	J_{max}(IO, h_{IO})	J_{max}(LS, h_{LS})	J_{max}(TC, h_{TC})	J_{max}(SLS, h_{SLS})	Time	type
1	<50.287	<54.95	<23.130	96.078(134)	83.997(134)	134	TC
2	<50.287	<54.95	<23.130	44.992(134)	125.785(134)	134	SLS
3	<50.287	90.342(134)	<23.130	<29.035	56.429(134)	134	IO
4	<50.287	113.89(134)	<23.130	<29.035	69.112(146)	134	IO
5	<50.287	<54.954	<23.130	<29.035	<47.600		
Critical.value	50.287	54.954	23.130	29.035	47.600		

These results are for the detection of outliers for monthly climatic data of Lahore at 5% critical value from January 2008 to July 2020, the number in parenthesis for joint tests is the corresponding time index.

In above results we have adopted an iterative procedure for outlier detection, once we have identified an outlier in the series then we adjusted that type of outlier in the series and repeated the process until when there is no outlier is declared.

We iterated the outlier detection procedure five times in 1st iteration we identified only two outliers TC and SLS on time point t=134, however by the rule of joint maximum test statistics J(TC, h_{TC}) outlier value is greater than the SLS therefore in iteration one we have declared TC outlier. Then we adjusted the data series with TC outlier as suggested above in chapter no 3, then repeated the same process in iteration 2, in 2nd iteration we declared a SLS at time point t=134. Then we adjusted the data series with seasonal level shift and repeat the same process on this adjusted series in iteration 3, in 3rd iteration we declared IO at time point t=134. Then we adjusted the data series with IO and repeat the same process on this adjusted series in iteration 4 then again, we declared IO in the series now again we adjust the series with IO as suggested above in chapter 3, again we repeated the process in 5th iteration now we did not identify any outlier. Now this series is considered as outlier free data series.

16TABLE 5.8. ω_{SLS} , ω_{AO}, ω_{LS}, ω_{TC} RESULTS

Iterations	ω_{SLS1}(t)	ω_{SLS2}(t)	ω_{SLS3}(t)
1	15.386(0.393)	11.026(0.406)	8.7409(0.199)
2	23.293(2.2902)	17.479(2.3079)	39.331(0.383)
3	-21.337(-1.406)	-8.648736(-1.2231)	17.9298(0.2492)
4	-40.48(-0.0086)	-31.59(-0.0101)	-70.59(-0.121)
5	14.92(0.131)	8.37(0.108)	19.45(0.122)
Iterations	ω_{AO1}(t)	ω_{AO2}(t)	ω_{AO3}(t)
1	-1.815(-1.65)	-3.212(-1.209)	6.629(1.6003)
2	-1.211(-1.205)	-2.383(-1.187)	0.0127(0.0158)
3	-0.5993(-0.562)	-2.035(-0.967)	-0.466(-0.499)
4	-0.709(-0.55)	-2.088(-0.81)	-0.482(-0.441)
5	-0.683(-0.617)	-2.015(-0.859)	-0.602(-0.652)
Iterations	ω_{LS1}(t)	ω_{LS2}(t)	ω_{LS3}(t)
1	0.696(0.276)	-2.749(-0.705)	-0.8711(-0.903)
2	-1.471(-1.674)	0.2009(0.1049)	-0.373(-0.401)
3	-1.154(-1.256)	-0.188(-0.086)	-0.403(-0.385)
4	-1.236(-1.149)	-0.0408(-0.015)	-0.497(-0.406)
5	-1.291(-1.396)	0.193(0.081)	-0.586(-0.614)
Iterations	ω_{TC1}(t)	ω_{TC2}(t)	ω_{TC3}(t)
1	16.362(2.47)	-20.651(-2.64)	1.315(0.953)
2	17.703(1.754)	-22.248(-1.877)	1.31340(0.623)
3	15.103(1.324)	-18.213(-1.359)	0.684(0.283)
4	15.801(1.231)	-19.153(-1.269)	0.656(0.2304)
5	7.68(0.69)	-6.128(-0.526)	0.041(0.017)
Iterations	ω_{IO1}(t)	ω_{IO2}(t)	ω_{IO3}(t)
1	-0.458	168.477	24.913
2	17.177	28.950	7.352
3	39.45	32.023	42.89
4	62.199	44.395	81.007
5	9.435	15.65	83.201

These results are for monthly climatic data of Lahore from January 2008 to July 2020, there is t-value of corresponding ω_{SLS} , ω_{AO}, ω_{LS}, ω_{TC} value in parenthesis respectively.

5.2.1 Residuals Standard Error before and after Outlier Adjustment

we have sample size 151 and use seasonal VAR (0)(1)12 model, residual standard error for 1st equation is equal to 1.936, for 2nd equation is 24.31, for 3rd equation 9.16.

By multivariate outlier detection in the data using joint test statistics, we identified here TC, IO & SLS after adjusting this outlier in the series residual standard error become smaller, residual standard error for 1st equation is equal to 2.953, for 2nd equation is 4.057, for 3rd equation 15.63. decline in standard error show the good performance of model.

5.3 Multivariate outlier detection for Karachi

In very initial stage we selected lags of VAR model through VAR select in R

17TABLE 5.9. LAG SELECTION

AIC.(n)	HQ.(n)	SC.(n)	FPE.(n)
12	11	4	12

Here AIC and FPE select the lag 12 for VAR model hence we preferred the results of AIC and took the seasonal VAR model with 12th lag.

18TABLE 5.10. REGRESSION RESULTS OF SEASONAL VAR MODEL

Before outlier adjustment					
Equation # 1 of model 5.4 for temperature		Equation # 2 of model 5.4 for rainfall		Equation # 3 of model 5.4 for humidity	
Coefficients	Estimates	Coefficients	Estimates	Coefficients	Estimates
$A_{.1,1,12}$	0.97275 (0.0336)	$A_{\cdot 2,1,12}$	-1.176 (0.343)	$A_{\cdot 3,1,12}$	0.384 (0.122)
$A_{\cdot 1,2,12}$	-0.01368 (0.01245)	$A_{\cdot 2,2,12}$	0.310 (0.127)	$A_{\cdot 3,2,12}$	0.048 (0.045)
$A_{\cdot 1,3,12}$	0.01726 (0.01707)	$A_{\cdot 2,3,12}$	0.694 (0.174)	$A_{\cdot 3,3,12}$	0.805 (0.062)
After outlier adjustment					
Equation # 1 of model 5.4 for temperature		Equation # 2 of model 5.4 for rainfall		Equation # 3 of model 5.4 for humidity	
Coefficients	Estimates	Coefficients	Estimates	Coefficients	Estimates
$A_{\cdot 1,1,12}$	0.973 (0.1101)	$A_{\cdot 2,1,12}$	-1.169 (0.464)	$A_{\cdot 3,1,12}$	0.320 (0.37)
$A_{\cdot 1,2,12}$	-0.013 (0.053)	$A_{\cdot 2,2,12}$	-1.169 (0.223)	$A_{\cdot 3,2,12}$	0.033 (0.17)
$A_{\cdot 1,3,12}$	0.0175 (0.062)	$A_{\cdot 2,3,12}$	0.695 (0.261)	$A_{\cdot 3,3,12}$	0.841 (0.208)

These are Parameters estimate of model 5.4 before and after multivariate outlier detection, the values in parenthesis are standard error.

19TABLE 5.11. RESULTS FOR MULTIVARIATE OUTLIER DETECTION FOR KARACHI

In the below results we have adopted an iterative procedure for outlier detection, once we have identified an outlier in the series then we adjusted that type of outlier in the series and repeated the process until when there is no outlier is declared.

	Joint maximum test results					outlier	
Iterations	J_{max}(AO, h_{AO})	J_{max}(IO, h_{IO})	J_{max}(LS, h_{LS})	J_{max}(TC, h_{TC})	J_{max}(SLS, h_{SLS})	Time	type
1	64.314(140)	<54.945	<23.130	<29.035	<47.600	140	AO
2	<50.287	<54.945	<23.130	<29.035	<47.600		
Critical.value	50.287	54.945	23.130	29.035	47.600		

These results are for monthly climatic data of Karachi at 5% significance level from January 2008 to July 2020, the number in parenthesis is for the joint tests is the matching time index.

We iterated the outlier detection procedure two times in 1st iteration we identified only one outliers AO outlier on time point t=140, Then we adjusted the data series with AO outlier as suggested above in chapter no 3, then repeated the same process in iteration 2, in 2nd iteration now we did not identified any type of outlier out of all four types of outlier in our study. Now this series can be treated as outlier free data series.

20TABLE 5.12. ω_{SLS} , ω_{AO}, ω_{LS}, ω_{TC} RESULTS

Iterations	ω_{SLS1}(t)	ω_{SLS2}(t)	ω_{SLS3}(t)
1	-5.111 (-0.0251)	5.9578(0.0716)	-2.656(-0.0062)
2	-7.927(-0.074)	-8.125(-0.184)	-25.571(-0.115)
Iterations	ω_{AO1}(t)	ω_{AO2}(t)	ω_{AO3}(t)
1	11.371(1.457)	0.1767(2.314)	-1.1625(-2.145)
2	3.454(0.331)	2.247(0.345)	-3.439(-0.483)
Iterations	ω_{LS1}(t)	ω_{LS2}(t)	ω_{LS3}(t)
1	-0.378(0.276)	-0.455(-0.916)	0.355(0.395)
2	-0.553(-0.873)	-0.491(-0.784)	0.155(0.136)
Iterations	ω_{TC1}(t)	ω_{TC2}(t)	ω_{TC3}(t)
1	-0.739 (-0.4601)	-1.10302(-0.8495)	0.555(0.464)
2	-0.8215(-0.393)	-1.536(-0.925)	0.816(0.533)
Iterations	ω_{IO1}(t)	ω_{IO2}(t)	ω_{IO3}(t)
1	1.664	158.275	4.102
2	-0.303	48.867	22.604

These are results for monthly climatic data of Karachi from January 2008 to July 2020, there is t-value of corresponding ω_{SLS} , ω_{AO}, ω_{LS}, ω_{TC} value in parenthesis respectively.

5.3.1 Residuals Standard Error before and after Outlier Adjustment

We have sample size 151 and use seasonal VAR (0)(1)12 model, residual standard error for 1st equation is equal to 1.751, for 2nd equation is 17.89, for 3rd equation 6.378.

By multivariate outlier detection in the data using joint test statistics we identified here AO outlier after adjusting this outlier in the series residual standard error become smaller, residual standard error for 1st equation is equal to 1.721, for 2nd equation is 7.258, for 3rd equation 5.868, decline in standard error show the good performance of model.

From all the results of standard error of residuals for real time series for three stations explained above, we have observed that outliers badly affected the standard error of residuals and standard error of estimators also, after adjusting outlier there is significant decline in standard error occurs, more lesser the standard error improve the model performance. It is noted that standard error for rainfall equation is observed very high as compare to other equations, this is because of highest level of rainfall in 2019 and 2020which cause the presence of outlier in the series.

In above all the results show that, if we analyze this data without considering these outliers in the model the results are totally changed and doubtful. We cannot rely on these results for any purpose like policy implication, future planning based on the current results because the unusual observation occurs after 2018 and analysis of data gives results as whole, did not take the unusual observations as separate and results are not reliable if we consider these outliers during our data analysis and adjust in the model this procedure will consider these unusual observation and gives results by considering this unusual observation.

In conclusion from all the results we have concluded that in the presence of MSLS estimated coefficients of temperature, rainfall and humidity with seasonal VAR model for monthly frequency contains that these are less affected by each other but according to theory and previous literature we can see that these variables are heavily depending on each other, however when we have adjusted the data series by adopting an iterative procedure and re-estimated the model on adjusted data series this gave much different results, mostly equations results shows that temperature, rainfall and humidity heavily depending on each other. Hence, we identified that if there is MSLS along with other types of outlier exist in the data the results are not reliable and cannot make any decision based on these results without adjusting MSLS in the model. MSLS also depends on the sample size, order and structure of the VAR model therefore if anyone wants to use empirical critical values obtained from these results for outlier detection with any other order of VAR model, this may be not reliable.

In the current study we suggested a modified method of outlier detection in multivariate time series by incorporating seasonal level shift in multivariate time series. Basically, our focus is not on forecasting, but our focus was introducing a modified multivariate outlier detection method and to check its performance in terms of power and size via simulation and also detect outlier in real world scenario. And we suggested the study for forecasting in case of covariate determination.

Here some studies are found for forecasting with outliers in time univariate series, Ledolter (1989) find out that, the additive outlier affect forecasts from ARIMA models. Shio etal. (1993) in their study found that, Reallocation outliers often have little effect on forecasts, except for a few time points for which forecasts give large weight to observations that are affected by the reallocation. According to Galeano et al. (2006) the existence of even few outliers usually leads to inaccurate models and not satisfactory forecasts, according to Carnero et al. (2007) outlier may deeply influence the estimates that classical methods propose. According to chen and liu (1993) an outlier occurring at the forecast origin has the greatest impact on forecasts. As an outlier occurs further away from the forecast origin, its effect on forecasts becomes smaller.

All the studies above mentioned confirms the effect of outliers on forecast, I this study we did not moved towards forecasting but in future we will make a study on forecasting with outliers in multivariate time series.

CHAPTER 6

CONCLUSION AND STRATEGY COMENDATION

6.1 Conclusion

The main focused of our study is that, to obtain a modified Tsay *et al*. (2000) multivariate outlier detection method by adding multivariate seasonal level shift (MSLS) in vector time series analysis using the seasonal VAR model for monthly time series data by including four other outliers used in Tsay *et al.* (2000) study for multivariate outlier detection in time series named AO, IO, LS and TC outlier. We have used simulation to obtain power, size of the proposed test statistics and investigate the influence of multivariate SLS along with other types of outliers on estimates and covariance matrix. we also calculated empirical critical values for SLS along with other four types of outliers for different sample size and order of the VAR model from simulation. We use sample size for seasonal VAR, n=150 and 200, we have detected all of outlier by using one real data example of Pakistan for climatic data of three stations Faisalabad, Lahore and Karachi including three variables temperature, rainfall, humidity.

In simulation results, we have observed that empirical power, empirical level of significance, β values, the empirical quantiles of the joint maximum test for MSLS depends on sample size, dimension and structure of the VAR model and other four outliers AO, IO, LS, TC have a significant impact on all of these results. With the sample size 150 using seasonal VAR for the monthly time series data, we have observed that the overall size of the joint maximum test statistic is 0.547, however if we see separately for $J_{SLS,}$ there we observed 0.147 empirical level of significance J(SLS, h_{SLS}). For Sample size 200 with VAR $(0)(1)_{12}$ overall value of α=0.493 and separately for J(SLS, h_{SLS}) α= 0.05, For Sample size 200 with VAR(1)$(1)_{12}$ overall value of α=0.294 and separately for J (SLS, h_{SLS}) α= 0.014. Power of the overall j-max test with VAR(1)$(1)_{12}$ for n=200 is 99.9% and specific for j_{SLS} = 78.5%, Power of the overall j-max test with VAR(0)$(1)_{12}$ for n=200 is 100% and specific for j_{SLS} = 89.9%, Power of the overall j-max test with VAR(0)$(1)_{12}$ for n=150 is 99.8% and specific for j_{SLS} = 62.2%. These all of the results concluded that α-value, empirical power and empirical critical value of j_{SLS} and overall joint-max test statistics depends on sample, dimension and structure of the model and as well as for all types of outliers. From this study we also concluded that multivariate SLS have clear and significant impact on estimates, covariance matrix and standard errors. From all of the simulation results we concluded that seasonal level shift along with other four types of outlier drastically affect our analysis and make results unreliable which may lead to defective future predictions also based on this analysis and give wrong estimate of the model. We also identified that the larger the outlier size contains large empirical critical value of the joint test statistics this also identified that j_{SLS} test not only depends on outlier size ω_{SLS} but also depend on the interaction between them and also depend on covariance matrix. $\widehat{\omega}_{SLS} = (\omega_{1SLS}, \omega_{2SLS}, \omega_{3SLS})$.

In real data series, we found that existence of seasonal level shift along with other types of outliers in the multivariate time series climatic data seriously affect the estimates and standard error of residuals and standard error of estimates. Which may lead to wrong decisions based on these results, erroneous weather forecast can create major problems specially for formers because Pakistan is mostly agriculture-based country. There must be need to detect and adjust the seasonal level shift in multivariate time series analysis along with other types of outliers.

In empirical analysis we have identified three outliers, IO,TC and SLS on time point 134 and 146 for Faisalabad, three outliers are identified at t=134 for Lahore in 1st iteration TC and in 2nd iteration SLS and in third iteration IO, one AO identified at t=140 for karachi, we can see durable difference in all the estimates with seasonal VAR model for all three stations after outlier adjustment. Standard error of estimates and residuals have significant change after outlier adjustment in the data series. SLS recorded for Faisalabad at t=134 that is feburary 2019, the reason behind the outlier on that time point is that in feburary 2019 there is extreme level of humidity was recorded because of which an extreme level of rainfall was recorded and the same pattern of rainfall and humidity was recorded in 2020, this conclude that extreme rainfall and high level of humidity in 2019 And 2020 in the same season make a seasonal level shift in the weather of Pakistan. this also clarify that outlier in one series may cause outlier in another series.

We also identified that larger value at time index of outlier in data series also caused the large $\widehat{\omega}$ (size of outlier). Positive and significant $\widehat{\omega}$ show that the outlier significantly affects the series. In addition, we have concluded in our study that SLS and innovative outlier is not confused with other types of outlier with seasonal VAR model for large sample size.

Hereafter, our study determined that extra material confined by these outliers which disturbs estimates and forecasting is beneficial for all statistical analysis. Main contribution of our study in literature is that we introduced multivariate seasonal level shift directly in vector time series in terms of both theory and practice. From all the results of simulation and real data analysis we observed that MSLS have a significant impact on estimates, covariance matrix, standard error of the residua of seasonal VAR model. Without detecting and adjusting MSLS along with other types of outlier's results are not reliable.

6.2 Strategy Commendation

On the basis of results following policy is recommend that, there is need to detect seasonal level shift in multivariate times series in all the fields which follow seasonal patterns, this will improve their analysis and make results reliable and help to make right decision on the basis of results and also it is necessary for improving analysis on climatic data which helps to Pakistan formers for making right decision about their land cultivation and all agriculture business.

This study also suggests forecasting with outliers in multivariate time series and we suggested the study for forecasting in case of covariate determination.

REFERENCES

- Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.
- Ahmed, B. (2018). Climate Change Effects and Relationship with Violent Conflicts in Pakistan. *Stratagem*, *1*(2), 73-88.
- Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, *55*, 278-288.
- Akram, N., & Hamid, A. (2015). Climate change: A threat to the economic growth of Pakistan. *Progress in Development Studies*, *15*(1), 73-86.
- Alvarez, E., Garcıa-Fernández, R. M., Blanco-Encomienda, F. J., & Munoz, J. F. (2014). The effect of outliers on the economic and social survey on income and living conditions. *World Academy of Engineering and Technology, International Journal of Social, Behavioral, Education, Economic, Business and Industrial Engineering*, *8*(10), 3268-3272.
- Asghar, Z., & Urooj, A. (2012). Structural Breaks, Automatic Model Selection and Forecasting Wheat and Rice Prices for Pakistan. *Pakistan Journal of Statistics and Operation Research*, *8*(1), 1-20.
- Asghar, Z., & Urooj, A. (2017). Analysis of seasonal level shift (SLS) detection in SARIMA models. *Communications in Statistics-Simulation and Computation*, *46*(9), 7264-7318.
- Baek, C., Davis, R. A., & Pipiras, V. (2017). Sparse seasonal and periodic vector autoregressive modeling. *Computational Statistics & Data Analysis*, *106*, 103-126.
- Balke, N. S. (1991). *Detecting level shifts in time series: misspecification and a proposed solution* (No. 9109). Federal Reserve Bank of Dallas.
- Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics*, *11*(1), 81-92.
- Balke, N. S., & Fomby, T. B. (1994). Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics*, *9*(2), 181-200.
- Baragona, R., & Battaglia, F. (2007). Outliers detection in multivariate time series by independent component analysis. *Neural computation*, *19*(7), 1962-1984.
- Baragona, R., Battaglia, F., & Cucina, D. (2013). Empirical likelihood for break detection in time series. *Electronic Journal of Statistics*, *7*, 3089-3123.
- Battaglia, R., Baragona, F., & Cucina, D. (2004). Estimating threshold subset autoregressive moving-average models by genetic algorithms. *Metron*, *62*(1), 39-61.
- Bauer, M., Gather, U., & Imhoff, M. (1999). *The identification of multiple outliers in online monitoring data* (No. 1999, 29). Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund.

- Birrell, C. L., Steel, D. G., & Lin, Y. X. (2011). Seasonal adjustment of an aggregate series using univariate and multivariate basic structural models. *Journal of Statistical Theory and Practice*, *5*(2), 179-205.
- Blazsek, S. I., Escribano Sáez, Á., & Licht, A. (2018). Seasonal quasi-vector autoregressive models for macroeconomic data.
- Bowden, R. S., & Clarke, B. R. (2016). Using multivariate time series to estimate location and climate change effects on historical temperatures employed in future electricity demand simulation.
- Bowden, R. S., & Clarke, B. R. (2017). Using multivariate time series methods to estimate location and climate change effects on temperature readings employed in electricity demand simulation. *Australian & New Zealand Journal of Statistics*, *59*(4), 413-431.
- Box, G. E., Hillmer, S., & Tiao, G. C. (1979). Analysis and modeling of seasonal time series. In *Seasonal analysis of economic time series* (pp. 309-346). NBER.
- Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). *Introduction to time series and forecasting* (Vol. 2). New York: springer.
- Chang, I., Tiao, G. C., & Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, *30*(2), 193-204.
- Charles, A., & Darné, O. (2005). Outliers and GARCH models in financial data. *Economics Letters*, *86*(3), 347-352.

- Charles, A., & Darné, O. (2005). Outliers and GARCH models in financial data. *Economics Letters*, *86*(3), 347-352.

- Charles, A., & Darné, O. (2006). Large shocks and the September 11th terrorist attacks on international stock markets. *Economic Modelling*, *23*(4), 683-698.
- Chen, C., & Liu, L. M. (1993). Forecasting time series with outliers. *Journal of Forecasting*, *12*(1), 13-35.

- Chen, C., & Liu, L. M. (1993). Forecasting time series with outliers. *Journal of Forecasting*, *12*(1), 13-35.

- Chen, C., & Liu, L. M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, *88*(421), 284-297.
- Cheng, H., Tan, P. N., Potter, C., & Klooster, S. (2009, April). Detection and characterization of anomalies in multivariate time series. In *Proceedings of the 2009 SIAM international conference on data mining* (pp. 413-424). Society for Industrial and Applied Mathematics.
- Chongde, L., & Tsingan, L. (2003). Multiple intelligence and the structure of thinking. *Theory & Psychology*, *13*(6), 829-845.
- Dielman, T. E., & Rose, E. L. (1994). Forecasting in least absolute value regression with autocorrelated errors: a small-sample study. *International Journal of Forecasting*, *10*(4), 539-547.
- Emerson, S. C., & Emerson, S. S. (2011). The importance of statistical theory in outlier detection.
- Finch, H. (2012). Distribution of variables by method of outlier detection. *Frontiers in psychology*, *3*, 211.

- Fok, D. D., Franses, P. H. P. H., & Paap, R. R. (2005). *Performance of seasonal adjustment procedures: Simulation and empirical results* (No. EI 2005-30).
- Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(3), 350-363.
- Fox, J. (2002). Time-series regression and generalized least squares. *An R and S-plus companion to regression. Sage, Thousand Oaks, CA*.
- Franses, P. H. P. H. (2008). *Outliers and judgmental adjustment of time series forecasts* (No. EI 2008-04, pp. 1-14).
- Furusjö, E., Svenson, A., Rahmberg, M., & Andersson, M. (2006). The importance of outlier detection and training set selection for reliable environmental QSAR predictions. *Chemosphere*, *63*(1), 99-108.
- Galeano, P., Peña, D., & Tsay, R. S. (2006). Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association*, *101*(474), 654-669.
- Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security*, *28*(1-2), 18-28.
- Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one*, *11*(4), e0152173.
- Gómez, V., & Maravall, A. (2001). Seasonal adjustment and signal extraction in economic time series. *A course in time series analysis*, 202-247.
- Greenaway-McGrevy, R. (2013). A multivariate approach to seasonal adjustment.
- Hänninen, H. (1995). Effects of climatic change on trees from cool and temperate regions: an ecophysiological approach to modelling of bud burst phenology. *Canadian Journal of Botany*, *73*(2), 183-199.
- Hänninen, J., Vuorinen, I., & Kornilovs, G. (2003). Atlantic climatic factors control decadal dynamics of a Baltic Sea copepod Temora longicornis. *Ecography*, *26*(5), 672-678.
- Helbling, J., & Cleroux, R. O. B. E. R. T. (2009). On outlier detection in multivariate time series. *Acta Math. Vietnamica*, *34*, 19-26.
- Helbling, J., & Cleroux, R. O. B. E. R. T. (2009). On outlier detection in multivariate time series. *Acta Math. Vietnamica*, *34*, 19-26.
- Hillmer, S. (1984). Monitoring and adjusting forecasts in the presence of additive outliers. *Journal of Forecasting*, *3*(2), 205-215.
- Hossain, M. J., & Majumder, A. K. (2018). Proposed Distance-Based Test for Testing Multivariate Multiple Regression Coefficients under Restricted Alternatives. *International Journal of Statistics and Probability*, *7*(2).
- Hosseinalizadeh, M., Rivaz, F., & Hedayatizadeh, R. (2011). Importance of outlier detection in spatial analysis of wind erosion. *Procedia Environmental Sciences*, *7*, 341-346.
- Hotta, L. K., Pereira, P. L. V., & Ota, R. (2004). Effect of outliers on forecasting temporally aggregated flow variables. *Test*, *13*(2), 371-402.
- Hussain, S. A., & Mustafa, U. Impact of Climate Change on Agricultural Land Values: An Application of Ricardian Model in Punjab Pakistan.

- Ikram, F., Afzaal, M., Bukhari, S. A. A., & Ahmed, B. (2016). Past and future trends in frequency of heavy rainfall events over Pakistan. *Pakistan Journal of Meteorology Vol*, *12*(24).
- Imran, A., Zaman, Q., Rasul, G., & Mahmood, A. (2014). An analytical study of the variations in the monsoon patterns over Pakistan. *Pakistan Journal of Meteorology Vol*, *10*(20).
- Irshad, H., & Hussain, A. Analysis of Ecological Efficiency and its Influencing Factors in Developing Countries.
- Jere, S., & Moyo, E. (2016). Modelling epidemiological data using Box-Jenkins procedure. *Open Journal of Statistics*, *6*(02), 295.
- Kaiser, R., & Maravall, A. (1999). Seasonal outliers in time series.
- Kaiser, R., & Maravall, A. (2001). Seasonal outliers in time series.
- Khan, S. R. (2001). *Does Climate Change Matter in Pakistan?*. Sustainable Development Policy Institute.
- Kodde, D. A., & Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica: journal of the Econometric Society*, 1243-1248.
- Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *International Journal of Forecasting*, *5*(2), 231-240.
- Lehmann, R., & Lösler, M. (2016). Multiple outlier detection: hypothesis tests versus model selection by information criteria. *Journal of surveying engineering*, *142*(4), 04016017.
- Li, R., & Lin, D. K. (2003). Analysis methods for supersaturated design: some comparisons. *Journal of Data Science*, *1*(3), 249-260.
- Lin, H. C., Chen, C. S., Xirasagar, S., & Lee, H. C. (2008). Seasonality and climatic associations with violent and nonviolent suicide: a population-based study. *Neuropsychobiology*, *57*(1-2), 32-37.
- Liu, N., Shin, D., & Hu, X. (2017). Contextual outlier interpretation. *arXiv preprint arXiv:1711.10589*.
- Lucchetti, J. (2015). The SVAR addon for gretl.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Marascuilo, L. A., & Serlin, R. C. (1988). *Statistical methods for the social and behavioral sciences*. WH Freeman/Times Books/Henry Holt & Co.
- Marczak, M., & Proietti, T. (2016). Outlier detection in structural time series models: The indicator saturation approach. *International Journal of Forecasting*, *32*(1), 180-202.
- Mazzoni, R., & Rezende, C. F. (2003). Seasonal diet shift in a Tetragonopterinae (Oateichthyes, Characidae) from the Ubatiba river, RJ, Brazil. *Brazilian Journal of Biology*, *63*(1), 69-74.
- McElroy, T. (2017). Multivariate seasonal adjustment, economic identities, and seasonal taxonomy. *Journal of Business & Economic Statistics*, *35*(4), 611-625.
- McQuarrie, A. D., & Tsai, C. L. (2003). Outlier detections in autoregressive models. *Journal of Computational and Graphical Statistics*, *12*(2), 450-471.

- Mercy, C., & Kihoro, J. (1998). Application of Vector Autoregressive (VAR) Process in Modelling Reshaped Seasonal Univariate Time Series.
- Micenková, B., McWilliams, B., & Assent, I. (2015). Learning representations for outlier detection on a budget. *arXiv preprint arXiv:1507.08104*.
- Mielke, P. W., & Berry, K. J. (2007). Regression Analysis, Prediction, and Agreement. In *Permutation Methods* (pp. 171-262). Springer, New York, NY.
- Mourão-Miranda, J., Hardoon, D. R., Hahn, T., Marquand, A. F., Williams, S. C., Shawe-Taylor, J., & Brammer, M. (2011). Patient classification as an outlier detection problem: an application of the one-class support vector machine. *Neuroimage*, *58*(3), 793-804.
- Nair, K. K., Kiremidjian, A. S., & Law, K. H. (2006). Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure. *Journal of Sound and Vibration*, *291*(1-2), 349-368.
- Nieto, F. H. (2013). Characterizing Common Seasonality in Multivariate Time Series. *Preprint*.
- Nieto, F. H., Pena, D., & Saboyá, D. (2016). Common seasonality in multivariate time series. *Statistica Sinica*, 1389-1410.
- Oluyomi, O. O., Ismail, S. O., & Adesola, A. K. (2016). Performance of Two Generating Mechanisms in Detection of Outliers in Multivariate Time Series. *American Journal of Theoretical and Applied Statistics*, *5*(3), 115-122.
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, *9*(6), 1-12.
- Pal, I., Anderson, B. T., Salvucci, G. D., & Gianotti, D. J. (2013). Shifting seasonality and increasing frequency of precipitation in wet and dry seasons across the US. *Geophysical Research Letters*, *40*(15), 4030-4035.
- Pan, J. X., Fung, W. K., & Fang, K. T. (2000). Multiple outlier detection in multivariate data using projection pursuit techniques. *Journal of Statistical Planning and Inference*, *83*(1), 153-167.
- Pedhazur, E. J., & Kerlinger, F. N. (1982). *Multiple regression in behavioral research*. Holt, Rinehart, and Winston.
- Peña, D., & Prieto, F. J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, *43*(3), 286-310.
- Penny, W., & Harrison, L. (2007). Multivariate autoregressive models. *Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, Amsterdam*, 534-540.
- Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, *27*(4), 1-32.
- Pratikno, B. (2012). *Test of hypotheses for linear regression models with non-sample prior information* (Doctoral dissertation, University of Southern Queensland).
- Rahman, A., & Hasan, M. M. (2017). Modeling and forecasting of carbon dioxide emissions in Bangladesh using Autoregressive Integrated Moving Average (ARIMA) models. *Open Journal of Statistics*, *7*(04), 560.

- Ramachandran, G., Adgate, J. L., Pratt, G. C., & Sexton, K. (2003). Characterizing indoor and outdoor 15 minute average PM 2.5 concentrations in urban neighborhoods. *Aerosol Science & Technology*, *37*(1), 33-45.
- Ravnik, R., & Žilić, I. (2011). The use of SVAR analysis in determining the effects of fiscal shocks in Croatia. *Financial theory and practice*, *35*(1), 25-58.
- Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer Science & Business Media.
- Schmidt, P. (2010). Rezension: Timothy A. Brown, 2006: Confirmatory factor analysis for applied research. *Methoden, Daten, Analysen (mda)*, *4*(1), 58-61.
- Sedman, R. (2018). Online Outlier Detection in Financial Time Series.
- Shah, S. M. A., & Ahmad, M. F. (2017). Minimum Temperature Analysis and Trends in Pakistan. *Pakistan Journal of Meteorology Vol*, *13*(26).
- Sidra, S., Ali, Z., Nasir, Z. A., & Colbeck, I. (2015). Seasonal variation of fine particulate matter in residential micro–environments of Lahore, Pakistan. *Atmospheric Pollution Research*, *6*(5), 797-804.
- Singh, K., & Upadhyaya, S. (2012). Outlier detection: applications and techniques. *International Journal of Computer Science Issues (IJCSI)*, *9*(1), 307.
- Smith, A. (2004). Forecasting in the presence of level shifts. *UC Davis Agricultural and Resource Economics Working Paper*, (04-014).
- Smith, A. (2005). Forecasting in the presence of level shifts. *Journal of Forecasting*, *24*(8), 557-574.
- Smyth, G. K. (2004). Statistical applications in genetics and molecular biology. *Linear models and empirical Bayes methods for assessing differential expression in microarray experiments*.
- Srivastava, M. S., & Khatri, C. G. (2009). An introduction to multivariate statistics.
- Streets, D. G., Yarber, K. F., Woo, J. H., & Carmichael, G. R. (2003). Biomass burning in Asia: Annual and seasonal estimates and atmospheric emissions. *Global Biogeochemical Cycles*, *17*(4).
- Tang, G., Pei, J., Bailey, J., & Dong, G. (2015). Mining multidimensional contextual outliers from categorical relational data. *Intelligent Data Analysis*, *19*(5), 1171-1192.
- Tanja, S., Berninger, F., Vesala, T., Markkanen, T., Hari, P., Mäkelä, A., ... & Laurila, T. (2003). Air temperature triggers the recovery of evergreen boreal forest photosynthesis in spring. *Global change biology, 9*(10), 1410-1426.
- Thury, G., & Wüger, M. (1992). Outlier detection and seasonal adjustment. *Empirica*, *19*(2), 245-257.
- Tiao, G. C., & Tsay, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *The Annals of Statistics*, 856-871.
- Trívez, F. J. (1995). Level shifts, temporary changes and forecasting. *Journal of Forecasting*, *14*(6), 543-550.
- Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, *81*(393), 132-141.

- Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of forecasting*, *7*(1), 1-20.
- Tsay, R. S., Pena, D., & Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika*, *87*(4), 789-804.
- Ullah, S. (2017). Climate change impact on agriculture of Pakistan-A leading agent to food security. *Int. J. Environmental Science and Natural Resources*, *6*, 1-4.
- Uriel, E. (2013). Hypothesis testing in the multiple regression model. *Universidad de Valencia: Department of economics*.
- Vaage, K. (2000). Detection of outliers and level shifts in time series: An evaluation of two alternative procedures. *Journal of Forecasting*, *19*(1), 23-37.
- Wagan, S. A., Jingdong, L., Shuangxi, X., Noonari, S., Memon, I. N., Memon, Q. U. A., & Wagan, G. H. (2016). Economic Assessment of Rice Production in Sindh, Pakistan. *American Journal of Environmental and Resource Economics*, *1*(1), 37-42.
- Wang, C., & Weiss, D. J. (2018). Multivariate hypothesis testing methods for evaluating significant individual change. *Applied psychological measurement*, *42*(3), 221-239.
- Wang, M. Y., & Zwilling, C. E. (2015). Multivariate computing and robust estimating for outlier and novelty in data and imaging sciences. *Advances in Bioengineering*, 317-336.
- Wang, X. (2011, July). Two-phase outlier detection in multivariate time series. In *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (Vol. 3, pp. 1555-1559). IEEE.
- Weng, X., & Shen, J. (2008). Detecting outlier samples in multivariate time series dataset. *Knowledge-based systems*, *21*(8), 807-812.
- Wilcox, V., & Zealand, S. N. (2000). *Seasonal Adjustment of QGDP: The Implications of Outliers*. Statistics New Zealand.
- Xia, Y., Cai, T. T., & Li, H. (2018). Joint testing and false discovery rate control in high-dimensional multivariate regression. *Biometrika*, *105*(2), 249-269.
- Yu, Y., Zhu, Y., Li, S., & Wan, D. (2014). Time series outlier detection based on sliding window prediction. *Mathematical problems in Engineering*, *2014*.
- Yunfei, F., Yihua, L., Liu, G., & Qiang, W. A. N. G. (2003). Seasonal characteristics of precipitation in 1998 over East Asia as derived from TRMM PR. *Advances in Atmospheric Sciences*, *20*(4), 511-529.
- Zhu, G. F., Li, X., Su, Y. H., Lu, L., & Huang, C. L. (2011). Seasonal fluctuations and temperature dependence in photosynthetic parameters and stomatal conductance at the leaf scale of Populus euphratica Oliv. *Tree Physiology*, *31*(2), 178-195.
- Zijlstra, G. A. R., Van Haastregt, J. C. M., Van Eijk, J. T. M., van Rossum, E., Stalenhoef, P. A., & Kempen, G. I. (2007). Prevalence and correlates of fear of falling, and associated avoidance of activity in the general population of community-living older people. *Age and ageing*, *36*(3), 304-309.
- Zijlstra, G. R., van Haastregt, J. C., Du Moulin, M. F., de Jonge, M. C., van der Poel, A., & Kempen, G. I. (2012). Effects of the implementation of an evidence-

based program to manage concerns about falls in older adults. *The Gerontologist*, *53*(5), 839-849.

- Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling Financial Time Series with S-Plus®*, 385-429.
- Zwilling, C. E., & Wang, M. Y. (2014, October). Multivariate voronoi outlier detection for time series. In *2014 IEEE Healthcare Innovation Conference (HIC)* (pp. 300-303). IEEE.

APPENDIX

APPENDIX A

This appendix related to chapter number 3

A VAR $_{(p)\ (P)\ s}$ model with r dimension and p number of lags in matrix form can be written as:

$X_{1t} = \alpha_{11} + A'_{1,1,1}. X_{1,t-1} + A'_{1,2,1}. X_{2,t-1} + A'_{1,3,1}. X_{3,t-1} + \ldots\ldots + A'_{1,r,1}. X_{r,t-1} + \ldots.. + A'_{1,1,p}. X_{1,t-p} + A'_{1,2,p}. X_{2,t-p} + A'_{1,3,p}. X_{3,t-p} + \ldots\ldots + A'_{1,r,p}. X_{r,t-p} + u_{1t}$

$X_{2t} = \alpha_{21} + A'_{2,1,1}. X_{1,t-1} + A'_{2,2,1}. X_{2,t-1} + + A'_{2,3,1}. X_{3,t-1} + \ldots.. + A'_{2,r,1}. X_{r,t-1} + \ldots.. + A'_{2,1,p}. X_{1,t-p} + A'_{2,2,p}. X_{2,t-p} + A'_{2,3,p}. X_{3,t-p} + \ldots.. + A'_{2,r,p}. X_{r,t-p} + u_{2t}$

$X_{3t} = \alpha_{31} + A'_{3,1,1}. X_{1,t-1} + A'_{3,2,1}. X_{2,t-1} + + A'_{3,3,1}. X_{3,t-1} + \ldots.. A'_{3,r,1}. X_{r,t-1} + \ldots.. + A'_{3,1,p}. X_{1,t-p} + A'_{3,2,p}. X_{2,t-p} + A'_{3,3,p}. X_{3,t-p} + \ldots\ldots + A'_{3,r,p}. X_{r,t-p} + u_{3t}$

:

$X_{rt} = \alpha_{R1} + A'_{r,1,1}. X_{1,t-1} + A'_{r,2,1}. X_{2,t-1} + + A'_{r,3,1}. X_{3,t-1} + \ldots.. A'_{r,r,1}. X_{r,t-1} + \ldots.. + A'_{r,1,p}. X_{1,t-p} + A'_{r,2,p}. X_{2,t-p} + A'_{r,3,p}. X_{3,t-p} + \ldots\ldots + A'_{r,r,p}. X_{r,t-p} + u_{rt}$

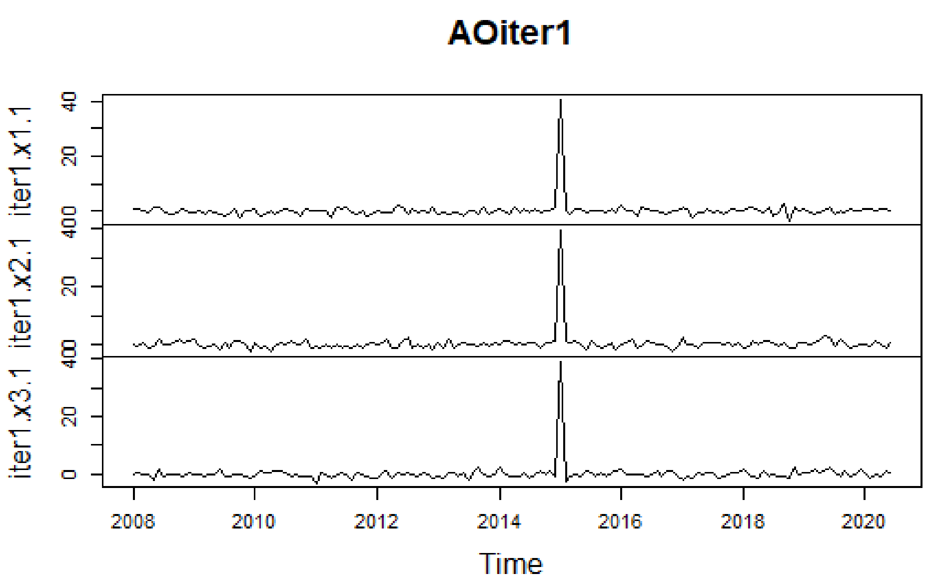
$$
\begin{bmatrix} X1t \\ X2t \\ X3t \\ . \\ . \\ Xrt \end{bmatrix} = \begin{bmatrix} \alpha 11 \\ \alpha 21 \\ \alpha 31 \\ . \\ . \\ \alpha r1 \end{bmatrix} + \begin{bmatrix} A1,1,1 & A1,2,1 & A1,3,1 \ldots & A1,r,1 \\ A2,1,1 & A2,2,1 & A2,3,1 \ldots & A2,r,1 \\ A3,1,1 & A3,2,1 & A3,3,1 \ldots & A3,r,1 \\ . & . & . & . \\ . & . & . & . \\ Ar,1,1 & Ar,2,1 & Ar,3,1 \ldots & Ar,r,1 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{3,t-1} \\ . \\ . \\ X_{r,t-1} \end{bmatrix} + - - -
$$

$$
+ \begin{bmatrix} A1,1,p & A1,2,p & A1,3,p \ldots & A1,r,p \\ A2,1,p & A2,2,p & A2,3,p \ldots & A2,r,p \\ A3,1,p & A3,2,p & A3,3,p \ldots & A3,r,p \\ . & . & . \ldots & . \\ . & . & . \ldots & . \\ Ar,1,p & Ar,2,p & Ar,3,p \ldots & Ar,r,p \end{bmatrix} \begin{bmatrix} X_{1,t-p} \\ X_{2,t-p} \\ X_{3,t-p} \\ . \\ . \\ X_{r,t-p} \end{bmatrix} + \begin{bmatrix} u1t \\ u2t \\ u3t \\ . \\ . \\ urt \end{bmatrix}
$$

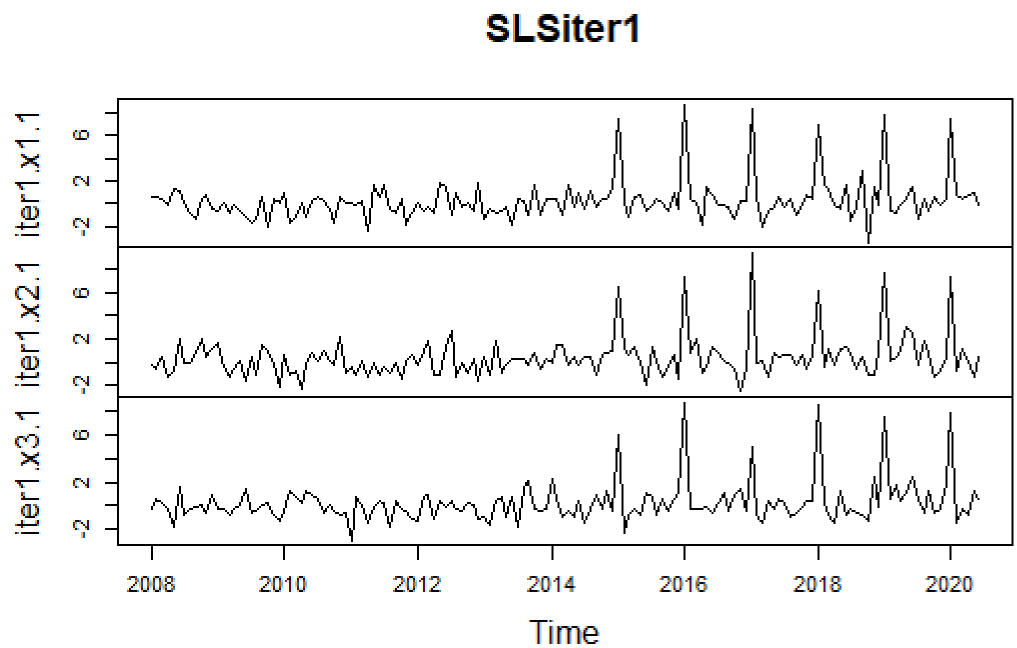
Appendix B

These are the figures of simulated data with outliers which is explained in chapter 4.

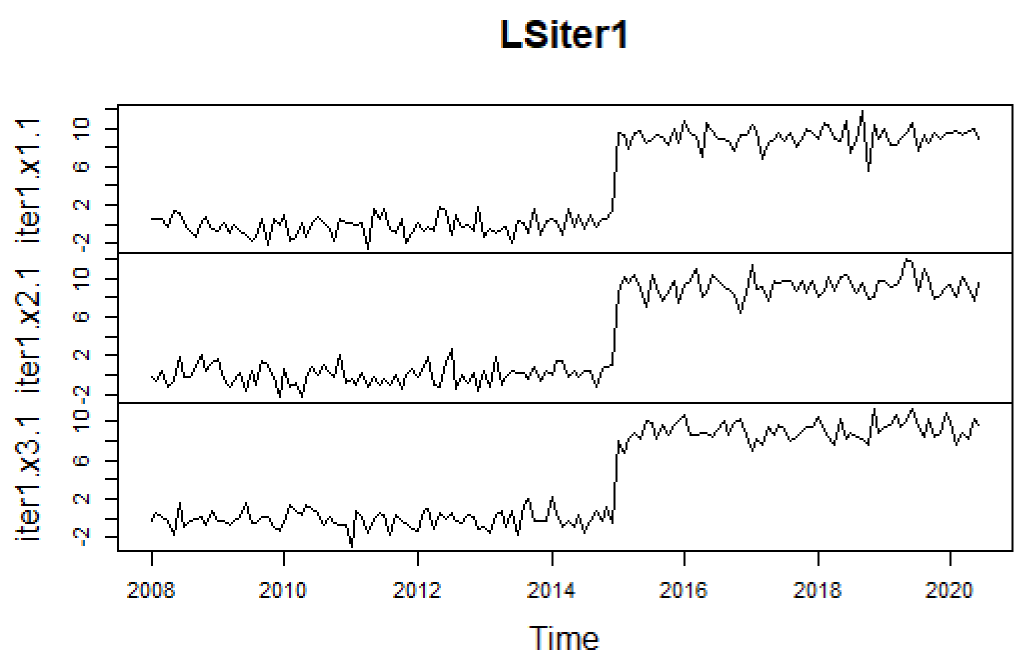
4FIGURE4.1. SERIES WITH ADDITIVE OUTLIER



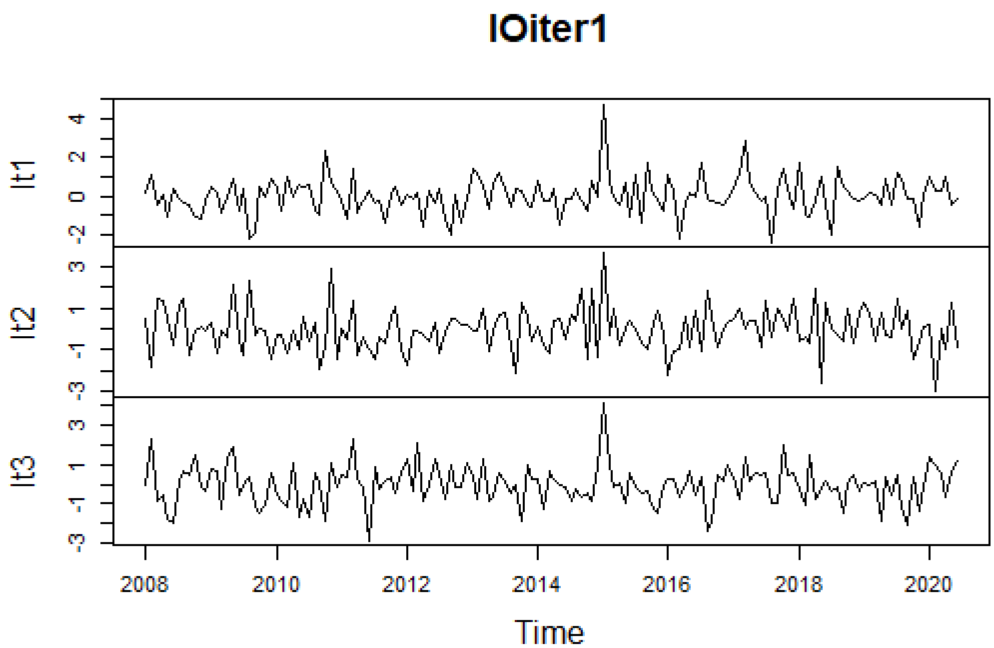
5FIGURE4.2. SERIES WITH SEASONAL LEVEL SHIFT



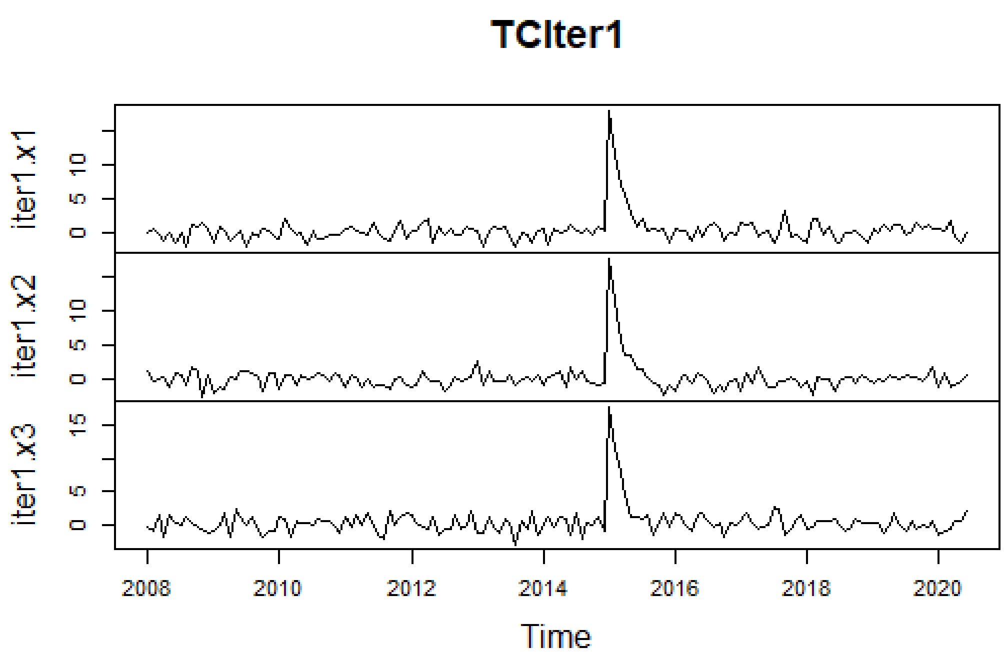
6FIGURE4.3. SERIES WITH LEVEL SHIFT



7FIGURE4.4. SERIES WITH INNOVATIVE OUTLIER



8FIGURE4.5. SERIES WITH TRANSIENT CHANGE



These are the figures of time series data of sample size 150 with five types of outliers named AO, SLS, LS, IO and TC respectively. These data series are generated from "rnorm" in a statistical package R-language and then we introduced all types of outlier at a time point t=85 in these series and then detected the outliers from these series by using a method which s explained in chapter 3 to check the performance of suggested procedure. Mostly we used five libraries in Rstudio named "vars", "matlib", "Hmisc", "nlme", "matrixstats".