Kurdistan Regional Government
Ministry of Higher Education and
Scientific Research
University of Sulaimani
College of Commerce

Comparison Between Classical and Spatial Regression Techniques using Fuzzy Logic

A Thesis Submitted to

The council of the College of Commerce University of Sulaimani in partial Fulfillment of the Requirements for the Degree of Master of Science in Statistics

By

Sham Azad Rahim

Supervised By

Assistant professor

Dr.Mohammad Mahmood Faqe Hussen

2016 M **2716 K** **1437H**

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

قَالُوا سُبْحَانَكَ لَا عِلْمَ لَنَا إِلَّا
مَا عَلَّمْتَنَا ۖ إِنَّكَ أَنْتَ الْعَلِيمُ
الْحَكِيمُ ﴿٣٢﴾

صَدَقَ اللَّهُ الْعَظِيمُ

سورة البقرة

آية (٣٢)

Linguistic Evaluation Certification

This is to certify I, Ranji Shorsh Rauf, have proofread this thesis entitled "***Comparison Between Classical and Spatial Regression Techniques using Fuzzy Logic***" by Sham Azad Rahim. After marking and correcting the mistakes, the thesis was handed again to the researcher to make the corrections in this last copy.

Signature:

Name: Ranji Shorsh Rauf Muhammad

Date : 16 / 10 / 2016

Department of English, School of Languages, Faculty of Humanities, University of Sulaimani.

Supervisor Certification

I certify that the preparation of thesis titled "***Comparison Between Classical and Spatial Regression Techniques using Fuzzy Logic***", accomplished by (***Sham Azad Rahim***) was prepared under my supervision in the College of Commerce, Statistic and Computer department as a partial fulfillment of the requirement for the degree of Master of Science in Statistics.

Signature:

Name: Dr. Mohammad Mahmood Faqe Hussen

Title: Asst. Professor

Date: 19/10/ 2016

Chairman Certification

In view of the available recommendation, I forward this thesis for debate by the examining committee.

Signature:

Name: Mahdi Mohammed Younis

Title: Asst. Lecturer

Date: 23/10 / 2016

Exam Committee Certificate

We are the exam committee certificate that we read this thesis having title: "***Comparison Between Classical and Spatial Regression Techniques using Fuzzy Logic***" and have examined the student (**Sham Azad Rahim**) in its contents, and that in our opinion, it is adequate as a thesis for the degree of Master of Science in Statistics.

Signature:

Name: Dr.Shawnm A.Mhedin

Title: Asst.Professor

Date: 20/10/2016

(Chairman)

Signature:

Name: Dr.Kawa M.Jamal Rashid

Title: Asst. Professor

Date: / / 19-10-2016

(Member)

Signature: Rez

Name: Dr. Rezan H. Kareem

Title: Lecturer

Date: 20/10/2016

(Member)

Signature:

Name: Dr. Mohammad M.Faqe

Title: Asst. Professor

Date: 19/10/2016

(Supervisor-Member)

Approved by the council of Commerce College

Signature:

Name: Dr. Narmen M. Ghafor

Title: Asst. Professor

Date: 24/10/2016

Dedication

I dedicate this thesis to:

- My dear **Parents lovely mother and sisters**
- My **father's mighty soul**
- My lovely **husband**
- My **one and only one daughter**
- My dear supervisor **(Dr.Mohammad Mahmood Faqe),**

  who taught me every time and always help me.

SHAM

Acknowledgements

First of all, praise and thank to Allah who facilitated this thesis. Also, I would like to thank the dean of College of Commerce, Dr. Narmin Marof Khafur, and the head of Statistics and Computer Department

I would like to acknowledge and express my great appreciation to my supervisor, Dr. Mohammad Mahmood Faqe, for his guidance. He has been of a great help for me to write this thesis, and I have learned so much from him.

And I would like to thanks dean of College of Administration and Economy, Dr.Kawa Muhamad jamal Rashid and special thanks to statistic department which always help me.

Then, I would like to thank my family, my lovely mother, my lovely husband and my nice sisters for their love and support. They have always encouraged me to achieve my goals and helped me to get through tough times. I would not have finished my study without them.

I would like to thanks my teachers Dr.Shawnm Abdulqader and Dr.Rezan Muhamad Rashid who always advised me.

I would like to express my deep gratitude to Dr.Wasfi Tahir Kahwachi, who taught me so much.

I also would like to thank those teachers who did their best to support me furthermore, library unit of college.

SHAM

Abstract

Spatial regression methods allow us account for dependence between observation, which often arise when observation are collected from points or regions located in space. Labor for spatial dependency or spatial parameter in analysis phenomenon is going to find important information instead of time because that we must find mathematical models which in of them spatial parameter and it is spatial regression models which show effect of explanatory variables on response variable where we have spatial effect for the neighbor places according the role for weigh matrix.In this thesis study the effect of explanatory variables air temperature,wind speed and relative humidity to the response variable Atmospheric Pressure.

The data were collected from 27 places or stations in Kurdistan region(Sulaimaniyah,Erbil,and Dhouk). Spatial Regression Models used to find the effect of neighbor place. Spatial Auto regressive Model (SAR) and Spatial Error Model (SEM) which two models have a spatial parameter and also use General Linear Model (GLM) .The different criteria or measures like (Adjusted determinations of coefficient (R^2_{adj}) , root mean square error (RMSE) , mean absolute percentage error (MAPE) , Akaike Information Criterion(AICc) used for finding the best fit model and for detecting the spatial dependency use Moran's test and also Lagrange tests used to select the best spatial model (SAR and SEM) with three weight matrixes rook,bishop and queen.

The important result in the practical part shows that the spatial regression model is better than the general linear model. The parameters of the Spatial Autoregressive Model(SAR) for queen and rook matrices is significant and

while the converting data into fuzzy and applying the models GLM,SAR and SEM and comparison between them depending on some criteria or measures like $R^2_{adj,,}$RMSE,MAPE,AIC_C the result show that the data converted to fuzzy is better than the data with unfuzzy and spatial regression with fuzzy data is better than the spatial regression with unfuzzy and the best appropriate model is SAR depending on rook weight matrix. Finally the most appropriate model was obtained from the analysis in the practical part as follows:

SAR Model depending on the queen matrix with raw data:

$$\hat{y}_i = 6.4880 + 0.1096\ A.T + 2.0739 R.H + 0.0240\ \lambda$$

SAR Model depending on the rook matrix with fuzzy data :

$$\hat{y}_i = 7.8610 - 0.3162 W.S + 0.0632\ A.T + 1.6837\ R.H - 0.0019\ \lambda$$

Table of Contents

List of Table

List of Figure

List of abbreviations

Abbreviation	Details
OLS	Ordinary Least Squares
MLE	Maximum Likelihood Estimator
WLS	Weighted Least Squares
BLUE	Best Linear Unbiased Estimation
SAR	Spatial autoregressive model
SEM	Spatial error model
SLM	Spatial Lag model
D.W	Durbin Watson
P-value	Probability value

A.P	Atmospheric Pressure
T.D	Temperature degree
R.H	Relative Humidity
W.S	Wind Speed
GWR	Geographically weight regression
SRT	Spatial regression test
AIC	Akaike Information Criterion
NCDC	National Climatic Data Center
HPRCC	High Plains Regional Climate Center
GIS	Geographically information system
CDC	Centers for Disease Control
NO	Neutrogena
SDM_s	Species distribution models
GLMM	Generalized linear mixed models
SATUM	Aware Technology Utilization Model
MSE	Mean square error
CAD	computer-aided design
SRID	Single Radial Immunodiffusion
W_R	Rook weight matrix
W_B	Bishop weight matrix
W_Q	Queen weight matrix

GLM	General linear model
LM	Lagrange multiplier
RMSE	Root Mean square error
MAPE	Mean absolute percentage error
R.A	Ronald Fisher
R.S	Remote Sensing
SAC	Scottish Attainment Challenge
EU	European
SDPD	Spatial dynamic panel data
FL	Fuzzy logic
VIF	Variance Inflation Factor
RAM	Random-access memory

Chapter One

Introduction , Literature Review and the Objective of the Study

Chapter One

Introduction and the Literature Review

1-1 Introduction

Model selection is the task of selecting a statistical model from a set of potential model given data. In its most basic forms, this is one of the fundamental tasks of scientific inquiry.

An important component of any linear modeling problem consists of determining an appropriate size and Form for the design matrix. Improper specification may substantially impact both estimators of the model parameters and predictors of the response variable: under specification may lead to results which are severely biased term which is called the under-fitting model, whereas over specification may lead to results with unnecessarily high variability term which is called the over-fitting model.

In statistical modeling, one of the main objectives is to select a suitable model from a candidate class to characterize the underlying data. Model selection criteria provide a useful tool in this regard. A selection criterion assesses whether a fitted model offers an optimal balance between goodness of fit and parsimony. Ideally, a criterion will identify candidate models which are either too simplistic to accommodate the data or unnecessarily complex.

Several model selection criteria have been used in computer vision, and many other have found popularity in the statistics literature. Due to their better accuracy in estimating the correct model, such as multiple linear regression and spatial regression model which F-statistic have become widely used. more recently, information theoretic model selection criteria have gained increasing popularity[18].

Regression analysis is a statistical tool for the implementing of relationships between variables. Usually, the implementer seeks to the causal effect of one variable upon another—the effect of a price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate. To explore such issues, the implementer assembles data on the underlying variables of interest and employs regression to estimate the quantitative effect of the causal variables upon the variable that they influence. The investigator also typically assesses the "statistical significance" of the estimated relationships, that is, the degree of confidence that the true relationship is close to the estimated relationship.
And you have two types of regression the first one is Simple linear regression it is a statistical method that allows us to abstract and study relationships between two continuous (quantitative) variables and it is defined to be $y_i=B_0+B_1x_1+e_i$ and The second is multiple linear regression it attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observe data. Every value of the independent variable *x* is associated with a value of the dependent variable *y*. The population regression line for *p* explanatory variables(x_1, x_2, ... , $x_{p)}$ is defined to be $y_i=B_0 + B_1x_1 + B_2x_2 + ... + B_px_{p+}e_i$. This line describes how the mean response y changes with the explanatory variables. The observed values for *y* vary about their means y and are assumed to have the same standard deviation. The fitted values b_0, b_1, ..., b_p estimate the parameters 0, 1, ..., p of the population regression line [22]

Spatial regression is methods capture spatial dependency, bypass statistical problems such as unsettled parameters and unreliable significance tests, as well as providing information on spatial relationships among the variables involved depending on the specific technique, spatial dependency can enter the regression model as relationships between the independent variables and the dependent, between the dependent variables and a spatial lag of itself, or in the error terms. Geographically weighted regression (GWR) is a local version of spatial regression that generates parameters disaggregated by the spatial units of analysis. This allows assessment of the spatial heterogeneity in the estimated relationships between the independent and dependent variables. and the history of spatial analysis is late 1950s and early 1960s stopped for the late 1960s and early 1970s. Not usually mentioned in the geographic literature is that the seeds planted in the quantitative revolution produced a steady produce of contributions that has now evolved into a vibrant field, both inside and outside the discipline of geography.[9] [24]

Spatial autoregressive model Some time is called mixed model or mixed regressive model because connect between ordinary least square and spatial lag model in a dependent variable.Insert depend variable that is spatialy deferent as one of explanatory variable (WY). [25]

In spatial error model the important term is independency of error term and in this model errors correlated spatially and the aim of this model is correct spatial error[8][15]

Spatial regression methods allow us to account for dependence between observations, which often arises when observations are collected from points or regions located in space. We might also have individual form establishment point locations indicate by latitude-longitude coordinates that can be found by applying geo-coding software to the postal address. It is commonly observed that sample data collected for regions or points in space are not independent, but rather spatially dependent, which means that observations from one location tend to exhibit values similar to those from near by locations[26]

Fuzzy logic starts with and builds on a set of user-supplied human language rules. The fuzzy systems convert these rules to their mathematical equivalents. This simplifies the job of the system designer and the computer, and results in much more accurate representations of the way systems behave in the real world .

The membership function of a fuzzy set is a generalization of the indicator function in classical sets. In fuzzy logic, it represents the degree of truth as an extension of valuation. Degrees of truth are often confused with probabilities. For any set X, a membership function on X is any function from X to the real unit interval [0,1] . [27]

1-2 Literature Review

Spatial regression is measured by spatial dependency , which is a property of data that arises whenever there is a spatial pattern in the values located on a map, as opposed to a random pattern that indicates no spatial autocorrelation. To measure the spatial pattern (spatial association and spatial dependency), some standard global and local spatial statistics have been developed. These include Moran's I, Geary's C and Getis statistics. Besides spatial dependence in the data, there can be spatial heterogeneity. This means that the underlying process being studied may vary systematically over space. This creates problems for regression and other econometric methods that do not accommodate spatial variation in the relationship. [28]

In 1998 Christian A. L. Hilber study Neighborhood Externality Risk and The Homeownership Status of Properties

In contrast to corporate and institutional investors, single owner-occupiers cannot adequately diversify housing investment risk. Consequently, homeownership should be relatively less likely in places with higher housing investment risk. Using the American Housing Survey, it is documented that neighborhood externality risk, a major component of housing investment risk, substantially reduces the probability that a housing unit is owner-occupied, even when controlling for housing type and numerous location and household specific characteristics. The effects are quantitatively meaningful and change-in-change estimates suggest that the effects are causal[30]

In 1999, Gustav Visser study Researcher positionality and political-temporal contingency in a post apartheid research environment

Since the mid 1990s South Africa has experienced a series of fundamental changes in its local government system. This paper suggests that these institutional changes underpin a "new" research environment, the dynamics of which have not been commented upon in academic methodological debate. Working in the context of both an established and an emerging geographical debate, the paper gives further voice to the impact of researcher positionality and political-temporal contingency in the research of local government elites[40]

In 2001, Smirnov, O and Anselin, L studies the Fast maximum likelihood estimation of very large spatial autoregressive models a characteristic polynomial approach Computational Statistics & Data Analysis.

This paper states that the maximization of the log-likelihood function used in spatial autoregressive models is computationally intensive and requires significant amounts of memory. This becomes problematic during analysis when very large spatial data sets are used. This papers contribution is a new method for evaluating the Jacobian term based on the characteristic polynomial of the spatial weights matrix W. Comparisons made between Cholesky factorization and this characteristic polynomial algorithm showed pronounced improvement when large data sets (n>50000) were examined. In addition, the Cholesky algorithm failed when large data set were used due to large memory requirements. Clearly the characteristic polynomial algorithm proposed by this paper is preferred when using large data sets. This algorithm also includes a tuning variable to vary the accuracy of the result. However, increasing the accuracy of the result also increased computation times. The proposed solution is O (n) solution for regular lattices and O (nlogn) for irregular lattices [29]

In 2001, Badi H.Baltagi study the Companion to Theoretical Econometrics

This new attention to specifying, estimating, and testing for the presence of spatial interaction in the mainstream of applied and theoretical econometrics can be attributed to two major factors. One is a growing interest within theoretical economics in models that move towards an explicit accounting for the interaction of an economic agent with other heterogeneous agents in the system. These new theoretical frameworks of “interacting agents” model strategic interaction, social norms, neighborhood effects, copy-catting, and other peer group effects, and raise interesting questions about how the individual interactions can lead to emerge collective behavior and aggregate patterns. Models used to estimate such phenomena require the specification of how the magnitude of a variable of interest (say crime) at a given location (say a census tract) is determined by the values of the same variable at other locations in the system (such as neighboring census tracts).[30]

In 2002 Luc Anselin study the Under the Hood Issues in the Specification and Interpretation of Spatial Regression Models

This paper reviews a number of conceptual issues pertaining to the implementation of an explicit "spatial" perspective in applied econometrics. It provides an overview of the motivation for including spatial effects in regression models, both from a theory-driven as well as from a data-driven perspective. Considerable attention is paid to the inferential framework necessary to carry out estimation and testing and the different assumptions, constraints and implications embedded in the various specifications available in the literature. The review combines insights from the traditional spatial econometrics literature as well as from geostatistics, biostatistics and medical image analysis. [31]

In 2002 A.S. Fotheringham, C. Brunsdon, M.E. Charlton and Wiley Chichester studies the Geographically Weighted Regression the Analysis of Spatially Varying Relationships.

Authors mentioned few techniques that incorporate local spatial relationships in to the regression framework, which is very popular and well known in the statistics community. Authors showed that hedonic price model to capture price variation in London housing market are incorrect. This is because of non-stationary property exhibited by the dataset. One method attempts to calibrate the geographic model based on established boundaries[29]

In 2003 Edmonton, Alberta study the Spatial Analysis and Timber Potential in the Deh Cho Territory

The Deh Cho Land Use Planning Committee is responsible for developing a land use plan for the Deh Cho territory, pursuant to the Deh Cho Interim Measures Agreement. To assist in the completion of this effort, the Committee commissioned PACTeam Canada and associates to prepare of a "Spatial Analysis and Literature Review of Timber Potential of the Deh Cho Territory" that will contribute to the information base to be used in the development of a land use plan for the area. In the context of this project, timber refers to trees of sawlog size only.[32]

In 2003 Paaß, G. and Kindermann,J studies the Bayesian regression mixtures of experts for georeferenced data.

This paper identifies the need for politicians, planners and social scientist to be provided the tools to clarify and manipulate spatial distributions to predict future developments. Bayesian statistics offers a way to estimate values of a variable at locations that are not sampled. The paper tries to address a case where Tobler's law is not applicable [29]

In 2004 Baris M. Kazar, Shashi Shekhar, David J. Lilja, Ranga Raju Vatsavai and R. Kelley Pace studies the Comparing Exact and Approximate Spatial Auto-regression Model Solutions for Spatial Data Analysis.GIScience.

Applications that use spatial auto-regression (SAR) for data mining are working with ever increasing sizes of geo-spatial databases. The explosive growth in databases coupled with the demand for exact solutions for estimating SAR parameters are both computationally expensive and memory intensive. This paper presents two candidate approximate-semi-sparse solutions of the SAR model based on the Taylor series expansion and Chebyshev polynomials. When accuracy of these new approximation algorithms and an exact algorithm were compared, both provided accurate results. However, the approximation algorithms outperformed the exact algorithm in both terms of computation and memory usage. It was also noted that the exact algorithm was unable to solve any problem with over 10K observation points. They performed experiments on satellite imagery. Authors suggested exploring better model based on this approach to get better prediction[29]

In 2004 dani gameman study the Multivariate spatial regression models

This paper describes the inference procedures required to perform Bayesian inference to some multivariate econometric models. These models have a spatial component built into commonly used multivariate models. In particular, the common component models are addressed and extended to accommodate for spatial dependence. Inference procedures are based on a variety of simulation-based schemes designed to obtain samples from the posterior distribution of model parameters. They are also used to provide a basis to forecast new observation [33]

In 2005 Nathaniel B. Guttman study the Spatial Regression as a technique for assessing the quality of tempera true data

The SRT methodology is being incorporated into the NCDC processing system. Because the SRT software developed by the HPRCC could not be "plugged and played" in the NCDC system, new code had to be written. In order to insure that the code was written as intended, parallel testing is being conducted on January through May, 2005 data for the lower 48 states. The NCDC version of the software is being run at the NCDC, and the HPRCC version is being run at the HPRCC, and results are being compared for one-to one correspondence [35]

In 2005 Arthur Getis study the screening for spatial dependence in regression analysis

A technique of analysis is presented that is designed to circumvent the problem of finding was y to estimate parameters of spatially stochastic independent variables. It is based on (1) a type of second-order analysis that describes the spatial association among weighed observations, and (2) a screening procedure that removes most of the spatial dependence in the dependent variable. The approach is illustrated by a study of the incidence of certain crimes in 49 districts of Columbus, Ohio. It is concluded that spatial just a position of observations plays a large role in regression analyses that are based on spatial series[38]

In 2006 Prasanna Man Shrestha study the Comparison of Ordinary Least Square Regression, Spatial Autoregression, and Geographically Weighted Regression for Modeling Forest Structural Attributes Using a Geographical Information System (GIS)/Remote Sensing (RS) Approach

The performances of three modeling techniques: (i) ordinary least square (OLS) regression, (ii) spatial autoregression (SAR) and (iii) geographically weighed regression (GWR) were compared for the task of predicting a key forest structural parameter crown closure – across a study area in west-central Alberta using a series of spectral and topographic variables.[49]

In 2006 Sohair F Higazi study the Application of Spatial Regression Models to Income Poverty Ratiosin Middle Delta Contiguous Counties in Egypt

Regression analysis depends on several assumptions that have to be satisfied. A major assumption that is never satisfied when variables are from contiguous observations are the independence of error terms. Spatial analysis treated the violation of that assumption by two derived models that put contiguity of observations into consideration.[46]

In 2007 Maria Carbolic, Mirand Cuffaro, and Peter Nijkamp studies the uses spatial econometric methods to analyze the relationship in unemployment rates among different regions of the Italian labor market

Determine a model that explains the spatial differences in unemployment rates among Italy's provinces through the use of both equilibrium and disequilibrium variables. The authors classify equilibrium variables as those that coincide with the equilibrium explanation of the spatial distribution of unemployment rates- that workers migrate to areas with new jobs until personal utility is constant across all provinces, and high unemployment in a specific province is balanced by other positive attributes of that province. Demographic measures are perfect examples of equilibrium variables; Cracolici's equilibrium variables seem to roughly correspond to Topa's sorting variables[42]

In 2007 Raymond J.G.M. Florax and Peter Nijkamp studies the Misspecification in Linear Spatial Regression Models

Spatial effects are endemic in models based on spatially referenced data. The increased awareness of the relevance of spatial interactions, spatial externalities and networking effects among actors, evoked the area of spatial econometrics. Spatial econometrics focuses on the specification and estimation of regression models explicitly incorporating such spatial effects. The multidimensionality of spatial effects calls for misspecification tests and estimators that are notably different from techniques designed for the analysis of time series. With that in mind, we introduce the notion of spatial effects, referring to both heterogeneity and interdependence of phenomena occurring in two dimensional space. Spatial autocorrelation or dependence can be detected by means of cross correlation statistics in univariate as well as multivariate data settings.[40]

In 2008 James p.lesage study the Spatial Regression-Based Model Specifications for Exogenous and Endogenous Spatial Interaction

Studying relationships between environmental factors and infectious diseases is an important topic in public health research. The existing studies have been focused on temporal correlations among environmental risks and infectious disease outbreaks. In this paper, we advocate the importance of spatial data analysis in infectious disease-related environmental analysis. Using data from the Beijing CDC, we have conducted spatial regression analysis to study correlation between Measles occurrences and the following environmental factors population density and proximities to railways, roads, and water systems. We report some preliminary findings concerning significant spatial autocorrelation identified from our analysis[34]

In 2009 Andhra Pradesh study the Spatial Regression Analysis Model for Temporal Data Mining in Estimation of House Hold Data Through Different States in India

In this work a model is going to be developed which helps in measuring household data distributed over a wide area. The model considers the assumption that the households data follows an ordered sequence. The house hold data at some states is considered from the census data. A grid point identifies each state. Each grid point is identified by a set of coefficients. These coefficients are represented in terms . Thus known house hold data from the census, a set of simultaneous equations will be developed by multivariate regression model. By solving these simultaneous equations, coefficients of the simultaneous equations will be calculated. These coefficients will be used to generate household data for any years between known data and also for any future house hold data analysis.[39]

In 2009 Luc Anselin study the Thirty Years of Spatial Econometrics

In this paper, I give a personal view on the development of the field of spatial econometrics during the past thirty years. I argue that it has moved from the margins to the mainstream of applied econometrics and social science methodology. I distinguish three broad phases in the development, which I refer to as preconditions, take off and maturity. For each of these phases I describe the main methodological focus and list major contributions. I conclude with some speculations about future directions.[43]

In 2010 Colin M.Beale study the Regression analysis of spatial data

Many of the most interesting questions ecologists ask lead to analyses of spatial data yet, perhaps confused by the large number of statistical models and fitting methods available, many ecologists seem to believe this is best left to specialists Here, we describe the issues that need consideration when analyzing spatial data and illustrate these using simulation studies. Our comparative analysis involves using methods including generalized least squares, spatial filters, wavelet revised models, conditional autoregressive models and generalized additive mixed models to estimate regression coefficients from synthetic but realistic data sets, including some which violate standard regression assumptions[36]

In 2010 Iowa. J Environ Manage study the GIS-based spatial regression and prediction of water quality in river networks

Nonpoint source pollution is the leading cause of the U.S.'s water quality problems. One important component of nonpoint source pollution control is an understanding of what and how watershed-scale conditions influence ambient water quality. This paper investigated the use of spatial regression to evaluate the impacts of watershed characteristics on stream NO(3) NO(2)-N concentration in the Cedar River Watershed, Iowa. An Arc Hydro geo database was constructed to organize various datasets on the watershed. Spatial regression models were developed to evaluate the impacts of watershed characteristics on stream NO(3)NO(2)-N concentration and predict NO(3)NO(2)-N concentration at unmonitored locations. Unlike the traditional ordinary least square (OLS) method, the spatial regression method incorporates the potential spatial correlation among the observations in its coefficient estimation[37]

In 2011 Jørgen Lauridsen& Reinhold Kosfeld studies the Spurious spatial regression and heteroscedasticity

A two-step Lagrange Multiplier test strategy has recently been suggested as a device to reveal spatial non stationary and spurious spatial regression. The present paper generalizes this procedure by incorporating control for unobserved heteroscedasticity. Using Monte Carlo simulation, the behavior of several relevant tests for nonstationarity and/or heteroscedasticity is investigated. The two-step Lagrange Multiplier test for spatial nonstationarity turns out to be robust towards heteroscedasticity. While several tests for heteroscedasticity prove inconclusive under certain circumstances, it is shown that a Lagrange Multiplier test for heteroscedasticity based on spatially differenced variables serves well as an indication of heteroscedasticity irrespective of stationarity status [44]

In 2011 J. Paul Elhorst study the Spatial panel models

This paper provides a survey of the existing literature on spatial panel data models. Both static and dynamic models will be considered. The paper also demonstrates that spatial econometric models that include lags of the dependent variable and of the independent variables in both space and time provide a useful tool to quantify the magnitude of direct and indirect effects, both in the short term and in long term[45]

In 2012 Alan Karl Swanson study the spatial regression methods capture prediction Uncertainty in species distribution model projections through time

Species distribution models (SDMs) relate observed locations of a species to climate, And are used for projecting the fate of a species under climate change scenarios. To be useful in a decision-making context, the uncertainty associated with these projections must be known. However, the uncertainty associated with SDM projections is largely ignored; perhaps because many current methods have been shown to produce biased estimates.Failure to account for spatial autocorrelation of residual error explains much of this bias. Generalized linear mixed models (GLMM) have the ability to account for SAC through the inclusion of a spatially structured random intercept, interpreted to accountfor the effect of missing predictors[47]

In 2013 Emanuela Marrocu, Raffaele Paci study the Knowledge production function and proximities Evidence from spatial regression models for the European regions

This paper aims at investigating the connections among regional innovation systems along several proximity dimensions. In particular, we assess if, and how much, the creation of new ideas in a certain region is the result of internal efforts as much as of knowledge flows coming from other regions which may be considered neighbors not only in the geographical space but also in the institutional, technological, social and organizational one. The analysis, based on spatial econometric techniques, is implemented for an ample dataset referring to 276 regions in 29 European countries (EU27 plus Norway, Switzerland) for the last decade.[48]

In 2013 Cellmer.R use Spatial autocorrelation to build regression models of transaction prices

This paper presents the principles of studying global spatial autocorrelation in the land property market, as well as the possibilities of using these regularities for the construction of spatial regression models. Research work consisted primarily of testing the structure of the spatial weights matrix using different criteria and conducting diagnostic tests of two types of models: the spatial error model and the spatial lag model. The paper formulates the hypothesis that the application of spatial regression models greatly increases the accuracy of transaction price prediction while forming the basis for the creation of cartographic documents including, among others, maps of land value. [41]

In 2014 Omar Abdul Mwhsen Ali with student Sawsan Qasm Hadi studies the Spatial Regression Models Estimation for the poverty Rates In the districts of Iraq

This paper shows spatial regression model and model possessory error in an attempt to provide a general guide Shows the importance of spatial loading, with particular on the importance of using spatial regression models,which Each of which includes spatial reliability testing and that is whether or not find tests the Moran, and ignore this Reliability may lead to the loss of information important for empowerment reflected in end up on the strength of estimate Statistical index extracted, these models are the link between the usual regression models with change models[20]

In 2014 James P. Lesage study the What regional scientists need to know about spatial Econometrics

Regional scientists frequently work with regression relationships involving sample Data that is spatial in nature. For example, hedonic house-price regressions relate selling prices of houses located at points in space to characteristics of the homes as Well as neighborhood characteristics. Migration, commodity, and transportation flow models relate the size flows between origin and destination regions to the distance between origin and destination as well as characteristics of both origin and destination regions. Regional growth regressions relate growth rates of a region to past period own and nearby-region resource inputs used in production. Spatial data typically violates the assumption that each observation is independent of other observations made by ordinary regression methods. [22]

In 2014 Camille Ternynck study the Spatial regression estimation for functional data with spatial dependency

We propose a nonparametric estimator of the regression function of a scalar spatial variable Yi given a functional variable Xi. The specificity of the proposed estimator is to depend on two kernels in order to control both the distance between observations and spatial locations. Mean square consistency of this estimator is obtained when the sample considered is an a-mixing sequence. Lastly, numerical results are provided to illustrate the behavior of our estimator.[50]

In 2015 James B. Pick , Avijit Sarkar, and Jessica Rosales studies a Spatial and Regression Analysis of Social Media in the United States Counties

The locational distribution and socio-economic determinants of social media are analyzed for the United States counties in 2015. A theory of determinants is presented that is modified from the Spatially Aware Technology Utilization Model (SATUM). SocioEconomic factors including demography, economy, education, innovation, and social capital are posited to influence social media factors, while spatial analysis is conducted including exploratory analysis of geographic distribution and confirmatory screening for spatial randomness. The determinants are identified through OLS regression analysis. Findings for the nation indicate that the major determinants are demographic factors, service occupations, ethnicities, and urban location. Further subsample analysis is conducted for the U.S. metropolitan, metropolitan, and rural subsamples. The subsamples differ most evidently in effects of ethnicities and construction occupations, and there are inverse effects of social capital at the metropolitan and rural levels. The regression findings are discussed in terms of the literature mostly of larger geographic units, and the few nationwide studies at the county level. The exploratory spatial analysis generally indicates similar national geographic patterns of use. Among the results is that although Twitter users are more heavily concentrated in southern California and have strong presence in the lower Mississippi region, users are highly concentrated in Colorado, Utah and adjacent Rocky Mountain States. Social media usage is the lowest in the Great Plains, lower Midwest, and South with the exceptions of Florida and the major southern cities such as Atlanta. The overall extent of spatial agglomeration is very high and is examined in detail for the nation and subsamples. The paper concludes by discussing the policy implications of the analysis at the county as well as the national levels[51]

In 2016 Hamid Saed Nwr with master student Swhad Ali Shahid for master project studies estimate spatial dynamic panel data model (SDPD) with fixed effects-stable state using the direct approach

Although spatial dynamic panel data model sparked a lot of attention in the last decade, however, the econometric analysis of spatial models and dynamic panel data rare so far, with capabilities there are no available take into account the study the lagged of dynamic spatial model of panel data for the presence of one or more of the endogenous variables (dependent) as explanatory variables, with lagged in time along or both. The presence of the dependent variable and endogenous lagged variables in a spatial lag model invalidate the use of known estimation methods such as(OLS) and (ML)[21]

In 2016 Philomine Roseline T, N. Ganesan and Clarence J M Tauro studies A Study of Applications of Fuzzy Logic in Various Domains of Agricultural Sciences

Fuzzy logic (FL) has emerged as an important branch of Expert system which has proved to provide solution to real life problems that had remained unsolvable otherwise. It has found wide range of applications in diversified areas. In this paper, we study how the methods of fuzzy logic have been effectively used to solve a myriad of problems in the field of agricultural sciences. This paper reviews a few of the applications of fuzzy logic integrated with expert systems which had been applied in the field of agricultural sciences. This study could be considered as a part of the literature survey done for research work in future for developing expert system for a particular crop for a given region in our country. It can serve as the baseline for further work to be carried out in this domain.[65]

1-3 Objective Of This Thesis

The objectives (aims) of this study are as follows:

1. Identify the factors that affected to the atmospheric pressures (A.P) in Kurdistan Region.
2. Construct a special weight matrix W reflecting the data special arrangement.
3. Test for statistical dependence via a series of diagnostic measures.
4. Use 2 and 3 to specify and estimate a special regression model (SAR and SEM) for both (raw and fuzzy)data.
5. Compare SAR and SEM to GLM by (AIC, RMSE, MAPE, R^2_{adj}) criteria for both (raw and fuzzy) data.

1-4 Layout Of This Thesis

Chapter One: It comprises the following: introduction, literature review and the aim of the thesis.

Chapter Two: This chapter is divided into two sections:

Section One: Comprises the following linear regression model, some classical methods and alternative method for estimation, problems concerning linear model.

Section Two: Details and background of spatial regression of those models that can be used in practiced points and fuzzy logic.

Chapter Three: This chapter covers applying the data practically to traditional and modern statistical methods for estimating regression parameters for fuzzy and un fuzzy data, and comparing the result there of between models to find the best model. Then information criterion is used to search among a collection of families for the fitted model which serves the best approximation to true model.

Chapter Four: This is the last chapter of this thesis which deals with some conclusions and recommendations for future work.

Chapter Two : Theoretical Part

Section One: Classical Regression Model

Section Two: Spatial Regression Model

`Chapter Two

Section One: Classical Regression Model

2-1-1 Introduction

The term "regression" was struck by Francis Galton in the nineteenth century to describe a biological phenomenon. The phenomenon was that the heights of descendants of tall ancestors tend to regress down towards a normal average (a phenomenon also known as regression toward the mean). For Galton, regression had only this biological meaning, but his work was later extended by (Udny Yule and Karl Pearson)to a more general statistical context. In the work of Yule and Pearson, the joint distribution of the response and explanatory variables is assumed to be Gaussian. This assumption was weakened by R.A. Fisher in his works of 1922 and 1925. Fisher assumed that the conditional distribution of the response variable is Gaussian, but the joint distribution need not be. In this respect, Fisher's assumption is closer to Gauss's formulation of 1821. In the 1950s and 1960s, economists used electromechanical desk calculators to calculate regressions. Before 1970, it sometimes took up to 24 hours to receive the result from one regression[4].

Regression analysis continue to be an area of active research. In recent decades, new methods have been developed for robust regression, regression involving correlated responses such as time series and growth curves, regression in which the predictor (independent variable) or response variables are curves, images, graphs, or other complex data objects, regression methods accommodating various types of missing data, nonparametric regression,Bayesian methods for regression, regression in which the predictor variables are measured with error, regression with more predictor variables than observations, and causal inference with regression and the regression analysis entered the social sciences in the 1870s with the pioneering work by Francis Galton. But "least squares" goes back at least to the early 1800s and the German mathematician Karl Gauss, who used the technique to predict astronomical phenomena. In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed. Less commonly, the focus is on a quintile, or other location parameter of the conditional distribution of the dependent variable given the independent variables. In all cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function which can be described by a probability distribution[11].

Regression shows us how variation in one variable occurs with variation in another. What regression cannot show is causation; causation is only demonstrated analytically, through substantive theory. For example, a regression with clothes size as an independent variable and people size as a dependent variable would show a very high regression coefficient and highly significant parameter estimates, but we should not conclude that higher clothes size causes higher people size. All that the mathematics can tell us is whether or not they are correlated, and if so, by how much. It is important to recognize that regression analysis is fundamentally different from ascertaining the correlations among different variables .As deferent between regression and correlation the correlation determines the strength of the relationship between variables, while regression attempts to describe that relationship between these variables in more detail. [52]

Linear regression attempts to model the relationship between variables by fitting a linear equation to observed data. One variable is considered to be explanatory (independent) variable, and the other is considered to be response (dependent) variable.

One of the most popular linear models is regression analysis. In simple regression analysis one assumes a relation of the type:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i, \qquad i = 1,2,\ldots,n. \qquad (2.1)$$

where

β_0 is the intercept, gives the value of Y when $X = 0$ and $\varepsilon_i = 0$.

β_1 is the slope, relates a change in Y to change in X (holding ε_i constant)

ε_i referred to as the error term or disturbance.

In simple regression, the observations are 2-D so they can be plotted. It is necessary to do the plotting first to see if any unusual features are present and to make sure that the data are roughly linear[4]

In general, the multiple regression models have the following general formulation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i, \quad i = 1,2,\ldots,n \qquad (2.2)$$

where $\beta_0, \beta_1, \ldots, \beta_k$ unknown parameters and the explanatory variables are $X_{11}, X_{21}, \ldots, X_{kn}$ which are fixed.

We can thus write the regression model in the matrices form as:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

Where

$\underline{Y} = (Y_1, Y_2, \ldots, Y_n)'$ is an $(n \times 1)$ vector of observation .

X is an $(n \times k)$ matrix of explanatory variable.

$\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ is an $(k \times 1)$ vector of unknown parameters.

$\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is an $(n \times 1)$ vector of random errors.

Dealing with several explanatory variables simultaneously in a regression analysis is considerably more difficult than dealing with a single explanatory variable, for the following reasons:[11]

1- It is more difficult to choose the best model, since several reasonable candidates may exist.

2- It is more difficult to visualize what the fitted model looks like (especially if there are more than two explanatory variables) since it is not possible to plot either the data or the fitted model directly in more than three dimensions.

3- It is sometimes more difficult to interpret what the best fitting model means in real life terms.

4- Computations are virtually impossible without access to a high speed computer and a reliable packaged computer program.

2-1-2 Review of Multiple Regressions

2-1-2-1 Multiple Regression Model Assumptions

These assumptions are summarized as follows: [11][9]

1. $\mu\{Y \mid X_1 = x_1, \ldots, X_k = x_k\} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$ (linearity)

2. $SD\{Y \mid X_1 = x_1, \ldots, X_k = x_k\} = \sigma$ (constant variance)

3. Distribution of Y for each subpopulation $X_1 = x_1, \ldots, X_k = x_k$ is normally distributed normality with mean(0) and variance (σ^2)

4. Observations are independent.

2-1-2-2Test for Multiple Regressions: [11]

1-Test Hypothesis :

$\mu\{Y \mid X_1 = x_1, \ldots, X_k = x_k\} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$.

Test of null hypothesis $H_0 : \beta_1 = 0$

vs. alternative hypothesis H_1: at least one of the $\beta_1, \ldots, \beta_k$ does not equal to zero

Uses t-statistic:$/t/ = /\dfrac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}/$, reject for greater than |t|. Interpretation of test: "Is there evidence that X_1 is a useful predictor (improves predictions) once $X_2, \ldots, X_k$ have been taken into account (held fixed)? or is X_1 associated with $X_2, \ldots, X_k$ once X_1 has been taken into account?

2-Overall usefulness of predictors: For multiple regression models(MRM) $\mu\{Y \mid X_1 = x_1, \ldots, X_k = x_k\} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, test whether any of the explanatory variables (predictors) are useful.

Null hypothesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

vs. alternative hypothesis H_a : at least one of $\beta_1,\ldots,\beta_k$ does not equal to zero.

Test (called overall F test) is carried out using Analysis of Variance table or (ANOVA) table.

$$F = \frac{MSR}{MSE}$$

where MSR is the (mean square regression) and MSE is the (mean square error). If the null hypothesis, H_0, is true then the statistic F_0 follows the F distribution with K degrees of freedom in the numerator and n-(K+1) degrees of freedom in the denominator. The nulhypothesis, H_0, is rejected if the calculated statistic, F_0, is such that[53]

$$F_0 > F_{\alpha,k,n-(k+1)}$$

When we satisfy the assumptions, it means that we have used all of the information available from the patterns in the data. When we violate an assumption, it usually means that there is a pattern to the data that we have not included in our model, and we could actually find a model that fits the data better[54]

3- Coefficient of determination R^2statistic :

R^2 is a measure of how good the predictions from the multiple regression model are compared to using the sample mean of Y, $\bar{Y}$ (i.e., use none of the predictors) to predict Y. Similar interpretation to simple linear regression, R-squared statistic is the proportion of the variation in Y explained by the multiple regression models[7] [69]

2-1-2-3 Diagnostics and Model Building

The diagnostics for checking assumptions and remedies for violations of assumptions are summarized as follows:

1. Test for linearity: Residual plots versus predicted values and versus explanatory variables $x_1,\dots,x_k$. If the model is correct, there should be no pattern in these plots. A pattern in the mean of the residuals indicates a violation of linearity.

2. Test for constant variance: Residual plots versus predicted values and versus explanatory variables $x_1,\dots,x_k$ .A pattern in the spread of the residuals indicates non constant variance.

3. Test for normality: Make plot or histogram of the residuals and see if it is approximately bell shaped (Jarque-Bera, Shapiro-Wilk W, Anderson-Darling, Martinez-Iglewicz, Kolmogorov-Smirnov) [11].

2-1-3 Some Classical Parameter Estimation Method

2-1-3-1 Ordinary Least Squares (OLS)

The parameters $(\beta_0, \beta_1,...,\beta_k)$ are estimated by minimizing the sum squares of the residuals: [56]

$$S(\beta) = \varepsilon'\varepsilon = \sum_{i=1}^{n} \varepsilon_i^2 = (Y - X\beta)'(Y - X\beta)$$

$$= Y'Y - Y'X\beta - \beta' X'Y + \beta' X'X\beta$$

$$= Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

The least square estimator of β must satisfy

$$\frac{\partial S}{\partial \beta} = -2X'(Y - X\beta) = 0$$

$$-2X'Y + 2X'X\beta = 0 .....(2.3)$$

Divide equation(2.3) by (2) we get

$$-X'Y + X'X\beta = 0 \qquad \text{.......(2.4)}$$

Divide equation (2.4) by($X'X$) we get

$$(\qquad \frac{X'X\beta}{X'X} = \frac{X'Y}{X'X}$$

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \text{....(2.5)}$$

Where Y and X take the following column vector and matrix

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ . & . & . & \cdots & \\ . & . & . & \cdots & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}_{n\times(k+1)}$$

$$\underline{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ . \\ . \\ \hat{\beta}_k \end{pmatrix}$$

One of the problems facing OLS is that it has a breakdown point which is a criteria showing a maximum degree of tolerance against contamination data in any sample, only 1/n ; that is a single point properly placed, can cause the OLS estimator to take virtually any value.

2-1-3-2 Weighed Least Square (WLS)

The weighed least squares method of analysis is a modification of standard regression analysis procedures that is used when a regression model is to be fit to a set of data for which the assumptions of variance homogeneity do not hold.

Weighed least squares analysis can be used when the variance of Y varies for different value of the independent variable, provided that these variances (i.e. σ_i^2 for the ith observation on Y) are known or can be assumed to be of the form $\sigma_i^2 = \sigma^2 / W_i$,where weighs $\{w_i\}$ are known.

The specific weighed least squares solution for the straight line regression case i.e.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

is given by the formulas

$$\hat{\beta}_1^* = \frac{\sum_{i=1}^{n} W_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} W_i (X_i - \overline{X})^2}$$

and

$$\hat{\beta}_0^* = \overline{Y}^* - \hat{\beta}_1^* \overline{X}^*$$

$$\overline{Y}^* = \frac{\sum_{i=1}^{n} W_i Y_i}{\sum_{i=1}^{n} W_i} \qquad \text{and} \qquad \overline{X}^* = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i}$$

Where W_i = weight assigned to the i_{th} observation

This procedure is known as weighted least squares and the resulting parameter estimates are called weighted least squares estimates. In ordinary least squares procedure a weight of $W_i = 1$ is assigned to each[56]

2-1-3-3 Maximum Likelihood Estimation (MLE)

The term maximum likelihood refers to a very general algorithm for obtaining estimators of population parameters; such estimators have excellent (large-sample) statistical properties. One major advantage of the ML method of estimating parameters is its applicability to a wide variety of situations. In particular, when a multiple linear regression model is fitted to normally distributed data, the least squares estimators of the regression coefficients are identical to the ML estimators.

To illustrate, suppose that we make the multiple regression model.

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}, \quad i = 1,2,\ldots,n, \quad \ldots\ldots\ldots(2.6)$$

Specifically, Let us assume that:

$$\underline{Y} \sim \mathrm{N}(X\underline{B}, \sigma^2) \qquad \mathrm{i} = 1,\ldots\ldots,\mathrm{n}$$

Employing the expression for the distribution (density function) of a normally distributed random variable, from (2.6) find that the distribution of Y_i is

$$f(\underline{Y}\ ; X\underline{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\underline{Y} - X\underline{\beta})^2\right\} \ \ldots\ (2.7)$$

The likelihood function from (2.7) is

$$L(\underline{Y}\ ; X\underline{\beta}, \sigma^2) = \prod_{i=1}^{n} f(\mathrm{Y}_i; X\beta, \sigma^2)$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})\right\} \qquad \ldots(2.8)$$

Taking the natural logarithm of this expression gives:

$$\ln L(Y\ ; X\beta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\underline{Y} - X\,\underline{\beta})'(\underline{Y} - X\,\underline{\beta})$$

By solving simultaneously the two ML equations

$$\frac{d}{d\beta}\left[\ln L(Y; X\beta, \sigma^2)\right] = 0 \qquad \text{and} \qquad \frac{d}{d\sigma^2}\left[\ln L(Y\ ; X\beta, \sigma^2)\right] = 0$$

Then it is clear to obtain from above that:-

$$\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{Y} \quad \ldots\ldots (2.9)$$

and $$\hat{\sigma}^2 = \frac{e'e}{n} \quad \ldots\ldots (2.10)$$

A specific algebraic expression for the maximized likelihood $L(Y\ ; X\hat{\beta}, \hat{\sigma}^2)$ can be specified by substituting equation (2.9) and (2.10) into equation (2.8) and then simplifying; the resulting maximized likelihood function can be written in the form [11]

$$L(Y\ ; X\hat{\beta}, \hat{\sigma}^2) = (2\pi\,\hat{\sigma}^2 e)^{-n/2} \quad \ldots\ldots\ (2.11)$$

2-1-4 Problems Concerning Linear Model

The accuracy of an estimated parameter basically depends on certain assumptions. If some of these assumptions are not met then the estimation of the parameters in the model will be unreasonable and will lead to inaccurate results and then to a bad model. The most important assumptions arenas mentioned in the ensuing subsections.

2-1-4-1 Hetroscedasticity Problem:

Hetroscedasticty usually occurs in most statistical studies especially in those that depend on sectional data. The dispersion of observations of such a data relating to response variable may differ significantly among explanatory variables which result in the hetroscedasticity of the variance in the form:

$$\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq .... \neq \sigma_n^2$$

To detect this problem test was used to test the following hypothesis:

$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$ (Homoscedasticity)

vs. H_a : at least one of σ_i^2 does not equal zero. (Hetroscedasticity)

To detection this problem several test can be use like (Breusch-pagan and Koenker-Basset , Goldfeld–Quandt , Park test, White test)

From the tests, the bruch-pagan is used to detect the problem[33][70]

Breusch-Pagan Test

In statistics, the Breusch–Pagan test, developed in 1979 by Trevor Breusch and Adrian Pagan is used to test for heteroskedasticity in a linear regression model. It was independently suggested with some extension by R. Dennis Cook and Sanford Weisberg. It tests whether the estimated variance of their from a regression are dependent on the values of the independent variables. In that case, heteroskedasticity is present

Steps to Find Breusch-Pagan and Koenker-Basset Test

Define the matrix **z** to be composed of the values of the variables listed in the Breusch= option, such that $z_{i,j}$ is the value of the j_{th} variable in the Breusch = option for the $\boldsymbol{i}_{th}$ observation. The null hypothesis of the Breusch-Pagan test is

$\sigma_i^2 = \sigma^2 (a_0 + a' z_i)$ H_0:a=0

where σ_i^2 is the error variance for the $\boldsymbol{i}_{th}$ observation and α_0 and a' are regression coefficients.

The test statistic for the Breusch-Pagan test is

$$\text{B.P} = \frac{1}{v}(u - \bar{u}_i)' Z (Z'Z)^{-1} Z' (u - \bar{u}_i) \ldots\ldots (2.12)$$

Where $u=(e_1^2, e_2^2, e_3^2 \ldots e_n^2)$, $\imath$ is a $n\times 1$ vector of ones, and

$$v= \frac{1}{n}\sum_{i=1}^{n}(e_i^2 - \frac{e\prime e}{n})^2$$

This is a modified version of the Breusch-Pagan test, which is less sensitive to the assumption of normality than the original test [57]

2-1-4-2 Autocorrelation (Serial Correlation) Problem:

The second problem arises as a result of violating one of the linear regression model assumptions, which is related to the behavior of the disturbance term (ε_i); where

$$\text{Cov}(e_i, e_j) = 0 \qquad \text{if } i \neq j$$

Autocorrelation is usually found among time series data rather than in grouped data when an error term in a period is related with another term before or after it. To detect this problem Durbin-Watson test was used.

Durbin-Watson Test

$$\text{D.W}= \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}\ldots\ldots(2.13)$$

Assumption for this test

$H_0 : \rho = 0$

H_1 : ρ does not equal to zero

where $e_i = y_i - \hat{y}_i$ and yi and $\hat{y}_i$ are, respectively, the observed and predicted values of the response variable for individual i.d becomes smaller as the serial correlations increase. Upper and lower critical values, dU and dL have been tabulated for different values of k (the number of explanatory variables) and n

And we can say that we have don't have this problem if

If d < dL reject $H_0 : \rho = 0$

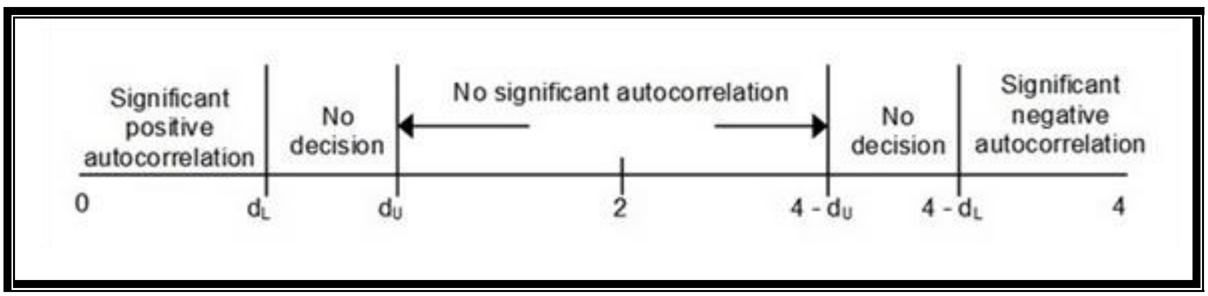
If d > dU do not reject H_1 : ρ does not equal to zero

If dL < d < dU test is inconclusive.

Or

$$D.L \leq D.W \leq 4-D.L$$

To solve this problem use Generalized Least Square, Cochrane-Orcutt Method or Iterative Method. [58]



Figure(2-1) Show durbin watson test

2-1-4-3 Multicollinearity Problem:

Often two or more of the explanatory variables used in the linear regression model produce redundant information. That is, the explanatory variables will be correlated with each other. In practice, it is not uncommon to observe correlations among the explanatory variables, since a few problems arise when serious multicollinearity is present in the regression analysis.

Variance Inflation Factor (VIF)

As the name suggests, a variance inflation factor (VIF) quantifies how much the variance is inflated. But what variance? Recall that we learned previously that the standard errors and hence the variances of the estimated coefficients are inflated when multicollinearity exists. So, the variance inflation factor for the estimated coefficient b_k denoted VIF_k is just the factor by which the variance is inflated

Let's be a little more concrete. For the model in which x_k is the only predictor:

$$y_i=\beta_0+\beta_k x_{ik}+e_i$$

it can be shown that the variance of the estimated coefficient b_k is

$$\mathrm{Var}(b_k)=\frac{\sigma^2}{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}\times\frac{1}{1-R_k^2}$$

Note that we add the subscript "min" in order to denote that it is the smallest the variance can be. Don't worry about how this variance is derived we just need to keep track of this baseline variance, so we can see how much the variance of b_k is inflated when we add correlated predictors to our regression model.

$$\mathrm{Var}(b_k)\mathrm{min}=\frac{\sigma^2}{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}$$

$$\mathrm{VIF}=\frac{\mathrm{Var(bk)}}{\mathrm{Var}(b_k)\mathrm{min}}$$

$$=\frac{1}{1-R_k^2}$$

while VIFs exceeding 4 are signs of serious multicollinearity requiring correction.[59]

2-1-5 Test for Normality

- **Jarque Bera Test**

The calculation of p-values for hypothesis testing typically is based on the assumption that the population distribution is normal. Therefore, a test of the normality assumption may be useful to inspect. A variety of tests of normality have been developed by various statisticians. One of these tests will be described here. To start, the calculation of descriptive statistics is reviewed. A data set has the numeric observations: x_1 , x_2 , ... , x_n. Familiar descriptive statistics are the sample mean:

H_0:The distribution of standard residual is very close to standard normal

H_1:The stand and residual are significantly different from the standard normal

$$K = \frac{1}{n} \times \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{(\hat{\sigma}^2)^2} \text{……(2.14)}$$

$$EK=K-3$$

$$JB=n\times [\frac{S^2}{6} + \frac{(EK)^2}{24}] \text{…..(2.15)}$$

It turns out that this test statistic can be compared with a χ^2 (chi-square) distribution with 2 degrees of freedom. The null hypothesis of normality is rejected if the calculated test statistic exceeds a critical value from the χ^2 distribution[60]

Chapter Two

Section Two: Spatial Regression Model

2-2-1 Introduction

Spatial data refers to all types of data objects or elements that are present in a geographical space or horizon. It enables the global finding and locating of individuals or devices anywhere in the world.

Spatial data is also known as geospatial data, spatial information or geographic information Spatial data consists of spatial objects made up of points, lines, regions, rectangles, surfaces, volumes and even data of higher dimension which includes time. Examples of spatial data include cities, rivers, roads, counties, states, crop cover ages, mountain ranges, parts in a CAD system, etc. Examples of spatial properties include the extent of a given river, or the boundary of a given county, etc. Often it is also desirable to attach non-spatial attribute information such as elevation heights, city names, etc. to the spatial data. Spatial databases facilitate the storage and efficient processing of spatial and non-spatial information ideally without favoring one over the other. Such databases are finding increasing use in applications in environmental monitoring, space, urban planning, resource management, and geographic information systems (GIS) (Buchmann et al. 1990, unther and Schek 1991). A common way to deal with spatial data is to store it explicitly by parametrizing it and thereby obtaining a reduction to a point in a possibly higher dimensional space. This is usually quite easy to do in a conventional database management system since the system is just a collection of records, where each record has many fields. In particular, we simply add a field (or several fields) to the record that deals with the desired item of spatial information. This approach is fine if we just want to perform a simple retrieval of the data.

Generally speaking, spatial data represents the location, size and shape of an object on planet Earth such as a building, lake, mountain or township. Spatial data may also include attributes that provide more information about the entity that is being represented. Geographic Information Systems (GIS) or other specialized software applications can be used to access, visualize, manipulate and analyze geospatial data[61].

2-2-2 Spatial Regression Models

The spatial regression model worked as time Series but in it lag refers to any place or far between places and not to time as time series and in it we can show depend of explanatory variable for depend variable where a place is neighbor or not and the models depend on weight matrix and we can use the spatial regression models if we have spatial dependency[68]

Spatial Cross-Sectional Models

Our starting point is the linear-in-parameters cross-sectional model

$$y = X\beta + u$$

with the error term u classical[68]

2-2-2-1 Spatial Autoregressive Model (SAR) or (SLM)

Spatial autoregressive (SAR) model captures as well substantial spatial dependencies like external effects or spatial interactions. It assumes that such dependencies manifests in the spatial lag WY of the dependent variable Y. Regional growth may be fostered by growth in neighborhood regions by flows of goods for example. In this case, spillover effects are not restricted to adjacent regions but propagated over the entire regional system. In accordance to the time-series analogue the pure spatial autoregressive model is also termed SPATIAL LAG model. In applications the model also incorporates a set of explanatory variables X_1, $X_{2,}$ …, $X_{k.}$ This extension is expressed by the term mixed regressive, spatial autoregressive model. In all instances OLS estimation will produce biased and of inconsistent parameter estimates. We will introduce the method of the maximum likelihood (ML method) as adequate estimation methods for that type of model. Because only the spatial lag Wy is relevant for the choice of an alternative estimation method to OLS, the term spatial lag model is often kept in cases where the model is extended by exogenous X-variables [7]

$$\underline{Y} = \lambda W\underline{Y} + X\underline{\beta} + e \ldots..(2.16)$$

Where

$e \sim N(0, \sigma^2 In)$

$\underline{Y}$: is a vector (n×1) for the observation depending variable.

W: is the spatial weights (n×n) matrix

λ:parameter of spatial auto regressive regression model

X: matrix (n× (k+1)) for the observation explanatory variables

$\underline{\beta}$: is a vector ((K+1) ×1) Parameter required estimation

$\underline{e}$: is the vector (n×1) for error term

$\underline{In}$: is identity (n×n) matrix

Value of parameter of spatial regression model $-1 < \lambda < 1$ if equal to 0 no have spatial correlation or (spatial dependency) and we go to classical regression model and if $\lambda>0$ we have positive spatial correlation and we can say the neighbor place is same with the other but if $\lambda<0$ we have negative spatial correlation and the neighbor is not same with other[14]

2-2-2-1-1 Maximum Likelihood Estimation for (SAR) Model [17][16][14]

Maximum likelihood estimation one of the most important ways because it gives the best estimate for parameter from among several possible estimates, when the time series data, in which the depend variable is different period of time, for example (Y_{t-1}). Where it does not cause this variable "time" for any problems for OLS if there is not any time Series correlation for the residuals in the regression model. Is facing the use estimate of least squares to estimate the spatial model the problem show where ε and WY is not independent from other , so use (MLE) for spatial regression models first by (Ord 1975)

$$\underline{Y} = X\beta + \lambda W\underline{Y} + e \ldots\ldots(2.17)$$

$$\underline{Y}=(I-\lambda W)^{-1}X\beta+(I-\lambda W)^{-1} e$$

$$e \sim N(0, \sigma^2)$$

$$e=(I-\lambda W)(\underline{Y}) - X\beta$$

Thus, the log likelihood function for y of the spatial lag model is obtained by adding the term ln |I-λW| to the log likelihood function of the standard regres sion model

$$\text{Ln } L(\beta, \lambda, \sigma^2/\underline{Y},X)= \frac{-n}{2}\text{Ln } 2\pi - \frac{n}{2}\text{Ln } \sigma^2 + \text{Ln}|(I-\lambda W)| - \frac{1}{2\sigma^2} e'e \ldots\ldots(2.18)$$

Where

$$e=\underline{Y} - \lambda W\underline{Y} - X\beta \ldots\ldots\ldots(2.19)$$

Putting equation (2.19) in equation (2.18) we get get

$$\text{Ln } L(\beta, \lambda, \sigma^2/\underline{Y},X) = \frac{-n}{2}\text{Ln } 2\pi - \frac{n}{2}\text{Ln } \sigma^2 + \text{Ln}|(I-\lambda W)| - \frac{1}{2\sigma^2}(\underline{Y} - \lambda W\underline{Y} - X\beta)'(\underline{Y} - \lambda W\underline{Y} - X\beta)$$

On account of this correction the MLE estimates will differ from the OLS estimates. They coincide for λ=0 where the spatial lag model approaches the standard regression model.

In this step we get the derivative for β, σ^2 in log of likelihood and equal to zero we get

$$b_{MLE} = (X'X)^{-1} X'A \underline{Y} \ldots\ldots(2.20)$$

Let

$$A=(I-\lambda W)$$

$$= (X'X)^{-1} X'(I-\lambda W)\underline{Y}$$

$$= (X'X)^{-1} X'\underline{Y} - \lambda(X'X)^{-1} X'W\underline{Y}$$

$$b_O = (X'X)^{-1} X'\underline{Y} \ldots\ldots(2.21)$$

$$b_L = (X'X)^{-1} X'W\underline{Y} \ldots\ldots\ldots(2.22)$$

$$b_{MLE} = b_O - \lambda\, b_L \ldots(2.23)$$

$$e_{MLE} = \underline{Y} - \lambda W\underline{Y} - Xb_{MLE}$$

$$e_{MLE} = \underline{Y} - \lambda W\underline{Y} - X(b_O - \lambda\, b_L) \ldots\ldots(2.24)$$

$$= \underline{Y} - Xb_O - \lambda(W\underline{Y} - Xb_L)$$

$$\underline{e}_O = \underline{Y} - X\, b_O \ldots(2.25)$$

$$\underline{e}_L = W\underline{Y} - X\, b_L \ldots\ldots(2.26)$$

$$\underline{e} = \underline{e}_O - \lambda\, \underline{e}_L \ldots(2.27)$$

according the first order condetion the error variance can be esemaion by

$$\hat{\sigma}^2_{MLE} = \frac{(e_o - \lambda e_L)'(e_o - \lambda e_L)}{n} \ldots\ldots(2.28)$$

$\underline{b_o}$: vector of the regression parameter $\underline{Y}$ for X

$\underline{b_L}$: vector of regression parameter $W\underline{Y}$ for X

λ:parameter of spatial regression model

$\underline{e_o}$: vector of another regression model $\underline{Y}$ for X

$\underline{e_L}$: vector of another regression model $W\underline{Y}$ for X

by putting the value of($\hat{\beta}_{MLE}$ $\hat{\sigma}^2_{MLE}$) in to the likelihood function(2.23) we get

$LC = -n/2 \, Ln \left[\frac{1}{n} (e_o - \lambda e_L)'(e_o - \lambda e_L)\right] + \ln |I - \lambda W| \ldots\ldots(2.29)$

2-2-2-2 Spatial Error Model (SEM): [63][64]

One of the most important violations that plague regression model is the independence of the error term , so it will be studied with this model . It is assumed that the error or (model errors are linked spatially) reversed the presumption of independence of errors. In the traditional model and aims of this model spatial error model (SEM) to spatial error correction .

$$\underline{Y} = X\underline{\beta} + \underline{u} \ldots..(2.30)$$

$$\underline{u} = (I - \theta W)^{-1} \underline{e}$$

$$\underline{e} \sim N(0,\sigma^2 I_n)$$

or we can write

$$\underline{Y} = X\underline{\beta} + (I - \theta W)^{-1} \underline{e}$$

Where

$\underline{Y}$: is a vector (n ×1) for the views depend variable.

W: is the spatial weights (n×n) matrix

X: matrix (n× (k+1)) the observation of explanatory variables

$\underline{\beta}$: vector ((K+1) ×1) Parameter required estimation

I_n: identity (n×n) matrix

θ :is the spatial parameter

In: is identity (n×n) matrix

$\underline{u}$: is a vector of (n×1) error term which spatial correlated

$\underline{e}$: is a vector of (n×1) random error term

If $\theta \neq 0$ we have a spatial dependency between the errors for neighbor observation but if θ=0 we go to the classic regression model and not have spatial dependency between errors for neighbor observation

2-2-2-2-1 Maximum Likelihood Estimation for (SEM) Model:

In spatial autoregressive model attention to (λ) it is parameter of autoregressive spatial regression model which reflects value of spatial effect from the nature of the spatial correlation between the depend variable values, and in the spatial error model attention to θ where show the correlation between the residuals :[12][13]

Then the Maximum likelihood estimation for (SEM) model is :

For this model

$$\underline{Y} = X\underline{\beta} + \underline{u} \quad \ldots(2.31)$$

$$\underline{u} = \theta W\underline{u} + \underline{e}$$

or

$$\underline{Y} = X\underline{\beta} + \theta W\underline{u} + \underline{e}$$

And put the value of error in a likelihood function

$$L(\beta,\ \theta,\ \sigma^2/Y,X)=\frac{-n}{2}\text{Ln}2\pi - \frac{n}{2}\text{Ln}\ \sigma^2 + \text{Ln}\ /I-\ \theta W\ /\ - \frac{1}{2\sigma^2}[(\underline{Y}-X\beta)'(I-\theta W)'(I-\theta W)(\underline{Y}-X\beta)] \ldots..(2.32)$$

In this step we get the derivative for β, σ^2 in log of likelihood and equal to zero we get

$$Y^*= (I - \theta W)(\underline{Y}) \text{ and } X^*=(I - \theta W)\ (X)$$

$$b_{MLE}= (X'^{*}X^{*})^{-1}\ X'^{*}Y^{*} \ldots.(2.33)$$

put the value for X* an d Y* we get

$$b_{MLE} = [\ (X'\ (I - \theta W)'\ (I - \theta W)\ X\ ]^{-1}\ \ X'\ (I - \theta W)'\ (I - \theta W)\ \underline{Y}\ \ ]\ldots..(2.34)$$

$$e =\underline{Y}-Xb_{MLE}$$

$$\sigma^{\wedge 2}{}_{MLE} = \frac{e'\,e}{n}\ldots..(2.35)$$

after putting the value of ($\sigma^{\wedge 2}{}_{MLE,}$ $\beta_{\ MLE}$) in likelihood function we get

$$Lc= -n/2\times \text{Ln}[\frac{1}{n}e'\ (I - \theta W)'\ (I - \theta W)e]+\ln/\ I - \theta W/\ \ \ldots\ldots(2.36)$$

2-2-3 The General Spatial Model

It is possible to combine the SAR and SEM models:

$$y = \lambda\ W1y + u$$

$$u = \theta W2u + v$$

where v is classical and both W1 and W2 are spatial weights matrices. (The description of this as The General Spatial Model is due to LeSage). One motivation for this is as follows: suppose we have estimated a SAR model. We then test the residuals for spatial autocorrelation using (say) Moran's test. If we cannot reject the hypothesis that the residuals are (still) spatially autocorrelated, then this model, which allows for both sources, may be appropriate [68]

2-2-4 Computational Considerations

It is important to understand that there are real computational difficulties in the case where data represents many regions. For example, suppose we have data by US counties, of which there are about 3000 .=(3×10^3). Then W is a matrix of about 9×10^6 (9 million) elements. Clearly it is going to be very difficult to work with arrays this size. However, it turns out that for this particular case (the US counties), and with a rook definition of contiguity, only about 12,500 elements are non-zero (LeSage and Pace (2009)). (A county in Ohio will not be a neighbor of a county in California, or indeed of counties in most other states). So if we could just keep track of which elements are not zero, we could save considerable space. That is what sparse-matrix representations do. Instead of keeping track of the entire

matrix, they record, for each non-zero element only, its row, its column, and its value. The result, for US counties, is that we need to keep track of only about 37,500 (= 12,500× 3) numbers, rather than 9 million. Of course, for all this to be practically useful, we need routines to work with sparse matrix representations, to be able to multiply or invert them, without having to expand them into their full (“dense”) forms.

A related problem is that even if W can be represented as a sparse matrix, (I - λW) is not sparse and is also $R \times R$. However, note from the likelihood functions, that all we really need is the log determinant of(I- λ W) . and not .(I - λW) itself. For small R one can compute the log determinant directly. For larger R one approach is to note that

$$\mathrm{Ln}|I-\lambda W| = \ln\left(\prod_{i=1}^{R}(1-\lambda\vartheta i)\right)$$

where the ϑi are the eigen values of W which are specific to W and do not change as estimates of λ change. The advantage of this is that the eigen values need to be computed only once. The disadvantage is that the computation can be time-consuming,even once: for example, Kelejian and Prucha (1999, footnote 12) report for a case of R = 1500 and with the average number of spatial neighbors (non-zero entries in a row) equal to 10, computing the eigen values took 22 minutes; however, this was on what is now considered very old and slow hardware. This slowness motivated a search for a method to compute the log determinant that did not need to compute the eigen values. Nowadays, with improved hardware and more RAM, this is less of a problem. Still, there is room for improvement: a popular approach is a Monte
Carlo method due to Barry and Pace (1999).[68]

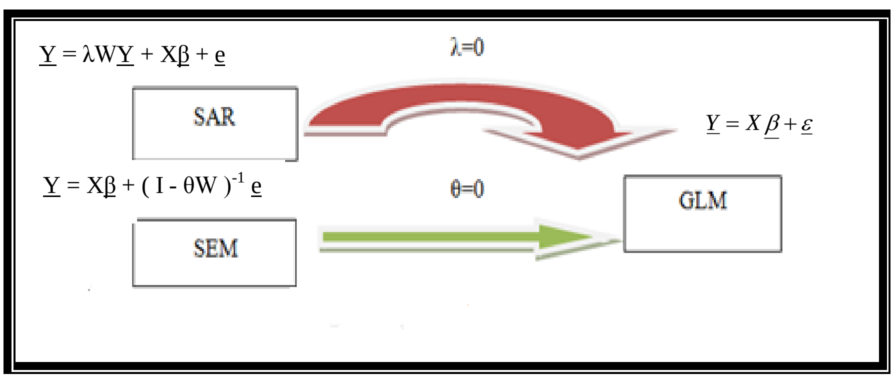


Figure (2-2) Show the relationship between the models

relation between spatial models where depend on spatial parameters (λ, θ) wher λ=0 then not have (SAR) model and where $\theta = 0$ not have (SEM) model and we go the classical regression (GLM) [30] .

2-2-5 Weight Matrix:

It is square matrix which it elements positive value and denoted by W and not necessary to be symmetric and create this matrix based on neighboring and relation neighboring from location for another location in same row in the rows of matrix and value for the diagonal usually equal to zero and chose weight matrix is very important for Determine the spatial effects so we must create a Appropriate weight matrix and there for some way to create this matrix[62].

2-2-5-1 Binary Contiguity Weights Matrix: [63]

The matrix is positive and square ($n \times n$) ,if i,j contiguous $w_{ij}=$ 1and if not contiguous $w_{ij}=0$

$$W_{ij} \begin{cases} 1 & \text{if } i \text{ neighbour } j \\ 0 & \text{otherwise} \end{cases}$$

2-2-5-2 Methods to Create Weight Matrix

1. **Rook Contiguity:**

The value of the element that take one value if the two areas neighbor by limited and have relation between the two area in any side and other is zero. This matrix is used so much than other

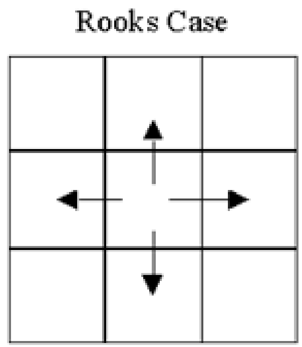


Figure (2-3) Show the rook weight matrix

2.Bishop Contiguity

Neighbor would occur if the two areas connected a point and this point is the connected limited between the two location is smallest connected limited and be the elements value is equal to the one and another element is equal to zero .

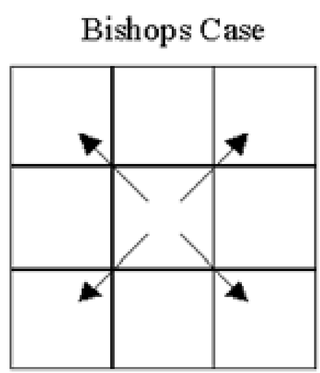


Figure (2-4) Show the bishop weight matrix

3.Queen Contiguity

This matrix get its elements from the sum of (rook) and (bishopes) matrix elements and neighbor in this matrix is based on connect point or connect limited. [64]

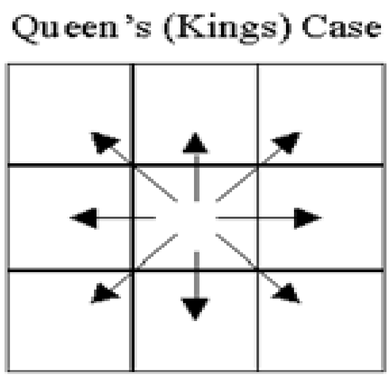


Figure (2-5) Show the queen weight matrix

For find weight matrix Depend on figure method example

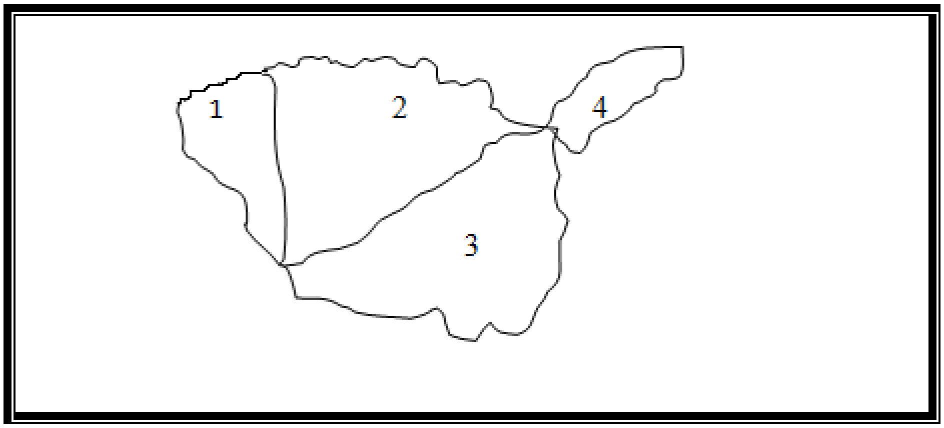


Figure (2-6) Show the example of the weighted matrix

1.Rook Contiguity

$$W_R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2.Bishop Contiguity

$$W_B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

3.Queen Contiguity

$$W_Q = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

2-2-5-3 Row - Standardized Weights Matrix:

In this matrix sum of row equal to 1 and this matrix depend on (Binary Contiguity Weights Matrix) [64]

$$W_{ij}^{std} = \left\{ \begin{array}{ll} \frac{W_{ij}}{\sum W_{ij}} \quad i \text{ neighbor } j & 0 < W_{ij}^{std} \leq 1 \\ 0 & \text{other wise} \end{array} \right\}$$

2-2-6 Tests for Spatial Regression

There are several tests to detect spatial dependence and between these tests

2-2-6-1 Moran's Test:

This test is measure to show that we have spatial dependemcy in data or not and it is a general measure and depends on the(GLM) model [$Y = X\beta + \varepsilon$] it is (called Moran coefficient)because Moran is the name of the Scientist that find the test.For measuring the similarity in neighboring phenomena , the idea depend on first Geographic role (Tobler) in 1970 Which indicates that "the near things " more of a relationship of " long things " any phenomenon related to each other phenomenon , but the phenomena are more convergent than divergent relationship phenomena Overall spatial autocorrelation coefficient measures at the same time the extent of the similarity between the spatial elements and Describe Distinction if value of Moran coefficient near 1 that is mean Existing Spatial autocorrelation coefficient.[1][3][70]

$$I_M = \frac{n\,(e'we)}{S_o\,(e'e)}$$

$S_o = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$:sum of every element in W matrix

W: weight matrix square ($n\times n$) matrix

n: sample size

e: vector Residual dimensions($n\times 1$)

where we using row – standardized where sum of row equal to 1 in this case ($n = S_o$) that is work to simplify the above formula for the follows

$$I_M = \frac{e'we}{e'e} \quad \ldots\ldots(2.37)$$

For now that value of Moran coefficient where it is Indication of Statistics in certain Degree of confidence we must use Moran (z) test As shown below

Hypotheses test for Moran's

Null hypotheses H_0: $\lambda = 0$, $\theta=0$ no have spatial dependency

Alternative hypotheses H_1: at least one of $\lambda \neq 0$ or $\theta \neq 0$ spatial dependency is exist

$$Z = \frac{I - E(I)}{\sqrt{Var(I)}} \;.....(2.38)$$

$$E(I) = \frac{(tr\,(MW))}{(n-k)} ......(2.39)$$

$$Var(I) = \frac{tr(MWMW') + tr(MWMW) + (tr(MW))^2}{(n-k)\,(n-k+2)} - (E(I))^2 .....(2.40)$$

$M = I - X\,(X'X)^{-1}\,X'$: it is Idempotent Matrix that is square and symmetric

tr: Sum diagonal element

k: Number of explanatory variables

And now comparison value of calculate Z with value of Z table for(α, two sided test) where Moran test is significant that is mean relation between geographic location that refers to use spatial regression model and not enough general linear model(GLM)and we have spatial dependency[10]

2-2-6-2 Lagrange Multiplier(LM) Lag Test for (SAR) Model:

Test of Lagrange multiplier test is more use than Moran test because Moran is use only to test spatial dependency does Exist or not and we cannot now by Moran test what is the alternative model for (GLM) model but

Lagrange is tell we which is the model is alternative (SAR) or (SEM) and assumption for this test is

Hypotheses test for Lagrange (SAR) model:

H_0: $\lambda = 0$ spatial dependence exist

H_1: λ at least one of the λ does not equal to 0 spatial dependence not exist

Where reject null hypotheses is mean spatial dependence not exist and accept alternative hypnoses where alternative hypnoses is mean exist spatial dependence mean model alternative is (SAR) model where it role show in down and some time denoted by (LM-SAR) or LM_λ

$$LM_\lambda = \frac{\left(\frac{e' W Y}{s^2}\right)^2}{D} \quad \text{.....(2.41)}$$

$$D = \frac{(W X b)' M (W X b)}{s^2} + tr\,(W'W + W W) \quad \text{.....(2.42)}$$

$$s^2 = \frac{e' e}{n} \quad \text{.....(2.43)}$$

s^2: Variance of error for general linear model regression

We compare the calculate value with table value for $\chi^2(1,\alpha)$after that we decide to the hypotheses[2][10].

2-2-6-3 Lagrange Multiplier (LM) error Test for (SEM) Model:

Hypotheses test for Lagrange (SEM) model

H_0: θ = 0 spatial dependence exist in error

H_1: at least one of the θ does not equal to zero 0

spatial dependence not exist in error

Where reject null hypotheses and accept alternative hypotheses that is mean spatial dependency does exist and alternative model is spatial error model (SEM) and the role is show in dawn where some time denoted by (LM-SEM) [2]

$$LM_\theta = \frac{\left(\frac{e'\,W\,e}{S^2}\right)^2}{T} \ldots\ldots(2.44)$$

$$T = tr\,[(w + w')w\,] \ldots\ldots(2.45)$$

s^2:variance of error for general linear model regression

Comparison between (LM_θ, LM_λ) with value of χ^2 table by once degree freedom and once level significant where Lagrange test for (SAR) or (SEM) significant or exist spatial dependency in each of them we must go to robust test and robust role for (SAR) and (SEM) model is

$$Robust - LM_\lambda = \frac{\left[\left[\frac{e'WY}{S^2}\right] - \left[\frac{e'We}{S^2}\right]\right]^2}{D - T} \ldots\ldots(2.46)$$

$$Robust - LM_\theta = \frac{\left[\left[\frac{e'We}{S^2}\right] - \left[\frac{T}{D}\right]\left[\frac{e'WY}{S^2}\right]\right]^2}{T - \left[\frac{T^2}{D}\right]} \ldots\ldots(2.47)$$

$$D = \frac{(W\,X\,b)'M\,(W\,X\,b)}{s^2} + tr\,(W'W + W\,W) \ldots\ldots(2.48)$$

$$T = tr\,(w'\,w + w\,w\,)$$

Comparison each $(Robust - LM_\lambda, Robust - LM_\theta)$ with table value for χ^2 by once degree freedom and once level significant [2][15]

Table (2-1) Show the roles of test parameter in spatial regression models

Test	**Formulation**	**Table value**	**Source**
MORAN	$I_M = \frac{n\,(e'we)}{S_o\,(e'e)}$	N(0,1)	Cliff and ord (1981)
LM-SAR	$LM_\lambda = \frac{\left(\frac{e'W\,Y}{S^2}\right)^2}{D}$ $D = \frac{(W\,X\,b)'M\,(W\,X\,b)}{s^2} + tr\,(W'W + W\,W)$ $s^2 = \frac{e'e}{n}$	$\chi^2(1,\alpha)$	Anselin (1988)
Robust LM-SAR	$Robust - LM_\lambda = \frac{\left[\left[\frac{e'WY}{S^2}\right] - \left[\frac{e'We}{S^2}\right]\right]^2}{D - T}$	$\chi^2(1,\alpha)$	Anselin etal(1996)
LM-SEM	$LM_\theta = \frac{\left(\frac{e'\,W\,e}{S^2}\right)^2}{T}$	$\chi^2(1,\alpha)$	Burridge (1988)
Robust LM-SEM	$Robust - LM_\theta = \frac{\left[\left[\frac{e'We}{S^2}\right] - \left[\frac{T}{D}\right]\left[\frac{e'WY}{S^2}\right]\right]^2}{T - \left[\frac{T^2}{D}\right]}$	$\chi^2(1,\alpha)$	Anselin (1988)

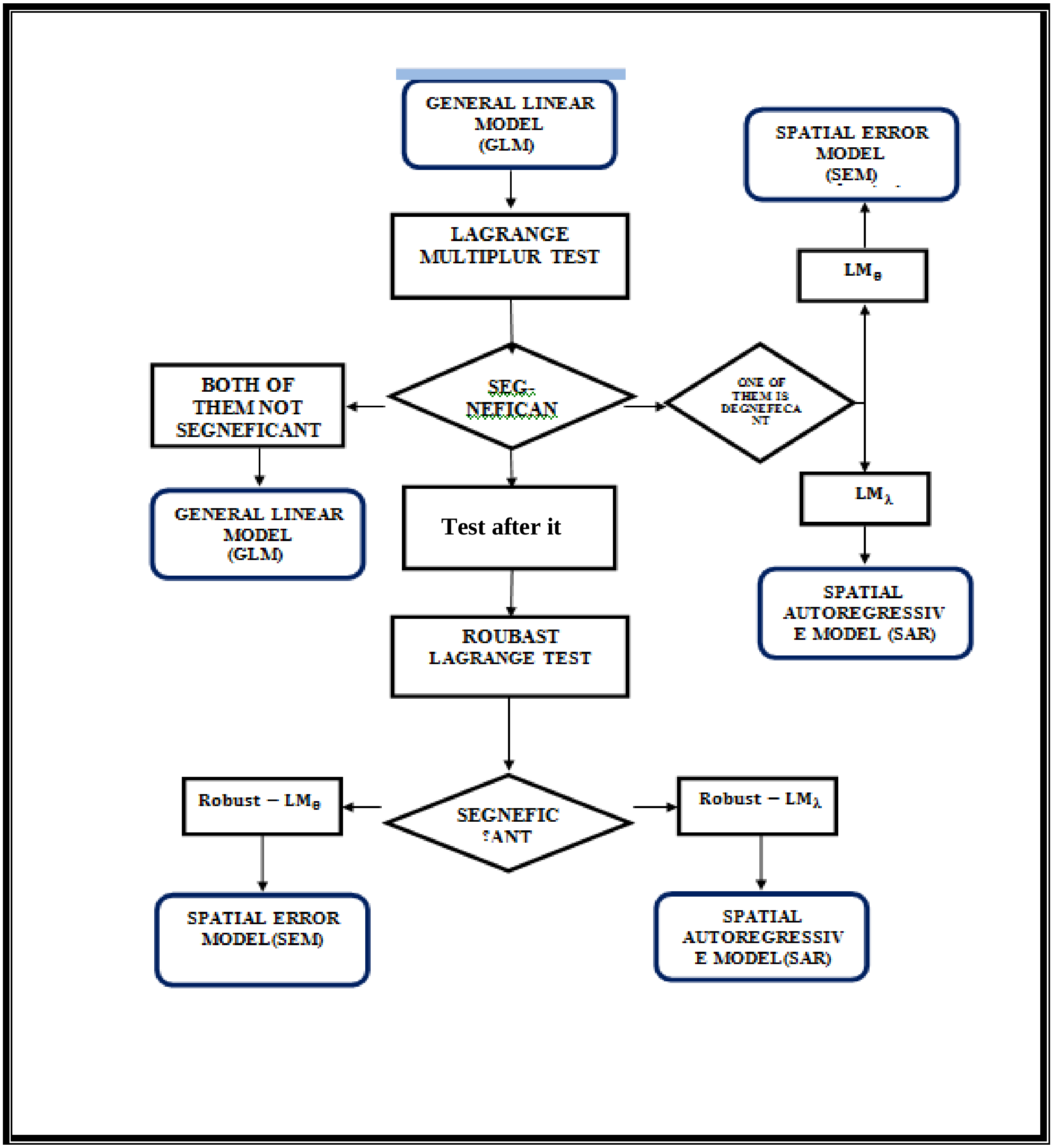


Figure (2-7) Show the idea choose alternative spatial regression model

2-2-7 Comparison Criteria for Choosing the Best Model

The choice of a particular model from a range of models style is an important aspect of the analysis of the data as it leads us to choose the best model , where the use of certain statistical criteria which are as described below

2-2-7-1 Root Mean Squares Error:

It is a square root of the sum of squares errors divided by the (n-k-1) is calculated for each models, the specimen which is the square root of the mean square value of the smaller mistakes is the best specimen . The formula for calculating this standard is[10]

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}} \quad \ldots\ldots(2.49)$$

2-2-7-2 Mean Absolute Percentage Error(MAPE):

It is calculated by dividing the sum of the absolute value of the error on the real value divided by the number of Views (n) and the formula for calculating this standard is [10]

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \ldots\ldots(2.50)$$

And the smaller value is the best model

2-2-7-3 Akaike Information Criterion(AIC):

This standard introduced for the first time in 1973 , of the world (Akaike's), and this standard is equal to twice the number of parameters minus twice the maximum likelihood function of the observations and is expressed mathematically as the following [6]

$$\text{AIC} = -2\text{Log}\,(L) + 2K \ldots\ldots(2.51)$$

L: Great value for the value of the logarithm (M.L.E)

K : Number for the parameters of the model

It has been corrected Akaki criterion (AIC) , which takes the formula described in the below. It is worth mentioning that if the sample size was small ((n / k)<40) it is better to use the standard debugger Akaki

$$AIC_C = AIC + \frac{2K(K+1)}{(n-K-1)} \ldots\ldots(2.52)$$

2-2-7-4 Adjusted Determinations of Coefficient(R^2_{adj}):

The coefficient of determination R^2 is not alone as a good indicator of how well the explanatory variables in explaining the values seen. And the value of the coefficient of determination increases with each additional variable falls within the specimen , so is calculated coefficient of determination average which enters into account the number of explanatory variable s , and is calculated using the following formula

$$R^2_{adj} = 1 - \left[\frac{\left(\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k}\right)}{\left(\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}\right)}\right] \ldots..(2.53)$$

For comparison models the best model is the models which have bigger value [71]

2-2-8-1 Fuzzy Set and Fuzzy Logic

Fuzzy set is very convenient method for representary some form of uncertainty

The idea of fuzzy logic was first advanced by Dr. Lotfi Zadeh of the University of California at Berkeley in the 1960s. Dr. Zadeh was working on the problem of computer understanding of natural language. Natural language (like most other activities in life and indeed the universe) is not easily translated into the absolute terms of 0 and 1. (Whether everything is ultimately descriable in binary terms is a philosophical question worth pursuing, but in practice much data we might want to feed a computer is in some state in between and so, frequently, are the results of computing)[66].

And fuzzy is a mathematical logic that attempts to solve problems by assigning values to an imprecise spectrum of data in order to arrive at the most accurate conclusion possible. Fuzzy logic is designed to solve problems in the same way that humans do: by considering all available information and making the best possible decision given the input [67].

2-2-8-2 Various Types of Membership Functions: Depending on the types of membership function different types of fuzzy set will be obtained proposed a series of member ship functions that could be classified into two groups .those made up of start lines being “linear” ones ,and the “cueved” or “non linear” . [65] we will now go on to look at some types of membership functions(Triangular,Singleton,L-Function,Gamma Function,Gaussian…...)

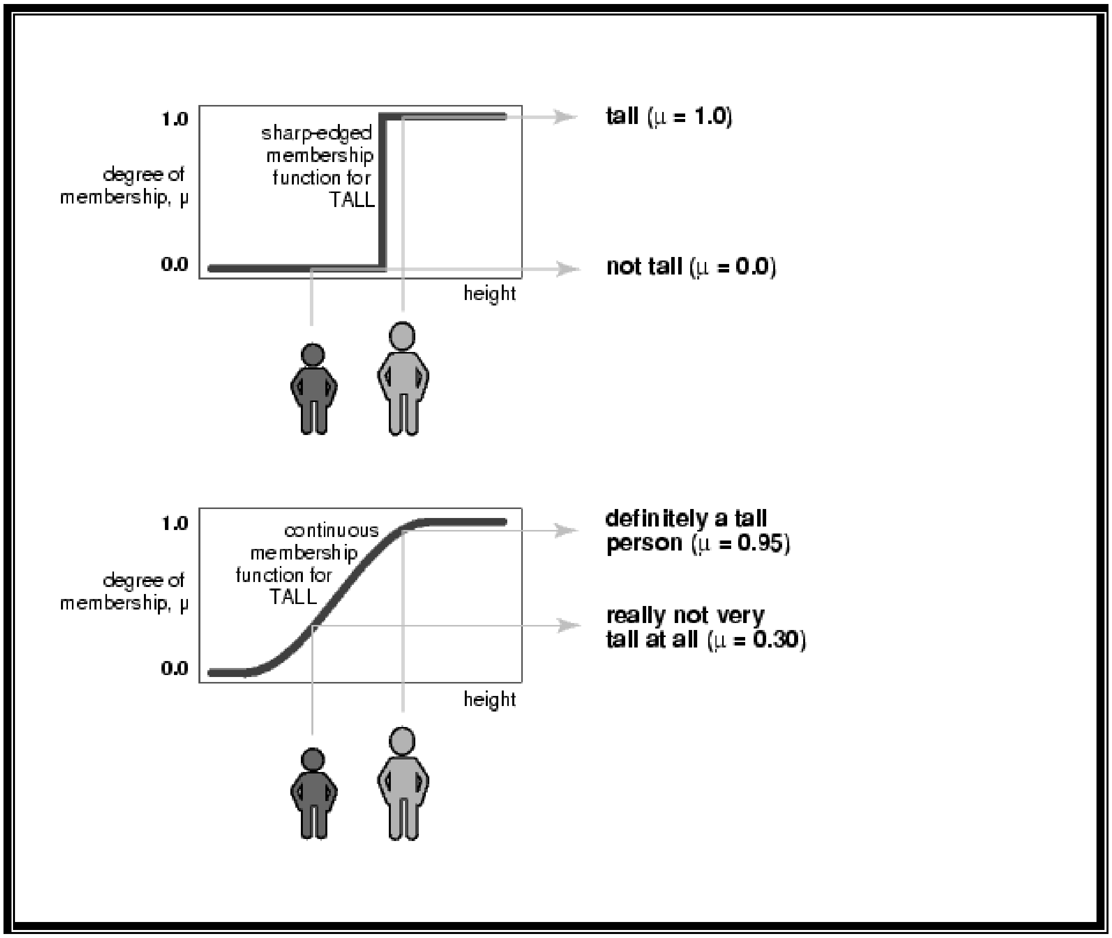


Figure (2-8) :Show an example of a fuzzy logic

2-2-8-3 Steps for Find Fuzzy Data

The following steps explain how the data converting to the fuzzy data by using the Gaussian membership function

Step 1 :We must at first test the data for normality and my research has a normal distribution because that we must use the roll of Gaussian normal distribution

Step 2: Find memberships by Gaussian member ship function

Specified by two parameters {m, σ^2} as follows:

$$\text{Gaussian } (x{:}\mu,\sigma^2) = e^{\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]} \quad \text{....(2.54)}$$

$$\text{Where Mean } (\mu) = \frac{\sum_{i=1}^{n} Xi}{n}$$

And

$$\text{Variance}(\sigma^2) = SD^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$$

Step 3 :After find memberships by this role multiple memberships with variable (x) which denoted by (S_y,S_{x1},S_{x2},S_{x3})

Step 4 : Find fuzzy Data [66][67]

$$Fx=\frac{\sum_{i=k}^{n} \mathrm{S_i\ Xi}}{\sum_{i=k}^{n} \mathrm{S_i}} \quad \text{.....(2.55)}$$

Chapter Three

Practical Part

Chapter Three

Practical Part

3-1 Introduction

This part covers the practical aspects of this thesis, applying the data practically to the traditional methods (OLS) and modern statistical Models SAR and SEM for estimating regression parameters using Raw data and Fuzzy data and comparing the results between different models.

3-2 Description of the Data

The data used in this thesis are taken from Ministry Rough waters weather.The sample consists of (27) places (observations) (every observations are a yearly average) about the (Atmospheric Pressure, Temperature degree, Relative Humidity, wind Speed) of Kurdistan Region(Sulaimaniyah,Erbil, and Dhouk) stations .

Statistical analysis was applied to find the best model to estimate the Atmospheric Pressure (A.P) in Kurdistan region

Atmospheric Pressure(A.P)(Y): Defined as the force per unit area exerted against a surface by the weight of the air above that surface.

Wind Speed(W.S)(x_1) :Is caused by air moving from high pressure to low pressure, usually due to changes in temperature. Wind speed affects weather forecasting, aircraft and maritime operations, construction projects, growth and metabolism rate of many plant species, and countless other implications

Air Temperature (A.T)(x_2): Is a measure of how hot or cold the air . It is the most commonly measured weather parameter. More specifically, temperature describes the kinetic energy, or energy of motion, of the gases that make up air. As gas molecules move more quickly, air temperature increases.

Relative Humidity(R.H)(x_3): The amount of water vapor in the air, expressed as a percentage of the maximum amount that the air could hold at the given temperature the ratio of the actual water vapor pressure to the saturation vapor pressure

A multiple linear regression model which is based on the data is expressed in the following form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Where

$Y_i =$ Atmospheric Pressure (%) (A.P.)

X_{i1}= Wind Speed (m/s) (W.S)

X_{i2}= Air Temperature (C^{o}) (A.T)

X_{i3}= Relative Humidity (%) (R.H)

Thus,

$$A.P = \beta_0 + \beta_1 W.S + \beta_2 A.T + \beta_3 R.H + \varepsilon_i \quad .....(3.1)$$

In matrix notation model (1) can be expressed as

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon},$$

Where $\underline{Y} = (Y_1, Y_2, ..., Y_{27})'$ is an (27×1) observation vector, $\underline{\beta} = (\beta_0, \beta_1, ..., \beta 3)'$ is an (4×1) vector of unknown parameters, X is an (27×4) matrix of full column rank and $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_{27})$ is an (27×1) random error vector

3-3 Software:

The following software and application used to analyze the practical

Part of this thesis:

1-Geoda

2-Minitab

3-Matlab

4-Excel

3-4 Regression Model for Raw Data

In practice part we need the data and the table show raw data for fit general linear model for raw data

Table (3-1) Show the Raw data

Location\Stations	A.P=Y(%)	W.S=X1	A.T=X2	R.H=X3 (%)
Bazian	9.2094	2.0607	19.6341	0.4619
Darbandexan	9.5305	2.7335	22.8866	0.4054
Dukan	9.4990	3.1633	21.6747	0.4324
Halabjae shahid	9.4295	1.0000	21.4600	0.4106
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Grdgan	9.4601	1.3017	20.5845	0.3375
Xwrmal	9.4188	1.0750	12.0750	0.6705

The table below present the descriptive Statistics for the variables(response & explanatory)

Table(3-2):Descriptive statistics for response and explanatory variables

Variables	A.P	W.S	A.T	R.H
Minimum	8.6312	0.8592	12.0750	0.2129
Maximum	10.1490	3.1633	24.8731	0.6705
Average	9.4075	1.7142	20.0776	0.4344
Standard division	0.4220	0.6036	2.4886	0.0907

Figure (3-1), the 27 stations of weather used in the analysis are shown on the map.Especially on the northern part (Kurdistan region) of the Iraq.

Figure (3-1) Show the Map of Kurdistan region

Estimating and testing the parameter of the Model :

The following tables explain the parameters estimated and testing

Table (3-3) Observed parameter and standard error for model I

Model	Unstandardized coefficients		Standardized coefficients	T	Sig. P value
	B	Std.Error	Beta		
Constant	6.083	0.781		7.791	0.000
W.S	-0.196	0.101	-0.281	1.948	0.064
A.T	0.121	0.027	0.712	4.501	0.000
R.H	2.847	0.754	0.612	3.774	0.001

Table (3-3) Show that the p-value of wind speed (W.S) is greater than level of significant ($\alpha = 0.05$) this mean that the explanatory variable (W.S) is not significant and the p-value of other explanatory variables (A.T and R.H) are less than ($\alpha = 0.05$) this mean that the explanatory variables (A.T and R.H) are statistically significant, so they remain in the model. To test the significant of the model,ANOVA table was made as follows:

Table (3-4) :ANOVA table for Model I using OLS Method

Model	Sum of Square	D.F	Mean Square	F	Sig.
Regression	2.600	3	0.867	9.815	0.000
Residual	2.031	23	0.088		
Total	4.632	26			

In the ANOVA table, the corresponding P-Value is less than ($\alpha = 0.05$), then the null hypothesis was rejected, i.e., this means that the relationship between Atmospheric Pressure(A.P) with (A.T and R.H) are highly significant (under $\alpha = 0.05$).

Table (3-5) Summary table for model I

R^2	Adjusted R^2	Durbin-Watson
0.561	0.504	1.698

Table (3-5) Shows that the adjusted R^2 - value is 0.504 indicating that the 50.4% of the variation in Atmospheric Pressure has been explained by the regressors (A.T and R.H)

3-4-1 Test for Problem Econometric and Assumption of Regression Model

3-4-1-1 Test of Normality

Now consider testing for the standard residuals using jarque-Bera test i.e

H_0: The distribution of residual standard is very close to normal standard

H_1: The standard residual are significantly different from the normal standard

The null hypothesis H_0 is not rejected if the P-value for jarque-Bera test is greater than 0.05 , otherwise rejects H_0 and fails to reject the alternative hypothesis H_a .

Table (3-6) Show test of normality

Test	D.F	Value	P-value
Jarque-Bera	2	1.0108	0.6032

Table (3-6) Shows that the P-value for residuals using (jarque-Bera)test is 0.6032 this value is greater than(α = 0.05), then we accept the null hypothesis. This indicates that residuals are normally distribution.

3-4-1 -2 Test of Hetroscedasticity for Model I

Now consider to test the heteroscedasiticity brush-pagan&Konker-basset

Test of hypotheses

$$H_0 : \sigma^2_{e_1} = \sigma^2_{e_2} = \ldots = \sigma^2_{e_k} \qquad \text{(Homoscedasticity)}$$

vs. H_a : at least one of σ_i^2 does not equal zero. (Hetroscedasticity)

the result of the test showing in the table below

Table (3-7) Diagnostics for heteroskadasticity random coefficients test

Test	D.F	Value	P-value
Breusch-pagan	3	1.1553	0.7637
Konker-Basset	3	1.7544	0.6249

Table (3-7) Show that the P-value of the test are greater than (α = 0.05),this indicate that the null hypothesis of homogeny is accept or the data will not have problem of hetroskadasticity

3-4-1-3 Test of Autocorrelation for Model I

Now consider to testing the autocorrelation using the Durbin-watson test

Test of hypotheses

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

The table below show that the result of Durbin Watson test:

Table (3-8) Show the interval accept for Durbin test

Model	D.W	D.U	4-D.U
OLS	1.698	1.65	2.35

The value of D.W statistical test is 1.698 then value is between 1.65≤1.698≤4-1.65=2.35 then the data has not the Auto correlation problem.The level of (D.W) is greater than (du=1.65) and less than(4-du) .This indicates that there is not any serious autocorrelation among the variables, and so the null hypothesis is accepted

3-4-1-4 Test of Multicollinearity for Model I

Now consider to test the multicollinearity using the variance inflation factor

Test of hypotheses

H_0: X_j Orthogonal

H_1:X_j not orthogonal

The table below show that the result of the variance inflation factor (VIF)

Table(3-9) Show the test of Multicollinearity

Collinearity Statistics		
Factors	Tolerance	Variance inflation factor
Wind Speed	0.919	1.089
Air Temperature	0.762	1.313
Relative Humidity	0.725	1.379

The output of Table (3-9) show that the (VIF) of the variables (W.S) and (A.T) and (R.H) are near to one, which indicate that there are no multicollinearity problem among the variables.

3-4-2 Fitting Linear Regression Estimation Using (OLS)

Here, fitting multiple linear regression models is used to describe the relationship between the response variable and the explanatory variables.

The fitted regression model is:

Model I A.P= 6.083+0.121 A.T+2.847 R.H…..(3.2)

Starting with Figure (3-2) we notice that it indicates the existence of statistical evidence that the explanatory variables in the model are related to the expected value of the response variables

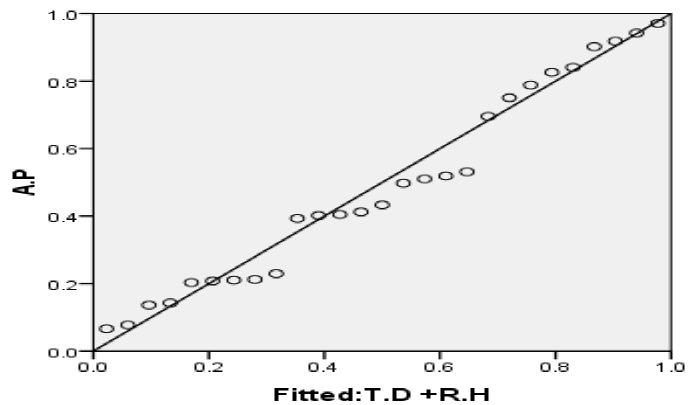


Figure (3-2) Observed A.P versus Fitted Regression Mosel I

3-5 Regression Model for Fuzzy Data

In this step we must find fuzzy data and work in it and the table below show procedures for find fuzzy data

Table (3-10) Show the procedures to calculate the fuzzy data

S	S_1	S_2	S_3	S_y	S_{x1}	S_{x2}	S_{x3}
0.895684	0.848092	0.9842442	0.955108	8.248689	1.747693	19.32472	0.441164
0.958446	0.240355	0.5288830	0.950204	9.134444	0.657018	12.10432	0.385213
0.97678	0.056068	0.8139055	0.999756	9.278422	0.177361	17.64112	0.432291
0.998643	0.496652	0.8570393	0.966182	9.416701	0.496652	18.39204	0.396714
0.38737	0.90816	0.7548803	0.993945	3.419005	1.797407	13.74734	0.42183
0.564129	0.739755	0.1562069	0.916034	5.561818	1.614824	3.885357	0.363116
0.970176	0.997496	0.8155506	0.995136	9.026164	1.752545	17.67038	0.423371
0.941411	0.999529	0.9859780	0.687729	8.718251	1.694868	20.20844	0.352743
0.346176	0.436036	0.6935224	0.993667	3.043825	1.086595	12.44766	0.441812
0.999988	0.994338	0.8665812	0.837651	9.40536	1.76846	18.55302	0.318636
0.964786	0.496652	0.8934368	0.945388	8.967189	0.496652	16.88254	0.439424

0.581035	0.813648	0.6850847	0.982287	5.210537	1.079311	12.27202	0.443556
0.6261	0.981776	0.6826604	0.920557	6.145771	1.569291	15.19068	0.433875
0.982764	0.984528	0.3697306	0.602063	9.168006	1.792633	6.125304	0.316565
0.761348	0.998051	0.3219832	0.977436	7.399681	1.673227	7.671014	0.405651
0.223636	0.695742	0.9976034	0.635273	2.26722	0.83489	20.20147	0.330869
0.254262	0.366743	0.9139547	0.960978	2.569568	0.315093	19.31488	0.44205
0.225475	0.512947	0.8310493	0.752964	2.285414	0.521497	17.94374	0.378553
0.726084	0.460418	0.8138877	0.907144	6.585428	1.135419	15.04101	0.430401
0.2137	0.97961	0.7877062	0.526547	2.168841	1.559213	14.46097	0.282843
0.831862	0.997642	0.9971380	0.050861	7.612697	1.668773	20.20802	0.01083
0.184265	0.985588	0.9082749	0.823257	1.590427	1.790872	17.24445	0.311036
0.999501	0.711702	0.9604032	0.960835	9.389479	0.865666	18.60317	0.392743
0.735064	0.640522	0.9379702	0.809818	6.671692	0.733014	19.66757	0.304059
0.669431	0.294084	0.9952731	0.072177	6.044538	0.781869	20.22381	0.016338
0.99227	0.791785	0.9794702	0.565301	9.386958	1.030666	20.1619	0.190784
0.999645	0.57087	0.0056840	0.033858	9.415405	0.613685	0.068634	0.022702

The table (3-10) show the applied role and step in equations (2.54) and (2.55) wich in this step find member ship and multiple it by raw data for find fuzzy data which in this step by member ship find weight for every elements

The fuzzy data show as follows

Table (3-11) Show the Fuzzy data

Location\Stations	A.P=Y(%)	W.S(X3)	A.T=X2	R.H=X3 (%)
Bazian	9.370397	1.645115	20.21879	0.438435
Darbandexan	9.378358	1.625695	20.24822	0.437307
Dukan	9.36986	1.610828	20.17488	0.43891
Halabjae shahid	9.362064	1.605953	20.10785	0.439273
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Grdgan	9.439338	1.206726	20.5354	0.356308
Xwrmal	9.41875	1.075	12.075	0.6705
Mean	9.3504	1.5027	19.8192	0.4295
Variance	0.0022	0.0163	2.4309	0.0034

Table (3-11) show fuzzy data witch found by the steps and in below do all test that applied for raw data

Table (3-12) Observed parameter and standard error for model II

Model	Unstandardized coefficients		Standardized coefficients	T	Sig. P value
	B	Std.Error	Beta		
Constant	7.990	0.299		26.685	0.000
W.S	-0.321	0.087	-0.867	3.690	0.001
A.T	0.059	0.014	1.942	4.341	0.000
R.H	1.575	0.297	1.934	5.300	0.000

Table (3-12) show the p-value of all variables explanatory variables are less than (α=0.05),this indicated that those variables are more significant and remaining in the model

Table (3-13) ANOVA table for Model II using OLS depending on fuzzy data

Model	Sum of square	D.F	Mean square	F	Sig.
Regression	0.034	3	0.011	10.573	0.000
Residual	0.024	23	0.001		
Total	0.058	26			

In the ANOVA table, the corresponding P-Value is less than 0.05, the null hypothesis was rejected, i.e., this means that the relationship between Atmospheric Pressure and W.S,A.T and R.H are highly significant (under $\alpha = 0.05$).

Table (3-14) Summary table for model II

R^2	Adjusted R^2
0.580	0.525

Table (3-14) shows that the adjusted R^2 - value is 0.525 indicating that the 52.5% of the variation in Atmospheric Pressure has been explained by the regressors.

3-5-1 Linear Regression Estimation Using (OLS) for Model II

Here, fitting multiple linear regression models is used to describe the relationship between the response variable and the explanatory variables when we have fuzzy data.

The fitted regression model is:

Model II A.P= 7.990-0.321 W.S+0.059 A.T+1.575 R.H……(3.3)

3-5-2 Regression Model with Raw and Fuzzy Data

The output in the table (3-15) show that the comparison between the regression model when using the Raw data and fuzzy data depending on some criteria or measure .

Table (3-15) Show the Regression model with Raw and Fuzzy Data

Regression model with	R^2_{adj}	AIC_C	RMSE	MAPE
Raw data	0.504	14.7703	0.2971	0.00084
Fuzzy data	0.525	-104.657	0.0325	0.00001

Table (3-15) Shows that the comparison between the Regression model with Raw data and fuzzy data, the result show that the Regression model with fuzzy data is taken the best result when we compared with regression model with Raw data depending on some criteria or measures

3-6 Test for finding Spatial Dependency for Raw Data

- **Moran's Test :**

Before construct create a model we must test the data for find spatial dependency of the data.

The general test for spatially dependency is Moran test

Table (3-16) Show the Moran test of spatial dependency for Rook,Bishop and Queen Matrices

Weight Matrix (W)		
Rook	Bishop	Queen
2.3119**	0.1332	2.3250**

Based on Table (3-16) it could be seen that the value of Moran's test using different weighted matrix rook, bishop and queen calculated and the value of test is greater than $Z_{0.025}$ =1.96;thus,it can be conclude that the test is significant where we use the rook and queen matrix and the otherwise is not significant when we use the Bishop matrix. If the test is significant meaning that the place has spatial dependency with each other

3-7 Spatial Regression Models with Raw Data

3-7-1 Spatial Regression Models (SAR)

3-7-1-1 SAR Model by using Rook Matrix

The table below shows the output of parameters estimation in SAR Model using the weight matrix rook. The positive spatial parameter (λ) indicated that the neighbor places is similar from each others, and in classical regression model the variable wind speed is not significant while in SAR model where using the rook weight matrix the variable wind speed is not significant too.

Table (3-17) Estimation the parameter of SAR model by using rook matrix

Model	Weight Matrix- Rook(WMR)			
	Coefficient	Std. Error	t-calculate	t- table
(Constant)	6.3189	0.7626	8.2852**	2.052
b1	-0.1834	0.09841	1.8635	
b2	0.1113	0.0262	4.2335**	
b3	2.5050	0.7369	3.39906**	
(λ)	0.0141			

The SAR model is estimated as:

$$\hat{y}_i = 6.3189+0.1113\ b_2+2.5050\ b_3+0.0141\ \lambda \ldots..(3.4)$$

Where λ is the average A.P in all neighboring countries, according to the weight matrix used, which is " Rook" in this application.

3-7-1-2 SAR Model by using Bishop Matrix

The table below shows the output of parameters estimation in SAR Model using the weight matrix bishop. The positive spatial parameter (λ) indicated that the neighbor places is similar from each others, and in classical regression model the variable wind speed is not significant while in SAR model where using the bishop weight matrix only the variable wind speed is not significant

Table (3-18) Estimation the parameter of SAR model by using bishop matrix

Model	Weight Matrix- Bishop(WMB)			
	Coefficient	Std. Error	t- calculate	t- table
(Constant)	6.0688	0.7792	7.7877**	2.052
b1	-0.1995	0.1005	1.9840	
b2	0.1221	0.0267	4.5587**	
b3	2.8149	0.7530	3.7381**	
(λ)	0.0047			

The SAR model is estimated as:

$$\hat{y}_i = 6.0688 + 0.1221\ b_2 + 2.8149\ b_3 + 0.0047\ \lambda \ldots\ldots(3.5)$$

Where λ is the average A.P in all neighboring countries, according to the weight matrix used,which is " Bishop" in this application.

3-7-1-3 SAR Model by using Queen Matrix

The table below show the output of parameters estimation in SAR Model using the weight matrix queen.The positive spatial parameter (λ) indicated that the neighbor places is similar from each others, and in classical regression model the variable wind speed is not significant while in SAR model where using the queen weight matrix only the variable wind speed is not significant too.

Table (3-19) Estimation the parameter of SAR model by using queen matrix

Model	Weight Matrix- Queen(WMQ)			
	Coefficient	Std. Error	t-calculate	t-table
(Constant)	6.4880	0.7384	8.7855**	2.052
b1	-0.1944	0.0952	2.0401	
b2	0.1096	0.0253	4.3180**	
b3	2.0739	0.7136	2.9062**	
(λ)	0.0240			

The SAR model is estimated as:

$$\hat{y}_i = 6.4880 + 0.1096\ b_2 + 2.0739\ b_3 + 0.0240\ \lambda \ldots\ldots(3.6)$$

Where λ is the average A.P in all neighboring countries,according to the weight matrix used,which is “ Queen” in this application.

3-7-2 Spatial Error Model (SEM)

3-7-2-1 SEM Model by using Rook Matrix

The table below shows the output of parameters estimation in SEM Model using the weight matrix rook.The positive spatial parameter (θ) indicated that the neighbor places is similar from each others, and in classical regression model the variable wind speed is not significant while in SEM model where using the rook weight matrix only the variable wind speed is not significant.

Table (3-20) Estimation the parameter of SEM model by using rook matrix

Model	Weight Matrix-Rook(WMR)			
	Coefficient	Std. Error	t-calculate	t- table
(Constant)	6.3462	0.7875	8.0585**	2.052
b1	-0.1599	0.1016	1.5736	
b2	0.1097	0.02706	4.0529**	
b3	2.5834	0.7609	3.3948**	
(θ)	0.2763			

The SEM model is estimated as:

$$\hat{y}_i = 6.3462 + 0.1096\ b_2 + 2.5834\ b_3 + 0.2763\ \theta \ldots\ldots(3.7)$$

Where θ is the average error of prediction in neighboring countries of station i,according to the weight matrix used ,which is “Rook” in this application.

3-7-2-2 SEM Model by using Bishop Matrix

The table below show the output of parameters estimation in SEM Model using the weight matrix bishop.The negative spatial parameter (θ) indicated that the neighbor places is not similar from the other, and in classical regression model the variable wind speed is not significant while in SEM model where using the bishop weight matrix only the variable wind speed is not significant.

Table (3-21) Estimation the parameter of SEM model by using bishop matrix

Model	Weight Matrix-Bishop(WMB)			
	Coefficient	Std. Error	t-calculate	t- table
(Constant)	6.0867	0.7806	7.7965**	2.052
b1	-0.1967	0.1007	1.9526	
b2	0.1206	0.0268	4.4945**	
b3	2.8453	0.7543	3.7716**	
(θ)	-0.0096			

The SEM model is estimated as:

$$\hat{y}_i = 6.0867 + 0.1206\ b_2 + 2.8453\ b_3 - 0.0096\ \theta \ldots.(3.8)$$

Where θ is the average error of prediction in neighboring countries of station i,according to the weight matrix used ,which is “Bishop” in this application.

3-7-2-3 SEM Model by using Queen Matrix

The table below show the output of parameters estimation in SEM Model using the weight matrix queen. The positive spatial parameter (θ) indicated that the neighbor places is similar from each others, and in classical regression model the variable wind speed is not significant while in SEM model where using the queen weight matrix only the variable wind speed is not significant.

Table (3-22) Estimation the parameter of SEM model by using queen matrix

Model	Weight Matrix-Queen (WMQ)			
	Coefficient	Std. Error	t- calculate	t-table
(Constant)	6.2555	0.7890	7.9275**	2.052
b1	-0.1479	0.1018	1.4526	
b2	0.1123	0.02712	4.1407**	
b3	2.6142	0.7624	3.4284**	
(θ)	0.2630			

The SEM model is estimated as:

$$\hat{y}_i = 6.2555 + 0.1123\ b_2 + 2.6142\ b_3 + 0.2630\ \theta \ldots\ldots(3.9)$$

Where θ is the average error of prediction in neighboring countries of station disaccording to the weight matrix used ,which is “Queen” in this application.

3-8 Tests for Finding the Best Model in Raw Data

3-8-1 Lagrange Test for SAR (LM λ):

This test is used for finding spatial dependency in spatial autoregressive model (SAR)

Table (3-23) Lagrange test for SAR Model

Test	Weight Matrix (W)		
	Rook	Bishop	Queen
LM λ	7.4654**	0.086381	9.65248**
Robust LM λ	6.7468**	0.087028	8.7016**

In W rook: The values of the two test (LM λ=7.4654, Robust LM λ=6.7468) are significant when we compared the value of the tests with the value of chi-square (1,α) degree freedom i.e $\chi^2_{(1,0.05)}$ =3.841

In W bishop: The values of the two test (LM λ=0.086381, Robust LM λ=0.087028) are non-significant when we compared the value of tests with the value of chi-square (1,α) degree freedom i.e $\chi^2_{(1,0.05)}$ =3.841 are not significant

In W queen: The values of the two test (LM λ=9.65248, Robust LM λ=8.7016) are significant when we compared the value of tests with the value of chi-square (1,α) degree freedom i.e $\chi^2_{(1,0.05)}$ =3.841

3-8-2 Lagrange Test for SEM (LM θ):

This test is for finding spatially dependency in spatial error model

Table (3-24) Lagrange test for SEM Model

Test	Weight Matrix (W)		
	Rook	Bishop	Queen
LM θ	1.203816	0.00027	1.035462
Robust LM θ	0.486513	0.000956	1.986976

In table (3-24), the values of the two test LM θ and robust LM θ according to the weight matrix rook, bishop and queen are not significant because the values of the two tests are less than the value of Chi-Square with degree of freedom $\chi^2_{(1,0.05)}$ =3.841 we can concluded that the SAR Model is better than the SEM model.

3-9 Calculate Different Criteria by Using SAR Model:

After create a Model by using three types of weighted matrices and test it. For finding spatial dependency we must use same criteria for finding the best model

Table (3-25) Show the calculate different criteria by using SAR Model

Criteria	SAR		
	Wrook	Wbishop	Wqueen
R^2_{adj}	0.5468	0.5266	0.5747
AIC_C	3.941577	3.956096	3.920615
RMSE	0.2903	0.2967	0.2811
MAPE	0.000803	0.000837	0.000748

Result and discussion:

1- R^2_{adj} This measure is based on the concept that how much variation in Y's stated by Syy is explained by SSreg.In this thesis the value of R^2_{adj} in SAR Model is a 0.5468,0.5266 and 0.5747 respectively of the variation of the response variable is explained by the model according to weight matrices (rook,bishop and queen)and the best and good value in queen weight matrix.

2- AIC is a measure of the relative quality of statistical models for a given set of data. Given a collection of models for the data in AIC_C the best model is the smaller value because in this thesis the best model is SAR model according to the queen weight matrix

3- RMSE is a measure used for differences between the values (sample and population) predicted by a model or an estimator and the values actually observed in RMSE the smaller value is better and in this thesis the best model is SAR model according to the queen weight matrix

4- MAPE is a measure of prediction accuracy of a forecasting method in statistics, in this measure the smaller value is better. In this thesis the best model is SAR model according queen weight matrices

When we compare between the models according to the weight matrixes by some criteria or measures in all measures accept to select SAR model with weight matrices queen is best model by raw data.

3-10 Test for finding Spatial Dependency in Fuzzy Data

Moran's Test:

After construct (create) model we must test the parameters of the model for find the spatial dependency of the Spatial parameter.

Moran's test is The first and general test for spatially dependency

Table (3-26): The test of spatial dependency for rook,bishop and queen matrices

Weight Matrix (W)		
Rook	Bishop	Queen
2.6600**	0.6990	2.7099**

Based on Table (3-26) it could be seen that the value of Moran's test using different weighted matrix rook, bishop and queen calculated and the value of test is greater than $Z_{0.025}$ =1.96;thus,it can be conclude that the test is significant where we use the rook and queen matrix but the otherwise test is not significant when we use the Bishop matrix. If the test is significant meaning that the place has spatial dependency with each other

3-11 Spatial Regression Models with Fuzzy Data

3-11-1 Spatial Autoregressive Model (SAR)

3-11-1-1 SAR Model by using Rook Matrix

The table below shows the output of parameters estimation in SAR Model using the weight matrix rook. The negative spatial parameter (λ) indicated that the neighbor places is not similar from each others, and in classical regression model all coefficient are significant while in SAR model where using the rook weight matrix all the coefficient are significant too.

Table (3-2) Estimation the parameters of SAR model by using rook matrix

Model	Weight Matrix-Rook (WMR)			
	Coefficient	Std. Error	t-calculate	t-table
(Constant)	7.8610	0.0826	95.1071**	2.052
b1	-0.3162	0.0106	29.6482**	
b2	0.0632	0.0028	22.2470**	
b3	1.6837	0.0798	21.0807**	
(λ)	-0.0019			

The SAR model is estimated as:

$$\hat{y}_i = 7.8610 - 0.3162\ b_1 + 0.0632\ b_2 + 1.6837\ b_3 - 0.0019\ \lambda \quad \text{….(3.10)}$$

Where λ is the average A.P in all neighboring countries, according to the weight matrix used, which is " Rook" in this application.

3-11-1-2 SAR Model by using Bishop Matrix

The table below show the output of parameters estimation in SAR Model using the weight matrix bishop. The positive spatial parameter (λ) indicated that the neighbor places is similar from each others, and in classical regression model all coefficient are significant while in SAR model where using the bishop weight matrix all the coefficient are significant too.

Table (3-28) Estimation the parameters of SAR model by using bishop matrix

Model	Weight Matrix -Bishop (WMB)			
	Coefficient	Std. Error	t- calculate	t-table
(Constant)	7.9768	0.0852	93.6133**	2.052
b1	-0.3204	0.01099	29.1409**	
b2	0.0594	0.0029	20.2822**	
b3	1.5740	0.08233	19.1161**	
(λ)	0.0007			

The SAR model is estimated as:

$$\hat{y}_i = 7.9768-0.3204\ b_1+0.0594\ b_2+1.5740\ b_3+0.0007\ \lambda \ldots\ldots(3.11)$$

Where λ is the average A.P in all neighboring countries, according to the weight matrix used, which is “ Bishop” in this application.

3-11-1-3 SAR Model by using Queen Matrix

The table below show the output of parameters estimation in SAR Model using the weight matrix queen. The negative spatial parameter (λ) indicated that the neighbor places is not similar from each others, and in classical regression model all coefficient are significant while in SAR model where using the queen weight matrix all the coefficient are significant too.

Table (3-29) Estimation the parameters of SAR model by using queen matrix

Model	Weight Matrix-Queen (WMQ)			
	Coefficient	Std. Error	t-calculate	t-table
(Constant)	7.9273	0.08430	94.0284**	2.052
b1	-0.3151	0.0108	28.9657**	
b2	0.0603	0.0028	20.8099**	
b3	1.6510	0.08146	20.2659**	
(λ)	-0.0013			

The SAR model is estimated as:

$$\hat{y}_i = 7.9273 - 0.3151\, b_1 + 0.0603\, b_2 + 1.6510\, b_3 - 0.0013\, \lambda \text{……(3.12)}$$

Where λ is the average A.P in all neighboring countries, according to the weight matrix used, which is “ Queen” in this application.

3-11-2 Spatial Error Model (SEM)

3-11-2-1 SEM Model by using Rook Matrix

The table below shows the output of parameters estimation in SEM Model using the weight matrix rook. The positive spatial parameter (θ) indicated that the neighbor places is similar from each others, and in classical regression model all coefficient are significant while in SEM model where using the queen weight matrix all the coefficient are significant too.

Table (3-30) Estimation the parameters of SEM model by using rook matrix

Model	Weight Matrix-Rook (WMR)			
	Coefficient	Std. Error	t-calculate	t-table
(Constant)	8.1554	0.08632	94.4683**	2.052
b1	-0.2974	0.01113	26.6983**	
b2	0.0524	0.00296	17.6600**	
b3	1.4101	0.08342	16.9034**	
(θ)	0.3252			

The SEM model is estimated as:

$$\hat{y}_i = 8.1554 - 0.2974\ b_1 + 0.0524\ b_2 + 1.4101\ b_3 + 0.3252\ \theta \ldots\ldots(3.13)$$

Where θ is the average error of prediction in neighboring countries of station i, according to the weight matrix used, which is “Rook” in this application.

3-11-2-2 SEM Model by using Bishop Matrix

The table below shows the output of parameters estimation in SEM Model using the weight matrix bishop. The positive spatial parameter (θ) indicated that the neighbor places is similar from each other's, and in classical regression model all coefficient are significant while in SEM model where using the queen weight matrix all the coefficient are significant too.

Table (3-31) Estimation the parameters of SEM model by using bishop matrix

Model	Weight matrix-Bishop			
	Coefficient	Std. Error	t-calculate	t-table
(Constant)	8.0255	0.08592	93.3970**	2.052
b1	-0.3299	0.01108	29.7539**	
b2	0.0577	0.0029	19.5369**	
b3	1.5766	0.08303	18.9875**	
(θ)	0.1349			

The SEM model is estimated as:

$$\hat{y}_i = 8.0255 - 0.3299\ b_1 + 0.0577\ b_2 + 1.5766\ b_3 + 0.1349\ \theta \ldots (3.14)$$

Where θ is the average error of prediction in neighboring countries of station i, according to the weight matrix used ,which is "Bishop" in this application.

3-11-2-3 SEM Model by using Queen Matrix

The table below shows the output of parameters estimation in SEM Model using the weight matrix queen. The positive spatial parameter (θ) indicated that the neighbor places is similar from each others, and in classical regression model all coefficient are significant while in SEM model where using the queen weight matrix all the coefficient are significant too

Table (3-32) Estimation the parameters of SEM model by using queen matrix

Model	Weight Matrix-Queen (WMQ)			
	Coefficient	Std. Error	t-calculate	t- table
(Constant)	8.1781	0.0867	94.2957**	2.052
b1	-0.3149	0.01119	28.1393**	
b2	0.0518	0.00298	17.3775**	
b3	1.4449	0.08380	17.2410**	
(θ)	0.2754			

The SEM model is estimated as:

$$\hat{y}_i = 8.1781 - 0.3149\ b_1 + 0.0518\ b_2 + 1.4449\ b_3 + 0.2754\ \theta \ldots\ldots(3.15)$$

Where θ is the average error of prediction in neighboring countries of station i, according to the weight matrix used ,which is “Queen” in this application.

3-12 Tests for Finding the Best Model in Fuzzy Data

3-12-1 Lagrange Test for SAR (LM λ):

This test is used for finding spatial dependency in spatial error model

Table (3-33): Lagrange test for SAR Model

Test	Weight Matrix (W)		
	Rook	Bishop	Queen
LM λ	11.1811**	0.1351	2.9363
Robust LM	11.3219**	0.1334	2.9970

In W rook: The values of the two test (LM λ=11.1811, Robust LM λ=11.3219) are significant when we compared the value of the tests with the value of chi-square (1,α) degree freedom i.e $\chi^2_{(1,0.05)}=3.841$

In W bishop: The values of the two test (LM λ=0.135137, Robust LM λ=0.1334) when we compared the value of the tests with the value of chi-square (1,α) degree freedom i.e $\chi^2_{(1,0.05)}=3.841$ are not significant

In W queen: The values of the two test (LM λ=2.9363, Robust LM λ=2.9970) when we compared the value of the tests with the value of chi-square (1,α) degree freedom i.e $\chi^2_{(1,0.05)}=3.841$ are not significant

3-12-2 Lagrange Test for SEM (LM θ):

This test is for finding spatially dependency in spatial error model

Table (3-34) Lagrange test for SEM Model

Test	Weight Matrix (W)		
	Rook	Bishop	Queen
LM θ	1.621807	0.15951	1.51688
Robust LM θ	1.78037	0.158977	1.468681

In table (3-34), the values of the two test LM θ and robust LM θ according to the weight matrix rook, bishop and queen are not significant because the values of the two tests is less than the value of Chi-Square with degree of freedom $\chi^2_{(1,0.05)}$ = 3.841 and in it we can concluded to that the SAR is better than the SEM Model.

3-13 Comparison Criteria with Fuzzy Data:

After create a model by using three types of weight matrices matrix and test it. For find spatial dependency we must use same criteria for finding the best model

Table (3-35) Show the Comparison between the Models (SAR and SEM) using different criteria or measures

Criteria	SAR		
	Wrook	Wbishop	Wqueen
R^2_{adj}	0.5700	0.5416	0.5511
AIC_C	3.082785	3.090796	3.087871
RMSE	0.03141	0.0324	0.0321
MAPE	0.000001	0.00001	0.00001

Result and discussion:

1- R^2_{adj} This measure is based on the concept that how much variation in Y's stated by Syy is explained by SSreg.In this thesis the value of R^2_{adj} 0.5700,0.5416 and 0.5511 respectively of the variation of the response variable is explained by the model according to weight matrices (rook, bishop and queen) and the best value given in rook weight matrix.

2- AIC is a measure of the relative quality of statistical models for a given set of data. Given a collection of models for the data in AIC the best model is the smaller value because that in this thesis the best model is SAR model according to the weight matrices rook

3- RMSE is a measure use for differences between the values (sample and population) predicted by a model or an estimator and the values actually observed in where the RMSE smaller value is better and in this thesis the best model is SAR model according to the weight matrix rook

4-MAPE is a measure of prediction accuracy of a forecasting or estimation method in statistics, and the best model is SAR according rook weight matrix

When we compare between the models according to the some weight matrix by the criteria or measures in all measures accept to select SAR model with weight matrices rook .

Table(3-36):Comparison between the models using (Raw data) and (Fuzzy data)

Data	Models	Weight Matrix	R^2_{adj}	AIC	RMSE	MAPE
Raw Data	GLM	_	0.504	14.7703	0.2971	0.00084
	SAR	Rook	0.5468	3.9415	0.2903	0.000803
		Bishop	0.5266	3.9560	0.2967	0.000837
		Queen	0.5747	3.9206	0.2811	0.000748
Fuzzy Data	GLM	_	0.525	-104.657	0.0325	0.00001
	SAR	Rook	0.5700	3.0827	0.03141	0.000001
		Bishop	0.5416	3.090796	0.0324	0.00001
		Queen	0.5511	3.0878	0.0321	0.00001

<u>Result and discussion:</u>

1. Adjusted coefficient of determination R^2_{adj} measure is best in the SAR model according to the Queen weight matrices for the raw data.

2. RMSE is another measure use to find the best model and worked in sample is better or not. The spatial regression model SAR is the best model according to the weighted matrices rook for fuzzy data.

3. According to the AIC measure the best model is GLM for fuzzy data because the GLM is lowest when we compared with the AIC of the Model

4. MAPE is a another measure use to find the best model and find best estemation. The spatial regression model SAR is the best and good model according to the weighted matrices rook for fuzzy data

Finally. The best and good model surely in fuzzy data according (RMSE,MAPE) it is the best and good by rook weight matrix and the best neighboring is in rook role weight matrix and not bishop and queen.

Chapter Four

Conclusions and Recommendations

4 Conclusions and Recommendation

4-1 Conclusions

After reviewing all the traditional and modern methods of estimating parameters of the regression model and used Information Criteria to produce the best fitted Model for A.P, the following conclusions are drawn

1. The place that neighbor by role for rook or queen matrices will be significant in test for spatial dependency. In SAR model when use the rook and queen weight matrices the place of some stations of Kurdistan region are dependent spatially and the effect of air temperate , relative humidity and wind speed on response variable Atmospheric pressure are similar too.

2. The most appropriate model is SAR model for queen weighted matrix with un fuzzy data but in fuzzy data the best appropriate model is SAR model for rook weighted matrix.

3. The explanatory variable wind speed is not significant in classic regression and spatial regression for Raw data but the variables air temperate and relative humidity are significant and for fuzzy data all parameters are significant in both models classical and spatial regression

4. Spatial regression with fuzzy data is more efficient than spatial regression with un fuzzy data and the regression model with fuzzy data is better than regression model with un fuzzy data.

4-2 Recommendations

The researcher recommends the following points:

1.Use this models in any domain of life such as the cancer disease according to places or study the rank of university according to the place and decide which university is most efficient than the others .

2. Use another type of weight matrix instead of rook bishop and queen such as linear weight matrix for finding the neighbor places.

3. Calculate the parameters by use the panel model that relation between time and place when use the panel data the researcher can work in time and place at once.

References

References

***Books*:**

[1] Anselin, L.(1988) "***Spatial Econometrics: Methods and Models"*** Kluwer Academic, Dordrecht Anselin, L.: Under the hood Issues in the specification and interpretation of spatial regression models. Agricul. Econ.

[2] Anselin, L. (2001) "***Spatial Econometrics*"** in Baltagi B.H. (ed), A Companion to Theoretical Econometrics, Blackwell Publisher, Oxford

[3] Anselin, L. and Bera, A. K. (1998) "***Spatial Dependence in Linear Regression Model With An Introduction to Spatial Econometrics***'' In: UIIah A, Giles DEA (eds) Handbook of applied economic statisties. Marcel Dekker, New York

[4] Bickel, P.J.and Rosenblatt, M. (1973). **"*On Some Global Measures of the Deviation Density Function Estimates*"**. Annals of Mathematical Statistics, **New York**

[5**]** Bivand, R., E.J. Pebesma and V. Gomez-Rubio**. (**1997) ''***Applied Spatial Data Analysis with R. New York: Springer*'' New York**

[6**]** Burnham, K.P. and Anderson, D.R. (2002) ''***Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach Second Edition***" Springer-Verlag, New York

[7] Celik , M., Kazar, B. M., Shekhar, S. and Boley, D. (2006) ''***Parameter Estimation For the Spatial Autoregression Model: A Rigorous Approach***''This Work was Partially Supported by (AHPCRC). USA

[8] Emills, Josh. (2010) ‘‘***Spatial Econometrics***’’ Institute of Transportation Engineers. New York

[9] Freund, Rudolf J. and William, J.Wilson. (1997). ‘‘***Statistical Methods. Academic Press***’’ New York.

[10**]** Gan, Jiansheng. (2009) ‘‘***Spatial Combination Interpolation Model Based on Panal Data and Its Empirical Study***’’ School of Management, Fuzhou university (or Fujian institute of Economics and Management) Fuzhou 350002, China.

[11] Kleinbaum, David G. and Nizam, Azhar. (1998). ‘‘***Applied Regression Analysis and other Multivariable Methods***’’. **3rd ed**’ An International Thomson Publishing, USA.

[12] Lesage, James P. (2004) *‘‘**Maximum Likelihood Estimation of Spatial Regression Model***’’ University of Toledo.

[13] Like, Y. and Zongyi, Z. (2007) ‘‘***A Spatial eEconometric Analysis On the Relationship Between Power Consumption and Regional Economic Development***’’ School of Economics and Business Administration, Chongqing University, 400044, China.

[14] Lesage, James P. (1997) ‘‘***Regression Analysis of Spatial Data***’’ University of Toledo, the Journal Regional Analysis and Policy 27.

[15] Maddala,G.S. (2003). “***Introduction to Econometrics***”. 3rd ed, John Wiley and Sons, New York.

[16] Observation. Chtterjee, Samprit and Bertam, Price. (1997) ‘‘***Regression Analysis by Example***” John Wily and Sons, New York.

[17] Ord, J.K.(1975) ‘‘***Estimation Methods for Models of Spatial Interaction***” china

[18] PSeghouane, A.K. (2006). “***Multivariate regression model selection from small samples using Kullback's symmetric divergence***”. Signal processing, University of toledo

[19] White, H. (2001). “***A Hetroscedasticity –Consistent Covariance Matrix Estimator and a Direct Test for Hetroscedasticity***”**.**Econometrica New York

Thesis:

[20**]** Ali.O.Abdul with sawsan qasm hade**(**2014) **“*Spatial Regression Model Estemation for the Poverty Rates in the Districts of Iraq*”,** university of baghdad

Journal:

[21] Hmid.N. saed (2016) “***Estimate Spatial Dynamic Panel Data Model*** ” the journal in Qairo confrance :confrance name “***The role of Statistic in the Management of crises in the Arab World*** ”

[22**]** James P LeSage (2014) **“*What Regional Scientists Need to Know About Spatial Econometrics*”** journal name **“*matmatical symbol*”** look at the web sit**e** [http://journal.srsa.org/ojs/index.php/RRS/article/view/44.1.2]

Web Site:

[23] Abstatsre (1998) "***Applied Regression Analysis and other Multivariable Methods***". Abstatsre [http://Abstatsre.com] Accessed on, Mar 17, 2016

[24] Arei (2001) " ***Regression analyses***" Arei [http://Arei.revues.org] Accessed on Apr 24,2016

[25] Astatsref (2011) ***"Statistical Analysis Handbook"*** Astatsref [http://www.Astatsref.com/HTML/index.html?sar_models.htm] Accessed on Aug 3, 2015

[26] Apeople (2010) "***Applied Spatial Statistics in R, Section 6***" Apeople [http://www.Apeople.fas.harvard.edu/~zhukov/Spatial6.pdf] Accessed on Aug 23,2015

[27] Avhatis.techtarget (2006) ''***Fuzzy Logic''*** Avhatis.techtarget [http://Avhatis.techtarget.com/definition/fuzzy-logic] Accessed on Jun 1, 2016

[28] Ausers.cs.umn (2003) ''***Spatial Prediction''*** ausers.cs.umn [http://www-ausers.cs.umn.edu/~cpatel/spatial/PapersSummary.pdf] Accessed on Feb 2,2016

[29] B.samples.sainsburys (2002) ''***Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models***'' B.samples.sainsburys [http://B.samples.sainsburysebooks.co.uk/9781444317985_sample_379330.pdf] Accessed on May 8,2016

[30] B.economicas/coro/docencia(2007)‘‘ ***SpatialDependenceModels :EstimationandTesting***’’b.economicas/coro/docencia [https://www.uam.es/personal_pdi/B.economicas/coro/docencia/doctorado/spateconUPC/Slides/Session7_SpatialDepModel_Slides.pdf] Accessed on Jun11, 2016

[31] Bideas.repec (2002 **)“*Under the Hood Issues in the Specification and Interpretation of Spatial Regression Models*”** Bideas.repec [https://Bideas.repec.org/a/ags/iaaeaj/177794.html] Accessed on Feb 21,2016

[32] Bdehcholands (2008)‘‘***Modeling Sighted Regression***’’ Bdehcholands [http://www.Bdehcholands.org/docs/reports/Contractor%20Reports/Timber%20Potential%20Final%20Report%20(3)/Timber%20Potential%20Final%20Report%20-%20Maps%20Not%20Inserted%20(3).pdf] Accessed on Mar 2 2016

[33] sciencedirect (2004) ‘***‘Multivariate Spatial Regression Model’’*** Bsciencedirect[http://www.sciencedirect.com/science/article/pii/S0047259X0400034X] Accessed on Jun 22,2016

[34] Conlinelibrary (2005) ‘‘***Spatial Dependence in Regression Analysis*’’**Conlinelibrary[http://Conlinelibrary.wiley.com/doi/10.1111/j.14355597.1990.tb01204.x/abstract] Accessed on Apr 1,2016

[35]Eciteseerx.ist.psu (2007) ‘‘***Spatial Econometric Methods”*** eciteseerx.ist.psu[http://eciteseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.571.9223&rep=rep1&type=pdf] Accessed on Feb10,2016

[36] degruyter (2013). “***Use Of Spatial Autocorrelation to Build Redression ModelsTransactionprices’’***Tdegruyter

[http://www.degruyter.com/downloadpdf/j/remav.2013.21.issue-4/remav-2013-0038/remav-2013-0038.xml] Accessed on Jan 22, 2016

[37] Economics/documents (2004)"***Seminars and Advanced Lectures*** '' economics/documents**[https://www.york.ac.uk/media/economics/documents/seminars/201112/Elhorst_November2011.pdf 4848] Accessed on Apr 3,2016**

[38] Geodacenter (1997) **''*GeoDa Center's New Home*''** geodacenter [https://geodacenter.asu.edu/system/files/Anselin0902.pdf] [Of time and space: The contemporary relevance of the Chicago school. Social Forces, 75:1149–1182] Accessed on May 1,2016

[39] Gate (1988) ''***Spatial Econometrics: Methods and Models. Dordrecht, Germany:Kluwer Academic Publishers***''egate [https://www.gate.cnrs.fr/IMG/pdf/Lesage2014.pdf] Accessed on Dec 23,2015

[40] Ncdc.noaa (2005) ''***The Quality Control of Long-Term Climatological Data Using Objective Data Analysis***'' Bncdc.noaa [ftp://ftp.ncdc.noaa.gov/pub/data/papers/200686ams9.6ngfree.pdf] Accessed on May 12,2016

[41] Openaire (2007) ''***Misspecification in Linear Spatial Regression Models***''eopenaire[https://www.openaire.eu/search/person?personId=dedup_wf_001::3f79e7f4ad59ddf85b745672c3178f8] Accessed on Oct12,2015

[42] Onlinelibrary (2010)**"*Spatial Regression and Prediction of Water Quality*''**G.Honlinelibrary [http://onlinelibrary.wiley.com/doi/10.1111/j.14610248.2009.01422.x/abstract] Accessed on Feb 2,2016

[43] people.ucalgary (2006) ‘‘***Comparison of Ordinary Least Square Regression, Spatial Autoregression, and Geographically Weighted Regression for Modeling Forest Structural Attributes Using a Geographical Information System (GIS)/Remote Sensing (RS) Approach”***Dpeople.ucalgary [http://people.ucalgary.ca/~mcdermid/Docs/Theses/Shrestha_2006.pdf] Accessed on Aug 21,2015

[44] pjsor (2006) ‘‘***Application of Spatial Regression Models to Income Poverty Ratios in Middle Delta Contiguous Counties in Egypt*’’** epjsor [http://www.pjsor.com/index.php/pjsor/article/viewFile/272/298] Accessed on Dec 3,2015

[45] Researchgate (2002) “***Spatial Regression and Prediction of Water Quality in RiverNetworks***” iresearchgate [https://www.researchgate.net/publication/44694419_GISbased_Spatial_Regression_and_Prediction_of_Water_Quality_in_River_Networks_A_Case_Study_in_Iowa] Accessed on Feb 2,2016

[46]Researchgate(2011)**“*Spurious Spatial Regression and Heteroscedasticity*’’**Iresearchgate [https://www.researchgate.net/publication/233445052_Spurious_spatial_regression_and_heteroscedasticity] Accessed on May18,2016

[47] Scholarworks (2012)‘‘***Spatial Regression Methods Capture Prediction Uncertainty in Species Distribution Model Projections throughTime*’’**kscholarworks [http://scholarworks.umt.edu/cgi/viewcontent.cgi?article=1924&context=e] Accessed on Jun 7,2016

[48] Searchproject (2004) "***Economic Literature***" Searchproject [http://www.ub.edu/searchproject/wp-content/uploads/2013/01/WP-4.1.pdf] Accessed on Oct 12,2015

[49] Treg (1996) "***Time Series and Forecasting, Springer***" ftreg [www.treg.com/TimeSeries.htm] Accessed on Mar 7,2016

[50] Uabo-Niang, S., Rachdi, M., and Yao (2011) "***Regression Estimation for Spatial Functional Random Variables***" Uabo-Niang, S., Rachdi, M., and Yao [http:// Uabo-Niang, S., Rachdi, M., and Yao, A.-F. (2011). Kernel regression estimation for spatial functional random variables. Far East Journal of Theoretical Statistics] Accessed on Jan 22, 2016

[51] Unredlands (2015)"***Spatial and Regressin Analysis of Social Media***" unredlands[http://www.unredlands.edu/globalassets/depts/schoolbusiness/gisab/workshops-conferences/james-pick-et-al.-spatial-and-regression-analysis-of-social-media_pre-icis-workshop-2015_12-3-15.pdf] Accessed on Feb 22,2016

[52] Udegruyter (2008) " ***The Drunkard's Walk: How Randomness Rules OurLivesinRegression***" udegruyter [http://www.Udegruyter.com/downloadpdf/j/remav.2013.21.issue-4/remav-2013-0038/remav-2013-0038.xml] Accessed on May 1,2016

[53]V.Mathworld (2012) ''***Least Square Estemation''*** V.mathworld [http://v.mathworld.wolfram.com/LeastSquaresFitting.html] Accessed on May 2,2016

[54] V.oltrr.arizona (2013) " ***Multiple Linear Regression''*** V.oltrr.arizona [http://www.v.oltrr.arizona.edu/~dmeko/notes_11.pdf] Accessed on Jan 10, 2016

[55]Vstat.washington (2013) "***Sufficiency and Unbiased Estimation***" vstat.washington[https://www.vstat.washington.edu/jaw/COURSES/580s/581/LECTNOTES/ch13.pdf] Accessed on Apr 3,2016

[56]Wtattrek (2016) "***Estimation in Statistic***" Wtattrek [http://Wtattrek.com/estimation/estimationinstatistics.aspx?Tutorial=AP] Accessed on Oct 18,2015

[57] Xidergisi.istanbul.edu (1979). "***A Simple Test for Heteroskedasticity and Random Coefficient Variation Brunch Pagan***". xidergisi.istanbul.edu [xidergisi.istanbul.edu.tr/sayi8/iueis8m2.pdf] Accessed on May22,2016

[58]X.math.nsysu (2001) "***Durbin-Watson Test***"x.math.nsysu [http://www.x.math.nsysu.edu.tw/~lomn/homepage/class/92/DurbinWatsonTest.pdf] Accessed on Jan 12,2016

[59]X.Multiple+Regression(1999). "***MultipleRegression***"x.Multiple+Regression [http://www.x.Multiple+Regression reference.com] Accessed on May 2 ,2016

[60] Xaculty.arts.ubc (2014) "***A Test of Normality***" Xfaculty.arts.ubc [http://Xaculty.arts.ubc.ca/dwhistler/325ClassNotes/chapNorTest.pdf] Accessed on Mar 2,2016

[61]Ymsdn.microso(2016) "***Spatial Data***" Ymsdn.microso [https://ymsdn.microsoft.com/en-us/library/bb964737.aspx] Accessed on Oct 12,2015

[62] Ypra.ub.uni-muenchen (2008) " ***Quality of Life in the Regions: An Exploratory Spatial Data Analysis for West German Labor Markets***" Ypra.ub.uni-muenchen[https:// ypra.ub.uni-muenchen] Accessed on Oct 2,2015

[63] Zacweb.knowlton.ohio-state (2011) " ***Some Notes on Spatial Statistics and Spatial Econometrics***'' Zacweb.knowlton.ohio-state [zacweb.knowlton.ohio-state.edu/pviton/courses2/crp8703/**spatial**.pdf] Accessed on Nov18,2015

[64] Z.spatial-econometrics. (1998) "***Spatial Econometrics***" z.spatial-econometrics. [http://www.z.spatial-econometrics.com/html/wbook.pdf] Accessed on Jan 18,2016

[65] Z**.**Fuzzy Logic with Engineering Applications (2004) **"** ***Fuzzy Logic with Engineering Applications***" z.Fuzzy Logic with Engineering Applications [http://Z.Fuzzy Logic with Engineering Applications: Timothy J. Ross] Accessed on Mar 21,2016

[66]Zolytech.univsavoie(1967)"***LFuzzySets***".zolytech.univsavoie [http://www.zolytech.univsavoie.fr/fileadmin/polytech_autres_sites/sites/listic/busefal/Papers/85.zip/85_06.pdf] Accessed on Feb 2,2016

[67]Zoxresearch (2014) ''***Gaussian Function***'' zoxresearch [http://zoxresearch.ijcaonline.org/icctac2015/number1/icctac2006.pdf] Accessed on Oct 2 ,2015

[68] Z.Notes-on-Spatial-Econometric-Models 2012 ‘‘***Notes on Spatial Econometric Models March***” z.Notes-on-Spatial-Econometric-Models [www.do-cu-cu.com/z.Notes-on-Spatial-Econometric-Models-Ohio-State.pdf] Accessed on Oct 2,2015

Arabic References:

69.داود، جمعة محمد (2012)" ***أسس التحليل المكاني في إطار نظم المعلومات الجغرافية GIS***" مكة المكرمة، المملكة العربية السعودية

70.كاظم ،اموري هادي (1988)." ***طرق القياس الاقتصادي*** ". مطبعة التعليم العالي والبحث العلمي جامعة بغداد,

71. كاظم ،اموري هادي و مسلم، باسم شليبة(2002) " ***القياس الاقتصادي المتقدم النظرية والتطبيق*** " مطبعة الطيف, بغداد،.

Appendix B: Matlab Code

Appendix B1 : Matlab Code for Find Spatial Parametr and Concentrated Likelihood Function For SAR Model

```
n = 27;
eO =[ input eO ];
eL = [ input eL];
eOt = eO';
eLt = eL';
I27 = eye(27);
W =[input weight matrix];
x=-1:0.0001:1;
LC=x;
rho=-0.9999;
for j=1:20000
   if j<10000
.
.
.
.

         LCmax=LC(1);
      end
      if LC(j)>LCmax
         rho=d;
         LCmax=LC(j);
      end
end

   plot(x,LC)
   grid on;
xlabel(' \rho');
ylabel('ln L( \rho)');
'rho = '; rho;'ln Lc(rho) = '; Lcmax
```

Appendix B2 : Matlab Code for Find Spatial Parametr and Concentrated Likelihood Function For SEM Model

```
n = 27;
eO =[ input eO ];
eL = [ input eL];
eOt = eO';
eLt = eL';
I27 = eye(27);
W =[input weight matrix];
x=-1:0.0001:1;
LC=x;
rho=-0.9999;
for j=1:20000
    if j<10000
.
.
.
.
.           LCmax=LC(1);
        end
        if LC(j)>LCmax
            rho=d;
            LCmax=LC(j);
        end
end

    plot(x,LC)
    grid on;
xlabel(' \rho');
ylabel('ln L( \rho)');
'rho = '; rho;'ln Lc(rho) = '; Lcmax
```

Appendix B3:Table show the value of(e_o,e_L)

Raw Data				Fuzzy Data			
e_L rook	e_L bishop	e_L queen	e_o	e_L rook	e_L bishop	e_L queen	e_o
3.7504	7.5828	2.5046	-0.1550	2.0683	8.3891	1.5435	0.0282
-5.8695	8.0950	2.5894	0.0662	-7.2512	8.4004	1.6593	0.0300
3.5044	-1.9455	1.9342	0.1885	2.2171	-1.0430	1.7013	0.0185
2.6281	-0.3321	2.9969	-0.2175	2.3771	-1.1045	1.7944	0.0125
5.5068	-2.0433	4.2896	-0.2756	2.3831	-1.1583	1.6884	0.0150
-7.4810	-0.0982	-7.2974	0.0725	-7.1813	-1.0949	-7.7442	0.0161
2.3952	-0.9076	2.0547	-0.2620	2.1844	-1.1500	1.6206	-0.0137
-7.8235	-1.7876	-9.0468	-0.4244	-7.0334	-1.2299	-7.6688	-0.0114
-3.7412	-2.6168	-5.6015	-0.2343	-6.7371	-1.2586	-7.3875	0.0002
4.3759	-0.6928	4.3243	0.0034	2.5001	-1.1833	2.0243	0.0012
3.2053	-1.4286	2.6233	-0.1972	2.4707	-1.2769	1.9285	-0.0097
-4.9485	-1.8503	-5.8968	-0.3035	-7.0303	-1.1962	-7.5226	0.0030
1.2508	-0.9377	0.7956	0.0179	2.0621	-1.0771	1.6864	0.0269
4.2403	-3.0836	2.0147	0.1058	2.5377	-1.1905	2.0286	0.0083
1.9213	-0.1651	2.1677	-0.0931	2.6655	-1.0865	2.3660	0.0076
1.4159	-1.5610	0.4941	0.3626	2.8454	-1.2018	2.4535	-0.0328
2.4053	-0.6586	2.4320	0.3302	3.2033	-1.1976	2.7450	-0.0359
1.1986	-0.9270	0.8457	0.2141	3.7113	-1.2832	3.0535	-0.0398
4.5513	-2.6476	2.5826	-0.1115	4.7446	-1.4142	3.8054	-0.0257
3.0580	-2.4873	1.2958	0.6324	5.3728	-1.2867	4.7607	-0.0082
-0.4243	0.1837	0.7158	0.3433	-3.1023	-1.0796	-3.4220	-0.0244
-3.4313	-1.3940	-3.9849	-0.4633	-3.0610	-1.1346	-3.3413	-0.0205
4.0842	8.3502	4.0385	0.0473	5.8897	8.4524	6.1998	-0.0466
-5.3124	-0.3376	-4.8839	-0.3829	-3.0745	-0.6554	-2.5444	-0.0485
0.1141	-0.5876	0.3338	0.3700	-1.8477	-0.7449	-1.6682	0.0752
-3.9926	-0.3024	-3.4705	0.1861	-2.2946	-0.7782	-1.5018	0.0673

Appendixes

الملـــخــص

الانحدار المكاني(الحيزي) طريقة لحساب الاعتمادية بين المشاهدات, والتي تظهر عندما يكون المشاهدات مشارك في فضاء بنقطة أو المسافة.اهتمام تاثير المكان او العامل الحيّزي في تحليل الظاهرة يؤدي الى ايجاد معلومات مهمة بدلاً عن الزمن .لذا يجب ايجاد النماذج الرياضية التي تسمح بتضمين عامل المكان والتي هي نماذج الانحدار الحيزي التي توضح تأثير المتغيرات التوضيحية على المتغير الاستجابة في ظل وجود التأثيرات الحيزية للمواقع المتجاورة.بالنسبة الي قانون مصوفات الوزن .في بحثنا يتم دراسة تاثير متغيرات التوضيحية(نسبة الرطوبة ,درجة حرارة, سرعة الرياح)علي المتغير الاستجابة ضغط الجوي .

البيانات جمعت من ٢٧ مواقع أو المحطات في اقليم الكردستان(السليمانية ,أربيل و دهوك) و نماذج الانحدار الحيزي يستخدم لايجاد تأثير المواقع المجاورة.وتم استخدام نماذج الأنحدار الذاتي الحيزي (SAR) ,وأنموذج الخطأ الحيزي (SEM) والتي تضم كل منها عامل الحيزية، واستخدمت الانموذج الخطي العام (GLM) و تم استخدام معايير المقارنة مثل :الجذرالتربيعي لمتوسط مربعات الخطأ (RMSE) ومتوسط النسبة المطلقة للخطأ(MAPE) ، ومعامل التحديد المعدل (R^2_{adj}) ,ومعيار أكاكي (. AIC) لاختيار أفضل نموذج و لايجاد الاعتماد الحيزي تم استخدام اختبارات موران واستخدام اختبار لاكرانج لايجاد افضل نموذج بين (SARوSEM) مع ثلاثة انواع من المصوفات الموزونة rook,bishopو queen

ومن اهم الاستنتاجات بعد أجراء التحليل في الجانب التطبيقي ، كانت نتائج تحليل نماذج الانحدار الحيزي هي افضل من نتائج تحليل أنموذج الانحدار الخطي العام . ان معلمات نموذج الانحدار الذاتي الحيزي (SAR) باستخدام مصفوفتين queen و rook معنوي. وعند تحويل البيانات الي بيانات المضببة وتطبيق النماذج SEM, SAR وGLM و عند مقارنة بين هذه النماذج و بأعتماد علي مجموعة من معايير او مقايس مثل R^2_{adj},,RMSE,MAPE,AIC_C و ظهر نتائج ان تحويل البيانات الي بيانات المضبب

يكون افضل من البيانات الخام وان الانحدار الحيزي مع البيانات المضبب أفضل من الانحدار الحيزي بدون التحويل و النموذج المناسب و الافضل هو النموذج الانحدار الذاتي الحيزي SAR باستخدام مصوفة الوزون rook .وأخيرا نموذج مناسب التي ظهر في تحليل من الجانب التطبيقي هو

النموذج الانحدار الذا تي الحيزي باعتماد علي مصوفة الوزن queen من البيانات الخام كالاتي :-

$$\hat{y}_i = 6.4880 + 0.1096\ A.T + 2.0739 R.H + 0.0240\ \lambda$$

النموذج الانحدار الذا تي الحيزي باعتماد علي مصوفة الوزن rook من البيانات المضبب كالاتي :-

$$\hat{y}_i = 7.8610 - 0.3162 W.S + 0.0632\ A.T + 1.6837\ R.H - 0.0019\ \lambda$$

المقارنة بين تقنية الا نحدار الكلاسيكي و الا نحدار الحيزي باستخدام المنطق المضبب

رسالة مقدمة

الى مجلس كلية التجارة في جامعة السليمانية كجزء من متطلبات نيل درجة

الماجستير في علوم الإحصاء

من قبل الطالبة

شه م ازاد رحيم

بإشراف

الاستاذ المساعد

د.محمد محمود فقي حسين

٢٠١٦م ٢٧١٦ك ١٤٣٧ ه

پــوختــه

لێژبوونەوەی شوێنی ڕێگەیەکە بۆ ئەژمارکردنی پشبەستن لە نێوان بەشداربوان وەکاتێك بوونی دەبێت کە ئەم کارئەکتەرانە هاوبەش بن بەشێوەی خاڵ یان دووری .گرنگی دان بە کاریگەری شوێن یان پارامیتەری تایبەت بە شوێن لە شیکردنەوەی دیاردەیەك دەبێتە هۆی دەست کەوتنی زانیاری گرنگ لەجیاتی کات . لەبەر ئەم هۆکارە پێویستە مۆدێلێکی بیرکاریانە بدۆزرێتەوە کە پارامیتەری شوێنی تیا بەکار هاتبێ کە ئەمانەش لێژبوونەوەکانن بەپێی شوێن کە لەم مۆدێلانەدا روون کردنەوە ئەدرێ لەسەر کاریگەری گۆڕاوەکانی ڕوون کردنەوە (سەربەست) لەسەر گۆڕاوەکانی پشت بەست بەبونی کاریگەری ئەو شوێنانەی دراوسێی یەکن بەپێی یاسای ماتریکسی هاوسەنگە .لەم توێژینەوەیەدا سێ گۆڕاوی سەربەست بەکار هاتووە کە ئەوانیش (ڕێژەی شێ, پلەی گەرمی, خێرای با) لە بەرامبەر گۆڕاوی پشت بەست کە بریتیە(پاڵە پەستۆی هەوا) .

وە داتاکان کۆکراوەتەوە لە ٢٧ شوێن یان وێستگە لە هەرێمی کوردستان (سلێمانی , هەولێر و دهۆك)وە لێژبوونەوەکانی شوێن بەکاردێت بۆ دۆزینەوەی کاریگەری ئەو شوێنانەی کە دراوسێن. لێژبوونەوەی خۆیی شوێن (SAR) وە لێژبوونەوەی هەڵەی شوێن (SEM) کەهەردوکیان پارامیتەری شۆنیان تێدایە وەلێژبوونەوەی گشتی (GLM) .وە بەکارهێنانی هەندێك پێوەری بەراورد (R^2_{adj},,RMSE,MAPE,AIC_C) بۆ دیاریکردنی باشترین مۆدێل. وە بۆدۆزینەوەی کاریگەری شوێن پشکنینی مۆران بەکارهاتووە وە بۆ دیاریکردنی باشترین مۆدیل لە نێوان (SAR,SEM) پشکنینی لاکرانج بەکارهاتووە لەگەڵ سێ جۆر لە ماتریکسی هاوسەنگ bishop rook, و queen

وە لەگرنگترین ئەو ئەنجامانەی کەلە بەشی پراکتیکیدا ئەوەیە کە لێژبوونەوە بە پێی شوێن باشترە وەك لە لیزبوونەوەی گشتی . لێژبوونەوەی کاتی خۆیی (SAR) بۆ ماتریکسەکانی rook و queen قبووڵ دەکرێت وە لە وەرگێڕانی داتا کەدا بۆ لۆجیکی تەماوی وە جێبەجێکردنی مۆدێلەکانی GLM,SARو SEM وە بەراورد کردن لەنێوانیاندا بە هەندێك پێوەری (R^2_{adj},,RMSE,MAPE,AIC_C) وە ئەنجامەکان

پشانی دەدەن کە وەرگێڕانی داتا کە بۆ فەزی(تەماوی) باشترە لە داتاکە بەبێ فەزی(تەماوی) وە لیژبوونەوەی شوێنی بە فەزی (تەماوی) داتا باشترە وەك لە لیژبوونەوەی شوێنی بەبێ فەزی (تەماوی) وە باشترین و گونجاو ترین مۆدیل بریتی یە لە SAR بە بەکارهێنانی ماتریکسی هاوسەنگی rook بە پێی داتای فەزی(تەماوی) .لە ئە نجامدا دەردەکەوێت باشترین و گونجاوترین مۆدیل بە پێی پشکنینەکانی بەشی پراکتیکی بریتییە لە

مۆدێلی لیژبوونەوەی خۆیی شوێنی SAR بە پێ ی ماتریکسی هاوسەنگی queen لە گەڵ داتای بنەڕەتدا :ی

$$\hat{y}_i = 6.4880 + 0.1096\ A.T + 2.0739 R.H + 0.0240\ \lambda$$

مۆدێلی لیژبوونەوەی خۆیی شوێنی SAR بە پێ ی ماتریکسی هاوسەنگی rook لە گەڵ داتای فەزی (تەماوی)دا :

$$\hat{y}_i = 7.8610 - 0.3162 W.S + 0.0632\ A.T + 1.6837\ R.H - 0.0019\ \lambda$$

بەراورد كردن لە نێوان تەكنيكى ليژبوونەوەى كلاسيكى و ليژبوونەوە بە پێى شوێن بە بەكارهێنانى لۆجيكى فەزى (تەماوى)

ئەم نامەيە پێشكەشە بە

ئە نجومەنى كۆلێجى بازرگانى زانكۆى سليمانى وەك بەشێك لە پێداويستيەكانى بە دەست هێنانى پلەى ماستەر لە زانستى ئامار

لە لايەن

شەم ئازاد رەحيم

بەسەرپەرشتى

پرۆفيسۆرى ياريدەدەر

د. محمد محمود فقى حسێن

١٤٣٧ ه ٢٧١٦ك ٢٠١٦م