Forecasting Foreign Direct Investment in South Africa using Non - Parametric Quantile Regression Models

By

Nyawedzeni Netshivhazwaulu
11583403

Dissertation submitted in fulfillment of the requirements for the Master of Science degree in Statistics

School of Mathematical and Natural Sciences
Department of Statistics
University of Venda
Thohoyandou, Limpopo
South Africa

Supervisor: **Dr C Sigauke**

Co-Supervisor: **Dr A Bere**

November 2018

Declaration

I, Nyawedzeni Netshivhazwaulu declare that the dissertation titled "Forecasting Foreign Direct Investment in South Africa using Non - Parametric Quantile Regression Models" for the Master of Science degree in Statistics at the University of Venda hereby submitted by me has not been previously submitted for a degree at this or any other university. I further declare that it is my own work in design and in execution and that all referenced material contained therein have been dully acknowledged.

Signature (Student):..

Date:..

Abstract

Foreign direct investment plays an important role in the economic growth process in the host country, since foreign direct investment is considered as a vehicle transferring new ideas, capital, superior technology and skills from developed country to developing country. Non-parametric quantile regression is used in this study to estimate the relationship between foreign direct investment and the factors influencing it in South Africa, using the data for the period 1996 to 2015. The variables are selected using the least absolute shrinkage and selection operator technique, and all the variables were selected to be in the models. The developed non-parametric quantile regression models were used for forecasting the future inflow of foreign direct investment in South Africa. The forecast evaluation was done for all models and the laplace radial basis kernel, ANOVA radial basis kernel and linear quantile regression averaging were selected as the three best models based on the accuracy measures (mean absolute percentage error, root mean square error and mean absolute error). The best set of forecast was selected based on the prediction interval coverage probability, Prediction interval normalized average deviation and prediction interval normalized average width. The results showed that linear quantile regression averaging is the best model to predict foreign direct investment since it had 100% coverage of the predictions. Linear quantile regression averaging was also confirmed to be the best model under the forecast error distribution. One of the contributions of this study was to bring the accurate foreign direct investment forecast results that can help policy makers to come up with good policies and suitable strategic plans to promote foreign direct investment inflows into South Africa.

Keywords: Foreign direct investment, least absolute shrinkage and selection operator, Non-parametric quantile regression, Local linear kernel.

Dedication

I am dedicating this dissertation to my late mother Mmboneni Florence Netshivhazwaulu who left a void never to be filled in our lives. I will make sure your memory lies on as long as I shall live. I love you and miss you beyond words.

Acknowledgement

I would like to express my deepest appreciation to all those who provided me the possibility to complete this study. A special gratitude goes to my supervisor, Dr. C. Sigauke, whose contribution in stimulating suggestions and encouragement, helped me to coordinate my project especially in writing this dissertation. I would also like to acknowledge Dr. A. Bere as the co - supervisor of this dissertation, and I am grateful to his valuable comments.

Finally, I must express my very great gratitude to my classmates (Ravele T, Mpfumali P. and Nevhungoni T.P), my family and to my husband (Mr. M.D. Nengovhela) for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this dissertation. This accomplishment would not have been possible without them.

Table of contents

List of Figures

List of Tables

Abbreviations

AIC	Akaike Information Criterion
AICC	Akaike Information Criterion Corrected
ANOVA	Analysis of Variance
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criterion
CPI	Consumer Price Index
CV	Cross Validation
Exp	Exports
FDI	Foreign Direct Investment
GDP	Gross Domestic Product
GFCF	Gross Fixed Capital Formation
ICRG	International Country Risk Guide
IVQR	Instrumental Variable Quantile Regression
LARS	The least angle regression
LASSO	Least Absolute Shrinkage and Selection Operator
LL	Lower Limit
LP	Labour Productivity
LPI	Labour Productivity Index
LQR	Linear Quantlie Regression

LQRA	Linear Quantile Regression Averaging
MAE	Mean Absolute Error
MAPE	Mean Absolute percentage Error
MLE	Maximum Likelihood Estimation
MSE	Mean of Squares Error
NQR	Non-parametric Quantile regression
OLS	Ordinary Least Square
OLSR	Ordinary Least Square Regression
OT	Oppeness to Trade
PS	Political Stability
PI	Prediction Interval
PICP	Prediction Interval Coverage Probability
PINAD	Prediction Interval Normal Average Deviation
PINAW	Prediction Interval Normalized average Width
PIW	Prediction Interval Width
QR	Quantile Regression
RMSE	Root Mean Square Error
SA	South Africa
SARB	South African Reserve Bank
UL	Upper Limit

Chapter 1

Introduction

Foreign Direct Investment (FDI) has grown everywhere throughout the world at a remarkable rate since the early 1980s and the world market for it has become more ambitious. The need for FDI came as a result of shortage in domestic funding sources to finance development projects in developing countries. According to Pugel (1983) FDI is the process whereby residents of one country acquire ownership of foreign assets for the purpose of controlling the production, distribution and other activities of a firm in another country. It is well documented in the growth literature that FDI plays an important role in the economic growth process in host countries, since FDI is considered as a vehicle to transfer new ideas, capital, superior technology and new skills from developed countries to developing countries. FDI therefore helps to reduce unemployment in host countries.

Whether or not FDI has a positive impact on the economy of a country depends on several conditional impact factors, such as government regulations and policies on investment, availability of raw materials and trade facilitation instruments, availability of appropriate human resources, economic growth of a country, political stability and security of a country, cost and ease of doing business (Markusen and Venables, 1999). South Africa (SA) has been pursuing a set of economic development strategies to promote national economic development. The driving force behind the strategies is to reduce the most challenging problems facing the SA economy; employment creation, economic growth, strengthening the program of radical economic transformation and reducing poverty (Bermo, 2009).

The transition to a multi racial democracy in 1994 presented difficult political, social, and economic challenges, since the new government was forced to concentrate on establishment of a credible and prudent fiscal stance, efforts to reduce inflation, and the needed reunification of the dual exchange rate system (Brooks et al., 2000). SA's interesting achievements in overcoming these challenges have been widely recognized and the policy perseverance shown over the past years has yielded tangible macro stabilization success and enhanced policy legitimacy.

1.1 Problem statement

The economic growth and employment challenges facing SA are cause for concern, FDI flows disappointing and this leaves SA at a disadvantage within an increasing competitive global environment. Many statistical models on determining the factors that explain the variation in FDI and the impact of FDI on economic growth have been built. Some of the researchers went on further to make forecasts using those best fit models.

The main objective is to come up with the best suitable strategic plans to promote FDI. This study uses quantile regression modelling to determine the important determinants of FDI in SA and investigates the relationship between FDI and those determinants such as economic growth (GDP), Consumer price index (CPI), political stability (PS), labour productivity index (LPI), gross fiscal capital formation (GFCF), etc.

1.2 Purpose of the study

SA has implemented various economic reforms to restructure the economy in order to archive higher economic growth and development. The proposed non - parametric quantile regression models will be used to forecast the FDI inflows for the years to come. The forecasts obtained from this study will help the government of SA to increase its FDI inflows in order to archive economic growth. The government can also know the intervention measures that can be launched to enhance the country's attractiveness for FDI.

1.3 Aim

The main aim of this study is to develop non-parametric quantile regression models for forecasting SA FDI.

1.4 Objectives

The main objectives of this study are to:

- identify the major determinants of Foreign Direct Investment in SA, and investigate on the nature of the relationship between each of the determinants and FDI,

- forecast the volume of FDI and evaluate the accuracy of the forecasts,
- to suggest measures to increase the inflow of FDI in SA.

1.5 Significance of the study

The results of this study will help formulate financial and investment policies, strategies and decisions to attract more FDI to South Africa.

1.6 Scope of the dissertation

Non-parametric quantile regression (NQR) will be used to predict the future inflow of FDI in SA, using the FDI data from SA Reserve Bank. This study will use the FDI measured from 1996 to 2015. Shrinkage methods such as least absolute shrinkage and selection operator (Lasso) will be used for variable selection and the developed NQR models will be used for forecasting the future inflow of FDI. The mathematical programming techniques such as linear programming and interior point algorithms method will be used to estimate the parameters. The forecast evaluation will be done for all the models that will be build in this study. The best models will be chosen based on the accuracy measures, root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The R packages, such as "quantreg" by Koenker and Park (1996), "kernlab" and "quantreg.nonpar" by Lipsitz et al. (2015) will be used.

1.7 Structure of the dissertation

Following the introduction, the study presents a literature review (chapter 2) with relevant theories and empirical studies, subsequently, the research methodology (Chapter 3) used in the study is presented. This is followed by a discussion on results and findings (Chapter 4), ending with a concluding Chapter 5.

Chapter 2

Literature review

2.1 Introduction

In the middle of the 1980s, the world foreign direct investment (FDI) started to increase sharply and has received a lot of attention (Agrawal and Khan, 2011). For so many years different models for FDI inflow were developed, applied, reviewed and published. This chapter provides an overview of FDI, factors affecting it as well as the summaries of some studies that used the proposed methodology (non-parametric quantile regression) to forecast FDI inflow in different countries.

2.2 An overview of FDI

FDI is one of the most dynamic international resources flows to developing countries. Görg and Greenaway (2004) highlighted that FDI can positively affect growth and development by complementing domestic investment in transfer of knowledge and technology, and facilitating trade. Each country needs FDI to leap itself to sustainable growth levels and fill the saving and foreign exchange gaps.

Accurate FDI forecasting results can play an important role in assisting policy makers and decision making to come up with good policies and suitable strategic plans to promote FDI. Accurate and efficient forecasts are necessary as they help in preventing things, like state capture, which does not benefit the recipient country, loss of tax and revenue, unemployment, destroying of local manufacturing industries, exploitation of raw materials, etc.

2.3 Factors influencing FDI

Previous studies has looked at the correlation of FDI with several determinants. Some such determinants that might be thought to have a connection to FDI flows are political stability and security of a country, market size and growth potential, government regulations and policies on investment, economic stability, clustering effect, openness to trade (OT), availability of appropriate human resources, wage rates, availability of raw materials, commodities, etc.

Larger host countries' markets have been associated with higher FDI due to higher potential demand and lower cost. Resmini (2000), finds that countries in central and eastern Europe with larger populations tend to attract more FDI. Bevan and Eastrin (2000) suggest that large economies attract more FDI. Researchers have identified clustering effect. By clustering with other firms, new investors benefit from positive spillovers from existing investors in the host country (Wheeler and Mody, 1992).

Political stability and security of a country determine the amount of FDI that a country attracts. It has been found that political stability is one of the key determinants of potential FDI inflows. Schneider and Frey (1985) argue that political instability significantly affects FDI inflows.

A weaker exchange rate is expected to increase FDI inflow, as firms take advantage of low prices in host markets to purchase facilities. Klimek (2011) shows that exchange rate depreciation in host countries tends to increase FDI inflow. Policies on OT might produce a significant impact in attracting FDI. OT could affect FDI inflows positively or negatively.

Chakrabarti (2001), finds a positive relationship between OT and FDI. Institutional quality is a likely determinant of FDI, for developing countries, since the good governance is associated with higher economic growth and should attract more FDI inflows. While the poor institutions that enable corruption tend to add to investment cost and reduce profits, it is not easy to measure institutional factors. Policies that are favourable to both local and foreign investors determine the amount of FDI that a country receives and more raw materials attract more FDI to a country.

2.4 Modelling FDI using quantile regression

Different models for FDI have been developed and used to forecast the future inflow of FDI in different countries with the aim of finding the model that will produce accurate results.

Quantile regression (QR) models the conditional relationship between a set of predictor variables at specific quantiles of the response variable (Koenker and Hallock, 2001). QR models are used in forecasting in many application areas. In a study by Chunying (2011) a QR model was applied to determine the impact of FDI on technology innovation from 1987 to 2009 in China. The main aim was to analyse the correlation between FDI and technology innovation. They found that FDI exerts significant positive effects only for low technological innovation level at different locations of conditional distribution.

Jyun-Yi et al. (2008) used the instrumental variable quantile regression (IVQR) on the country level data to examine the role of absorptive capacity in contributing to the growth effects of FDI at different lower quantiles of the income distribution. They found that the IVQR estimates positive growth effects of FDI at higher quantiles of the distribution, while the estimates of FDI for QR are insignificant and small in magnitude across different quantiles.

A more recent study is that of Paniagua et al. (2015), who developed a QR method for bilateral FDI panel data. The aim of the study was to estimate the individual firms' effect on FDI flows among 161 countries from 2003 to 2012. They found that FDI's determinants vary across quantiles, where the effect of individual projects on FDI flow increases in the upper quantiles.

In another study by Girma et al. (2005), conditional QR was used to find the different effects of FDI on establishments located at different quantiles of productivity distribution. The purpose of the study was to find the role of absorptive capacity in determining whether or not domestic firms benefit from productivity spillovers from FDI using level data for the United Kingdom. They found that both absorptive capacity and distance matter for productivity spillover benefits, while there is substantial heterogeneity in results across sectors and quantiles.

2.5 Conclusions from literature

One of the contributions the dissertation aims to make is to add the use of the proposed methodology to the existing literature focusing on its application to modelling FDI inflows in South Africa.

Chapter 3

Research Methodology

3.1 Introduction

This chapter discusses the models and the methods used for analysis. The main models will be built under the non-parametric quantile regression, where the focus is on the locally polynomial quantile regression and the penalty smoothing. A comparative analysis of the developed models will be done with two benchmark models, the ordinary least squares (OLS) and linear quantile regression models.

3.2 Ordinary least squares regression

Ordinary least squares regression will be used as one of the benchmark models. Ordinary least squares regression is a statistical method of analysis that estimates the relationship between one or more independent variables and the mean of a dependent variable. The method estimates the relationship by minimizing the sum of the squares of the difference between the observed and predicted values of the dependent variable (Craven and Islam, 2011). In this study the multiple regression will be used. Below, the multiple linear regression model is discussed.

Suppose the data consists of n observations. Each observation i includes a scalar response y_i and a vector of values of p predictors x_{ij} for $j = 0,1,...,p$. In a linear regression model the dependent variable is a linear function of the regressors with x_0 corresponding to the constant term:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i \tag{3.2.1}$$

or in a vector form:

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i, \tag{3.2.2}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters; ε_i's are unobserved scalar random variables (errors) which account for the discrepancy between the observed responses y_i and the predicted outcomes $\mathbf{X}_i^T \boldsymbol{\beta}$; and T denotes matrix transpose.

3.3 Quantile regression

Quantile regression (QR) was first introduced by Koenker and Bassett (1978), as a consolidated statistical methodology for estimating models of conditional quantile functions. QR is the extension of the median regression. The method of OLSR estimates the relationship between one or more predictor variables and the conditional mean of the response variable, while QR models the relationship between predictor variables and the conditional quantiles of response variable rather than just the conditional median of the response variable. A QR model gives a more comprehensive picture of the effect of the predictor variables on the response variable and actually contains more information than can be presented in OLSR (McMillen, 2013).

QR is flexible in modelling data with heterogeneous conditional distribution and richer in characterization and description of the data, as it can show different effects of predictor variables on the response variable across the whole spectrum of the response variable. QR estimates are robust to non - normal errors but the main interest of QR goes beyond that.

The QR model adapts readily to non-parametric estimation procedure, as the non-parametric estimation turns out to be more easy to implement in a QR framework and the results can be presented in a quite straight forward way in a set of graphs (Mcmillen, 2013).

Quantile regression model specification

The τ - quantile of a random variable Y, with cumulative distribution function $F_{Y|\mathbf{X}}$ is the conditional Q_τ of order $\tau \in [0,1]$ of Y knowing $\mathbf{X}$ is defined as the generalized inverse of $F_{Y|\mathbf{X}}$ (Koenker and Hallock, 2001):

$$Q_\tau(Y|\mathbf{X}) = F^{-1}_{Y|\mathbf{X}}(\tau) = inf\{y \in \mathbb{R} : F_{Y|\mathbf{X}}(y) \geq \tau\}. \tag{3.3.1}$$

QR assumes a model in the form of a deterministic distribution, as well as the distribution of the error term is not assumed, because it is non parametric estimation. A QR model is as follows:

$$Q_\tau(Y|\mathbf{X}) = \beta_0(\tau) + \beta_1(\tau)x_1 + ... + \beta_p(\tau)x_p + \varepsilon_\tau, \tag{3.3.2}$$

where $\beta_i(\tau)$ represents the parameter that corresponds to τ and $i = 0, 1, 2, ..., p$. Due to impact of error term to different quantiles, the model parameters will change when τ changes. The interpretation of the coefficients in the model is similar to the general linear model, but is not limited to the conditional median as different quantile positions are considered. This study focuses on non - parametric quantile regression.

3.4 Linear quantile regression

Linear quantile regression (LQR) will be used as one of the benchmark models. LQR is related to linear least - square regressions as both of them are interested in studying the linear relationship between a response variable and one or more explanatory variables. Whereas least - squares regression is concerned with modelling the conditional mean of the response variable, the LQR models the conditional τth quantile of the response variable, for some value of $\tau \epsilon$ (0, 1) (Koenker and Xiao, 2002). The linear quantile regression model is specified below.

Let y_i:$i = 1, ..., n$ be a univariate random sample on a random variable Y having a distribution function F. For $0 < \tau < 1$, the τth sample quantile may be defined as any solution to the minimization problem:

$$\min_{\beta \varepsilon \mathbb{R}} [\sum_{i:y_i > \beta} \tau |y_i - \theta| + \sum_{i:y_i < \beta} (1 - \tau)|y_i - \theta|]. \tag{3.4.1}$$

Let $x_i^T : i = 1, ..., n$ denote a sequence of p- vectors of a known design matrix $\boldsymbol{X}$ and suppose that $y_i : i = 1, ..., n$ is a random sample on the regression process, having distribution function F. Let $\theta = \mathbf{x}_i^T \boldsymbol{\beta}$. Then the τth regression quantile ($0 < \tau < 1$) is defined as a solution to the minimisation problem

$$\min_{\beta \epsilon \mathbb{R}^p} [\tau \sum_{i:y_i > \mathbf{X}_i^T \boldsymbol{\beta}} |y_i - \mathbf{X}_i^T \boldsymbol{\beta}| + (1-\tau) \sum_{i:y_i < \mathbf{X}_i^T \boldsymbol{\beta}} |y_i - \mathbf{X}_i^T \boldsymbol{\beta}|] \tag{3.4.2}$$

LQR can be derived by specifying the τth conditional quantile as $Q_y(\tau|X) = \mathbf{X}^T \boldsymbol{\beta}(\tau)$ and estimating $\beta(\tau)$ as the solution to

$$\min_{\beta \epsilon \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau (y_i - \mathbf{X}_i^T \boldsymbol{\beta}), \tag{3.4.3}$$

where ρ_τ is a linear loss function.

3.5 Non-parametric quantile regression

Non-parametric regression is a form of regression analysis in which none of the predictors take predetermined forms with the response but are constructed according to information derived from the data. Non-parametric quantile regression is a viable alternative to avoid restrictive parametric assumption. It gives much more accurate predictions and is quite easy to estimate parameters.

3.5.1 Kernel regression

Kernel regression is a non - parametric technique in statistics which is used to estimate the conditional expectation of a random variable using a weighted filter to the data. Its objective is to find a non - linear relation between a pair of random variables.

$$y_t = r(x_t) + \varepsilon_t, \quad E(\varepsilon_t) = 0, \tag{3.5.1}$$

where r is a smooth function, x_t are the covariates, y_t is the response variable and ε_t is an error term. We estimate $r(x_t)$ by averaging nearby values of y_t.

$$\hat{r}(x) = \sum_{t=1}^{n} w_t(x) y_t, \quad \sum_{t=1}^{n} w_t(x) = 1 \tag{3.5.2}$$

$\hat{r}(x)$ is a kernel smoother if weights are defined by a kernel function.

Assigning weights to the kernel function results in:

$$w_t(x) = cK\left(\frac{x_t - x}{h}\right), \tag{3.5.3}$$

where c is a constant and h is the bandwidth,

then

$$\sum_{t=1}^{n} w_t(x) = 1 \quad \Rightarrow \quad \sum_{t=1}^{n} cK\left(\frac{x_t - x}{h}\right) = 1 \tag{3.5.4}$$

$\Longrightarrow$

$$c = \left[\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right)\right]^{-1} \tag{3.5.5}$$

then

$$\hat{r}(x_t) = \frac{\sum_{t=1}^{n} K\left(\frac{|x_t - x|}{h}\right) y_t}{\sum_{t=1}^{n} K\left(\frac{|x_t - x|}{h}\right)} \tag{3.5.6}$$

$$\hat{y}_t = \frac{\sum_{t=1}^{n} K\left(\frac{|x_t - x|}{h}\right) y_t}{\sum_{t=1}^{n} K\left(\frac{|x_t - x|}{h}\right)} + \varepsilon_t. \tag{3.5.7}$$

We need h which minimises

$$CV(h) = \frac{1}{n} \sum [\hat{r}_t(x_t - y_t)]^2, \tag{3.5.8}$$

where $CV(h)$ is the cross validation of the bandwidth and $\hat{r}_t(x_t)$ uses all the data except (x_t, y_t). In this study local linear kernel quantile regression discussed in Yu and Jones (1998) will be used.

3.5.2 Locally linear kernel quantile regression

Having the model:

$$y_i = \beta_0 + \beta_1(X_i - x) + \varepsilon_i, \tag{3.5.9}$$

where

$$\varepsilon_i = y_i - \beta_0 - \beta_1(X_i - x). \tag{3.5.10}$$

we want to minimize the following function by Yu and Jones (1998):

$$\min_{\beta_0,\beta_1 \epsilon \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \beta_0 - \beta_1(X_i - x))K\left(\frac{x - X_i}{h}\right), \tag{3.5.11}$$

where K is the kernel function, h is the bandwidth, β_0 and β_1 are constants and ρ_τ is the pinball loss function defined as:

$$\rho_\tau(Y|X) = \begin{cases} \tau(y - q_\tau(Y|X) & \text{if} \quad y > q_\tau(Y|X) \\ (1-\tau)(y - q_\tau(Y|X) & \text{if} \quad y < q_\tau(Y|X) \end{cases}. \tag{3.5.12}$$

Expression (3.5.11) then reduces to

$$\begin{aligned} &\min_{\beta_0,\beta_1 \epsilon \mathbb{R}} [\sum_{i=1}^{n} \tau(y_i - \beta_0 - \beta_1(X_i - x))K\left(\frac{x - X_i}{h}\right) + \\ &\sum_{i=1}^{n} (1-\tau)(y_i - \beta_0 - \beta_1(X_i - x))K\left(\frac{x - X_i}{h}\right)]. \end{aligned} \tag{3.5.13}$$

3.5.3 Bandwidth

The bandwidth h is the maximum distance from the target point of any observation receiving weight. The quality of the curve estimates depends sensitively on the choice of h. This study will use h which minimises

$$CV(h) = \frac{1}{n}\sum[\hat{r}_t(x_t - y_t)]^2, \tag{3.5.14}$$

where $\hat{r}_t(x_t)$ uses all the data except (x_t, y_t) and $\hat{r}(x_t)$ is as defined in equation (3.5.2). After defining weights, all that is necessary to estimate a non-parametric quantile regression model is to provide "weight" option in the R package quantreg.

3.5.4 Different types of kernel functions

The idea is to approximate the results locally with series of quantile regressions that are estimated using a subset of the observations that are close to a set of target values, with more weight placed on observations that are close to the target points. For each target point, we define a set of weights that decline with distance, up to some maximum. At larger distances, the weight is set to zero. Then any kernel weight function can be used. A kernel function K calculates the inner product of two vectors $\boldsymbol{x}$,$\boldsymbol{x}^{'}$. Where $\boldsymbol{x}$ is the training input and $\boldsymbol{x}^{'}$ is the unlabeed input. This study will use the following kernel functions, discussed in Karatzogho et al.(2016).

Linear kernel function

The linear kernel is defined as:

$$K(\boldsymbol{x}, \boldsymbol{x}^{'}) = \langle \boldsymbol{x}, \boldsymbol{x}^{'} \rangle. \tag{3.5.15}$$

The linear kernel is the simplest kernel function, which is useful especially when dealing with large sparse data vectors of $\boldsymbol{x}$.

Gaussian kernel function

The Gaussian kernel is defined as:

$$K(x,x^{'}) = \exp(-\sigma\|x-x^{'}\|^{2}). \tag{3.5.16}$$

The adjustable parameter sigma plays a major role in the performance of the kernel. The Gaussian kernel is the most widely used kernel since its linear combinations can approximate any continuous function and the corresponding feature space has a finite dimension. For any given labelled data set there exists a linear hyperplane which correctly separates the data in the Gaussian feature space and this makes the expression power basically unlimited.

Bessel kernel function

Bessel kernel function is defined as:

$$K(x,x^{'}) = \frac{Bessel^{n}_{(\nu+1)}(\sigma\|x-x^{'}\|)}{(\|x-x^{'}\|)^{-n(\nu+1)}}. \tag{3.5.17}$$

Bessel kernel function is a general purpose kernel and is typically used when there is no further prior knowledge is available and mainly popular in the Gaussian process community.

Laplace radial basis kernel function

Laplace radial basis kernel is defined as:

$$K(x,x^{'}) = \exp(-\sigma\|x-x^{'}\|). \tag{3.5.18}$$

It is the radial basis function kernel that is less sensitive to changes in the sigma parameter. The observations made about the sigma parameter for Gaussian kernel also apply to Laplace kernel.

ANOVA radial basis kernel function

ANOVA radial basis kernel function is defined by:

$$K(\boldsymbol{x}, \boldsymbol{x}^{'}) = \left(\sum_{k=1}^{n} \exp(-\sigma(x^k - x^{'k})^2) \right)^d \tag{3.5.19}$$

where x^k is the k^{th} component of $\boldsymbol{x}$ and d is the degree. It is also a radial basis kernel function such as the Gaussian and Laplace kernels. It is said to perform well in multidimensional regression problems (Hofmann, 2008).

This study will use the R package 'kernlab' developed by Takeuchi et al. (2006) for estimating parameters of the locally linear quantile regression models under each of the discussed kernel functions.

3.6 Variable selection

Variables selection will often be put forward when building statistical models. It is not helpful to study the problems that the variables in the model are more or less than the actual variables. In the process of optimizing the models, most explanatory and influential subset of variables need to be found, in order to make the model more reasonable and forecast accurately.

There are many different methods available that can be used to select important variables in a study. The most common used method is the subset selection which includes techniques such as the stepwise criterion and many more. However this study will focus on one of the shrinkage methods. The shrinkage selection methods includes elasticnet, least absolute shrinkage and selection operator (Lasso) Tibshirani (1996). This study will use one of the variants of the Lasso, which is the adaptive - Lasso which is given in equation 4.2 and discussed in Zou (2006):

$$\min_{\beta}(\sum_{i=1}^{n}\rho_{\tau}(y_i - x_i^{\tau}\beta_{\tau}) + \sum_{j=1}^{p}\lambda|\beta_j|), \tag{3.6.1}$$

with $\rho_{\tau}(\mu) = \mu(\tau - I(\mu < 0))$ which is the error measure generally known as the pinball loss function. Lasso method regards absolute coefficient function as a penalty term to compress the coefficients of the model, and coefficients whose absolute value is relative smaller than others, so as to achieve the purpose of variable selection and parameter estimation.

There are many advantages in using the Lasso method, as it can provide a very good prediction accuracy, because shrinking and removing the coefficients can reduce variance without a substantial increase of the bias. This is especially useful when you have a small number of observations and a large number of variables.

3.7 Parameter estimation

There are several methods commonly used for estimating quantile regression parameters, such as Bayesian estimation, Maximum likelihood estimator (MLE), optimal quantization and mathematical programming techniques (i.e linear programming, interior point algorithms, etc.). This study will use linear programming algorithms to estimate the regression parameters.

Estimating linear quantile regression

The aim of linear programming is to find a vector minimising or maximizing the value of a given linear function among all vectors that satisfy a given system of linear equations and inequalities. The linear programming standard minimisation problem is given by

$$\min_{y \epsilon \mathbb{R}^m} y^T b, \tag{3.7.1}$$

subject to the constraints

$$y^T A \geq c^T, \tag{3.7.2}$$

and $y_1 \geq 0, ..., y_m \geq 0$, where A is $m \times n$ matrix, b $\epsilon \mathbb{R}^m$, c $\epsilon \mathbb{R}^n$. The maximisation problem $Ax \leq$b and $x \geq 0$. The quantile regression model can be rewritten as

$$y_i = X_i^T \beta(\tau) + e_i = X_i^T \beta(\tau) + (u_i - v_i), \tag{3.7.3}$$

where $u_i = e_i I(e_i > 0), v_i = |e_i| I(e_i < 0)$. Therefore,

$$\min_b \sum_{i=1}^{n} \rho_\tau (y_i - \boldsymbol{x}_i^T \boldsymbol{b}) \Leftrightarrow \min_{b,u,v} \tau \boldsymbol{1}_n^T \boldsymbol{u} + (1-\tau) \boldsymbol{1}_n^T \boldsymbol{v}, \tag{3.7.4}$$

$y - \boldsymbol{X}^T \boldsymbol{b} = u - v$ $\boldsymbol{b}$ $\epsilon \mathbb{R}^p$, $\boldsymbol{u} \geq 0$, $\boldsymbol{v} \geq 0$, which is the standard linear programming (minimization) program (Koenker, 2005). The error function for minimisation can be transformed into standard linear programming type dual problems for minimizing and maximisation, then a linear programming algorithms can be applied to solve the parameters for QR.

Estimating local linear kernel quantile regression

The optimisation problem for the local linear quantile regression model given in equation (3.5.11) can be expressed as a minimiser of the empirical risk plus a regulariser (Takeuchi et al., 2006) minimise:

$$R_{reg}[f] := \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y - f(x))K\left(\frac{x - X_i}{h}\right) + \frac{\lambda}{2}\|\beta\|_H^2, \tag{3.7.5}$$

where $f = \beta + b$ and $b\varepsilon\mathbb{R}$, $f(x) = \beta_0 + \beta_1(X_i - x)$. $\|.\|_H$ is the Reproducing Kernel Hilbert Space (RKHS) norm and we require $\beta\varepsilon H$. For efficient numerical implementation the dual optimization problem to equation (3.7.5) is then used. This is solved using an interior point quadratic programming solver given in the r package "kernlab" developed by Karatzoglou et al. (2016).

3.8 Model selection

The focus in the study is on forecasting, hence the predictive power of the model is very important. Based on this the model will be selected based on cross validation (CV), Akaike Information Criterion (AIC), adjusted R^2, AIC corrected AIC_c and Bayesian Information Criterion (BIC), (Behl, 2013). The AIC was first developed by Akaike (1983) and refined by Wagenmakers and farrell (2004). AIC estimates the expected distance between the model and the true value. AIC score aims to find approximate model whereas the BIC score tries to find the true model. AIC and BIC scores are calculated using the following equations (Yamaoka et al., 1978):

$$\text{AIC} = 2k - 2\log(\ell(\theta)) \tag{3.8.1}$$

$$\text{BIC} = k\log(n) - 2\log(\ell(\theta)), \tag{3.8.2}$$

respectively where, $\ell(\theta)$ is the likelihood function of the parameters in the model, and θ the vector of parameters is evaluated at the maximum likelihood estimate, k is the number of parameters to be estimated and n is the number of observations.

AIC$_c$ is AIC with a correction for finite sample sizes. The formula for AIC_c depends upon the statistical model:

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1}. \tag{3.8.3}$$

Goodness - of - fit test for quantile regression is motivated by the familiar R^2 of classical least squares regression. Consider a linear model for conditional quantile function,

$$Q_{yi}(\tau|\mathbf{x}) = \boldsymbol{x}'_{i1}\boldsymbol{\beta}_1(\tau) + \boldsymbol{x}'_{i2}\boldsymbol{\beta}_2(\tau), \tag{3.8.4}$$

and let $\hat{\beta}(\tau)$ denote the minimiser of

$$\hat{V}(\tau) = \min_{b\epsilon\mathbb{R}^p} \sum \rho_\tau(y_i - \boldsymbol{x}'_i\boldsymbol{b}) \tag{3.8.5}$$

and $\tilde{\beta}(\tau) = (\tilde{\beta}'_1(\tau), 0')$ denote the minimizer for the corresponding constrained problem, with

$$\tilde{V}(\tau) = \min_{b_1\epsilon\mathbb{R}^{p-q}} \sum \rho_\tau(y_i - \boldsymbol{x}'_{i1}\boldsymbol{b}_1), \tag{3.8.6}$$

that is $\hat{\beta}(\tau)$ and $\tilde{\beta}(\tau)$ denote the unrestricted and restricted quantile regression estimates. Then the goodness - of - fit criterion

$$R^1(\tau) = 1 - \hat{V}(\tau)/\tilde{V}(\tau), \tag{3.8.7}$$

which is the natural analog of R^2 (Koenker and Machado, 1999).

CV is one of the most commonly used methods of evaluating predictive performance of a model. It is mainly used in settings where the goal is prediction and one wants to estimate how accurately a predictive model will perform in practice, Zhang and Yang (2015). There are different types of cross validation. These include among others, leave-on-out cross validation (Loocv), leave-p-out cross validation, holdout method, k-fold cross validation, Monte Carlo cross validation. This study will use Loocv as it takes less time than computing the residual error and it is a much better way to evaluate models. CV idea is to randomly divide the data in K equal sized parts, and leave out part k, fit the model to the other $K-1$ parts, and then obtain predictions for the left out k^{th} part. This is done in turn for each part $k = 1, 2, ..., K$, and then the results are combined. Let the K parts be $C_1, C_2, ..., C_k$, where C_k denotes the indices of the observations in part k: if n is a multiple of K, then $n_k = n/K$.

$$CV_{(k)} = \sum_{i=k}^{K} \frac{n_k}{n} \text{MSE}_k, \tag{3.8.8}$$

where MSE_k (Mean of squares of errors) $= \sum_{i \epsilon C_k} (y_i - \hat{y}_i)^2 / n_k$, and $\hat{y}_i$ is the fit for the observation i, obtained from the data with part k removed, Zhang (2015).

3.9 Model diagnostics

Before any further analysis the model will be assessed to check if the assumptions of the residuals will not be violated. For instance, we will check if the residuals are normally distributed. Other than normality we will also check if the residuals will be independent and that there is no serial autocorrelation left in them. Residuals are defined as the difference between the actual and the estimated values of a response variable and statistically are represented by equation (3.9.1):

$$\varepsilon_i = y_i - \hat{y}_i, \tag{3.9.1}$$

where, y_i is the actual values of a response variable and $\hat{y}_i$ is estimated values of a response variable (Moons et al., 2004).

3.10 Forecasting performance

The accuracy of the forecasts are assessed by using the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). Generally, the three measures mentioned above are statistically represented by the equations in (3.10.1), (3.10.2) and (3.10.3) respectively:

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|\varepsilon_i| \tag{3.10.1}$$

$$\text{MAPE} = \frac{100}{m}\sum_{i=1}^{m}|\frac{\varepsilon_i}{\hat{y}_i}| \tag{3.10.2}$$

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}{\varepsilon_i}^2}, \tag{3.10.3}$$

where, m is the number of observations in the test data set, $\hat{y}_i$ is the estimated values of a response variable and $\varepsilon_i = y_i - \hat{y}_i$ is the residual of the i^{th} observation (Hyndman and Koehler, 2006).

3.10.1 Prediction interval

Prediction interval (PI) is a useful tool for uncertainty modeing by its nature. PI consists of lower and upper bounds that cover the future unknown target value with a certain probability $(1-\alpha)\%$ called the confidence level. PIs are more suitable and give useful information than point forecasts for decision makers (Khosvari et al. 2011). Prediction interval width (PIW) is defined by:

$$PIW_t = UL_t - LL_t, \tag{3.10.4}$$

where UL_t and LL_t are the upper and lower limits, respectively. The typical indicators that evaluate the performance of PIs of the models are prediction interval coverage probability (PICP) and prediction interval normalized average width(PINAW). PICP examine the reliability of the constructed PIs and is a fundamental feature of PIs (Heskes, 1997)). PICP is useful as it measures the probability of targets lie that in the constructed PIs. PICP is calculated as follows (Christoffersen, 1998):

$$\mathrm{PICP}_t = \frac{1}{m}\sum_{t=1}^{m} I_t, \tag{3.10.5}$$

where m is the number of samples and

$$I_t = \begin{cases} 1 & \text{if} \quad y_t \epsilon [LL_t; UL_t] \\ 0 & \quad otherwise. \end{cases}$$

PICP is valid, if it is greater or equal to the nominal confidence level of PIs (PICP$\geq 1-\alpha$). PICP=100% if the target values are within the PIs. PINAW is the significant evaluation index of the PIs. PINAW is the higher quality of PIs, which evaluates the overall width of PIs. PINAW is given as (Quan et al., 2014):

$$\text{PINAW} = \frac{1}{mR}\sum_{t=1}^{m}(UL_t - LL_t), \tag{3.10.6}$$

where R is the range of the underlying targets. Prediction interval normalized average deviation (PINAD) states the deviation degree from the real values to the PIs. PINAD is given as (Khosravi et al., 2010):

$$\text{PINAD} = \frac{1}{mR}\sum_{t=1}^{m} d_t \tag{3.10.7}$$

where d_t denotes the deviation degree and is given by:

$$d_t = \begin{cases} LL_t - y_t & \text{if} \quad y_t < LL_t \\ 0 & \text{if} \quad LL_t \leq y_t \leq UL_t \\ y_t - UL_t & \text{if} \quad y_t > UL_t \end{cases} .$$

3.10.2 Forecast error distribution

The measures of central tendency (mean and the median) together with the lower and upper quartiles of the residuals are used in the characterization of forecast error distribution. They provide important information about the distribution. The skewness and kurtosis can also provide more information about the distribution. Skewness measures the probability distribution's asymmetry, while the kurtosis measures the magnitude of the distribution's peak. The summary statistics will be obtained for the residuals of the best models, to emphasise the model that best forecast the FDI.

3.11 Conclusion

The chapter discussed the use two benchmark models, OLS and LQR. The main models will be built under the non-parametric quantile regression, considering the locally linear kernel quantile regression. The different types of kernel functions that will be used are, linear, Gaussian, Bessel, Laplace radial basis and ANOVA radial basis.

Chapter 4

Analysis of Data and Discussion

4.1 Introduction

This chapter presents data description and data analysis using the methods and models discussed in chapter 3.

4.2 Exploratory data analysis

4.2.1 Data sources

The study uses data, obtained from South African Reserve Bank (SARB) and International Country Risk Guide (ICRG). SARB is the central bank of the Republic of South Africa. The primary purpose of the SARB is to achieve and maintain price stability in the interest of balanced and sustainable economic growth in South Africa. ICRG is part of the political risk services group and has provided an international clientele with ratings affecting political risk, economic risk and financial risk. It offers analyses of events that affect the risk rating along with the economic and financial data underlying financial and economic risk ratings. The ICRG includes a Political Risk Index, which in turn consists of 12 components measuring various dimensions of the political and business environment facing firms operating in a country. The data used was measured for the economy of South Africa for the period 1996 to 2015.

4.2.2 Data characteristics

In this study dependent variable is the foreign direct investment (FDI), which is an investment made by a company or individual in one country in business interests in another country, in the form of either establishing business operations or acquires foreign business assets, including establishing ownership or controlling interest in a foreign company. The inward FDI stock is the value of foreign investors' equity in and net loans to enterprise resident in the reporting economy. FDI stocks measure the total level of direct investment at a given point in time, usually at the end of a quarter or of a year. This study will use the FDI measured at the end of each year from 1996 to 2015. The data for the period 1996 to 2010 will be used for training the models while the data for the period 2011 to 2015 will be used for validation.

The independent variables are defined as:

- GDP - is the gross domestic product, which is the total value of everything produced by all the people and companies in the country. It doesn't matter if they are citizens or foreign - owned companies, if they are located within the country's boundaries the government counts their production as GDP. GDP is considered as one of the best ways to measure a country's economy.

- CPI - is the consumer price index, which is the measure that examines the weighted average of prices of a basket of consumer goods and services. It is calculated by taking price changes for each item in the predetermined baskets of goods and averaging them. CPI is one of the most frequently used statistics for identifying periods of inflation.

- LP - is the labour productivity, which measures the amount of goods and services produced by one hour of labour. It is concerned with the amount or volume of output that is obtained from each employee and it is a measure of economic growth within a country.

- GFCF - is the gross fixed capital formation, which measures the net increase in fixed capital, including spending on land improvements, building of roads and industrial buildings.

- Exp - is the exports, which is sending of goods and services to another country for sale. It is important for development and growth of the national economies.

- PS - is the political stability which is the measure of the perceptions regarding the probability that the government will be destabilized by unconstitutional or violent means, including domestic violence and terrorism. The levels are given in political and economic factors to measure the political stability. The unstable political environment can reduce investment, reduce speed of economic development, increase the probability of a government collapse or result in political unrest. The higher the political index the more stable the economy.

- noltrend - is the nonlinear trend variable.

4.2.3 Summary statistics

Table 4.1 presents the summary statistics for all the variables that will be used in this study. From the year 1996 to the year 2015, SA received an average FDI inflow of R3,856,486,223.00, having the maximum of R9,885,001,293.00 and the minimum inflow of R550,338,596.00. The minimum FDI inflow was obtained in the year 1998. SA also had the average political stability of 0.71 between those years, with the minimum of 0.65 and a maximum of 0.79.

The Exp, LP, noltrend and GDP have negative skewness, while other variables are positively skewed. The kurtosis of the CPI is 3.92 which is more than 3, indicating that its distribution is leptokurtic, meaning its central peak is higher and sharper compared to the normal distribution. All other variables have kurtosis of less than 3, indicating that their distributions are specified as platykurtic when compared with normal distribution.

Table 4.1: Summary statistics.

Variable	Mean	Q1	Q2	Q3	Min	Max	Skewness	Kurtosis
FDI	3.9e+09	9.3e+08	3.8e+09	6.5e+09	5.5e+08	9.9e+09	0.426	1.83
GFCF	436734	289816	435706	562394	262458	639383	0.05	1.34
GDP	50015	44990	50333	54484	43720	56469	-0.003	1.33
Exp	736213	658529	738834	827752	542552	911366	-0.109	1.72
CPI	6.025	4.925	5.700	7.100	1.400	11.500	0.47	3.92
LP	90.33	79.58	91.70	101.10	64.8	110.50	-0.18	1.80
PS	0.71	0.69	0.69	0.73	0.65	0.79	0.45	2.50
noltrend	0.948	0.564	0.565	0.517	0.274	0.519	-0.374	1.000

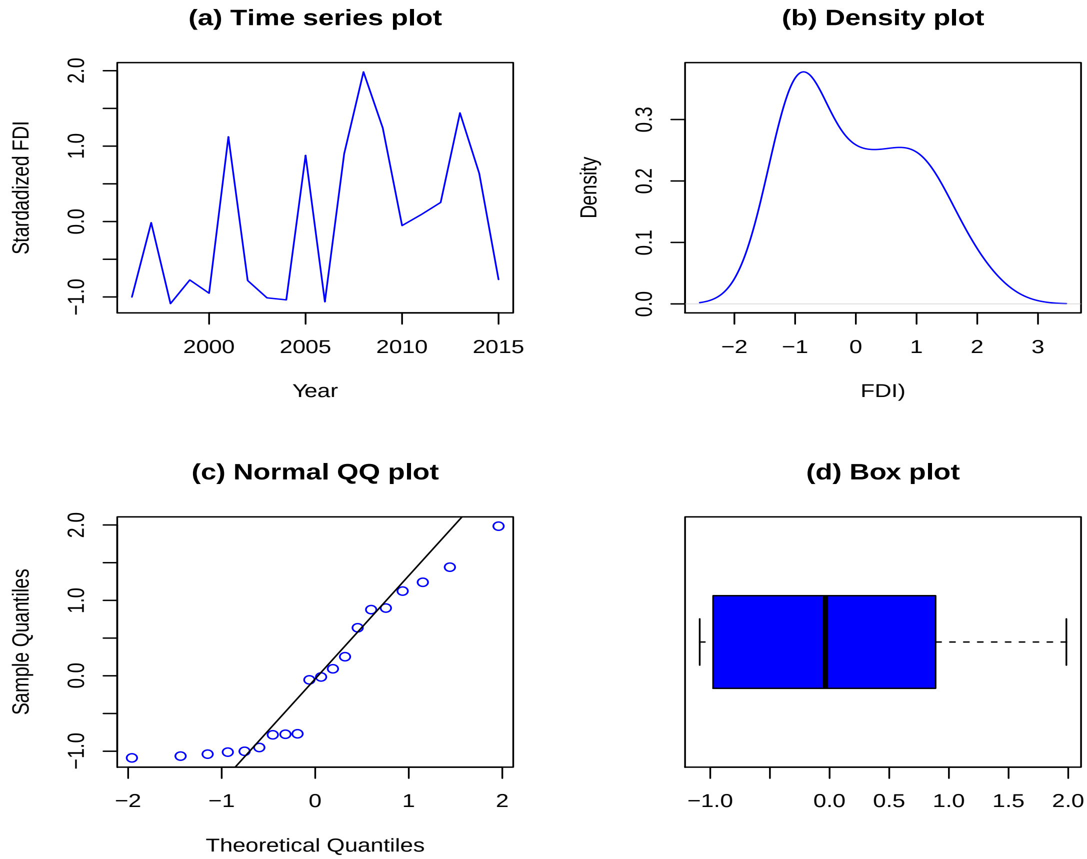

Figure 4.1: FDI time series, density, normal and box plots.

4.2.4 Dataset description

The dataset contains 8 variables with 20 observations. The data was initially standardised. Initially FDI is analysed using plots. Looking at Figure 4.1(a) on the time series plot it can be seen that our dependent variable (FDI) has a seasonal pattern. Figure 4.1(a) shows that since 1994, there is a positive slightly upward trend in FDI inflows into SA. This major increase in the attraction of FDI was due to democratisation of SA, where there was subsequent openness to trade. In 1997, however, there was a significant increase of FDI, due to the partial privatisation of Telkom and South African Airlines (Thomas et al., 2005). A remarkable increase occurred in 2001. In 2002, FDI inflows decreased and continued until fall in 2004. In 2005 there was another increase, since SA was the largest FDI recipient in Africa. In 2006, the inflows declined drastically, but made a comeback in 2007.

There was a steady increase until 2008 and 2009. In 2010, the inflows decreased as the world was experiencing a contraction in demand due to the 2008 financial crisis, even though there was an increase in the tourism industry's FDI inflows as a result of the 2010 World Cup, hosted by South Africa. In 2012, the country faced one of the biggest incident of labour unrest in its history from the mining industry, and as a result, FDI inflows slumped. In 2014, FDI inflows decreased again towards 2015.

An assessment of normality of data is a prerequisite for many statistical tests as normal data is one of the assumptions in parametric testing. In order to determine normality graphically, the normal QQ, density, box and whisker plots. Figure 4.1 is also showing the density plot, normal QQ plot and box and whisker plot. Reading the normal QQ plot from left to right in Figure 4.1(c), it can bee seen that the circles start out on the left hand side of the line, then below it, then rise above it, then go below again. This pattern means that the tails of the distribution of the data is too narrow for normality.

The density plot in Figure 4.1(b) shows that the data is bimodal since there are two peaks, having the highest concentration of points at negative 1 with the density peak of more than 0.3. The box and whisker plot in Figure 4.1(d) is also showing positive skewness of the data, since there is a longer whisker at the right of the box.

To decide which variables should be included in the final model Lasso shrinkage method for variable selection will be used.

4.3 Lasso regression

4.3.1 Introduction

This section looks at the different features of the Lasso regression for variable selection. The goal of this analysis is to find which independent variables are most relevant to predict the dependent variable FDI and in order to do so the Lasso method will be used.

4.3.2 Variable selection using Lasso

The data is split into two sets: a training set, from year 1996 to the year 2011, and a testing set which is from year 2012 to year 2015. The training dataset is used to build a model to predict the dependent variable FDI. To choose which variables to use in the model, the 'glmnet' package in R developed by Friedman et al., (2009) is used. Glmnet returns a sequence of different models for different values of λ.

The next step is to choose the most appropriate optimal value of λ. The R function 'cv.glmnet' helps us to select the most appropriate value of λ, choosing the number of n-fold validation. The relevant plot is shown in Figure 4.2. In order to choose the most appropriate value for λ the Lasso method extracts different values of λ, such as λ_{min} (first vertical dotted line) that gives minimum mean cross-validated error and λ_{1se} (second vertical dotted line), that gives a model such that the error is within one standard error of the minimum. For both the values of λ all seven variables are selected to be in the model. All solutions were found after 9 iterations. Solution path as shown in Figure A3 in the appendix, can be seen from the results corresponding to each step in the path of the estimation of model parameters. Lasso variable selection results showed that the GFCF, CPI, Exp, PS and noltrend have a significant positive effect on FDI, while GDP and LP has a significant negative effects on FDI. Table 4.2 shows variable selection using Lasso. From Table 4.2 all the variables were selected which shows that all the independent variables have an influence on the model. For comparison

Table 4.2: Variable selected using Lasso.

(Intercept)	1.1888
GFCF	3.4081
CPI	8.8239
LP	-6.6944
Exp	1.5286
GDP	-1.0792
PS	4.1839
noltrend	1.9708

purposes 'Lars' R package developed by Friedman et al. (2010) is used and

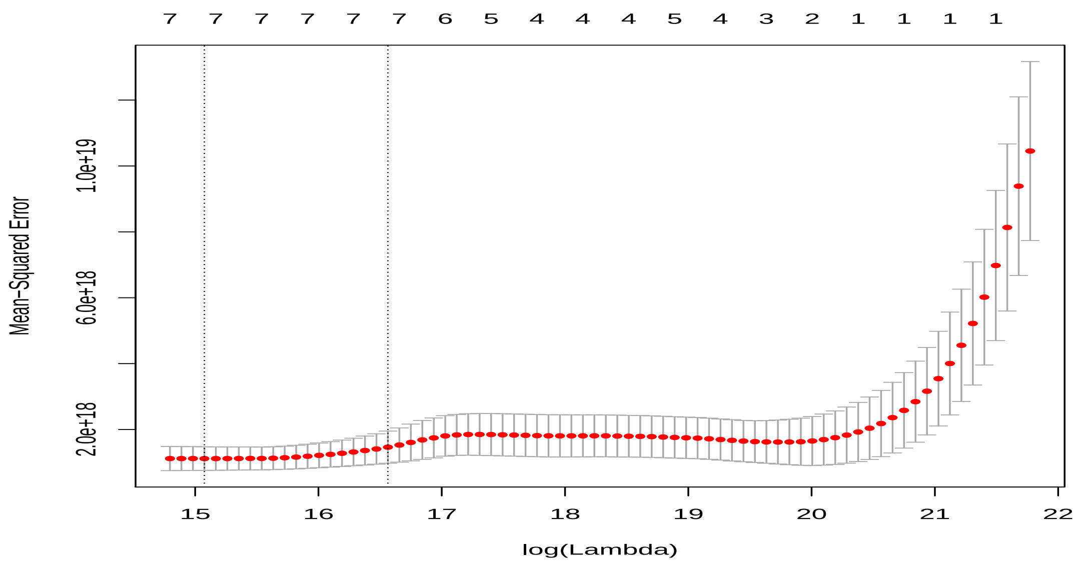

Figure 4.2: Cross - validation.

the results are the same as those from using the 'glmnet' package. Based on λ_{min} all variables were selected to be in the model.

4.4 Ordinary least squares

The ordinary linear regression model was fitted on the training dataset to see how well it models the observed data. This model denoted by M1 will be used as one of the benchmark models. The multiple linear regression model is given in equation 4.4.1. The model estimates the value of the intercept and each predictor's slope.

$$\begin{aligned}\text{FDI}_t = \beta_0 + \beta_1\text{GFCF}_t + \beta_2\text{GDP}_t + \beta_3\text{Exp}_t + \\ \beta_4\text{CPI}_t + \beta_5\text{LP}_t + \beta_6\text{PS}_t + \beta_7\text{noltrend}_t + \varepsilon_t,\end{aligned} \tag{4.4.1}$$

where ε_t is the error term at time t and all the other variables are as defined in section 4.2.2.

The intercept is the average expected FDI value for the average value across all predictors. The value for each slope estimate will be the average increase or decrease in FDI associated with one unit increase or decrease in each predictor value, holding the other predictors constant.

Table 4.3: Coefficients of OLS model.

	Estimate	Std. Error	t value	P-value
Intercept	0.013	0.085	0.150	0.8847
GFCF	1.657	0.613	2.704	0.0305
GDP	-1.817	0.715	-2.542	0.0386
Exports	0.593	0.356	1.664	0.1400
CPI	0.001	0.099	0.005	0.996
LP	-0.330	0.4876	-0.678	0.519
PS	0.044	0.097	0.448	0.668
noltrend	0.874	0.083	10.525	0.0001

Table 4.4: Residuals M1.

Residuals	Min	1Q	Median	3Q	Max
	-0.314	-0.076	0.036	0.065	0.343

From the model 4.4.1 output in Table 4.3, it can be seen that the GDP increase by one unit, the FDI is expected to decline by 1.817, provided the other predictors remain constant. If the LP increase by one unit, FDI is expected to decrease by 0.330. If the PS increase by one unit, we expect the FDI to increase by 0.044. The increase in one unit of GFCF, Exp or CPI will make the FDI to increase. Looking at p-values of the model, it can be seen that only GFCF, GDP and noltrend are significant to the model. All the variables are included since they were all selected to be in the model by the Lasso variable selection and the focus is on the predictive power of the model. Therefore our OLS model can be given by:

$$\begin{aligned} \mathrm{FDI}_t &= 0.01 + 1.66\mathrm{GFCF}_t - 1.82\mathrm{GDP}_t + 0.59\mathrm{Exp}_t + \\ &\quad 0.001\mathrm{CPI}_t - 0.33\mathrm{LP}_t + 0.04\mathrm{PS}_t + 0.87\mathrm{noltrend}_t + \varepsilon_t \end{aligned} \tag{4.4.2}$$

Residual standard error is a measure of the quality of a linear regression fit, is the average amount that the dependent variable FDI will deviate from the true regression line. We get the residual standard error of 0.251 on average. The R-squared statistic provides a measure of how well the model is fitting the actual data. For multiple regression adjusted R-squared is the preferred measure as it adjusts for the number of variables considered. An R-squared of 0.9437 is obtained, or about 94% of the variation found in the dependent variable (FDI) can be explained by the independent variables.

F-statistic is a good indicator of whether there is a relationship between dependent variable and independent variables. The further the F-statistic is from 1 the better it is. The F-statistic value from our model is 34.54 and seven degrees of freedom, which is relatively larger than 1 given the size of our data. This shows that, the dependent variable (FDI) had good relationship with the independent variables. Looking at Figure 4.3 top panel which is the residual plot, it is clear that there is no autocorrelation, since all the lags are within the intervals. The ARIMA(0,0,0) is also confirming that this model does not have autocorrelation. Figure 4.3 bottom panel shows actual

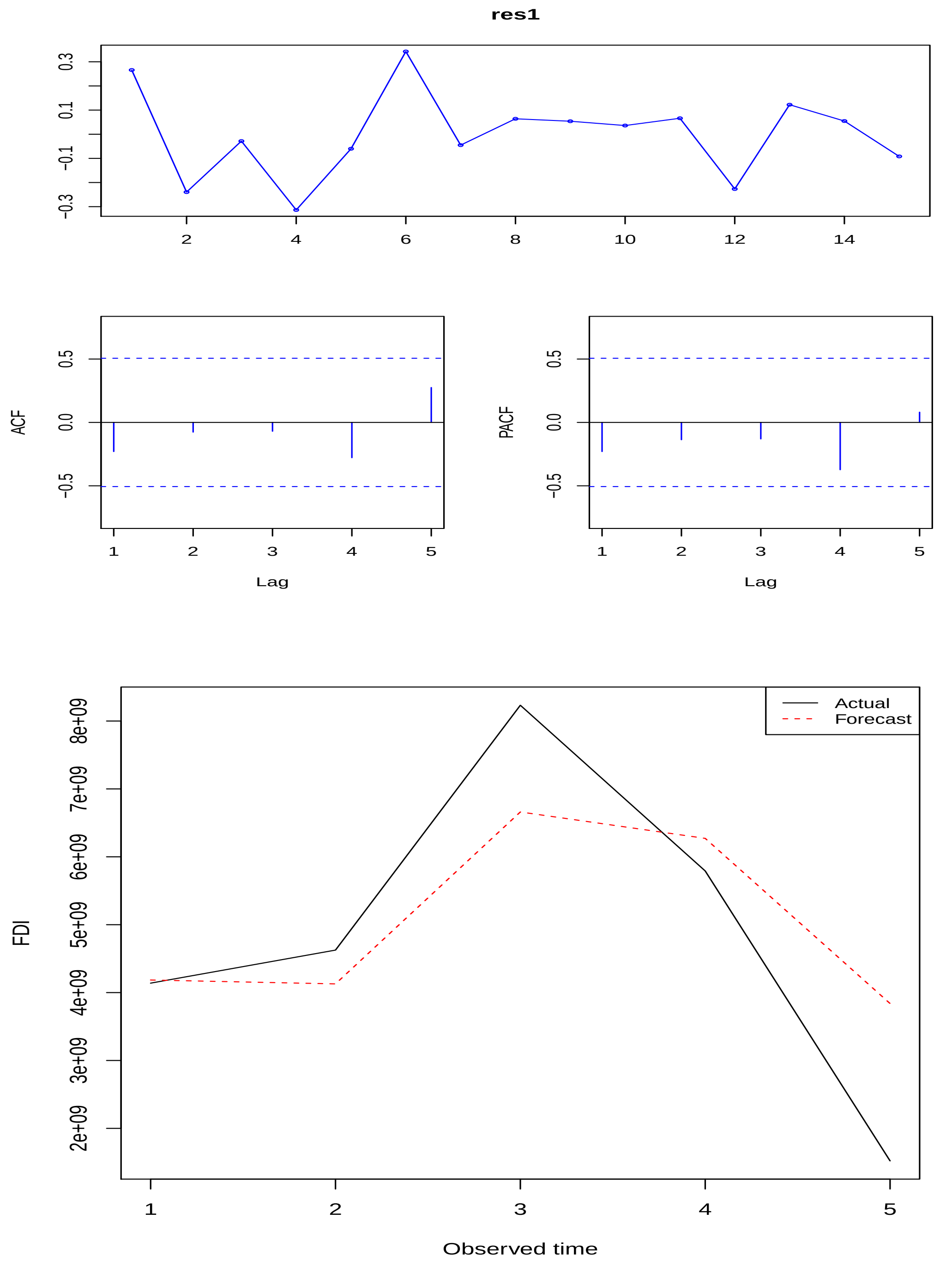

Figure 4.3: Top panel: Residual plot (OLS). Bottom panel: Time series plot for actual FDI(solid line) and its forecast (OLS: dotted line).

FDI (solid line) and FDI forecasts (dotted line) for model M1. It can be seen that the forecasted values for model M1 do not follow the actual FDI remarkably well. Table 4.5 is showing the forecasting accuracy for model M1,

Table 4.5: OLS forecast accuracy.

RMSE	MAE	MAPE
1290245860	982713622	38.32525

which is measured by RMSE, MAE and MAPE.

4.5 Linear quantile regression

The linear quantile regression model was built with the quantile level of τ of 0.5. Table 4.6 presents a summary of the estimates of the parameters. The linear quantile regression model which is denoted as M2 and the second benchmark model is given in equation 4.5.1:

$$\begin{aligned} Q_\tau(\text{FDI}_\tau) = \beta_0 + \beta_1 \text{GFCF}_\tau + \beta_2 \text{GDP}_\tau + \beta_3 \text{Exp}_\tau + \\ \beta_4 \text{CPI}_\tau + \beta_5 \text{LP}_\tau + \beta_6 \text{PS}_\tau + \beta_7 \text{noltrend}_\tau + \varepsilon_\tau. \end{aligned} \tag{4.5.1}$$

The residual plot in Figure 4.4 top panel shows no autocorrelation, since all the lags are within the intervals. The ARIMA (0,0,0) is also confirming that this model does not have autocorrelation. This was obtained from the Box - Ljung test. This model was chosen to be the best model by AIC = -6.32, since AIC was less compared to AICC of -6.01 and BIC of -5.61. Figure 4.4 bottom panel shows the actual FDI (solid line) and FDI forecasts (dotted line) for model M2. It can be seen that the forecast values for M2 do not follow the actual FDI remarkably well. From Table 4.7 and Table 4.5,

Table 4.6: Coefficients of linear quantile regression model 4.5.1.

	Coefficients	Std. Error	t value	P-value
Intercept	0.010	2.124	0.005	0.996
GFCF	1.815	8.948	0.203	0.845
GDP	-1.989	5.947	-0.335	0.748
Exp	0.669	12.699	0.053	0.959
CPI	0.028	1.871	0.015	0.988
LP	-0.362	14.116	-0.026	0.980
PS	0.099	1.699	0.058	0.955
noltrend	1.012	0.083	0.858	0.419

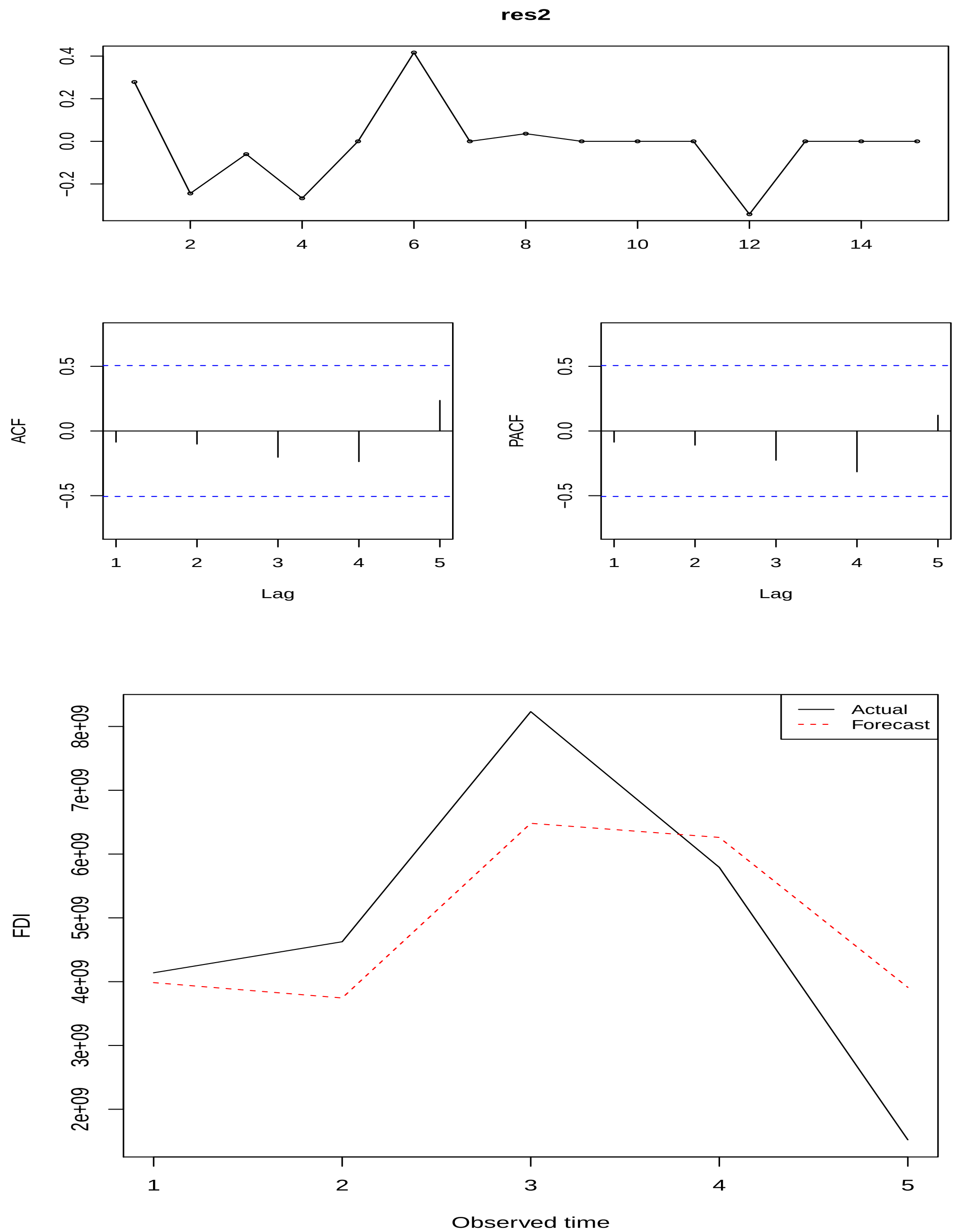

Figure 4.4: Top panel: Residual plot (Linear quantile regression). Bottom panel: Time series plot of FDI(solid line) and linear quantile regression forecasts(dotted line).

it is clear that the RMSE and MAE of model M1 are less compared to the RMSE and MAE of model M2. Based on these accuracy measures M1 is not a better model compared to M2.

Table 4.7: LQR model Accuracy.

RMSE	MAE	MAPE
1399749495	1129035054	41.84898

4.6 Non - parametric quantile regression models

This section presents the results of the kernel quantile regression models based on the following kernel functions: Linear, Gaussian, Bessel, Laplace radial basis and ANOVA radial basis. The Linear kernel is denoted as M3. Gaussian kernel is denoted as M4, Bessel kernel is denoted as M5 and Laplace radial basis and ANOVA radial basis are denoted as M6 and M7 respectively. Table 4.8 and Table 4.9 are showing the 95% prediction intervals for all kernel quantile regression models.

The forecast for the lower bound of the prediction interval was obtained with the quantile of τ of 0.25. The forecast for the upper bound of the prediction interval was obtained with the quantile of τ of 0.975. All the forecasts of models M3, M4, M5, M6 and M7 are inside the prediction intervals since there is no any forecast value that is below or above lower bound and upper bound.

Table 4.8: Prediction intervals of Linear, Gaussian and Bessel kernel functions.

Linear kernel function				
Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$
2011	4139289123	3601483119	4457091910	6312521517
2012	4626029122	3459780039	4295305779	6754326509
2013	8232518816	4727292115	6291891470	8232518819
2014	5791659020	4277708536	5791659046	7645834982
2015	1521139945	1521139931	2884559217	4246327276
Gaussian kernel function				
Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$
2011	4139289123	2327241043	4139289150	7493224082
2012	4626029122	2392803728	4626029130	7896707171
2013	8232518816	2509221920	6280533976	8232518850
2014	5791659020	2360959675	5791659018	8101312070
2016	1521139945	1521139935	1521139954	7192630913
Bessel kernel function				
Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$
2011	4139289123	2487178389	4362045845	7503534252
2012	4626029122	2505392846	4626029123	8005525466
2013	8232518816	2618284541	6152270374	8232518843
2014	5791659020	2461073480	5791659014	8138562732
2016	1521139945	1521139935	1521139949	7125854659

Table 4.9: Prediction intervals of Laplace and ANOVA radial basis kernel functions.

Laplace radial basis kernel function				
Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$
2011	4139289123	4139289119	4139289124	7658408711
2012	4626029122	4615340098	4626029126	7965387527
2013	8232518816	5246630120	7253495221	8232518860
2014	5791659020	4862831701	5791659021	8135666559
2016	1521139945	1840239977	1521139968	7652734546
ANOVA radial basis kernel function				
Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$
2011	4139289123	4139288915	4139289107	6359577977
2012	4626029122	4626029093	4626029127	6486153500
2013	8232518816	5002751144	8232518379	8232518826
2014	5791659020	4859770963	5823624433	7577278212
2016	1521139945	1521139935	1521139942	3594033866

4.7 Linear quantile regression averaging

The linear quantile regression averaging model was built using the linear quantile regression with all the non - parametric quantile regression models (M3, M4, M5, M6 and M7) as the independent variables. The linear quantile regression averaging model which is denoted as M8 is given by equation 4.7.1.

$$Q_\tau(\mathrm{FDI}_t) = \beta_{0,\tau} + \beta_{1,\tau}\mathrm{fM3} + \beta_{2,\tau}\mathrm{fM4} + \beta_{3,\tau}\mathrm{fM5} + \\ \beta_{4,\tau}\mathrm{fM6} + \beta_{5,\tau}\mathrm{fM7} + \varepsilon_{t,\tau}, \quad (4.7.1)$$

where fM_k is the forecast of the models M_k, for $k = 3, ..., 7$. Figure A5 is showing that the forecast values for model M8 follow the actual FDI remarkably well. Table 4.10 presents a summary of the estimates of the parameters. The coefficient of models M3, M4, M5 and M6 are all zero, showing that they are all not necessary in model M8. M7 has the coefficient of one and the p-value of 0.03684, this makes it to be the only variable that is significant on predicting FDI in model M8. Table 4.11 is showing the 95% prediction

Table 4.10: Coefficients of linear quantile regression averaging model.

	Coefficients	Std. Error	t value	P-value
Intercept	0.0000	0.0002	0.0000	1.0000
M3	0.0000	0.0020	0.0001	0.9999
M4	0.0000	0.62698	0.0000	1.0000
M5	0.0000	0.87009	0.0000	1.0000
M6	0.0000	0.46844	0.0000	1.0000
M7	1.0000	0.43347	2.3072	0.03684

intervals for model M8. The forecast for the lower bound of the prediction interval was obtained with the quantile of τ of 0.25. The forecast for the upper bound of the prediction interval was obtained with the quantile of τ of 0.975. All the forecasts of model M8 are inside the prediction intervals.

Table 4.11: Prediction intervals: Linear quantile regression averaging.

Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$
2011	4139289123	4119050821	4139289108	4139289123
2012	4626029122	4598699833	4626029126	4626029127
2013	8232518816	8232518815	8232518815	8232518815
2014	5791659020	5791659020	5823624444	5823624444
2016	1521139945	1509122141	1521139942	1521139945

4.8 Forecast evaluation for all models

The measures RMSE, MAE and MAPE are used to evaluate the forecasting accuracy of best prediction models. The model with smallest RMSE, MAE and MAPE is the best model. For all the models that have been built, from Table 4.12 that the three best models are M6, M7 and M8, since their RMSE, MAE and MAPE are less compared to those of all other models. Where model M7 is the best of them all. Figure 4.5 shows the time series

Table 4.12: Models forecast evaluation.

Models	RMSE	MAE	MAPE
M1	1290245860	982713622	38.32525
M2	1399749495	1129035054	41.84898
M3	990412816	710514555	20.34699
M4	872954158	390396977	4.742133
M5	935633917	460601035	6.130039
M6	548418213	245260093	2.979163
M7	14295367	6393175	0.1103855
M8	14295370	6393088	0.1103845

plots between the actual FDI and four different types of kernels. For all these four kernels, it can be seen that their forecasts (dotted line) don't follow the actual FDI (solid line) very well, since most forecasts deviate from the actual FDI. On the graph of the linear kernel function, it can be seen that the forecast goes below and above the actual FDI, meaning that this model has both under and over predictions. The Bessel, Gaussian and Laplace kernels have more under predictions. Figure 4.6 and Figure 4.7 are showing the time series plots of the actual FDI and the forecast of the models ANOVA radial

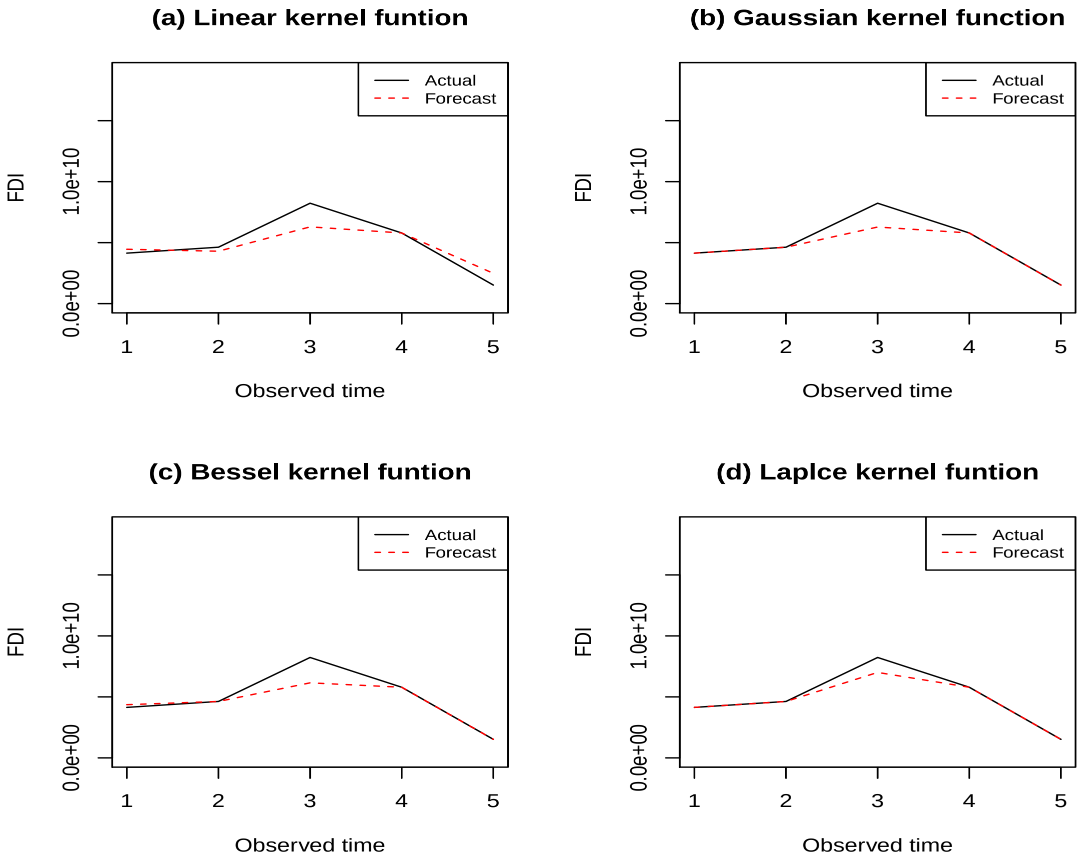

Figure 4.5: Time series plot (kernel functions forecast).

basis kernel function and linear quantile regression averaging. The graphs show that the forecast for these two models follow the actual FDI remarkably well, as there is no any visible deviations.

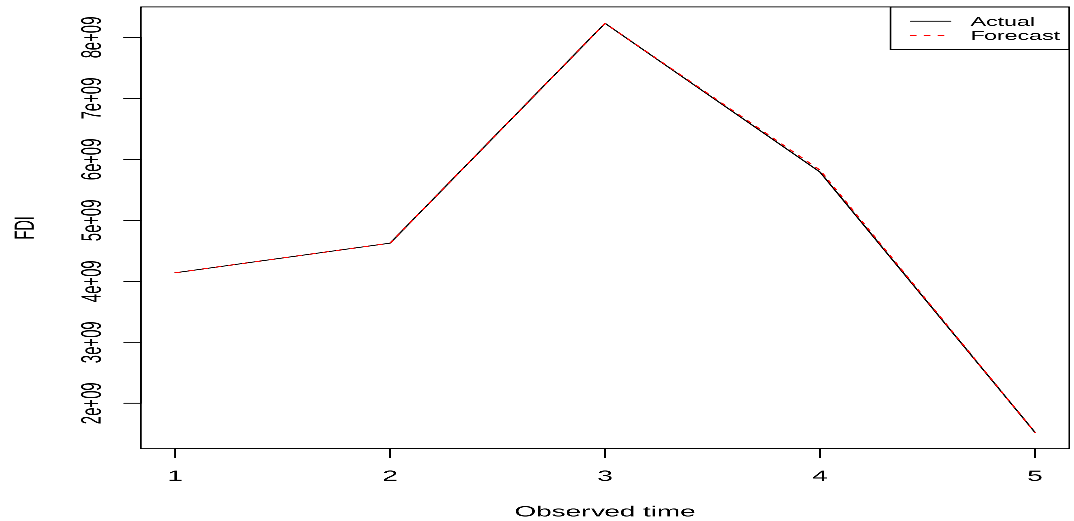

Figure 4.6: Time series plot (ANOVA radial basis kernel function forecast), M7.

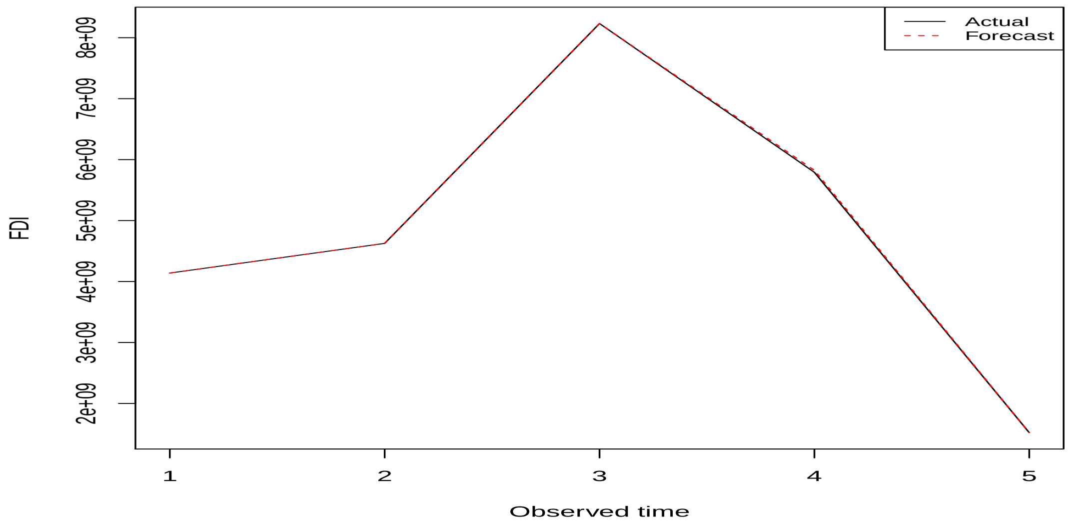

Figure 4.7: Time series plot (LQRA), M8.

4.9 Performance comparison of best models

4.9.1 Prediction intervals

This section presents the comparison of the performance of the models, for the best three chosen models (M6, M7 and M8). Tables 4.13, 4.14 and 4.15 are showing the prediction intervals width (PIW) and residuals for models M6, M7 and M8. All the prediction intervals width were constructed under the 95% confidence level. The PIWs were constructed by taking the difference between the upper bound of 0.975 and lower bound of 0.25 and the residuals where obtained from the difference between the actual FDI and its forecast ($\tau = 0.5$).

Table 4.13: Prediction intervals widths: M6.

Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$	PIW	Residuals
2011	4139289123	4139289119	4139289124	7658408711	3519119592	-1.313
2012	4626029122	4615340098	4626029126	7965387527	3350047429	-3.599
2013	8232518816	5246630120	7253495221	8232518860	2985888740	979023595
2014	5791659020	4862831701	5791659021	8135666559	3272834858	-0.900
2016	1521139945	1840239977	1521139968	7652734546	5812494569	-22.695

Table 4.14: Prediction intervals width: M7.

Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$	PIW	Residuals
2011	4139289123	4139289043	4139289107	6359577977	2220288934	15.6874
2012	4626029122	4626029122	4626029127	6486153500	1860124378	-4.5999
2013	8232518816	7545980351	8232518379	8232518826	686538475	436.6208
2014	5791659020	5791659022	5823624433	7577278212	1785619190	-319654.22
2016	1521139945	1521139941	1521139942	3594033866	2072893925	3.3053

Table 4.15: Prediction intervals widths: M8.

Date	Actual FDI	$Q_{0.25}$	$Q_{0.5}$	$Q_{0.975}$	PIW	Residuals
2011	4139289123	4119050821	4139289108	4139289123	20238302	14.7
2012	4626029122	4598699833	4626029126	4626029127	27329294	-3.6
2013	8232518816	8232518815	8232518815	8232518815	0	0.6
2014	5791659020	5791659020	5823624444	5823624444	31965424	-31965419
2016	1521139945	1509122141	1521139942	1521139945	12017804	3.3

The accuracy check of the prediction intervals was done using PICP, PINAW, PINAD, over prediction count and under prediction count. It can be seen from Table 4.16 that the prediction interval coverage probability

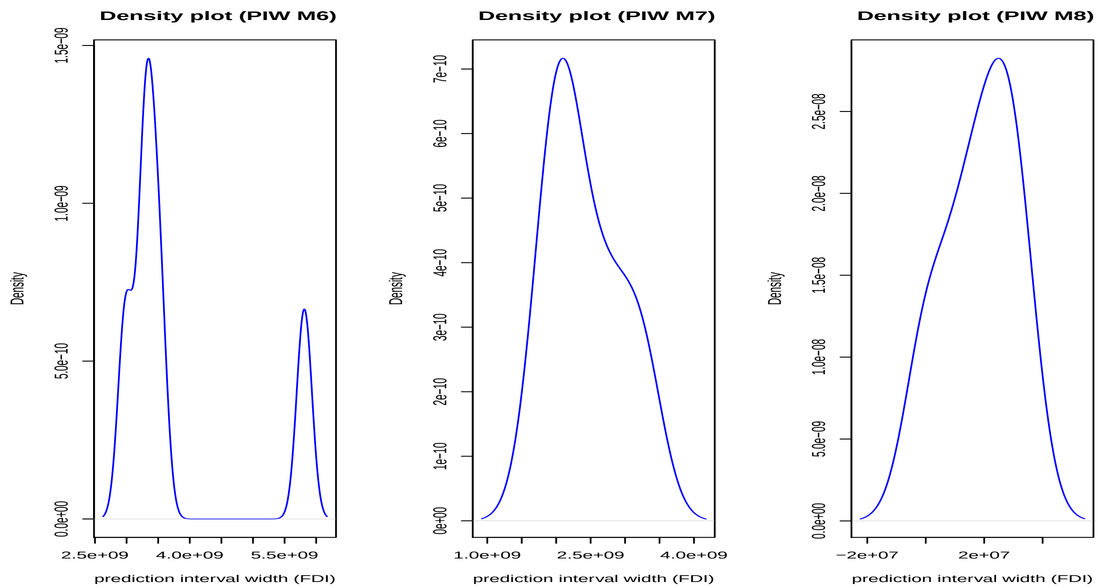

Figure 4.8: Density plot for PIW (M6, M8 and M7, respectively).

(PICP) of model M8 is 100%, meaning that for M8 the prediction interval is covering all the target values. For model M8 all the target values are covered completely. Models M6 and M7 are both having the PICP of 80%, meaning that the models M6 and M7 have 80% coverage. The PICPs of M6 and M7 are not valid since they are less than 95%. This means that for the model M8, the constructed PIs cover the target values with high probability. The PINAW values of the models M6 and M7 are wider than the PINAW value of model M8. The PINAD values of the models M6 and M7 are also wide,

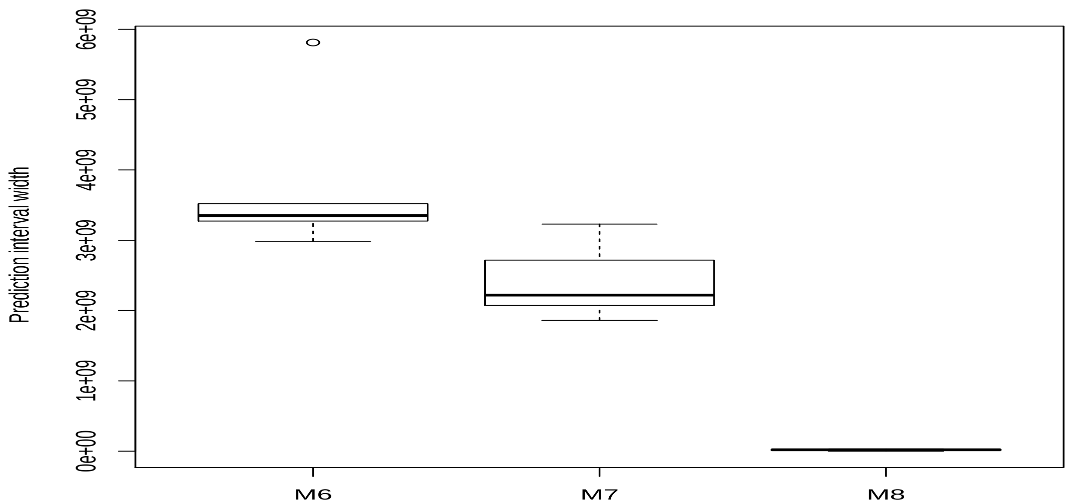

Figure 4.9: Box and whisker plot of PIW (M6, M8 and M7, respectively).

Table 4.16: Comparisons of M6, M7, and M8.

	PICP(%)	PINAW (%)	PINAD (%)	Over prediction	Under prediction
M6	80	56.4	$9.509\times^{-3}$	0	1
M7	80	25.7	$5.662\times^{-11}$	0	1
M8	100	0.27	0	0	0

compared to the value of model M8. Model M8 is considered to be the best model, compared to models M6 and M7. Since model M8 has 100% coverage and narrow PINAW and PINAD. Model M8 was also confirmed to be the superior model by the under and over predictions criteria. Model M8 has zero under and above predictions.

Figure 4.8 and Figure 4.9 are showing the density plots and box and whisker plots for prediction intervals widths for the Models M6, M7 and M8, respectively. The density plot of model M8 seems to be the best compared to the density plot of models M6 and M7, since the density of model M8 is narrow. This also confirms that model M8 is the best.

4.10 Forecast error distribution

Table 4.17 gives summary statistics of the residuals of the best three models. Model M8 has the lowest mean forecast error compared to the models M6 and M7 and also has the lowest standard deviation error. This also shows that model M8 is still the best model. All these three models have the same kurtosis and skewness values, except that, the skewness of model M6 is positive.

Table 4.17: Summary statistics for the residuals of forecast of M6, M7 and M8.

	Mean	Median	Min	Max	Kurtosis	Skewness	Std deviation
M6	195804713	-1	-23	979023595	-0.92	1.073313	437832665
M7	-6392992	3	-31965413	437	-0.92	-1.073313	14295418
M8	-6393081	1	-31965420	15	-0.92	-1.073313	14295372

Chapter 5

Conclusion

5.1 Introduction

The dissertation presented and analysed the non - parametric quantile regression models in forecasting FDI in South Africa over the period of 20 years (years 1996 to 2015). The data for the period 1996 to 2010 will be used for training the models while the data for the period 2011 to 2015 will be used for validation. Forecasts of FDI inflows are important to decision makers in government who have to ensure economic growth in South Africa. A detailed literature survey was presented in Chapter 2, with Chapter 3 discussing the models used in modelling and forecasting FDI.

5.2 Summary and concluding remarks

This section presents a brief summary of analyses and discussion done in Chapter 4. According to the findings in Section 4.2.3 (see Figure 4.1(a)), the FDI inflow of SA seemed to be unstable, since from 1996 to 2015 it has been increasing and decreasing. Lastly being decreasing in 2014 towards 2015 which is not good for the economic growth of SA. In Section 4.3 the Lasso was used for variable selection in order to avoid unnecessary predictors and over fitting. The goal of this analysis was to find which independent variables were most relevant to predict the dependent variable FDI. The Lasso selected all the variables to be the predictors of FDI. This means that the GDP, GFCF, LP, CPI, PI, Exp and noltrend can be regarded as the major determinants of FDI in SA. The noltrend, GDP, GFCF and Exp have more influence on the inflow of FDI compared to the other variables (see Tables 4.2 and 4.3).

Sections 4.4 and 4.5 showed how the two benchmark models forecast the FDI. It was seen that the forecast values for models M1 and M2 do not follow the actual FDI remarkably well (see Figure 4.3 and Figure 4.4). Based on the accuracy measures, model M2 was better compared to model M1 on forecasting FDI (see Table 4.5 and Table 4.7). Model M2 was also chosen to be the best model (on benchmark models) based on the Akaike information criterion and Bayesian information criterion.

Section 4.6 presented the nonparametric local linear kernel quantile regression models. The 95% prediction intervals for all the five kernels were presented (see Tables 4.8 and 4.9). All the actual and forecast values of models M3 to M7 were found to be inside the prediction intervals, even though some of the prediction intervals are wider than others. Even though the prediction intervals for models M3, M4, M5 and M6 cover all the forecast values, the times series plot shows that their forecast values do not follow the actual FDI very well (see Figure 4.5). The forecast values of Model M7 did follow the actual FDI very well, this made model M7 to be the best nonparametric kernel quantile regression model on forecasting FDI. Based on the accuracy measures, model M7 was also confirmed to be the best model, since it had the lowest values of MAPE, MAE and RMSE compared to M3, M4, M5 and M6.

The linear quantile regression averaging (M8) was also built under the nonparametric kernel quantile regression models. Model M8 forecast values follow the actual FDI values very well as shown in Figure 4.7 and all the forecast values were inside the prediction intervals.

The forecast evaluation was done for all the models that were built in this study. Based on the accuracy measures, the best three models were M6, M7 and M8, since their RMSE, MAE and MAPE were less compared to those of all other models. A comparison for the best three chosen models was also done. An analysis of the prediction intervals was done using PICP, PINAW, PINAD, over prediction count and under prediction count. Model M8 was considered to be the best fitting model, compared to models M6 and M7, since it had 100% coverage and narrow PINAW and PINAD. Model M8 was also confirmed to be the superior model for forecasting FDI as it had no under and over predictions.

The forecast error distribution also confirmed that model M8 was the best model for forecasting the FDI. Having the model that can accurately forecast the FDI inflow can help South Africa to increase FDI inflow. This can also help in promoting FDI inflow and reach high economic development.

5.3 Limitations of the study

Yearly data of FDI was used rather than using quarterly data. Quarterly data was not available for some of the variables included in this study.

5.4 Future research

Future research should use quarterly FDI data (to increase the number of observations) and also investigate the effect of FDI in some sectors of the economy such as in industries and services sectors. Future research should also include other predictor variables that are not included in this study and also use other kernels of non-parametric kernel quantile regression models for forecasting FDI, such as the hyperbolic tangent kernel and polynomial kernel.

References

G. Agrawal and M.A. Khan. Impact of fdi on gdp: A comparative study of china and india. *International Journal of Business and Management*, 6 (10):71, 2011.

P. Behl, G. Claeskens, and H. Dette. Focussed model selection in quantile regression. *Statistica Sinica*, 20(2):601–624, 2013.

N. Bermeo. Does electoral democracy boost economic equality? *Journal of Democracy*, 20(4):21–35, 2009.

A.A Bevan and S. Estrin. The determinants of foreign direct investment in transition economies. Technical Report 342, William Davidson Institute at the University of Michigan, 2000.

K. Brooks, J. van Zyl, and R. Townsend. Improving opportunities for the rural poor in South Africa. *World*, 2000.

A. Chakrabarti. The determinants of foreign direct investments: Sensitivity analyses of cross-country regressions. *kyklos*, 54(1):89–114, 2001.

P.F. Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998.

Z. Chunying. A quantile regression analysis on the relations between foreign direct investment and technological innovation in china. In *Information Technology, Computer Engineering and Management Sciences (ICM), 2011 International Conference on*, volume 4, pages 38–41. IEEE, 2011.

B.D. Craven and S.M. Islam. *Ordinary least-squares regression*. Sage Publications, 2011.

J. Friedman, T. Hastie, and R. Tibshirani. glmnet: Lasso and elastic-net regularized generalized linear models. *R package version*, 1(4), 2009.

J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

S. Girma and H. Görg. Foreign direct investment, spillovers and absorptive capacity: Evidence from quantile regressions. Technical Report 1248, Kiel Institute for the World Economy (IfW), 2005.

S. Girma, D. Greenaway, and R. Kneller. Export market exit and performance dynamics: a causality analysis of matched firms. *Economics letters*, 80(2):181–187, 2003.

H. Görg and D. Greenaway. Much ado about nothing? do domestic firms really benefit from foreign direct investment? *The World Bank Research Observer*, 19(2):171–197, 2004.

T. Heskes. Practical confidence and prediction intervals. In *Advances in neural information processing systems*, pages 176–182, 1997.

T. Hofmann, B. Schölkopf, and A.J. Smola. Kernel methods in machine learning. *The annals of statistics*, pages 1171–1220, 2008.

R.J. Hyndman and A.B. Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

W. Jyun-Yi, Hsu C., et al. Does foreign direct investment promote economic growth? evidence from a threshold regression analysis. *Economics Bulletin*, 15(12):1–10, 2008.

A. Karatzoglou, A. Smola, K. Hornik, and M. A. Karatzoglou. Package kernlab. *Google Scholar*, 2016.

A. Khosravi, S. Nahavandi, and D. Creighton. Construction of optimal prediction intervals for load forecasting problems. *IEEE Transactions on Power Systems*, 25(3):1496–1503, 2010.

A. Khosravi, S. Nahavandi, D. Creighton, and A.F. Atiya. Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on neural networks*, 22(9):1341–1356, 2011.

A. Klimek. Greenfield foreign direct investment versus cross-border mergers and acquisitions: the evidence of multinational firms from emerging countries. *Eastern European Economics*, 49(6):60–73, 2011.

R. Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.

R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

R. Koenker and K. Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

R. Koenker and J.A.F Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310, 1999.

R. Koenker and B.J. Park. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71(1):265–283, 1996.

R. Koenker and Z. Xiao. Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612, 2002.

F. Lequiller, D.W. Blades, D. Blades, et al. *Understanding national accounts*. OECD publishing, 2006.

M. Lipsitz, A. Belloni, V. Chernozhukov, and I. Fernández-Val. Nonparametric series quantile regression in r: A vignette. 2015.

J.R. Markusen and A.J. Venables. Foreign direct investment as a catalyst for industrial development. *European economic review*, 43(2):335–356, 1999.

D.P. McMillen. Linear and nonparametric quantile regression. In *Quantile Regression for Spatial Data*, pages 13–27. Springer, 2013.

K.G.M Moons, A. Rogier T. Donders, E.W. Steyerberg, and F.E. Harrell. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. *Journal of clinical epidemiology*, 57(12):1262–1270, 2004.

J. Paniagua and J. Sapena. Do credit constraints reduce foreign jobs? a note on foreign direct employment. *Applied Economics Letters*, 22(3):195–198, 2015.

J. Paniagua, E. Figueiredo, and J. Sapena. Quantile regression for the fdi gravity equation. *Journal of Business Research*, 68(7):1512–1518, 2015.

T.A. Pugel. Foreign investment in the us food-marketing system: Discussion. *American Journal of Agricultural Economics*, 65(2):423–425, 1983.

H. Quan, D. Srinivasan, and A. Khosravi. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE transactions on neural networks and learning systems*, 25(2):303–315, 2014.

L. Resmini. The determinants of foreign direct investment in the ceecs: new evidence from sectoral patterns. *Economics of transition*, 8(3):665–689, 2000.

F. Schneider and B.S. Frey. Economic and political determinants of foreign direct investment. *World development*, 13(2):161–175, 1985.

I. Takeuchi, Q.V. Le, T.D. Sears, and A.J. Smola. Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7(Jul):1231–1264, 2006.

L. Thomas, J. Leape, M. Hanouch, and R. Rumney. Foreign direct investment in south africa: The initial impact of the trade. *Development, and Cooperation Agreement between South Africa and the European Union, CREFSA, London School of Economics*, 2005.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

E.J. Wagenmakers and S. Farrell. Aic model selection using akaike weights. *Psychonomic bulletin & review*, 11(1):192–196, 2004.

D. Wheeler and A. Mody. International investment location decisions: The case of us firms. *Journal of international economics*, 33(1-2):57–76, 1992.

K. Yamaoka, T. Nakagawa, and T. Uno. Application of akaike's information criterion (aic) in the evaluation of linear pharmacokinetic equations. *Journal of Pharmacokinetics and Pharmacodynamics*, 6(2):165–175, 1978.

K. Yu and M.C. Jones. Local linear quantile regression. *Journal of the American statistical Association*, 93(441):228–237, 1998.

Y. Zhang and Y. Yang. Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1):95–112, 2015.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.

Appendix

Appendix A: Graphs

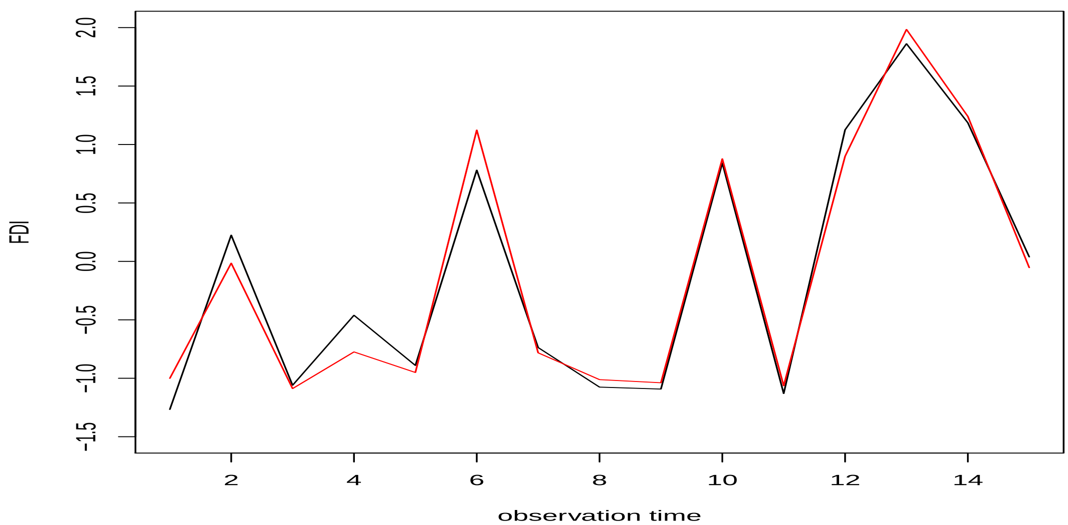

Figure A1: Time series plot (OLS).

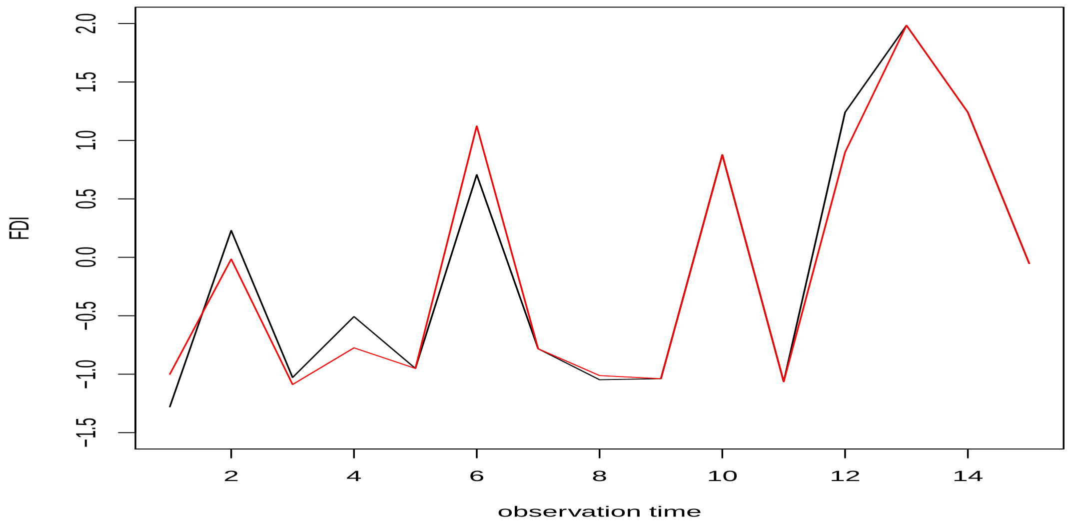

Figure A2: Time series plot(Linear quantile regression).

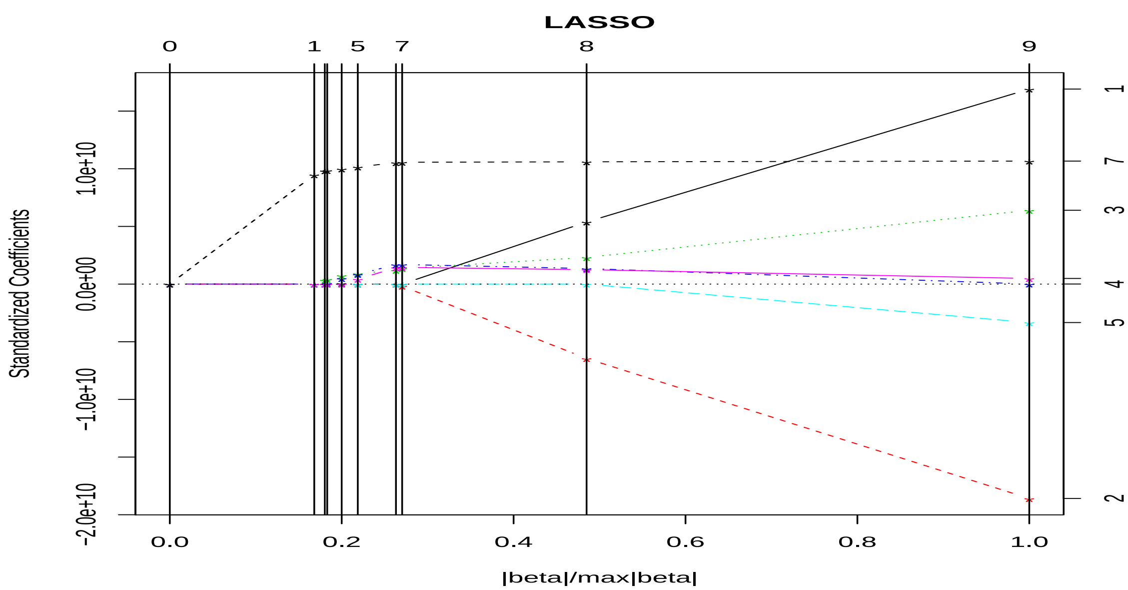

Figure A3: Variable selection: Lasso.

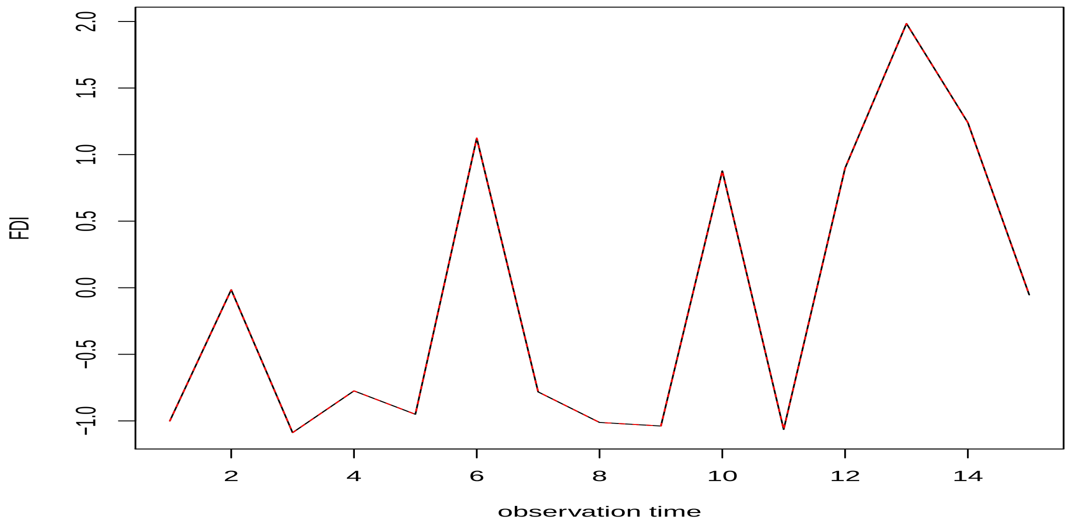

Figure A4: Time series plot (ANOVA radial basis kernel function).

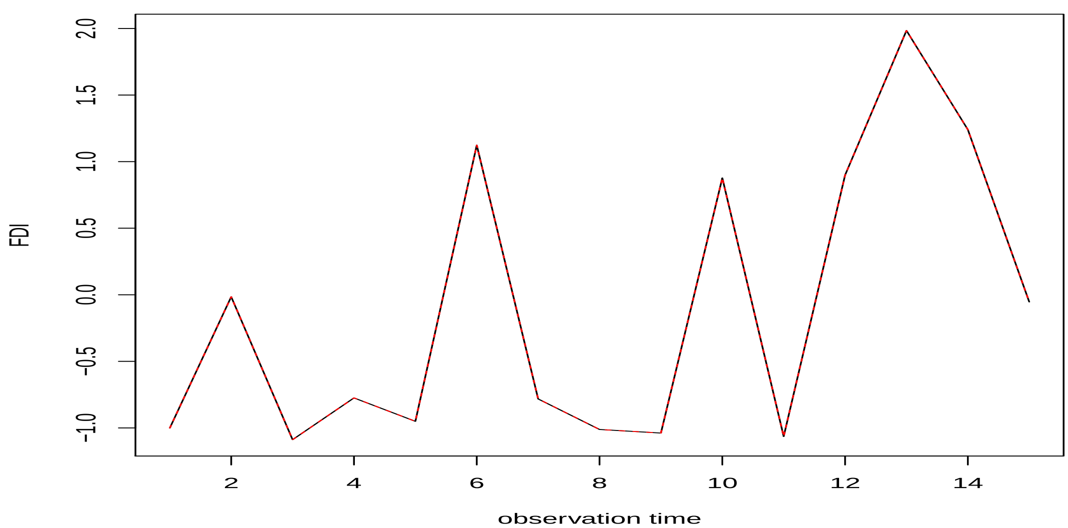

Figure A5: Time series plot (LQRA).

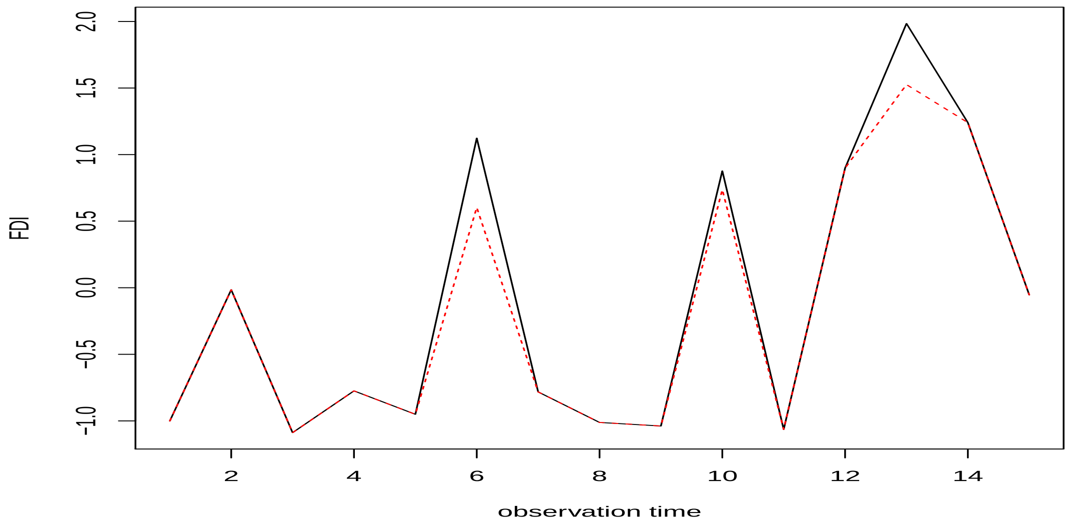

Figure A6: Time series plot (Laplace kernel function).

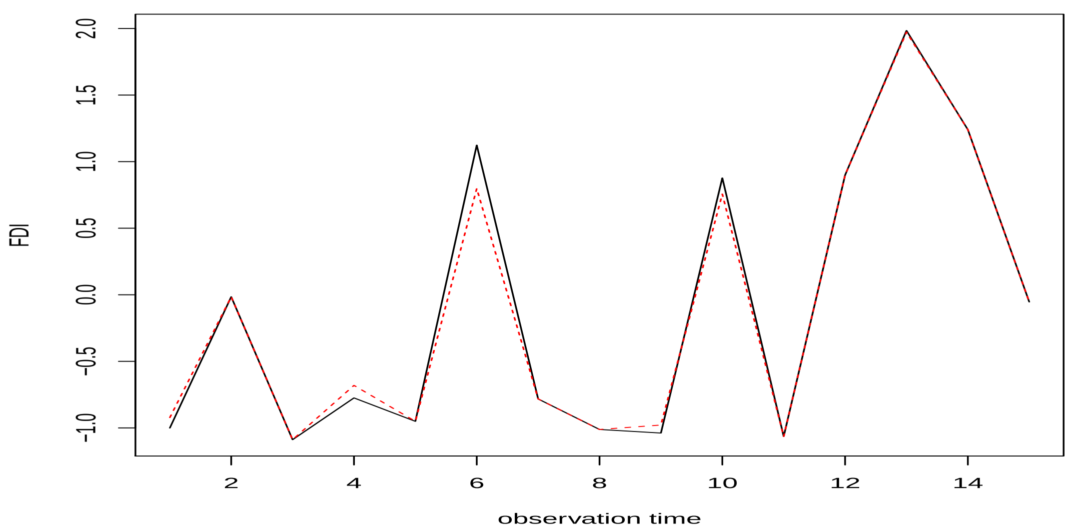

Figure A7: Time series plot (Bessel kernel function).

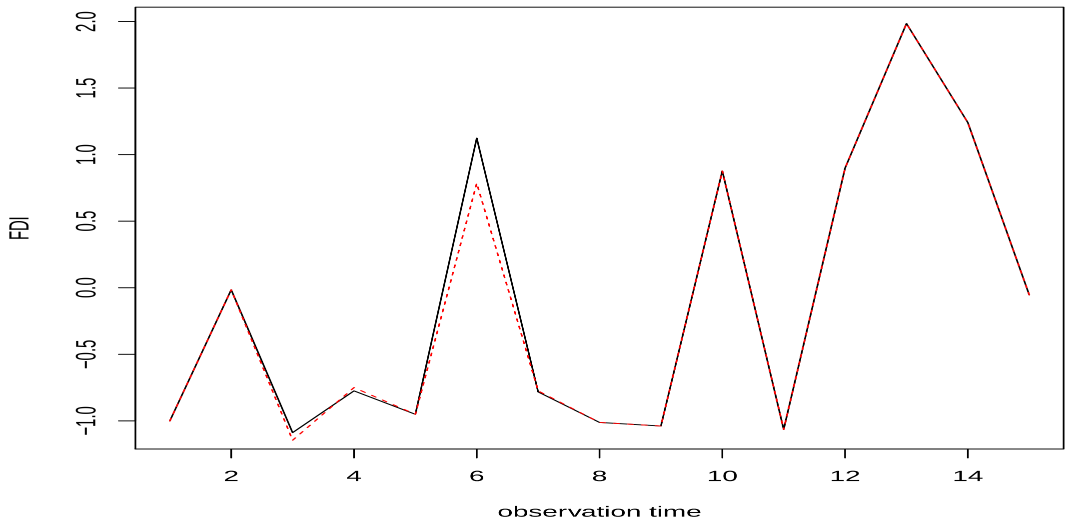

Figure A8: Time series plot (Gaussian kernel function).

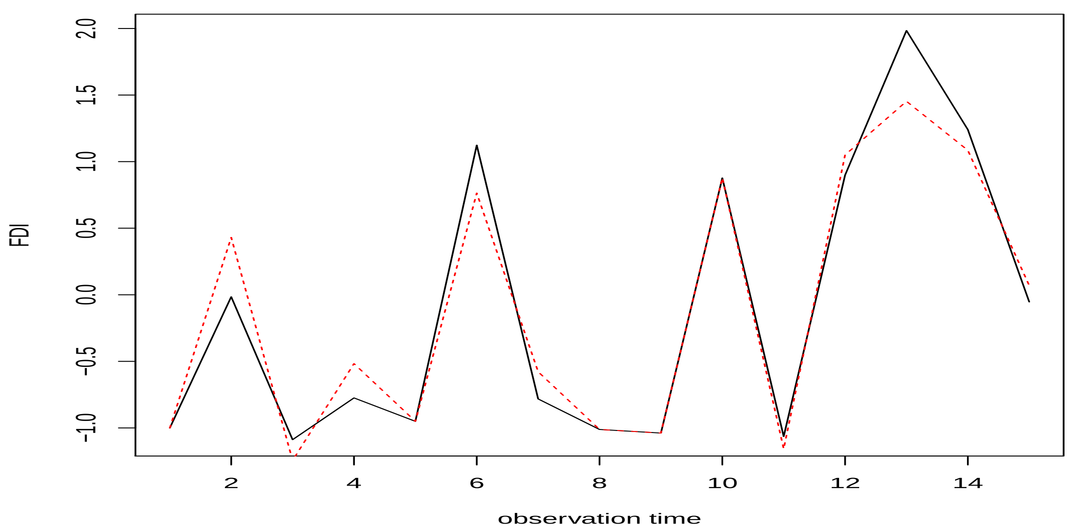

Figure A9: Time series plot (Linear kernel function).

Appendix B

```
#################################################################
## The following packages were used in this dissertation:#######
## mgcv, quantreg, forecast, gefcom2017, kernlab, e1071. #######
#################################################################



##################################################################
################### NOLTREND #####################################
##################################################################

win.graph()
x=ts(FDI)
plot(x)
z=(smooth.spline(time(FDI), FDI))
z
lines(smooth.spline(time(FDI), FDI, spar=1.494635),
      col="red",lwd=3) # spar = 0.01
dpdfits = fitted((smooth.spline(time(FDI), FDI, spar=1.494635)))
write.table(dpdfits,"~/nolfits.txt",sep="\t")

##################################################################
###################### VARIABLE SELECTION #######################
##################################################################

attach(datatrain)
```

```
head(datatrain)
win.graph()
library(glmnet)

x = cbind(GFCF,GDP,Exports,CPI,LP,PS,noltrend)
y=FDI

set.seed(2468)
cvfitlasso=cv.glmnet(x,y,alpha = 1, intercept = TRUE, type.measure = "mse")
cvfitlasso
plot(cvfitlasso)
coef(cvfitlasso, s=1503622) # s= Lambda

##########################################################################
############# FITTING MODELS #############################################
##########################################################################

attach(analyticdataST)
head(analyticdataST)
win.graph()

standardize <- function(x){(x-mean(x))/sd(x)}
z0 <- standardize(FDI)
z1 <- standardize(GFCF)
z2 <- standardize(GDP)
z3 <- standardize(Exports)
```

```
z4 <- standardize(CPI)
z5 <- standardize(LP)
z6 <- standardize(PS)
z7 <- standardize(noltrend)
z8 <- standardize(t)
z =cbind(z0,z1,z2,z3,z4,z5,z6,z7)
write.table(z,"~/standardized.txt",sep="\t")

y <- ts(z0,start=1996, freq=1)

par(mfrow=c(2,2))
plot(y,ylab="Stardadized FDI",col="blue",xlab="Year",
     main="(a) Time series plot")
plot(density(y),xlab="FDI)", col="blue",main="(b) Density plot")
qqnorm(y, main="(c) Normal QQ plot",col="blue")
qqline(y)
boxplot(y, main="(d) Box plot",horizontal = TRUE,col="blue")

#############################################################
##########  CORRELATION MATIX ###############################
#############################################################

x = cbind(z0,z1,z2,z3,z4,z5,z6,z7)
cor(x, method = "pearson", use = "complete.obs")

par(mfrow=c(3,2))
```

```
plot(z0,z1,xlab="FDI", ylab="GFCF")
plot(z0,z2, xlab="FDI", ylab="GDP")
plot(z0,z3, xlab="FDI", ylab="Exports")
plot(z0,z4, xlab="FDI", ylab="CPI")
plot(z0,z5, xlab="FDI", ylab="LP")
plot(z0,z6, xlab="FDI", ylab="PS")

#plot(FDI,noltrend)
#t=c(1:length(FDI))
#t
#plot(FDI,t)
##########################################################
######## NOLTREND  #######################################
##########################################################

win.graph()
x=ts(z0)
plot(x)
z=(smooth.spline(time(x), x))
z
lines(smooth.spline(time(x), x, spar=0.01),col="red",lwd=3) # spar = 0.01
dpdfits = fitted((smooth.spline(time(FDI), FDI, spar=1.494635)))
write.table(dpdfits,"~/nolfits.txt",sep="\t")


##########################################################
```

```
##################### MODELS  ############################
##########################################################

win.graph()
x=ts(z0)
plot(x)

data_test <- 16:nrow(analyticdataST)
data_train <- analyticdataST[-data_test, ]
data_test <- analyticdataST[data_test, ]

###########################################################
#######  LINEAR MODEL  ####################################
###########################################################

g = lm(z0~z1+z2+z3+z4+z5+z6+z7,data=data_train)
summary(g)

f=ts(fitted(g))
plot(f,ylim=c(-1.5,2))
y=data_train$z0
lines(y,col="red")

res1 = residuals(g)
#plot(res1)
```

```
library(forecast)
library(tseries)
tsdisplay(res1,col="blue")
Box.test(res1, lag=36, fitdf=2, type="Ljung")
auto.arima(res1)

g.forecast <- predict(g, newdata = data_test)
g.forecast
write.table(g.forecast,"~/forecastLM.txt",sep="\t")

####################################################################
##################### Quantile regression  ########################
####################################################################

library(quantreg)
qr.g = rq(z0~z1+z2+z3+z4+z5+z6+z7,data=data_train, tau=0.5)

#summary(qr.g,se = "nid")
#summary.rq(qr.g,se="ker" )
summary.rq(qr.g,se="boot" )

library(forecast)
res2=residuals(qr.g)
tsdisplay(res2)
Box.test(res2, lag=36, fitdf=2, type="Ljung")
auto.arima(res2)
```

```
g.forecast <- predict(qr.g, newdata = data_test)
g.forecast
write.table(g.forecast,"~/forecastLQR.txt",sep="\t")

######################################################################
######################################################################

attach(forecastsAM)
head(forecastsAM)
library(forecast)

accuracy(fLM, FDI)
accuracy(fLQR, FDI)

####################################################################
############## PINBALL LOSS FUNCTION  ##############################
####################################################################

install.packages("devtools")
library(devtools)
install_github("camroach87/gefcom2017")

attach(forecasts)
head(forecasts)
win.graph()
```

```
library(gefcom2017)
## tau: integer 1, 2, ... 99. Quantile to calculate pinball loss score for.
## y: numeric. Observed value.
## q: numeric. Predicted value for quantile tau.
#' Calculates the pinball loss score for a given quantile.
pinball_loss <- function(tau, y, q) {
  pl_df <- data.frame(tau = tau,
                      y = y,
                      q = q)
  pl_df <- pl_df %>%
    mutate(L = ifelse(y>=q,
                      tau/100 * (y-q),
                      (1-tau/100) * (q-y)))
  return(pl_df)
}
tau= 50
y= DPED
q= fgbm # fqgam, fqgamI, fgbm, mix.opera

z = pinball_loss(tau, y, q)
z
#write.table(z,"~/pinballfplLassoI.txt",sep="\t")

qloss =z$L
a=ts(qloss)
```

```
plot(a)
mean(qloss)


######################################################################
#################### NONPARAMETRIC QUANTILE REGRESSION ###############
######################################################################
#################### KERNEL FUNCTIONS  ###############################
######################################################################

attach(analyticdataST)
head(analyticdataST)
win.graph()
#y=ts(z0, start=1996, freq=1)
#plot(y, xlab="Year", ylab="Standardized FDI")

library(kernlab)
y=z0
x = cbind(z1,z2,z3,z4,z5,z6,z7)

# first calculate the median

qrm <- kqr(x,y, tau = 0.5,kernel="rbfdot", C= 3.8)
#vanilladot, rbfdot, besseldot, laplace, anovadot
qrm
win.graph()
```

```
# predict and plot
plot(z0,type="l", lty=1,ylab="FDI", xlab="observation time")
ytest <- predict(qrm, x)
ytest
lines(ytest, col="red",type="l", lty=2)

write.table(ytest,"~/rbf.txt",sep="\t")

################################################################
attach(analyticdataST)
head(analyticdataST)
win.graph()

#y=ts(z0, start=1996, freq=1)
#plot(y, xlab="Year", ylab="Standardized FDI")

library(kernlab)
y=z0
x = cbind(z1,z2,z3,z4,z5,z6,z7)

z = cbind(lk,gk,bk,laK,anova)

# first calculate the median
qrm <- kqr(x,y, tau = 0.5,kernel="rbfdot", C= 3.8)
#vanilladot, rbfdot, besseldot, laplace, anovadot
```

```
qrm

# predict and plot
plot(z0,type="l", lty=1,ylab="FDI", xlab="observation time")
ytest <- predict(qrm, x)
ytest
lines(ytest, col="red",type="l", lty=2)

write.table(ytest,"~/rbf.txt",sep="\t")

#####################################################################

attach(Book11)
head(Book11)
win.graph()
a=ts(z0)
b=ts(rbf)
plot(a,type="l", lty=1,ylab="FDI", xlab="observation time")
lines(b, col="red",type="l", lty=2)

###################################################################
## linear quntile regression averaging ############################
###################################################################

attach(LLQRA)
head(LLQRA)
```

```
win.graph()
plot(z0,type="l", lty=1, ylab = "FDI", xlab = "Observed time")
library(quantreg)

fit = rq(z0~lk+gk+bk+laK+anova,data=LLQRA, tau=0.5)

#summary(qr.g,se = "nid")
#summary.rq(qr.g,se="ker" )
summary.rq(fit,se="boot" )
fittedQRA =fitted(fit)
lines(fittedQRA, col="red",type="l", lty=2)

write.table(fittedQRA,"~/forecast7QRA.txt",sep="\t")

###########################################################################
attach(forecastsAM)
head(forecastsAM)
win.graph()
par(mfrow=c(2,2))
plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time",
     main = "(a) Linear kernel funtion",ylim = c(0,19000000000))
lines(kqr_vanilla, col="red",type="l", lty=2)
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)

plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time",
```

```
     main = "(b) Gaussian kernel function",ylim = c(0,19000000000))
lines(kqr_rbf, col="red",type="l", lty=2)
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)

plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time",
     main = "(c) Bessel kernel funtion",ylim = c(0,19000000000))
lines(kqr_bessel, col="red",type="l", lty=2)
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)

plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time",
     main = "(d) Laplce kernel funtion",ylim = c(0,19000000000))
lines(kqr_laplace, col="red",type="l", lty=2)
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)

win.graph()
plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time")
lines(kqr_anova, col="red",type="l", lty=2)
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)

win.graph()
plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time")
lines(LQRA, col="red",type="l", lty=2)
```

```
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)



win.graph()
plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time")
lines(fLM, col="red",type="l", lty=2)
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)

win.graph()
plot(FDI,type="l", lty=1, ylab = "FDI", xlab = "Observed time")
lines(fLQR, col="red",type="l", lty=2)
legend("topright", legend = c("Actual ", "Forecast"), col = c("black","red"),
       lty = 1:2, cex = 0.8)

win.graph()
par(mfrow=c(1,3))

plot(density(M6),xlab="prediction interval width (FDI)",
     col="blue",main=" Density plot (PIW M6)")
plot(density(M7),xlab="prediction interval width (FDI)",
     col="blue",main=" Density plot (PIW M7)")
plot(density(M8),xlab="prediction interval width (FDI)",
     col="blue",main=" Density plot (PIW M8)")
```

```
win.graph()
par(mfrow=c(1,3))
boxplot(M6, ylab="prediction interval width (FDI)",
        main=" Box plot (PIW M6)",vertical = TRUE,col="blue")
boxplot(M7, ylab="prediction interval width (FDI)",
        main=" Box plot (PIW M7)",vertical = TRUE,col="blue")
boxplot(M8, ylab="prediction interval width (FDI)",
        main=" Box plot (PIW M8)",vertical = TRUE,col="blue")

win.graph()
par(mfrow=c(1,3))

plot(density(Rlap),xlab="Residuals (FDI)", col="blue",
     main=" Density plot (Res M6)")
plot(density(Rano),xlab="Residuals (FDI)", col="blue",
     main=" Density plot (Res M7)")
plot(density(RLQ),xlab="Residuals (FDI)", col="blue",
     main=" Density plot (Res M8)")


win.graph()
par(mfrow=c(1,3))
boxplot(Rlap, ylab="Residuals (FDI)", main=" Box plot (Res M6)",
        vertical = TRUE,col="blue")
boxplot(Rano, ylab="Residuals (FDI)", main=" Box plot (Res M7)",
        vertical = TRUE,col="blue")
```

```
boxplot(RLQ, ylab="Residuals (FDI)", main=" Box plot (Res M8)",
        vertical = TRUE,col="blue")

##############################################################

summary (Rano)
summary(Rlap)
summary(RLQ)
sd (Rlap)
sd (Rano)
sd (RLQ)
win.graph()
PIW <- c("M6", "M7", "M8")
boxplot(M6, M7, M8, names= PIW, horizontal = FALSE,
        main="",ylab=" Prediction interval width",ylim = c(0,9000000000))

par(mfrow=c(1,3))
plot(density(residAQRI,bw="SJ"),xlab="Forecast error(MW)",
     col="blue", main="AQRI")
plot(density(residConvex, bw="SJ"),xlab="Forecast error(MW)",
     col="blue", main="Convex")
plot(density(residQRA,bw="SJ"),xlab="Forecast error(MW)",
     col="blue", main="QRA")

###################################################################
```

```
attach(forecastsAM)
head(forecastsAM)
library(forecast)

accuracy(fLM, FDI)
accuracy(fLQR, FDI)
accuracy(kqr_vanilla, FDI)
accuracy( kqr_rbf, FDI)
accuracy(kqr_bessel, FDI)
accuracy(kqr_anova, FDI)
accuracy(kqr_laplace, FDI)
accuracy(LQRA, FDI)

library (e1071)
skewness(Rlap)
skewness(Rano)
skewness(RLQ)
kurtosis(Rano)
kurtosis(RLQ)
kurtosis(Rlap)
```