Deanship of Graduate Studies

Al-Quds University

Linear Regression Models Assuming a Stable Distribution with Applications

Layla Khalid Mahmoud Lahaleeh

M.Sc. Thesis

Jerusalem-Palestine

1441/2020

Linear Regression Models Assuming a Stable Distribution with Applications

Prepared by:

Layla Khalid Mahmoud Lahaleeh

B.Sc: Computer Science-Al-Quds University/ Palestine

Supervisor: Dr. Khalid Salah

A thesis submitted in partial fulfillment of requirements for the degree of Master of Mathematics, Department of Mathematics / Graduate Studies / Al-Quds University.

1441\2020

Thesis Approval

Linear Regression Models Assuming a Stable Distribution with Applications

Prepared by: Layla Khaled Mahmoud Lahaleeh

Registeration No.: 21711994

Supervisor: Dr. Khalid Salah

Master Thesis submitted and accepted, Date: 31 / 8 / 2020

The names and the signatures of the examining committee members are as

1- Head of Committee: **Dr. Khalid Salah.** Signature:

2- Internal Examiner: **Dr. Jamil Jamal** Signature: Jamil Jamal

3- External Examiner: **Dr. Mahmoud AlManassra** Signature:

Jerusalem-Palestine

1441/2020

Dedication

To my father, my family, my respected doctors, my colleagues at Al-Quds University and all math lovers.

Layla lahaleeh

Declaration

I certify that this submitted for the degree of master is the result of my own research, except where otherwise acknowledge. And that this (or any part of the same) has not been submitted to a higher degree to any other university or institution.

Signature: *Layla K. Lahaleeh*

Student's name: Layla Khalid Mahmoud Lahaleeh

Date: 31 / 8 / 2020

Acknowledgement

First of all, I thank God for reaching this stage. Peace and blessing upon our first teacher and educator, Prophet Mohammed, who has taught the whole world.

Big thanks to my supervisors Dr. Khalid Salah for all his support, guidance, and encouragement to successfully finish this thesis despite the difficult circumstances we live in. I also extend my sincere thanks to the examiners who contributed to improving the message through their constructive comments and advice.

I thank the President of the University, Dr. Imad Abu Kishek, for giving me all the support.

I thank my colleagues at Al-Quds University, especially Ramzi Jafar, for his continued support always.

Abstract

A linear regression is a form of mathematical model that reflects results in a straight-line relationship between two variables (X and V), A regression model is given: $X = d_0 + d_1V + \epsilon$ Usually ϵ considered to be normally distributed with mean of 0 and variance σ^2 Moreover, in regression follows the normal distribution because it is inherited from.

In practical applications, we need to check if the normality assumption is satisfied, if the assumption of normality is violated, or outliers are present, then the linear regression goodness of fit test may not be the most powerful or informative test available, and this could mean the difference between detecting a linear fit or not. In this situation, we could use nonparametric regression models or assume more robust probability distributions for the data. One possibility is to assume that the random variable X has a stable distribution $X{\sim}S_\alpha(\beta, \delta, \gamma)$.

It is well known that, in general, there is no closed form for the probability density function of stable distributions. However, under a Bayesian approach, the use of a latent or auxiliary random variable gives some simplification to obtain any posterior distribution when related to stable distributions. To illustrate the usefulness of calculations, the method is applied to two applications: one is related to a standard linear regression model assuming a normal distribution, and the other is related to the same model assuming a stable distribution. Using MCMC (Markov Chain Monte Carlo) method and r software, interesting posterior summaries were obtained.

Table of Contents

Declaration i

Acknowledgment ii

Abstract iii

Table of Contents iv

List of Tables vi

List of Figures vii

1 Introduction 1

1.1 Background 3

1.2 Objective of the Thesis 3

1.3 Scope and Outline of the Thesis

2 Linear Regression Analysis 6

2.1 Linear Regression Models 6

2.2 Sample Regression Function 9

2.3 Parameters Estimation 12

2.3.1 The Ordinary Least Squares (OLS) Estimates 13

2.3.2 The Maximum Likelihood (ML) Estimates 14

2.4 The Bayesian Estimates 16

2.4.1 Marginal Posterior Distribution of d_1 20

2.4.2 Marginal Posterior Distribution of d_0 22

2.4.3 Marginal Posterior Distribution of σ^2 24

3 Stable Distribution 26

3.1 Introduction 26

3.2 Definitions, features, and graphics of stable distribution 26

3.3 Stable density and distribution functions 37

3.4 Logarithmic moments 41

3.5 methods of parameter estimation of stable distribution 42

3.5.1 The quantiles method 42

3.5.2 Empirical characteristic function method 43

3.5.3 Logarithmic moments method 46

3.5.4 Maximum likelihood method 48

4 Linear Regression Models Assuming a Stable Distribution 50

4.1 Introduction 50

4.2 Linear Regression Models Assuming a Stable Distribution 56

4.3 Bayesian Analysis for Linear Regression Models Assuming a Stable Distribution 57

4.3.1 The Gibbs Sampler 59

4.3.2 Random Variate Generation From $\pi(\alpha|\beta,\delta,\gamma,x,y)$ 60

4.3.3 Random Variate Generation From $\pi(\beta|\alpha,\delta,\gamma,x,y)$ 61

4.3.4 Random Variate Generation From $\pi(\delta|\alpha,\beta,\gamma,x,y)$ 61

4.3.5 Random Variate Generation From $\pi(\gamma|\alpha,\beta,\delta,x,y)$ 62

5 Applications 64

5.1 Data Analysis 64

5.2 Bayesian Approach 67

5.3 Bayesian Approach Assuming Stable Distribution 70

5.4 Maximum Likelihood Approach 75

5.5 Discussion 77

References 79

Appendices 82

Arabic Abstract 86

List of Tables

Table No.	Table’s name	page
3.1	Comparison of tail probabilities for standard normal, Cauchy and Levy distributions.	32
5.1	A sample data set of car Prices and Sales in thousands.	64
5.2	ANOVA table of linear regression model.	64
5.3	Coefficients of linear regression model.	65
5.4	Sales variable descriptive statistics.	66
5.5	Monte Carlo Summary statistics of the parameters estimate.	69
5.6	Stable Distribution Summary Statistics of the Parameters Estimate.	72
5.7	Model Parameters’ Estimates for Different Distributions.	75

List of Figure

Figure's No.	Figure's Name	Page
2.1	The population regression function (PRF).	10
2.2	The Scatter Diagram.	10
2.3	The population regression function and the scatter diagram.	11
2.4	The sample regression function and the scatter diagram	12
3.1	Heavy tailed.	27
3.2	A semi-logarithmic plot of symmetric (β= 0) stable densities for four value of α. Note, the distinct behavior of the Gussian α = 2 distribution.	27
3.3	A plot of stable densities for α = 1.2 and four values of β.	28
3.4	Dependence on alpha.	28
3.5	Tails of stable laws	29
3.6	Graphs of standarizad normal N(1,0), Cauchy (1,0) and levy (1,0) destination	32
3.7	Dependence on beta.	35
3.8	Gaussian, Cauchy, and Levy distributions.	35
3.9	S parameterization	36
3.10	S^0 parameterization	37
5.1	Density Plot of the Dependent Variable (Sales)	65
5.2	Standardized residuals plot	66
5.3	Normal P-P Plot of Regression Standardized Residual	66
5.4	Box-and-Whiskers Plot of dependent variable.	67
5.5	Posterior Time Series, Density, Average Values Plot of All Parameters	69
5.6	Plots of Observed, Fitted Means Considering Least Squares Model.	73
5.7	Plots of Observed, Fitted Means Considering Bayesian Model.	73
5.8	Plots of Observed, Fitted Means Considering Stable Model.	73

Chapter 1

Introduction

1.1 Background

The concept of linear regression was first proposed by Sir Francis Galton in 1894. Linear regression is a statistical test applied to a data set to define and quantify the relation between the considered variables. In biomedical or clinical research, the researcher often tries to understand or relate two or more independent (predictor) variables to predict an outcome or dependent variable. This may be understood as how the risk factors or the predictor variables or independent variables account for the prediction of the chance of a disease occurrence, i.e., dependent variable. Risk factors (or dependent variables) associate with biological (such as age and gender), physical (such as body mass index and blood pressure [BP]), or lifestyle (such as smoking and alcohol consumption) variables with the disease. The goal of the regression model is to determine the relationship of the straight line connecting x and v.

A regression model containing only one predictor variable is called a simple regression model $x = d_0 + d_1 v + \epsilon$. After the model has been defined and the data have been collected, the next task is to estimate the parameters of the model based on the collected data. The techniques used for parameter estimation are called estimators. In this study approaches will be considered: Least squares method, Maximum Likelihood Estimation, the other estimation method that we consider in this research is the Bayesian method.

A Bayesian analysis of stable distributions is introduced by [16] using Markov Chain Monte Carlo (MCMC) methods. The use of Bayesian methods with MCMC simulation can have great flexibility by considering latent variables, where samples of latent variables are simulated in each step of the Gibbs or Metropolis-Hastings algorithms. The appearance of outliers will absolutely affect the regression model under standard normality assumptions. The ideal results not affected by outliers could be obtained using the stable distribution. There are now reliable computer programs to compute stable densities, distribution functions and quantiles. With these programs, it is possible to use stable models in a variety of practical problems. We will provide a wide range of distributions that encompasses the Gaussian one is given by the class of stable distributions. This large class defines location-scale families that are closed under convolution.

This distribution family is described by four parameters α, β, δ and γ. The $\alpha \in (0, 2]$ parameter defines the "fatness of the tails", and when $\alpha = 2$ this class reduces to Gaussian distributions. The $\beta \in [1, 1]$ is the skewness parameter and for $\beta = 0$ there are symmetric distributions. The location and scale parameters are, respectively, $\delta \in (-\infty, \infty)$ and $\gamma \in (0, \infty)$. Stable distributions are usually denoted by $S_\alpha(\beta, \delta, \gamma)$. If a random variable $X \sim S_\alpha(\beta, \delta, \gamma)$, then $Z = \frac{x-\delta}{\gamma} \sim S_\alpha(\beta, 0, 1)$, whenever $\alpha \neq 1$ (see [13],[16]). The difficulty associated with stable distributions $S_\alpha(\beta, \delta, \gamma)$ is that, in general, there is no simple closed form for their probability density functions. However, it is known that the probability density functions of stable distributions are continuous. Also the support of all stable distributions is given in $(-\infty, \infty)$, except for $\alpha < 1$ and $\mid \beta \mid= 1$ when the support is $[\delta, \infty)$ for $\beta = 1$ and $(-\infty, \delta]$ for $\beta = -1$ see [16]. It known that asset returns are not normally distributed. Rather, the empirical observations exhibit fat tails. This heavy tailed or leptokurtic character of the distribution of price changes has been repeatedly observed in various markets and may be quantitatively measured by the kurtosis in excess of 3, a value obtained for the normal distribution. Returns are the cumulative outcome of a vast number of pieces of information and individual decisions arriving almost continuously in time ([14],[15]). The strongest statistical argument for it is based on the Central Limit Theorem, which states that the sum of a large number of independent, identically distributed variables from a finite variance distribution will tend to be normally distributed. Since stable distributions can accommodate the fat tails and asymmetry, they often give a very good fit to empirical data. In particular, they are valuable models for data sets covering extreme events.

1.2 Objective of the Thesis

The purpose of this study can be summarized as follows:

- To study and investigate the nature and assumptions of linear regression, more specifically, when the normality assumption is violated.
- To study the properties of stable distributions and how and when to use it in linear regression.
- To investigate different approaches for parameter estimation, such as, Least squares, MLE, and Bayesian.
- To derive required formulas, necessary for parameter estimations, like, likelihood functions, and posterior distributions.
- To execute an application to understand the accuracy and validity of the linear model related to the stable distribution through simulation and real data sets. In addition, the nature of the inspection data can be adapted to this type of model. This can be done by writing the program code using suitable software.

1.3 Scope and Outline of the Thesis

This thesis deals exclusively with linear regression when the model fitting criteria are violated or not met. The existence of outliers may lead to non-normal assumptions, in this case an alternative approach or distribution must be considered. A strong emphasis is placed on model fit, parameter estimation and distribution throughout the thesis because this is the main objective of many regression analyses. However, the interpretation of the observed distribution pattern is meaningful, especially the interpretation related to abnormal data, which must be dealt with here. This should not be seen as a narrow use of data obtained by abnormal populations, but should focus on the appropriate distribution of such data.

Since the main goal of this research is the development and application of linear regression assuming stable distribution and other programs, different parameter estimation methods are considered. In this regard, some mathematical methods of Bayesian analysis include stable distribution, using latent or auxiliary When related to a stable distribution, random variables help to obtain any posterior distribution.

To show the usefulness of the computational aspects, the methodology is applied to linear regression models. Posterior summaries of interest are obtained using the r software. For many applications in data analysis, using stable distributions may be a good choice because they are highly adaptable and therefore have powerful inference results. With the use of Bayesian methods and MCMC simulation algorithms, it is possible to get inferences for the model despite of the nonexistence of an analytical form for the density function, to point out that the computational work in the sample simulations the joint posterior distribution of interest can be greatly simplified using standard free software like the r software and Stable regression program (stablereg.exe).

For most illustrations, different observations are distinguished and represented by two types of data. Normal are considered solitary and rarly obtained in real data life, while the abnormal high or low usually presented in most observations. Although this may appear to limit a full discussion of results, generalization to other distributions of data requires only slight modifications in most cases.

This thesis is divided into five chapters. The first chapter is an introductory chapter that introduces the main themes of the study. The second chapter introduces the background information of linear regression, the definition of linear regression analysis, linear regression models and parameter estimation techniques.

A general introduction to the stable distribution, clarifying the definition, attributes, graphics, and methods of estimating the parameters of the stable distribution introduced in Chapter 3.

Linear regression models assuming a stable distribution for response data and a Bayesian analysis for parameter estimating including prior and posterior derivation will be introduced in the fourth chapter.

The last chapter will introduce the application of linear regression assuming a stable distribution, including simulation studies, analysis of real data sets and comparative studies, and then summarize and conclude.

Chapter 2

Linear Regression Analysis

In statistical analysis, *regression analysis* is a set of statistical processes for estimating the relationships between two or more variables. *Explanatory Variables,* also known as the independent or predictor variables, independent variable explains variations in the response variable; in an experimental study, it is manipulated by the researcher, moreover, Independent variables are controlled inputs. *Response variables*, also known as dependent variables, these variables represent the output or outcome resulting from altering these inputs, dependent variable's value is predicted or its variation is explained by the explanatory variable; in an experimental study, this is the outcome that is measured following manipulation of the explanatory variable.

A linear regression is a form of mathematical model that reflects results in a straight-line relationship between two variables (*V* and *X*) rather than a curve as in the case of a non-linear regression. In this research the linear regression only be considered.

2.1 Linear Regression Models

Regression analysis is a conceptually simple method for investigating functional relationships among variables. **A** real estate appraiser may wish to relate the sale price of a home from selected physical characteristics of the building and taxes (local, school, county) paid on the building. We may wish to examine whether cigarette consumption is related to various socioeconomic and demographic variables such as age, education, income, and price of cigarettes. The relationship is expressed in the form of an equation or a model connecting the ***response*** or ***dependent*** variable and one or more ***explanatory*** or ***predictor*** variables. In the cigarette consumption example, the response variable is cigarette consumption (measured by the number of packs of cigarette sold in a given state on a per capita basis during a given year) and the explanatory or predictor variables are the various socioeconomic and demographic variables. In the real estate appraisal example, the response variable is the price of a home and the explanatory or predictor variables are the characteristics of the building and taxes paid on the building.

We denote the response variable by X and the set of predictor variables by $V_1, V_2, \dots, V_m$, where *m* denotes the number of predictor variables. The true relationship between X and $V_1, V_2, \dots, V_m$, can be approximated by the regression model

$$X = f(V_1, V_2, \dots, V_m) + \epsilon \tag{2.1}$$

where ϵ is assumed to be a random error representing the discrepancy in the approximation. It accounts for the failure of the model to fit the data exactly. The function $f(V_1, V_2, \dots, V_m)$ describes the relationship between X and $V_1, V_2, \dots, V_m$. An example is the linear regression model

$$X = d_0 + d_1 V_1 + d_2 V_2 + \cdots + d_m V_m + \epsilon \tag{2.2}$$

where $d_0, d_1, d_2, \dots, d_m$ **,** called the *regression parameters* or *coefficients***,** are unknown constants to be determined (estimated) from the data. The predictor or explanatory variables are also called by other names such as *independent* variables, *covariates***,** *regressors***,** *factors***,** and *carriers***.** The name independent variable, though commonly used, is the least preferred, because in practice the predictor variables are rarely independent of each other.

The form of the model that is thought to relate the response variable to the set of predictor variables can be specified initially by the experts in the area of study based on their knowledge or their objective and or subjective judgments. The hypothesized model can then be either confirmed or refuted by the analysis of the collected data. Note that the model need to be specified only in form, but it can still depend on unknown parameters. We need to select the form of the function $f(V_1, V_2, \dots, V_m)$ in (2.1). This function can be classified into two types: *linear* and *nonlinear*. Model in (2.2) is an example of a linear function, another example of a linear function is the simple linear regression model

$$X = d_0 + d_1 V_1 + \epsilon \tag{2.3}$$

while a nonlinear function is

$$X = d_0 + \ln{(d_1)} V_1 + \epsilon \tag{2.4}$$

Note that the term *linear (nonlinear)* here does not describe the relationship between X and $V_1, V_2, \dots, V_m$,. It is related to the fact that the regression parameters enter the equation linearly (nonlinearly). A regression model containing only one predictor variable is called a *simple regression model* (2.3) . A model containing more than one predictor variable is called a *multiple regression model* (2.2).

In certain applications the response variable can actually be a set of variables, $X_1, X_2, \dots, X_m$, say, which are thought to be related to the same set of predictor variables, $V_1, V_2, \dots, V_m$,. When we deal only with one response variable, regression analysis is called *univariate* regression and in cases where we have two or more response variables, the regression is called *multivariate* regression.

In the case of *multivariate* linear regression if we let

$$\boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} 1 & v_{11} & v_{12} & \cdots & v_{1m} \\ 1 & v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & v_{n1} & v_{n2} & \cdots & v_{nm} \end{bmatrix}, \qquad \boldsymbol{d} = \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d_m \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Then, the Multivariate Linear Model can be expressed in terms of matrix form as

$$\boldsymbol{X} = \boldsymbol{V}\boldsymbol{d} + \boldsymbol{\varepsilon}. \tag{2.5}$$

After the model has been defined and the data have been collected, the next task is to estimate the parameters of the model based on the collected data. This is also referred to as *parameter estimation* or *model fitting*. The most commonly used method of estimation is called the *least squares* method. Under certain assumptions, least squares method produce estimators with desirable properties. In this research we will deal mainly with least squares method and its variants. In some instances (e.g., when one or more of the assumptions does not hold) other estimation methods may be superior to least squares. The other estimation method that we consider in this research is the *Bayesian* method.

2.2 Sample Regression Function

Consider the simple linear regression model in (2.3), The variable ϵ represents factors other than *v* that affect *x*. It is denominated error or random disturbance. The disturbance term can also capture measurement error in the dependent variable. The disturbance is an unobservable variable. The parameters d_0 and d_1 are fixed and unknown. On the right hand side of (2.3) we can distinguish two parts: the systematic component $d_0 + d_1 v$ and the random disturbance ϵ. Calling μ_x to the systematic component, we can write:

$$\mu_x = d_0 + d_1 v \tag{2.6}$$

This equation is known as the *population regression function* (*PRF*) or *population line*. Therefore, as can be seen in figure 2.1, μ_x is a linear function of v with intercept d_0 and slope

d_1. The linearity means that a one-unit increase in x changes the *expected value* of $x - \mu_x = E(x) -$ by d_1 unit.

Assume we have a random sample of size n, $(x_i, v_i), i = 1, \dots, n$ from the population. In figure 2.2 the scatter diagram, corresponding to these data, have been displayed.

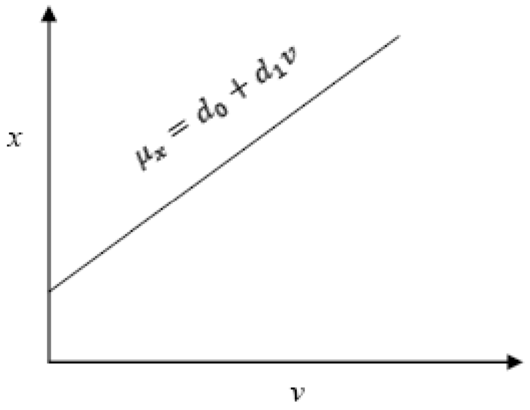

Figure 2.1. The population regression function (PRF)

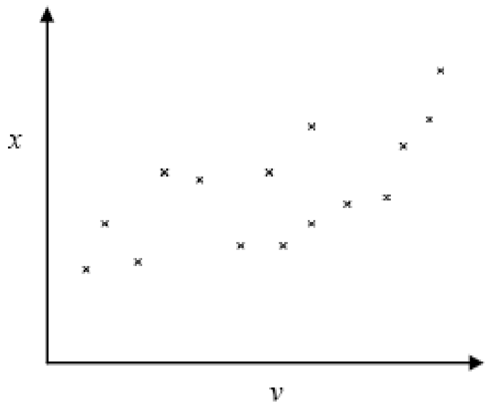

Figure 2.2. The Scatter Diagram

We can express the population model for each observation of the sample:

$$x_i = d_0 + d_1 v_i + \varepsilon_i, \quad i = 1, \dots, n \tag{2.7}$$

In Figure 2.3 the population regression function and the scatter diagram are put together, but it is important to keep in mind that although d_0 and d_1 are fixed, they are unknown. According to the model, it is possible to make the following decomposition from a theoretical point of view:

$$x_i = \mu_{x_i} + \varepsilon_i, \quad i = 1, \dots, n \tag{2.8}$$

which is represented in figure 2.3 for the i^{th} observation. However, from an empirical point of view, it is not possible because d_0 and d_1 are unknown parameters and ε_i is not observable.

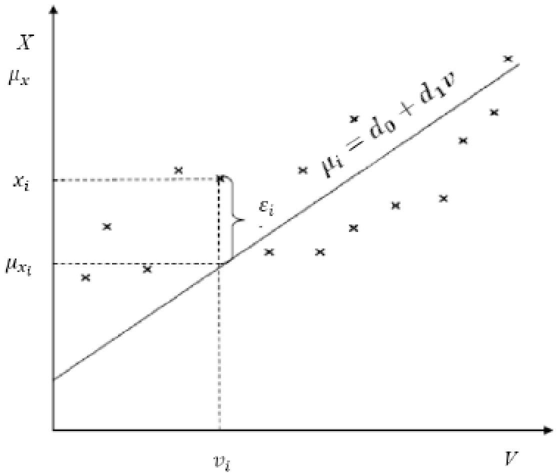

Figure 2.3. The population regression function and the scatter diagram.

The basic idea of the regression model is to estimate the population parameters, d_0 and d_1 from a given sample. The *sample regression function* (*SRF*) is the sample counterpart of the population regression function (*PRF*). Since the *SRF* is obtained for a given sample, a new sample will generate different estimates.

The *SRF*, which is an estimation of the PRF, given by

$$\hat{x}_i = \hat{d}_0 + \hat{d}_1 v_i \tag{2.9}$$

allows us to calculate the *fitted value* ($\hat{x}_i$) for *x* when $v = v_i$. In the *SRF* $\hat{d}_0$ and $\hat{d}_1$ estimators of the parameters d_0 and d_1. For each v_i we have an observed value (x_i) and
a fitted value ($\hat{x}_i$). The difference between x_i and $\hat{x}_i$ is called the residual $\hat{\varepsilon}_i$:

$$\hat{\varepsilon}_i = x_i - \hat{x}_i = x_i - \hat{d}_0 - \hat{d}_1 v_i \tag{2.10}$$

In other words, the residual $\hat{\varepsilon}_i$ is the difference between the sample value x_i and the fitted value of $\hat{x}_i$, as can be seen in figure 2.4. In this case, it is possible to calculate the decomposition:

$$x_i = \hat{x}_i + \hat{\varepsilon}_i \tag{2.11}$$

for a given sample.

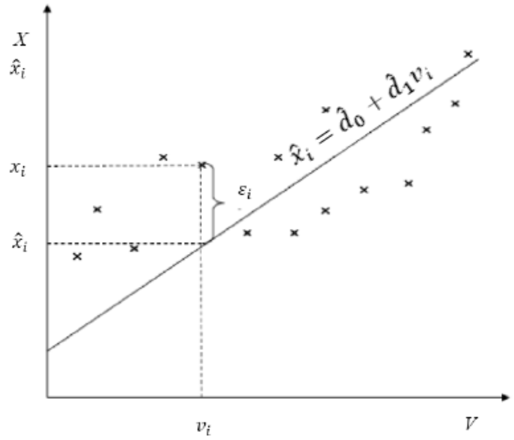

Figure 2.4. The sample regression function and the scatter diagram.

To sum up $\hat{d}_0, \hat{d}_1, \hat{x}_i$ and $\hat{\varepsilon}_i$ are the sample counterpart of $d_0, d_1\mu_{x_i}$ and ε_i respectively. It is possible to calculate $\hat{d}_0$ and $\hat{d}_1$ for a given sample, but the estimates will change for each sample. On the contrary, d_0 and d_1 are fixed, but unknown.

2.3 Parameters Estimation

The techniques used for parameter estimation are called estimators. In this study the following approaches will be considered:

- Rank Regression (Least Squares): A method of finding parameter values that minimizes the sum of the squares of the residuals.

- Maximum Likelihood Estimation: A method of finding parameter values that, given a set of observations, will maximize the likelihood function.

- Bayesian Estimation Methods: A family of estimation methods that tries to minimize the posterior expectation of what is called the utility function. In practice, what this means is that existing knowledge about a situation is formulated, data is gathered, and then posterior knowledge is used to update our beliefs.

2.3.1 The Ordinary Least Squares (OLS) Estimates

The objective of this approach is to minimize the residual sum of the squares (*S*), given by:

$$S = \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2 \tag{2.12}$$

By expressing *S* as a function of the estimators, using (2.10). Therefore, we must

$$\underset{\hat{d}_0,\hat{d}_1}{\text{Min}}\, S = \underset{\hat{d}_0,\hat{d}_1}{Min} \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2 = \underset{\hat{d}_0,\hat{d}_1}{Min} \sum_{i=1}^{n} \left(x_i - \hat{d}_0 - \hat{d}_1 v_i\right)^2 \tag{2.13}$$

To minimize *S*, we differentiate partially with respect to $\hat{d}_0$ and $\hat{d}_1$:

$$\frac{\partial \mathrm{S}}{\partial \hat{d}_0} = -2 \sum_{i=1}^{n} \left(x_i - \hat{d}_0 - \hat{d}_1 v_i\right)$$

$$\frac{\partial \mathrm{S}}{\partial \hat{d}_1} = -2 \sum_{i=1}^{n} \left(x_i - \hat{d}_0 - \hat{d}_1 v_i\right) v_i$$

The *LS* estimators are obtained by equaling the previous derivatives to 0:

$$-2 \sum_{i=1}^{n} \left(x_i - \hat{d}_0 - \hat{d}_1 v_i\right) = 0 \tag{2.13}$$

$$-2 \sum_{i=1}^{n} \left(x_i - \hat{d}_0 - \hat{d}_1 v_i\right) v_i = 0 \tag{2.14}$$

Equations (2.13) and (2.14) are denominated *normal equations* or *LS first order conditions*. Operating with the normal equations, we have

$$\sum_{i=1}^{n} x_i = n\hat{d}_0 + \hat{d}_1 \sum_{i=1}^{n} v_i \tag{2.15}$$

$$\sum_{i=1}^{n} x_i\, v_i = \hat{d}_0 \sum_{i=1}^{n} v_i + \hat{d}_1 \sum_{i=1}^{n} v_i^2 \tag{2.16}$$

Dividing both sides of (2.15) by *n*, we have

$$\overline{x} = \hat{d}_0 + \hat{d}_1 \overline{v} \tag{2.17}$$

Therefore

$$\hat{d}_0 = \overline{x} - \hat{d}_1\overline{v} \tag{2.18}$$

Substituting the value of $\hat{d}_0$ in the second normal equation (2.16), we have

$$\sum_{i=1}^{n} x_i\, v_i = \left(\overline{x} - \hat{d}_1\overline{v}\right)\sum_{i=1}^{n} v_i + \hat{d}_1\sum_{i=1}^{n} v_i^2$$

$$\sum_{i=1}^{n} x_i\, v_i = \overline{x}\sum_{i=1}^{n} v_i - \hat{d}_1\overline{v}\sum_{i=1}^{n} v_i + \hat{d}_1\sum_{i=1}^{n} v_i^2$$

Solving for $\hat{d}_1$ we have:

$$\hat{d}_1 = \frac{\sum_{i=1}^{n} x_i\, v_i - \overline{x}\sum_{i=1}^{n} v_i}{\sum_{i=1}^{n} v_i^2 - \overline{v}\sum_{i=1}^{n} v_i}$$

By simplifying the above expression we get

$$\hat{d}_1 = \frac{S_{xv}}{S_{vv}} \tag{2.19}$$

where $S_{xv} = \sum_{i=1}^{n}(x_i - \overline{x})(v_i - \overline{v})$, and $S_{vv} = \sum_{i=1}^{n}(v_i - \overline{v})^2$. Once $\hat{d}_1$is calculated, then we can obtain $\hat{d}_0$ by using (2.18). Hence these two estimators will minimize the function in (2.12) that what we are looking for.

2.3.2 The Maximum Likelihood (ML) Estimates

We introduced the method of maximum likelihood for simple linear regression by start with the statistical model in (2.3), which is the Gaussian-noise simple linear regression model, defined as follows:

1. The distribution of *V* is arbitrary (and perhaps *V* is even non-random).
2. If $V = v$, then $X = d_0 + d_1 v + \epsilon$, for some constants (“coefficients”, “parameters”) d_0 and d_1, and some random noise variable ϵ.
3. $\epsilon \sim N(0, \sigma^2)$ and is independent of *V*.
4. ϵ is independent across observations.

A consequence of these assumptions is that the response variable X is independent across observations, conditional on the predictor V. The first two assumptions are the same, but we are now assuming much more about the noise variable ϵ: it's not just mean zero with constant variance, but it has a particular distribution (Normal), and everything we said was uncorrelated before we now strengthen to independence. Because of these stronger assumptions, the model tells us the conditional pdf of X for each v, $P(x|V = v, d_0, d_1, \sigma^2)$. (This notation separates the random variables from the parameters.) Given any data set $(x_1, v_1), (x_2, v_2), \dots, (x_n, v_n)$, we can now write down the probability density, under the model, of seeing that data:

$$\prod_{i=1}^{n} P(x_i|v_i, d_0, d_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \operatorname{Exp}\left(\frac{-\big(x_i - (d_0 + d_1 v_i)\big)^2}{2\sigma^2}\right) \tag{2.20}$$

This is the likelihood, a function of the parameter values. It's just as informative, and much more convenient, to work with the log-likelihood,

$$\begin{aligned} L(d_0, d_1, \sigma^2) \quad &= \ln\left\{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} Exp\left(\frac{-\big(x_i - (d_0 + d_1 v_i)\big)^2}{2\sigma^2}\right)\right\} \\ &= \frac{-n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(x_i - (d_0 + d_1 v_i)\big)^2 \end{aligned} \tag{2.21}$$

In the method of maximum likelihood, we pick the parameter values which maximize the likelihood, or, equivalently, maximize the log-likelihood in (2.21). After some calculus, this gives us the following estimators:

$$\begin{aligned} \hat{d}_1 \quad &= \frac{S_{xv}}{S_{vv}} \\ \hat{d}_0 \quad &= \hat{d}_0 = \overline{x} - \hat{d}_1\overline{v} \\ \hat{\sigma}^2 \quad &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \left(\hat{d}_0 + \hat{d}_1 v_i\right)\right)^2 \end{aligned} \tag{2.22}$$

where $S_{xv} = \sum_{i=1}^{n}(x_i - \overline{x})(v_i - \overline{v})$, and $S_{vv} = \sum_{i=1}^{n}(v_i - \overline{v})^2$.Note, the same estimators we got as in LS method.

2.4. The Bayesian Estimates

In the Bayesian viewpoint, we formulate linear regression using probability distributions rather than point estimates. The response, x, is not estimated as a single value, but is assumed to be drawn from a probability distribution. The model for Bayesian Linear Regression with the response sampled from a normal distribution is:

$$x \sim N(d_0 + d_1 v, \sigma^2) \tag{2.23}$$

The output, x is generated from a normal (Gaussian) Distribution characterized by a mean and variance given in (2,23). The aim of Bayesian Linear Regression is not to find the single "best" value of the model parameters, but rather to determine the *posterior distribution* for the model parameters. Not only is the response generated from a probability distribution, but the model parameters are assumed to come from a distribution as well. Let $\Lambda = \{d_0, d_1, \sigma^2\}$ be the set of all model parameters, the posterior probability of the model parameters is conditional upon the training inputs and outputs:

$$P(\mathit{\Lambda}|x, v) = \frac{P(x|\mathit{\Lambda}, v) \times P(\mathit{\Lambda}|\ v)}{P(x|\ v)} \tag{2.24}$$

Here, $P(\mathit{\Lambda}|x, v)$ is the *posterior probability* distribution of the model parameters given the inputs and outputs. This is equal to the likelihood of the data, $P(x|\mathit{\Lambda}, v)$, multiplied by the prior probability of the parameters $P(\mathit{\Lambda}|\ v)$ and divided by a normalization constant $P(x|\ v)$. This is a simple expression of Bayes Theorem, the fundamental underpinning of Bayesian Inference:

$$Posterior = \frac{Likelihood \times Prior}{Normalization} \tag{2.25}$$

Let's stop and think about what this means. In contrast to LS, we have a posterior *distribution* for the model parameters that is proportional to the likelihood of the data multiplied by the *prior* probability of the parameters. Here we can observe the two primary benefits of Bayesian Linear Regression.

1. **Priors:** If we have domain knowledge, or a guess for what the model parameters should be, we can include them in our model, unlike in the frequentist approach which assumes everything there is to know about the parameters comes from the data. If we don't have any estimates ahead of time, we can use non-informative priors for the parameters such as a normal distribution.
2. **Posterior:** The result of performing Bayesian Linear Regression is a distribution of possible model parameters based on the data and the prior. This allows us to quantify our uncertainty about the model: if we have fewer data points, the posterior distribution will be more spread out.

As the amount of data points increases, the likelihood washes out the prior, and in the case of infinite data, the outputs for the parameters converge to the values obtained from OLS.

Our goal is to update the distributions of the unknown parameters $\Lambda = \{d_0, d_1, \sigma^2\}$, based on the data , $(x_1, v_1), (x_2, v_2), \dots, (x_n, v_n)$, where *n* is the number of observations.

Under the assumption that the errors ε_i are normally distributed with constant variance σ^2, we have for the random variable of each response X_i, conditioning on the observed data $v_i$and the parameters $\Lambda = \{d_0, d_1, \sigma^2\}$, is normally distributed:

$$X_i | v_i, \mathit{\Lambda} \sim N(d_0 + d_1 v_i, \sigma^2), \;\; i = 1, \dots, n \tag{2.26}$$

That is, the likelihood of each X_i given $v_i, \mathit{\Lambda}$ is given by

$$P(x_i | v_i, \mathit{\Lambda}) = \frac{1}{\sqrt{2\pi\sigma^2}} Exp\left(-\frac{\left(x_i - (d_0 + d_1 v_i)\right)^2}{2\sigma^2}\right) \tag{2.26}$$

And hence, the likelihood of $X_1, X_2, \dots, X_n$ is given as in (2.20).

We first consider the case under the reference prior, which is our standard noninformative prior. Using the reference prior, we will obtain familiar distributions as the posterior distributions of $d_0, d_1,$ and σ^2, which gives the analogue to the frequentist results. Here we assume the joint prior distribution of $d_0, d_1,$ and σ^2 to be proportional to the inverse of σ^2

$$P(d_0, d_1, \sigma^2) \propto \frac{1}{\sigma^2} \tag{2.27}$$

Using the *hierachical* model framework, this is equivalent to assuming that the joint prior distribution of d_0 and d_1 under σ^2 is the uniform prior, while the prior distribution of σ^2 is proportional to $\frac{1}{\sigma^2}$. That is

$$P(d_0, d_1 |\, \sigma^2) \propto 1, \qquad P(\,\sigma^2) \propto \frac{1}{\sigma^2}$$

Combining the two using conditional probability, we will get the same joint prior distribution (2.27).

Then we apply the Bayes' rule to derive the joint posterior distribution after observing data $x_1, x_2, \dots, x_n$. Bayes' rule states that the joint posterior distribution of $d_0, d_1,$ and σ^2 is proportional to the product of the likelihood and the joint prior distribution:

$$\begin{aligned} P(d_0, d_1, \sigma^2 | x_1, \dots, x_n) \quad &\propto \left[\prod\nolimits_{i=1}^{n} P(x_i | v_i, d_0, d_1, \sigma^2)\right] P(d_0, d_1, \sigma^2), d_1, \sigma^2) \\ &\propto \left[\prod\nolimits_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} Exp\left(\frac{-\left(x_i - (d_0 + d_1 v_i)\right)^2}{2\sigma^2}\right)\right] \frac{1}{\sigma^2} \\ &\propto \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{\sum_{i=1}^{n}\left(x_i - (d_0 + d_1 v_i)\right)^2}{2\sigma^2}\right) \end{aligned}$$

To obtain the marginal posterior distributions of $d_0, d_1,$ and σ^2 we introduce the following quantities derived from the formula of $\overline{x}, \overline{v}, \hat{d}_0$ and $\hat{d}_1$ to simplify our calculations.

$$\sum\nolimits_{i=1}^{n} (x_i - \overline{x}) = 0, \qquad \sum\nolimits_{i=1}^{n} (v_i - \overline{v}) = 0, \qquad \sum\nolimits_{i=1}^{n} (x_i - \hat{x}) = 0$$

$$\sum_{i=1}^{n}(v_i-\overline{v})\,(x_i-\hat{x})=0, \quad \sum_{i=1}^{n}v_i^2=\sum_{i=1}^{n}(v_i-\overline{v})^2+n\overline{v}^2=S_{vv}+n\overline{v}^2$$

We also further simplify the numerator inside the exponential function in the formula of $P(d_0,d_1,\sigma^2|x_1,\dots,x_n)$:

$$\sum_{i=1}^{n}(x_i-d_0-d_1v_i)^2=\sum_{i=1}^{n}\left(x_i-\hat{d}_0-\hat{d}_1v_i-\left(d_0-\hat{d}_0\right)-(d_1-\hat{d}_1)v_i\right)^2$$

$$=\sum_{i=1}^{n}(x_i-\hat{d}_0-\hat{d}_1v_i)^2+\sum_{i=1}^{n}\left(d_0-\hat{d}_0\right)^2+\sum_{i=1}^{n}\left(d_1-\hat{d}_1\right)^2v_i^2-2\sum_{i=1}^{n}\left(d_0-\hat{d}_0\right)(x_i-\hat{d}_0-\hat{d}_1v_i)$$

$$-2\sum_{i=1}^{n}\left(d_1-\hat{d}_1\right)(x_i)\left(x_i-\hat{d}_0-\hat{d}_1v_i\right)+2\sum_{i=1}^{n}\left(d_0-\hat{d}_0\right)\left(d_1-\hat{d}_1\right)(v_i)$$

$$=SSE+n\left(d_0-\hat{d}_0\right)^2-\left(d_1-\hat{d}_1\right)^2\sum_{i=1}^{n}v_i^2-2\left(d_0-\hat{d}_0\right)\sum_{i=1}^{n}(x_i-\hat{x}_i)-2\left(d_1-\hat{d}_1\right)\sum_{i=1}^{n}v_i(x_i-\hat{x}_i)$$

$$+2\left(d_0-\hat{d}_0\right)\left(d_1-\hat{d}_1\right)n\overline{v}$$

It is clear that

$$\left(d_0-\hat{d}_0\right)\sum_{i=1}^{n}(x_i-\hat{x}_i)=0 \qquad \text{and} \qquad \left(d_1-\hat{d}_1\right)\sum_{i=1}^{n}v_i(x_i-\hat{x}_i)=0$$

Finally, we use the quantity that $\sum_{i=1}^{n}v_i^2=\sum_{i=1}^{n}(v_i-\overline{v})+n\overline{v}^2$ to combine the terms $n\left(d_0-\hat{d}_0\right)^2$, $2\left(d_0-\hat{d}_0\right)\left(d_1-\hat{d}_1\right)\sum_{i=1}^{n}v_i$ and $\left(d_1-\hat{d}_1\right)^2\sum_{i=1}^{n}v_i^2$ together.

$$\therefore \; \sum_{i=1}^{n}(x_i-d_0-d_1v_i)^2 \; =SSE+n\left(d_0-\hat{d}_0\right)^2+\left(d_1-\hat{d}_1\right)^2\sum_{i=1}^{n}(v_i-\overline{v})^2$$

$$+\left(d_1-\hat{d}_1\right)^2n\overline{v}^2+2\left(d_0-\hat{d}_0\right)\left(d_1-\hat{d}_1\right)n\overline{v}$$

$$=SSE+\left(d_1-\hat{d}_1\right)^2S_{vv}+n\left[\left(d_0-\hat{d}_0\right)+\left(d_1-\hat{d}_1\right)\overline{v}\right]^2$$

where

$$SSE=\sum_{i=1}^{n}(x_i-\hat{x})^2=\sum_{i=1}^{n}(\varepsilon_i)^2$$

Therefore, the posterior joint distribution of $d_0, d_1,$ and σ^2 can be simplied as

$$
\begin{aligned}
P(d_0, d_1, \sigma^2 | x_1, \dots, x_n) \quad & \propto \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{\sum_{i=1}^{n}\left(x_i - (d_0 + d_1 v_i)\right)^2}{2\sigma^2}\right) \\
& = \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 S_{vv} + n\left[\left(d_0 - \hat{d}_0\right) + \left(d_1 - \hat{d}_1\right)\overline{v}\right]^2}{2\sigma^2}\right)
\end{aligned}
$$

2.4.1 Marginal Posterior Distribution of d_1

To get the marginal posterior distribution of d_1, we need to integrate out d_0 and σ^2 from $P(d_0, d_1, \sigma^2 | x_1, \dots, x_n)$:

$$
\begin{aligned}
P(d_1 | x_1, \dots, x_n) \quad & = \int_0^\infty \int_{-\infty}^\infty P(d_0, d_1, \sigma^2 | x_1, \dots, x_n)\, d\, d_0\, d\sigma^2 \\
& = \int_0^\infty \left(\int_{-\infty}^\infty \left(\frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 S_{vv} + n\left[\left(d_0 - \hat{d}_0\right) + \left(d_1 - \hat{d}_1\right)\overline{v}\right]^2}{2\sigma^2}\right)\right) dd_0 \right) d\sigma^2 \\
& = \int_0^\infty P(d_1, \sigma^2 | x_1, \dots, x_n)\, d\sigma^2
\end{aligned}
$$

We first calculate the inside integral, which gives us the joint posterior distribution of d_1 and σ^2

$P(d_1, \sigma^2 | x_1, \dots, x_n)$

$$
= \int_{-\infty}^{\infty} \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2 + n\left[\left(d_0 - \hat{d}_0\right) + \left(d_1 - \hat{d}_1\right)\overline{v}\right]^2}{2\sigma^2}\right) dd_0
$$

$$
= \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} Exp\left(-\frac{n\left[\left(d_0 - \hat{d}_0\right) + \left(d_1 - \hat{d}_1\right)\overline{v}\right]^2}{2\sigma^2}\right) dd_0
$$

Here

$$
Exp\left(-\frac{n\left[\left(d_0 - \hat{d}_0\right) + \left(d_1 - \hat{d}_1\right)\overline{v}\right]^2}{2\sigma^2}\right)
$$

can be viewed as part of a normal distribution of d_0, with mean $\hat{d}_0 - \left(d_1 - \hat{d}_1\right)\overline{v}$, and variance σ^2/n. Therefore, the integral from the last line above is proportional to $\sqrt{\sigma^2/n}$. We get

$$P(d_1, \sigma^2 | x_1, \dots, x_n) \quad \propto \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2}{2\sigma^2}\right) \times \sqrt{\frac{\sigma^2}{n}}$$

$$\propto \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2}{2\sigma^2}\right)$$

We then integrate σ^2 out to get the marginal distribution of d_1. Here we first perform change of variable and set $d_1 = \frac{1}{\phi}$. Then the integral becomes

$$P(d_1 | x_1, \dots, x_n) \quad \propto \int_0^{\infty} \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2}{2\sigma^2}\right) d\sigma^2$$

$$\propto \int_0^{\infty} (\phi)^{(n-3)/2} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2}{2}\phi\right) d\sigma^2$$

$$\propto \left(\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2}{2}\phi\right)^{-\frac{(n-2)+1}{2}} \int_0^{\infty} (S)^{(n-3)/2} e^{-S}\, dS$$

Here we use another change of variable by setting $S = \frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2}{2}\phi$, and the fact that $\int_0^{\infty}(S)^{(n-3)/2} e^{-S}\, dS$ gives us the Gamma function $\Gamma(n-2)$, which is constant. We can rewrite the last line from above to obtain the marginal posterior distribution of d_1. This marginal distribution is the Student's *t* distribution with degrees of freedom $n-2$, centerred at $\hat{d}_1$, and scale parameter of $\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2}$.

$$P(d_1 | x_1, \dots, x_n) \quad \propto \left(1 + \frac{1}{n-2} \frac{\left(d_1 - \hat{d}_1\right)^2}{\frac{SSE}{n-2} / \sum_{i=1}^{n}(v_i - \overline{v})^2}\right)^{-\frac{(n-2)+1}{2}}$$

$$= \left(1 + \frac{1}{n-2} \frac{\left(d_1 - \hat{d}_1\right)^2}{\hat{\sigma}^2 / \sum_{i=1}^{n}(v_i - \overline{v})^2}\right)^{-\frac{(n-2)+1}{2}}$$

where $\hat{\sigma}^2/\sum_{i=1}^{n}(v_i - \overline{v})^2$ is exactly the square of the standard error of $\hat{d}_1$ from the frequentist OLS model.

2.4.2 Marginal Posterior Distribution of d_0

A similar approach will lead us to the marginal distribution of d_0. We again start from the joint posterior distribution and integrate d_1 and σ^2 out to get the marginal posterior distribution of d_0. We first compute the integral

$$P(d_0, \sigma^2 | x_1, \dots, x_n) \quad = \int_{-\infty}^{\infty} \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2 + n\left[\left(d_0 - \hat{d}_0\right) + \left(d_1 - \hat{d}_1\right)\overline{v}\right]^2}{2\sigma^2}\right) dd_1$$

Here we group the terms with $d_1 - \hat{d}_1$ together, then complete the square so that we can treat is as part of a normal distribution function to simplify the integral

$$n\left[\left(d_0 - \hat{d}_0\right) + \left(d_1 - \hat{d}_1\right)\overline{v}\right]^2 + \left(d_1 - \hat{d}_1\right)^2 \sum_{i=1}^{n}(v_i - \overline{v})^2$$

$$= \left(d_1 - \hat{d}_1\right)^2 \left(\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2\right) + 2n\overline{v}^2\left(d_0 - \hat{d}_0\right)\left(d_1 - \hat{d}_1\right) + n\left(d_0 - \hat{d}_0\right)^2$$

$$= \left(\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2\right)\left[\left(d_1 - \hat{d}_1\right) + \frac{n\overline{v}\left(d_0 - \hat{d}_0\right)}{\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2}\right]^2 + n\left(d_0 - \hat{d}_0\right)^2\left[\frac{\sum_{i=1}^{n}(v_i - \overline{v})^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2}\right]$$

$$= \left(\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2\right)\left[\left(d_1 - \hat{d}_1\right) + \frac{n\overline{v}\left(d_0 - \hat{d}_0\right)}{\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2}\right]^2 + \frac{\left(d_0 - \hat{d}_0\right)^2}{\frac{1}{n} + \frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2}}$$

When integrating, we can then view

$$Exp\left(-\frac{\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2}{2\sigma^2}\left[\left(d_1 - \hat{d}_1\right) + \frac{n\overline{v}\left(d_0 - \hat{d}_0\right)}{\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2}\right]^2\right)$$

as part of a normal distribution function, and get

$$P(d_0, \sigma^2 | x_1, \ldots, x_n) \quad \propto \frac{1}{(\sigma^2)^{(n+2)/2}} Exp\left(-\frac{SSE + \left(d_0 - \hat{d}_0\right)^2 \Big/ \left(\frac{1}{n} + \frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2} \right)}{2\sigma^2} \right)$$

$$\times \int_0^\infty Exp\left(-\frac{\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2}{2\sigma^2} \left(d_1 - \hat{d}_1 + \frac{n\overline{v}\left(d_0 - \hat{d}_0\right)}{\sum_{i=1}^{n}(v_i - \overline{v})^2 + n\overline{v}^2} \right)^2 \right) dd_1$$

$$\propto \frac{1}{(\sigma^2)^{(n+1)/2}} Exp\left(-\frac{SSE + \left(d_0 - \hat{d}_0\right)^2 \Big/ \left(\frac{1}{n} + \frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2} \right)}{2\sigma^2} \right)$$

To get the marginal posterior distribution of d_0, we again integrate σ^2 out. using the same change of variable $\sigma^2 = \frac{1}{\phi}$, and $S = \frac{SSE + \left(d_0 - \hat{d}_0\right)^2 \Big/ \left(\frac{1}{n} + \frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2} \right)}{2} \phi$

$$P(d_0 | x_1, \ldots, x_n) \quad \propto \int_0^\infty (\phi)^{(n-3)/2} Exp\left(-\frac{SSE + \left(d_0 - \hat{d}_0\right)^2 \Big/ \left(\frac{1}{n} + \frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2} \right)}{2} \phi \right) d\phi$$

$$\propto \left(SSE + \left(d_0 - \hat{d}_0\right)^2 \Big/ \left(\frac{1}{n} + \frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2} \right) \right)^{-\frac{(n-2)+1}{2}} \int_0^\infty (S)^{(n-3)/2} e^{-S}\, dS$$

$$\propto \left(1 + \frac{1}{n-2} \frac{\left(d_0 - \hat{d}_0\right)^2}{\frac{SSE}{n-2} \left(\frac{1}{n} + \frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i - \overline{v})^2} \right)} \right)^{-\frac{(n-2)+1}{2}}$$

$$= \left(1 + \frac{1}{n-2} \left(\frac{d_0 - \hat{d}_0}{s.e.(d_0)} \right)^2 \right)^{-\frac{(n-2)+1}{2}}$$

In the last line, we use the same trick as we did for d_1 to derive the form of the Student's *t* distribution. This shows that the marginal posterior distribution of d_0 also follows a Student's *t* distribution, with $n-2$ degrees of freedom. Its center is $\hat{d}_0$, the estimate of d_0 in the frequentist OLS estimate, and its scale parameter is $\hat{\sigma}^2\left(\frac{1}{n}+\frac{\overline{v}^2}{\sum_{i=1}^{n}(v_i-\overline{v})^2}\right)$, which is the square of the standard error of $\hat{d}_0$.

2.4.3 Marginal Posterior Distribution of σ^2

To show that the marginal posterior distribution of σ^2 follows the inverse Gamma distribution, we only need to show the precision $\phi=\frac{1}{\sigma^2}$ follows a Gamma distribution. It is clear that the prior distribution of σ^2 proportional to $\frac{1}{\sigma^2}$ is equivalent to taking the prior distribution of ϕ proportional to $\frac{1}{\phi}$.

$$P(\sigma^2)\propto\frac{1}{\sigma^2}\;\Longrightarrow\;P(\phi)\propto\frac{1}{\phi}$$

Therefore, under the parameters d_0, d_1 , and the precision ϕ, we have the joint prior distribution as

$$P(d_0,d_1,\phi\,)\propto\frac{1}{\phi}$$

and the joint posterior distribution as

$$P(d_0,d_1,\phi|x_1,\dots,x_n)\propto\phi^{\frac{n}{2}-1}Exp\left(\frac{\sum_{i=1}^{n}(x_i-d_0-d_1v_i)}{2}\phi\right)$$

Using the partial results we have calculated previously, we get

$$\begin{aligned}P(d_1,\phi|x_1,\dots,x_n)\quad&=\int_{-\infty}^{\infty}P(d_0,d_1,\phi|x_1,\dots,x_n)\,dd_0\\&\propto\phi^{\frac{n}{2}-1}Exp\left(-\frac{SSE+\left(d_1-\hat{d}_1\right)^2\sum_{i=1}^{n}(v_i-\overline{v})^2}{2}\phi\right)\end{aligned}$$

Intergrating over d_1, we have

$$
\begin{aligned}
P(\phi|x_1,\dots,x_n) \quad & \propto \int_{-\infty}^{\infty} \phi^{\frac{n-3}{2}} Exp\left(-\frac{SSE+\left(d_1-\hat{d}_1\right)^2\sum_{i=1}^{n}(v_i-\overline{v})^2}{2}\phi\right) dd_1 \\
& = \phi^{\frac{n-3}{2}}\text{Exp}\left(-\frac{SSE}{2}\phi\right)\int_{-\infty}^{\infty}\text{Exp}\left(-\frac{\left(d_1-\hat{d}_1\right)^2\sum_{i=1}^{n}(v_i-\overline{v})^2}{2}\right) dd_1 \\
& \propto \phi^{\frac{n-4}{2}}\text{Exp}\left(-\frac{SSE}{2}\phi\right) \\
& = \phi^{\frac{n-2}{2}-1}\text{Exp}\left(-\frac{SSE}{2}\phi\right)
\end{aligned}
$$

This is a Gamma distribution with shape parameter $\frac{n-2}{2}$ and rate parameter $\frac{SSE}{2}$. Therefore, the updated σ^2 follows the inverse Gamma distribution

$$\phi = \frac{1}{\sigma^2} \mid x_1,\dots,x_n \sim \text{Gamma}\left(\frac{n-2}{2},\frac{SSE}{2}\right)$$

That is

$$P(\phi|\text{Data}) \propto \phi^{\frac{n-2}{2}-1}\text{Exp}\left(-\frac{SSE}{2}\phi\right)$$

Chapter 3

Stable Distribution

3.1 Introduction

Stable distribution are general family of probabilities distributions that share certain properties.They were first described by Paul Levy (1925) and so are also sometimes informally called Levy distributions. This can cause confusion as ”Levy Distribution” is actually a specific member of the Stable Distribution family. Most of these distributions do not have a distinct probability dinsity function, with the exception of the Cauchy Distribution, Levy Distribution and Normal Distribution, but they do share certain properties, like skewness and heavy tails or fat tails.

3.2 Definitions, features, and graphics of stable distribution

A distribution that is heavy tailed goes to zero slower than on without heavy tails.Heavy tailed distributions tend to have many outliers with very high values.

The stable distribution requires four parameters. The index of stability $\alpha \in (0, 2]$, also called the tail index,tail exponent or characteristic exponent, determines the rate at which the tails of the distribution taper off, see the figure 3.2.

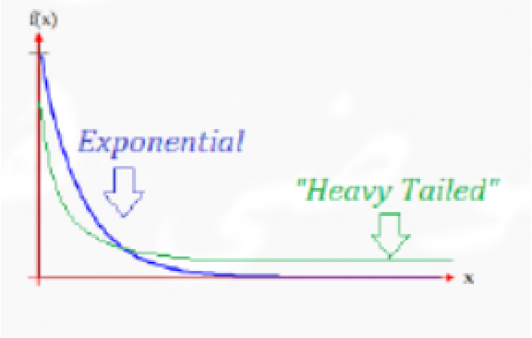

Figure 3.1: Heavy tailed.

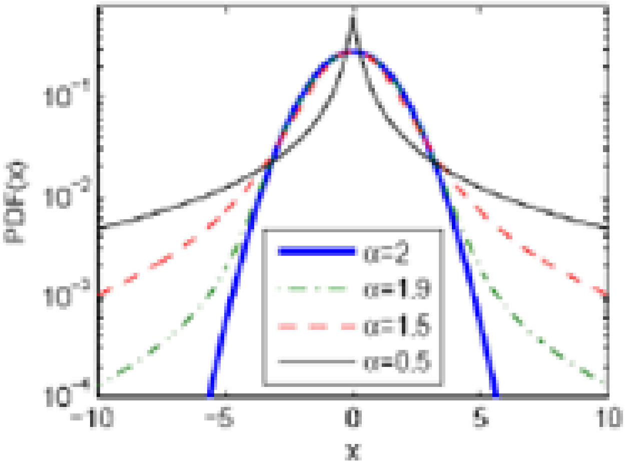

Figure 3.2: A semi-logarithmic plot of symmetric ($\beta = 0$) stable densities for four value of α. Note, the distinct behavior of the Gussian $\alpha = 2$ distribution.

The skewness parameter $\beta \in [-1, 1]$ defines the asymmetry. When $\beta > 0$, the distribution is skewed to the right, i.e. the right tail is thicker, see figure 3.3. When $\beta < 0$, it is skewed to the left.When $\beta = 0$, the distribution is symmetric about the mode (the peak) of the distribution.

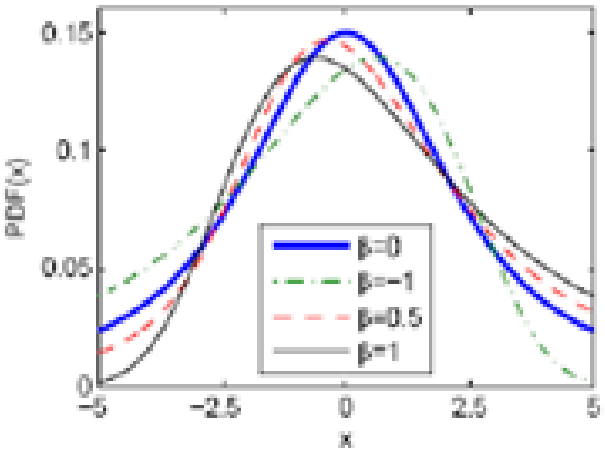

Figure 3.3: A plot of stable densities for $\alpha = 1.2$ and four values of β

The last two parameters, $\delta \in (-\infty, \infty)$ and $\gamma \in (0, \infty)$, are the location and scale parameters respectively. Sable distributions are usually denoted by $S_\alpha(\beta, \delta, \gamma)$, then $Z = \frac{X-\delta}{\gamma} \sim S_\alpha(\beta, 0, 1)$, (see [13], [16]).

Figure 3.4: A semilog plot of symmetric ($\beta = \delta = 0$) α-stable probability density functions (pdfs) for $\alpha = 2$ (black), 1.8 (red), 1.5 (blue dashed line) and 1 (green long-dashed line). The Gaussian ($\alpha = 2$) density forms a parabola and is the only α-stable density with exponential tails.

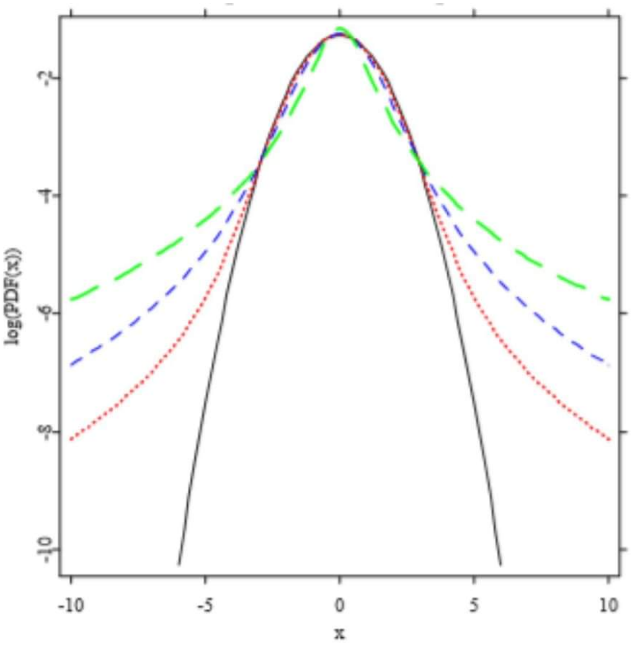

Figure 3.4: Dependence on alpha

Figure 3.5: Right tails of symmetric α-stable cumulative distribution functions (cdfs) for $\alpha = 2$ (black), 1.95 (red), 1.8 (blue dashed line) and 1.5 (green long-dashed line) on a double logarithmic paper. For $\alpha < 2$ the tails form straight lines with slope $-\alpha$.

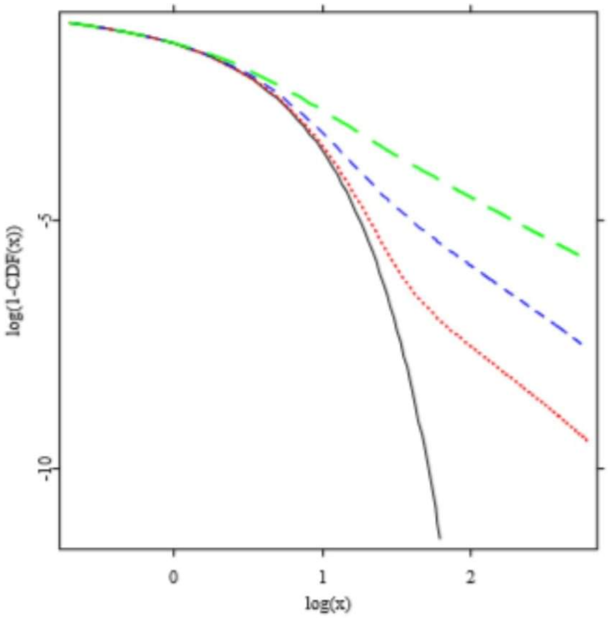

Figure 3.5: Tails of stable laws

using a central limit theorem type argument it can be shown that [11]

$$\begin{cases} \lim_{x\to\infty} x^{\alpha} P(X > x) = D_{\alpha}(1+\beta)\gamma^{\alpha}, \\ \lim_{x\to\infty} x^{\alpha} P(X < -x) = D_{\alpha}(1+\beta)\gamma^{\alpha}, \end{cases} \tag{3.1}$$

Where :

$$D_{\alpha} = \left(2\int_0^{\infty} x^{-\alpha}\sin(x)dx\right)^{-1} = \frac{1}{\pi}\Gamma(\alpha)\sin\frac{\pi\alpha}{2} \tag{3.2}$$

The convergence to a power-law tail varies for different α's and, as can be seen in the Figure 3.5, is slower for larger values of the tail index. The tails of α-stable distribution functions exhibit a crossover from an approximate power decay with exponent $\alpha > 2$ to the true tail with exponent α. This phenomenon is more visible for large α's.

An important property of normal or Gaussian random variables is that the sum of two of them is itself a normal random variable. One consequence of this is that if Z is normal, then for Z_1 and Z_2 independent copies of Z and any positive constants a and b,

$$aZ_1 + bZ_2 \overset{\mathrm{d}}{=} eZ + h \tag{3.3}$$

or some positive e and some $h \in R$. The symbol $\overset{\mathrm{d}}{=}$ means equality in distribution, i.e. both expressions have the same probability law. equation (3.3) says that the shape of Z is preserved up to scale and shift under addition.

Definition 3.1 A random variable Z is stable or stable in the broad sense if for Z_1 and Z_2 independent copies of Z and any positive constants a and b, (3.3) holds for some positive e and some $h \ \in R$. The random variable is strictly stable or stable in the narrow sense if (3.3) holds with $h = 0$ for all choices of a and b. A random variable is symmetric stable if it is stable and symmetrically distributed around 0, e.g. $Z \overset{\mathrm{d}}{=} -Z$, [16].

The addition rule for independent normal random variables says that the mean of the sum is the sum of the means and the variance of the sum is the sum of the variances.

Suppose $Z \sim N(\mu, \sigma^2)$, then the terms on the left hand side above are$N(a\mu, (a\sigma)^2)$ and $N(b\mu, (b\sigma)^2)$ respectively, while the right hand side is $N(e\mu + h, (e\sigma)^2)$. By the addition rule one must have $e^2 = a^2 + b^2 and \quad h = (a + b - e)\mu$. Expressions for e and h in the general stable case are given below. The word stable is used because the shape is stable or unchanged under sums of the type (3.3). Some authors use the phrase sum stable to emphasize the fact that (3.3) is about a sum and to distinguish between these distributions and max-stable, min-stable, multiplication stable and geometric stable distributions. Also, some older literature used slightly different terms: stable was originally used for what we now call strictly stable,quasi-stable was reserved for what we now call stable.

Two random variables Z and Y are said to be of the same type if there exist constants $A > 0$ and $B \in \mathbb{R}$ with $Z \overset{\mathrm{d}}{=} AY + B$. The definition of stability can be restated as $aZ_1 + bZ_2$ has the same type as Z. There are three cases where one can write down closed form expressions for the density and verify directly that they are stable - normal, Cauchy and Levy distributions.

Normal or Gaussian distributions $Z \sim N(\mu, \sigma^2)$, if it has a density [16]

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right), \ -\infty < z < \infty \tag{3.4}$$

The cumulative distribution function, for which there is no closed form expression, is $F(z) = P(Z \leq z) = \varphi((z-\mu)/\sigma)$, where $\varphi(z)$ = probability that a standard normal random variable. is less than or equal z.

Cauchy distributions $Z \sim$ Cauchy (δ, γ) , if it has density [16]

$$f(z) = \frac{1}{\pi}\frac{\gamma}{\gamma^2 + (z-\delta)^2} \quad -\infty < z < \infty \tag{3.5}$$

Levy distributions $Z \sim$ Levy (δ, γ) , if it has density [16]

$$f(z) = \sqrt{\frac{\gamma}{2\pi}}\frac{1}{(z-\delta)^{3/2}} exp\left(-\frac{\gamma}{2(z-\delta)}\right), \ \delta < z < \infty \tag{3.6}$$

Note that some authors use the term Levy distribution for all sum stable laws; Figure (3.6) shows a plot of these three densities. Both normal distributions and Cauchy distributions are symmetric, bell-shaped curves. The main qualitative distinction between them is that the Cauchy distribution has much heavier tails, see Table 3.1. In particular,there is a tiny amount of probability above 3 for the normal distribution, but a significant amount above 3 for a Cauchy. In a sample of data from these two distributions, there will be (on average) approximately 100 times more values above 3 in the Cauchy case than in the normal case. This is the reason stable distributions are called heavy tailed. In contrast to the normal and Cauchy distributions, the Levy distribution is highly skewed, with all of the probability concentrated on $x > 0$, and it has even heavier tails than the Cauchy.

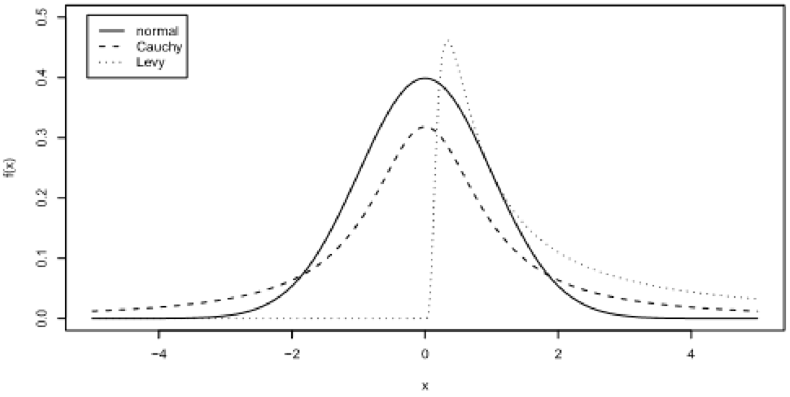

Figure 3.6: Graphs of standarizad normal $N(1, 0)$, Cauchy $(1, 0)$ and levy $(1, 0)$ destination

Table 3.1: Comparison of tail probabilities for standard normal, Cauchy and Levy distributions

$P(Z > c)$			
c	Normal	Cauchy	Levy
0	0.5000	0.5000	1.0000
1	0.1587	0.2500	0.6827
2	0.0228	0.1476	0.5205
3	0.001347	0.1024	0.4363
4	0.00003167	0.0780	0.3829
5	0.0000002866	0.0628	0.3453

General stable distributions allow for varying degrees of tail heaviness and varying degrees of skewness. Other than the normal distribution, the Cauchy distribution, the Levy distribution, and the reflection of the Levy distribution, there are no known closed form expressions for general stable densities and it is unlikely that any other stable distributions have closed forms for their densities. [21] shows that in a few cases stable densities or distribution functions are expressible in terms of certain special functions. This may seem to doom the use of stable models in practice, but recall that there is no closed formula for the normal cumulative distribution function. There are tables and accurate computer algorithms for the standard normal distribution function, and people routinely use those values in normal models. We now have computer programs to compute quantities of interest for stable distributions, so it is possible to use them in practical problems.

Definition 3.2 Non-degenerate Z is stable if and only if for all $n > 1$,there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$, by such that [16]
$Z_1 + ... + Z_n \overset{\mathrm{d}}{=} c_n Z + d_n$, where $Z_1, ..., Z_n$ are independent, identical copies of Z. Z is strictly stable if and only if $d_n = 0$ for all n.

The only possible choice for the scaling constants is $c_n = n^{\frac{1}{\alpha}}$ for some $\alpha \in (0, 2]$. Both the original definition of stable and the one above use distributional propertiesof Z, yet another distributional characterization is given by the Generalized Central Limit Theorem, While useful, these conditions do not give a concrete way of parameterizing stable distributions. The most concrete way to describe all possible stable distributions is through the characteristic function or Fourier transform. (For a random variable Z with distribution function $F(z)$, the characteristic function is defined by $\varphi(x) = E \exp(ixZ) = \int_{-\infty}^{\infty} \exp(ixz) dF(z)$. The function $\varphi(x)$ completely determines the distribution of Z and has many useful mathematical properties, The sign function is used below, it is defined as

$$sign(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases} \tag{3.7}$$

the $\alpha = 1$ case, $0.log0$ is always interpreted as $\lim_{z \downarrow 0} z \log z = 0$.

Due to the lack of closed form formulas for densities for all but three distributions, see the Figure 3.8, the α-stable law can be most conveniently described by its characteristic function $\varphi(t)$ the inverse Fourier transform of the probability density function. However, there are multiple parameterizations for α-stable laws and much confusion has been caused by these different representations. see Figure 3.9, figure 3.10. The variety of formulas is caused by a combination of historical evolution and the numerous problems that have been analyzed using specialized forms of the stable distributions.

The most popular parameterization of the characteristic function of $X \sim S_\alpha(\beta, \delta, \gamma)$, i.e. an α-stable random variable with parameters α, β, δ and γ, is given by [16] parameterization.

$$\log(\varphi(t)) = \begin{cases} i\delta t - \gamma^\alpha \mid t \mid^\alpha \{1 + i\beta \sin(t) \tan \frac{\pi\alpha}{2}[(\mid \gamma t \mid)^{1-\alpha} - 1]\}, & for\ \alpha \neq 1, \\ i\delta t - \gamma \mid t \mid \{1 + i\beta \frac{2}{\pi} \sin(t) \log(\gamma \mid t \mid)\}, & for\ \alpha = 1. \end{cases} \tag{3.8}$$

When $\alpha > 1$, the mean of the distribution exists and is equal to δ. In general,the pth moment of a stable random variable is finite if and only if $p < \alpha$. When the skewness parameter β is positive, the distribution is skewed to the right,i.e. the right tail is thicker, see the Figure 3.7, Figure 3.8. When it is negative, it is skewed to the left. When $\beta = 0$, the distribution is symmetric about δ. As α approaches $2, \beta$ loses its effect and the distribution approaches the Gaussian distribution regardless of β. The last two parameters, γ and δ, are the usual scale and location parameters, i.e. γ determines the width and δ the shift of the mode (the peak) of the density. For $\gamma = 1$ and $\delta = 0$ the distribution is called standard stable.

see figure 3.7 Stable pdfs for $\alpha = 1.2$ and $\beta = 0$ (black), 0.5 (red), 0.8 (blue dashed line) and 1 (green long-dashed line). Figure 3.8: Closed form formulas for densities are known only for three distributions – Gaussian ($\alpha = 2$; black), Cauchy ($\alpha = 1$; red) and Levy ($\alpha = 0.5, \beta = 1$; blue dashed line). The latter is a totally skewed distribution, i.e. its support is $\mathbb{R}_+$. In general, for $\alpha < 1$ and $\beta = 1(-1)$ the distribution is totally skewed to the right (left).

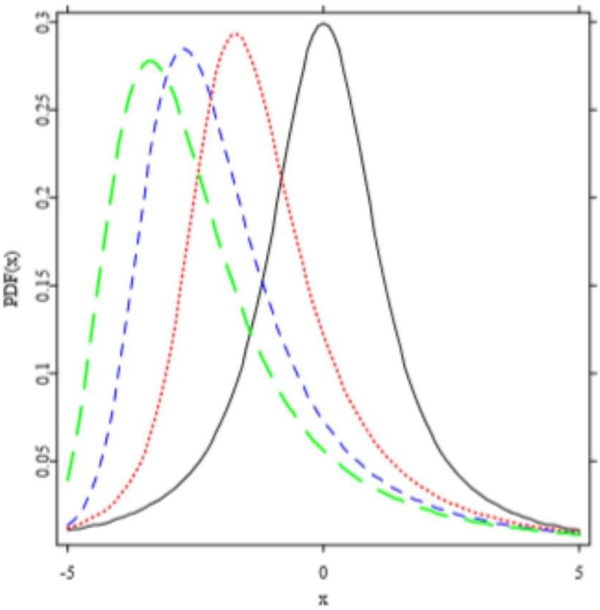

Figure 3.7: Dependence on beta

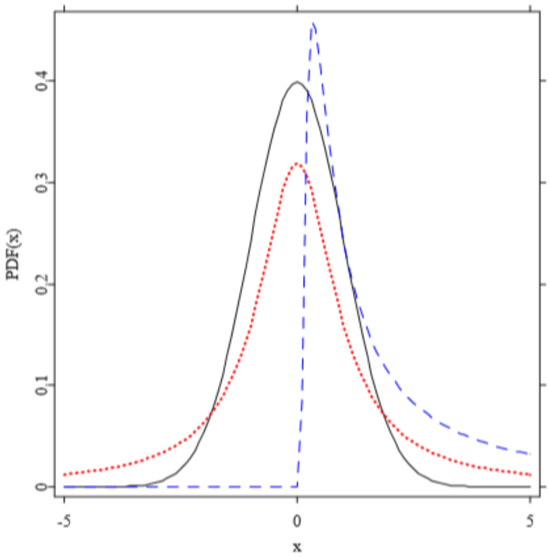

Figure 3.8: Gaussian, Cauchy, and Levy distributions

see Figure 3.9, figure 3.10. The variety of formulas is caused by a combination of historical evolution and the numerous problems that have been analyzed using specialized forms of the stable distributions.

The $S_{\alpha}^{0}\left(\beta, \delta_0, \gamma\right)$ parameterization is a variant of Zolotariev's (M)-parameterization [21] , with the characteristic function and hence the density and the distribution function jointly continuous in all four parameters, see the Figure 3.10. In particular, percentiles and convergence to the power-law tail vary in a continuous way as α and β vary. The location parameters of the two representations are related by $\delta = \delta_0 - \beta\gamma \tan \frac{\pi\alpha}{2}$ for $\alpha = 1$ and $\delta = \delta_0 - \beta\gamma \frac{2}{\pi} \log \gamma$ Note also, that the traditional scale parameter γ of the Gaussian distribution defined by:

$$f_G(x) = \frac{1}{\sqrt{2\pi}\gamma} exp\left\{-\frac{(x-\delta)^2}{2\gamma_G^2}\right\}, \gamma_G = \sqrt{2}\gamma. \qquad (3.9)$$

Figure 3.9 , 3.10: Comparison of S and S^0 parameterizations: α-stable pdfs for $\beta = 0.5$ and $\alpha = 0.5$ (black line), 0.75 (red line), 1 (blue short-dashed line), 1.25 (green dashed line) and 1.5 (cyan longdashed line).

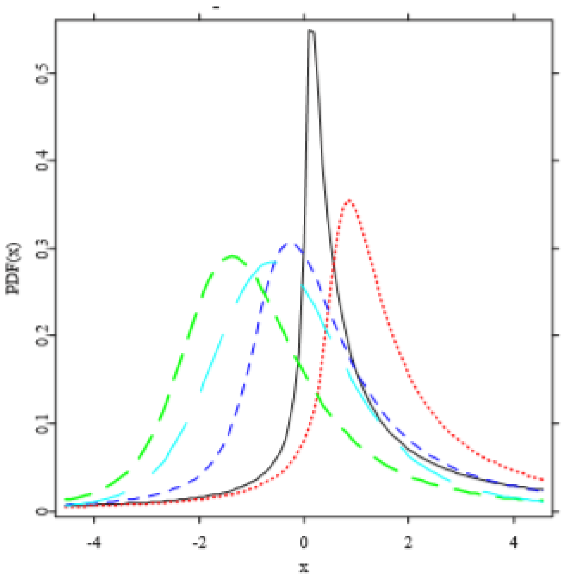

Figure 3.9: S parameterization

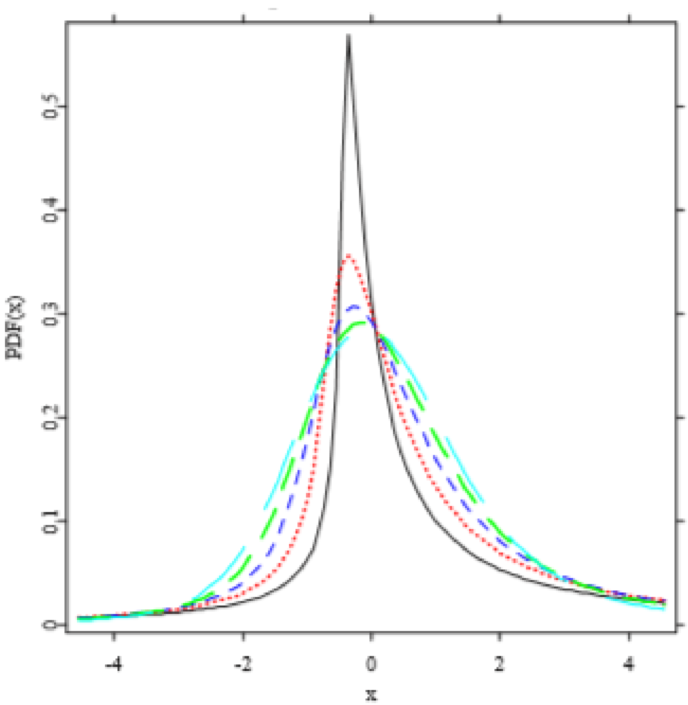

Figure 3.10: S^0 parameterization

3.3 Stable density and distribution functions

The lack of closed form formulas for most stable densities and distribution functions has negative consequences. For example, during maximum likelihood estimation computationally burdensome numerical approximations have to be used. There generally are two approaches to this problem. Either the fast Fourier transform (FFT) has to be applied to the characteristic function [15] or direct numerical integration has to be utilized [16]. For data points falling between the equally spaced FFT grid nodes an interpolation technique has to be used. Taking a larger number of grid points increases accuracy, however, at the expense of higher computational burden. The FFT based approach is faster for large samples, whereas the direct integration method favors small data sets since it can be computed at any arbitrarily chosen point. [15] report that for $N = 2^{13}$ the FFT based method is faster for samples exceeding 100 observations and slower for smaller data sets. Moreover, the FFT based approach is less universal – it is efficient only for large α's and only for pdf calculations. When computing the cdf the density must be numerically integrated.

In contrast, in the direct integration method [21] formulas either for the density or the distribution function are numerically integrated.

Set $\zeta = -\beta \tan \frac{\pi\alpha}{2}$.Then the density $f(x;\alpha,\beta)$ of a standard α-stable random variable in representation S^0, i.e. $X \sim S^0_\alpha(1,\beta,0)$ [4].

- when $\alpha \neq 1$ and $x > \zeta$:

$$f(x;\alpha,\beta) = \frac{\alpha(x-\zeta)^{\frac{1}{\alpha-1}}}{\pi \mid \alpha - 1 \mid} \int_{-\xi}^{\frac{\pi}{2}} V(\theta;\alpha,\beta) exp\left\{-(x-\zeta)^{\frac{\alpha}{\alpha-1}} V(\theta;\alpha\beta)\right\} d\theta, \tag{3.10}$$

- when $\alpha \neq 1$ and $x = \zeta$:

$$f(x;\alpha,\beta) = \frac{\Gamma(1+\frac{1}{\alpha})\cos(\xi)}{\pi(1+\zeta^2)^{\frac{1}{2\alpha}}}, \tag{3.11}$$

- when $\alpha \neq 1$ and $x < \zeta$:

$$f(x;\alpha,\beta) = f(-x;\alpha,-\beta), \tag{3.12}$$

- when $\alpha = 1$:

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{2|\beta|} e^{\frac{\pi x}{2\beta}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} V(\theta;1,\beta) exp\left\{-e^{-\frac{\pi x}{2\beta}} V(\theta;1,\beta)\right\} d\theta, & \beta \neq 0, \\ \frac{1}{\pi(1+x^2)}, & \beta = 0, \end{cases} \tag{3.13}$$

where

$$\xi = \begin{cases} \frac{1}{\alpha} arctan(-\zeta), & \alpha \neq 1, \\ \frac{\pi}{2}, & \alpha = 1, \end{cases} \tag{3.14}$$

and

$$V(\theta;\alpha,\beta) = \begin{cases} (\cos\alpha\xi)^{\frac{1}{\alpha-1}} \left(\frac{\cos\theta}{\sin\alpha(\xi+\theta)}\right)^{\frac{\alpha}{\alpha-1}} \frac{cos\{\alpha\xi+(\alpha-1)\theta\}}{\cos\theta}, & \alpha \neq 1, \\ \\ \frac{2}{\pi}\left(\frac{\frac{\pi}{2}+\beta\theta}{\cos\theta}\right) exp\left\{\frac{1}{\beta}(\frac{\pi}{2}+\beta\theta)\tan\theta\right\}, & \alpha = 1, \beta \neq 0. \end{cases} \tag{3.15}$$

The distribution $F(x;\alpha,\beta)$ of a standard α-stable random variable in representation S_0 can be expressed as:

- when $\alpha \neq 1$ and $x > \zeta$:

$$F(x;\alpha,\beta) = d_1(\alpha,\beta) + \frac{\sin(1-\alpha)}{\pi}\int_{-\xi}^{\frac{\pi}{2}} exp\left\{-(x-\zeta)^{\frac{\alpha}{\alpha-1}}V(\theta;\alpha,\beta)\right\} d\theta,$$

where

$$d_1(\alpha,\beta) = \begin{cases} \frac{1}{\pi}(\frac{\pi}{2}-\xi), \alpha < 1, \\ \\ 1, \quad \alpha > 1 \end{cases} \tag{3.16}$$

when $\alpha \neq 1$ and $x = \zeta$:

$$F(x;\alpha,\beta) = \frac{1}{\pi}(\frac{\pi}{2}-\xi), \tag{3.17}$$

when $\alpha \neq 1$ and $x < \zeta$:

$$F(x;\alpha,\beta) = 1 - F(-x;\alpha,-\beta), \tag{3.18}$$

when $\alpha = 1$:

$$F(X;1,\beta)=\begin{cases}\frac{1}{\pi}\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}}\exp\{-e^{-\frac{\pi x}{2\beta}}V(\theta;1,\beta)\}d\theta,\ \beta>0,\\ \frac{1}{2}+\frac{1}{\pi}\arctan x,\ \beta=0,\\ 1-F(x,1,-\beta),\ \beta<0,\end{cases}\tag{3.19}$$

Formula (3.10) requires numerical integration of the function $g(\cdot)\exp\{-g(.)\}$, where $g(\theta;x,\alpha,\beta)=(x-\zeta)^{\frac{\alpha}{\alpha-1}}V(\theta;\alpha,\beta)$. The integrand is 0 at $-\xi$, increases monotonically to a maximum of $\frac{1}{e}$ at point θ^* for which $g(\theta^*$;x,$\alpha,\beta)=1$, and then decreases monotonically to 0 at $\frac{\pi}{2}$ [16]. However, in some cases the integrand becomes very peaked and numerical algorithms can miss the spike and underestimate the integral. To avoid this problem we need to find the argument θ^* of the peak numerically and compute the integral as a sum of two integrals: one from $-\xi$ to θ^* and the other from θ^* to $\frac{\pi}{2}$.

Theorem 3.1 All (non-degenerate) stable distributions are continuous distributions with an infinitely differentiable density, [16].

To distinguish between the densities and cumulative distribution functions in different parameterizations, $f(x\mid\alpha,\beta,\delta,\gamma;k)$ will denote the density and $F(x\mid\alpha,\beta,\delta,\gamma;k)$ will denote the d.f. of a $S(\alpha,\beta,\delta,\gamma;k)$ distribution. When the distribution is standardized, i.e. scale $\gamma=1$, and location $\delta=0$, $f(x\mid\alpha,\beta;k)$ will be used for the density, and $F(x\mid\alpha,\beta;k)$ will be used for the d.f..
Since all stable distributions are shifts and scales of some $Z\sim(\alpha,\beta;0)$, we will focus on those distributions here. The computer program STABLE, was used to compute the probability density functions (pdf) and (cumulative) distribution functions (d.f.) below to illustrate the range of shapes of these distributions. Stable densities are supported on either the whole real line or a half line. The latter situation can only occur when $\alpha<1$ and ($\beta=+1$ or $\beta=-1$). Precise limits are given by the following lemma.

Lemma 3.1 The support of a stable distribution in the different parameterizations is [16]

$$support\ f(x|\alpha,\beta,\gamma,\delta;0) = \begin{cases} [\delta - \gamma\tan\frac{\pi\alpha}{2}, \infty),\ \alpha < 1\ and\ \beta = 1 \\ (-\infty, \delta + \gamma\tan\frac{\pi\alpha}{2}]\ , \alpha < 1\ and\ \beta = -1 \\ (-\infty, +\infty),\ otherwise \end{cases} \tag{3.20}$$

$$support\ f(x|\alpha,\beta,\gamma,\delta;1) = \begin{cases} [\delta, \infty),\ \alpha < 1\ and\ \beta = 1 \\ (-\infty, \delta],\ \alpha < 1\ and\ \beta = -1 \\ (-\infty, +\infty),\ otherwise \end{cases} \tag{3.21}$$

The constant tan $\frac{\pi\alpha}{2}$ appears frequently when working with stable distributions, so it is worth recording its behavior. As $\alpha \uparrow 1$, tan $\frac{\pi\alpha}{2} \uparrow +\infty$, the expression is undefined at $\alpha = 1$, and when $\alpha \downarrow 1$,tan $\frac{\pi\alpha}{2} \downarrow -\infty$ .This essential discontinuity at $\alpha = 1$ is sometimes anuisance when working with stable distributions, but here it is natural: if $|\beta| = 1$ then as $\alpha \uparrow 1$, the support in Lemma 1.4 grows to Rin a natural way.

3.4 Logarithmic moments

This approach is as a result of the challenges encountered when using the FLOM method which requires computing Gamma functions, the inversion of the sinc function and it only works for some p. The current method suggests computing derivatives with respect to the moment order p resulting in moments of the logarithms of the stable process. We illustrate in the following.

Lemma 3.2 Let S denote a symmetric stable random variable and let $p \in \mathbb{R}$. Then [10]

$$M_n = E[(\log \mid S \mid)^n] = \lim_{p \to 0} \frac{d^n}{dp^n} E[\mid S \mid^p], \quad n = 1, 2, ... \tag{3.22}$$

The moments follow readily for $n = 1, 2, ...$

$$M_1 = E[\log \mid S \mid] = \phi_0(1 - \frac{1}{\alpha}) + \frac{1}{\alpha} \log \mid \frac{\gamma}{\cos\theta} \mid \tag{3.23}$$

$$M_2 = E[(\log \mid S \mid - E[\log \mid S \mid])]^2 = \phi_1(\frac{1}{2} + \frac{1}{\alpha^2}) - \frac{\theta^2}{\alpha^2} \tag{3.24}$$

$$M_3 = E[(\log \mid S \mid - E[\log \mid S \mid])^3] = \phi_3(1 - \frac{1}{\alpha^3}) \tag{3.25}$$

where $\theta = \arctan(\frac{\beta \tan \alpha\pi}{2})$ terms ϕ_k are given by $\phi_0 = -0.57721566$, $\phi_1 = \frac{\pi^2}{6}$, $\phi = 1.2020569$ derived from the polygamma function

$$\phi_{k-1} = \frac{d^k}{dx^k} \log \Gamma(x) \mid_{x=1} \tag{3.26}$$

3.5 methods of parameter estimation of stable distribution

The four common methods for estimating parameters of stable processes include: quantiles method, the logarithmic moments method, the empirical characteristics method and the ML method. We will investigate their accuracy in the following.

3.5.1 The quantiles method

Was much more appreciated through [14] after its extension to include asymmetric distributions and for cases where $\alpha \in [0.6, 2]$ unlike the approach that restricts it to $\alpha \geq 1$.

Suppose $\hat{h}$ is a given data sample then the estimates for α and β are given by $\hat{\alpha} = \Theta_1(\hat{\theta}_\alpha, \hat{\theta}_\beta)$ and $\hat{\beta} = \Theta_2(\hat{\theta}_\alpha, \hat{\theta}_\beta)$ where

$$\hat{\theta}_\alpha = \frac{\hat{h}_{0.95} - \hat{h}_{0.05}}{\hat{h}_{0.75} - \hat{h}_{0.25}}, \quad \hat{\theta}_\beta = \frac{\hat{h}_{0.95} + \hat{h}_{0.05} - 2\hat{h}_{0.05}}{\hat{h}_{0.95} - \hat{h}_{0.05}} \tag{3.27}$$

The notation $\hat{h}_q$ represents the qth quantile of sample $\hat{h}$ and, $\hat{\alpha}$ and $\hat{\beta}$ are obtained by functions $\Theta_1(\hat{\theta}_\alpha, \hat{\theta}_\beta)$ and $\Theta_2(\hat{\theta}_\alpha, \hat{\theta}_\beta)$ given in Tables III and IV in [14] through linear interpolation. Consequently, the scale parameter is given by

$$\hat{\gamma} = \frac{\hat{h}_{0.75} - \hat{h}_{0.25}}{\Theta_3(\hat{\alpha}, \hat{\beta})} \tag{3.28}$$

where $\Theta_3(\hat{\alpha}, \hat{\beta})$ is given by Table V in [14]. The consistent estimator γ is then obtained through interpolation.

Finally the location parameter δ is estimated through a new parameter defined by

$$\xi = \begin{cases} \delta + \beta v \tan \frac{\pi\alpha}{2}, & \alpha \neq 1 \\ \delta, & \alpha = 1 \end{cases} \tag{3.29}$$

Moreover, ξ is estimated by

$$\hat{\xi} = \hat{h}_{0.5} + \hat{\gamma}\Theta_5(\hat{\alpha}, \hat{\beta}), \tag{3.30}$$

where $\Theta_5(\hat{\alpha}, \hat{\beta})$ is obtained from Table VII in [14] by linear interpolation. The location parameter is estimated consistently by

$$\hat{\delta} = \hat{\xi} + \hat{\beta}\hat{\gamma} \tan \frac{\pi\hat{\alpha}}{2} \tag{3.31}$$

3.5.2 Empirical characteristic function method

Suppose a set of observable data $\{h_1, h_2, ..., h_N\}$ follows a stable distribution. Then we can approximate the characteristic function of this data by applying a basic

Monte Carlo approach based on the law of large numbers i.e.

$$\varphi(u) = E[e^{iuh_j}] \approx \hat{\varphi}(u) = \frac{1}{N}\sum_{j=1}^{N} e^{iuh_j} \tag{3.32}$$

We can express the characteristic function (3.8) in terms of the cosine and sine function from basic trigonometric principles, i.e.

$$\varphi(u) = e^{-|\gamma u|^{\alpha}}(\cos\eta + i\sin\eta), \tag{3.33}$$

where

$$\eta = \gamma u - \mid \gamma u \mid^{\alpha} \beta \sin(u)\omega(u,\alpha)$$

$$\omega(u,\alpha) = \begin{cases} \tan\frac{\pi\alpha}{2}, \ \alpha \neq 1 \\ \frac{2\log|u|}{\pi}, \ \alpha = 1 \end{cases}$$

As a result, we observe that

$$\mid \varphi(u) \mid = e^{-|\gamma u|\alpha} \tag{3.34}$$

The estimated characteristic function relates to the model parameters by:

$$\log \mid \hat{\varphi}(u_k) \mid = \gamma^{\alpha} \mid u_k \mid^{\alpha}, \ for\ k = 1, 2, \ u_k > 0, \ \alpha \neq 1. \tag{3.35}$$

Solving this system leads to the estimation representation formulas for the stability and variance parameters:

$$\hat{\alpha} = \frac{\log \frac{\log|\hat{\varphi}(u_1)|}{\log|\hat{\varphi}(u_2)|}}{\log \mid \frac{u_1}{u_2} \mid}$$

$$\log\hat{\gamma} = \frac{\log\mid u_1\mid\log(-\log\mid\hat{\varphi}(u_2)\mid) - \log\mid u_2\mid\log(-\log\mid\hat{\varphi}(u_1)\mid)}{\log\mid\frac{u_1}{u_2}\mid}$$

The real and imaginary parts of the characteristic function (3.33) provide estimates for $\hat{\beta}$ and $\hat{\delta}$:

$$\arctan\frac{Im(\varphi(u))}{Re(\varphi(u))} = \delta u - \mid\gamma u\mid^{\alpha}\beta\sin(u)\omega(u,\alpha) \tag{3.36}$$

Suppose $\Upsilon(u) = \arctan(\frac{Im(\varphi(u))}{Re(\varphi(u))})$ and choose another set of positive numbers $x_k, k = 3, 4$ together with $\hat{\alpha}$ and $\hat{\gamma}$ then the estimates of the location and skewness parameters are given respectively by

$$\hat{\delta} = \frac{u_4^{\hat{\alpha}}\Upsilon(u_3) - u_3^{\hat{\alpha}}\Upsilon(u_4)}{u_3u_4^{\hat{\alpha}} - u_3u_4^{\hat{\alpha}}} \tag{3.37}$$

$$\hat{\beta} = \frac{u_4\Upsilon(u_3) - u_3\Upsilon(u_4)}{\hat{\gamma}^{\hat{\alpha}}tan\frac{\pi\hat{\alpha}}{2}(u_4u_3^{\hat{\alpha}} - u_3u_4^{\hat{\alpha}})} \tag{3.38}$$

Notice, it can be deduced from Equation (3.33) that

$$\log(-\log(\mid\varphi(u)\mid^2)) = \log(2\gamma^{\alpha}) + \alpha\log(u)$$

This provides an alternative way to envision the regression estimation method:

$$x_i = d_0 + d_1v_i + \epsilon_i,\ i = 1, 2, ..., n,$$

where $x_i = \log(-\log(\mid\hat{\varphi}(u_i)\mid^2)), d_0 = \log(2\gamma^{\alpha}), v_i = \log(u_i)$ and ϵ_i is an error term. The stability parameter d_1 and the scale parameter γ can be estimated by selecting $u_i = \frac{\pi i}{25}, i = 1, 2, ..., n$; of real data see [10]. The estimates $\hat{\alpha}$ and $\hat{\gamma}$ are then used to estimate β and δ using the following relation

$$z_k = \eta_k + \xi_k, \quad k = 1, 2, ..., Q.$$

where $z_k = \Upsilon_n(u_k) + \pi l_n(u_k), \eta_k = \hat{\gamma}_k u - \mid \hat{\gamma}_k u \mid^{\hat{\alpha}} \beta \sin(u)\omega(u, \hat{\alpha})$ and ξ_k is some random error. The proposed real data set for Q (see [10], Table II) is $u_k = \frac{\pi k}{50}, \; k = 1, 2, ..., Q$.

3.5.3 Logarithmic moments method

The key innovation with this method is that there is no need of computing Gamma functions and the sinc function as in the FLOM. Secondly, techniques of parameter estimation for symmetric stable random variables (i.e. $\beta = 0$) can be applied to skewed stable random variables (i.e. $\beta \neq 0$) and, techniques of parameter estimation for centered stable random variables (i.e. $\gamma = 0$) to non-centered ones (i.e. $\gamma \neq 0$) through centro-symmetrization. However, this comes at a cost of losing almost half of the sample data. Therefore to obtain better estimates one has to use large sample data sets.

Centro-symmetrization of stable random data sets

Let S_k be a sequence of n independent stable random variables distributed according to $S_k \sim S(\alpha, \beta, \delta, \gamma)$. Then the distribution of a weighted sum of the above sequence with weights a_k can be estimated using their characteristic function:

$$Z = \sum_{k=1}^{n} a_k S_k \sim S \left[\alpha, \quad \frac{\sum_{k=1}^{n} a_k^{<\alpha>}}{\sum_{k=1}^{n} \mid a_k \mid^{\alpha}} \quad \beta, \sum_{k=1}^{n} a_k \delta, \sum_{k=1}^{n} \mid a_k \mid^{\alpha} \gamma \right] \tag{3.39}$$

where the *pth* power of a number x is defined by $X^{<P>} = \sin(X) \mid X \mid^{P}$.

As a result, it is easy to obtain sequences of independent stable random variables with zero δ zero β as well as both zero δ and zero β for $\alpha \neq 1$. This yields the centred, deskewed, and symmetrized sequences:

$$S_k^C = S_{3k} + S_{3k-1} - 2S_{3k-2} \sim S(\alpha, [\frac{2-2^\alpha}{2+2^\alpha}]\beta, [2+2^\alpha]\gamma, 0) \tag{3.40}$$

$$S_k^D = S_{3k} + S_{3k-1} - 2^{\frac{1}{\alpha}} S_{3k-2} \sim S(\alpha, 0, 4\gamma, [2-2^{\frac{1}{\alpha}}]\delta \tag{3.41}$$

$$S_k^S = S_{2k} - S_{2k-1} \sim S(\alpha, 0, 2\gamma, 0) \tag{3.42}$$

Parameter estimation

Suppose S_k is a data set assumed to be drawned from $S(\alpha, \beta, \delta, \gamma)$ .Then the exponent parameter α is estimated by setting $\theta = 0$ in (3.24), and the log moment M_2 is estimated from the obverted data (3.42). That is,

$$\hat{\alpha} = (\frac{M_2}{\phi_1} - \frac{1}{2})^{\frac{-1}{2}} \tag{3.43}$$

The estimated $\hat{\alpha}$ is used to estimate θ using (3.23) where M_1 is estimated from the obverted data (3.41). That is,

$$\mid \hat{\theta} \mid = ((\frac{\phi_1}{2} - M_2)\hat{\alpha}^2 + \phi_1)^{\frac{1}{2}} \tag{3.44}$$

From the definition of θ, $\mid \beta_0 \mid$ can be estimated by

$$\hat{\beta_0} = \frac{\tan \hat{\theta}}{\tan \frac{\hat{\alpha}\pi}{2}} \tag{3.45}$$

Centering see (3.40) requires $\mid \hat{\beta_0} \mid$ to be multiplied by $\frac{2+2^\alpha}{2-2^\alpha}$ to obtain $\mid \hat{\beta} \mid$ of the original data where the sign of β is determined by

$K = \sin(\mid S_{max} - S_{md} \mid - \mid S_{min} - S_{md} \mid)$, such that $\hat{\beta} = K \mid \hat{\beta} \mid$.

where S_{max}, S_{md}, S_{min} is the maximum, median and minimum of the original data. Next we estimate the scale parameter $\hat{\gamma_0}$ using (3.23) where M_1 is estimated from the obverted data (3.40). That is

$$\hat{\gamma_0} =\mid \cos\hat{\theta} \mid \exp(M_1 - \phi_0)\hat{\alpha} + \phi_0) \tag{3.46}$$

Again centering see (3.40) gives the parameter estimate $\hat{\gamma}$ of the original data by $\hat{\gamma} = \hat{\gamma}_0(2 - 2^{\frac{1}{\alpha}})^{-1}$. Finally, the location parameter δ is estimated by

$$\hat{\delta} = \hat{\delta}_0(2 - 2^{\frac{1}{\alpha}})^{-1} \tag{3.47}$$

where δ_0 is the median or mean of the obverted data .

3.5.4 Maximum likelihood method

The ML method is the most favored parameter estimation method in economic and financial applications. The method relies on the density function which in the case of stable distributions poses a closed form representation problem.The method of maximum likelihood chooses as estimates those values of the parameters that are most consistent with the sample data.

$$If \quad X_i \sim F(\Theta), \quad i = 1, ..., n$$

then the likelihood function is

$$Ł(\{X_i\}_{i=1}^n, \Theta) = \prod_{i=1}^{n} F(X_i, \Theta) \tag{3.48}$$

The likelihood function can be maximized w.r.t. the parameter(s) Θ, doing this one can arrive at estimators for parameters as well.

$$Ł(\{X_i\}_{i=1}^n, \Theta) = \prod_{i=1}^{n} F(X_i, \Theta)$$

To do this, find solutions to (analytically or by following gradient)

$$\frac{d\text{Ł}(\{X_i\}_{i=1}^n, \Theta)}{d\Theta} = 0 \tag{3.49}$$

almost maximize the likelihood function, maximize the log likelihood function instead.

$$\log(\text{Ł}(\{X_i\}_{i=1}^n, \Theta)) = \log(\prod_{i=1}^n F(X_i, \Theta)) = \sum_{i=1}^n \log(F(X_i, \Theta)) \tag{3.50}$$

Quite often the log of the density is easier to work with mathematically.

Likelihood function :

$$\text{Ł}(d_0, d_1, \sigma^2) = \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(X_i - d_0 - d_1 V_i)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - d_0 - d_1 V_i)^2}$$

$$\hat{\sigma}^2 = \frac{\sum_i (X_i - \hat{X}_i)^2}{n}$$

Note that maximum likelihood estimator is biased as s^2 is unbiased and

$$s^2 = MSE = \frac{n}{n-2}\hat{\sigma}^2$$

Chapter 4

Linear Regression Models Assuming a Stable Distribution

4.1 Introduction

Stable distributions have long been regarded as important generalizations of the normal distribution, being defined as the class of distributions whose location-scale families are closed under convolution. In a more practical setting, stable distributions have attracted considerable interest, be- cause they can allow for skew and for arbitrarily larger tails than can the normal distribution.

Stable distributions are described by four parameters $(\alpha, \beta, \delta, \gamma)$ with $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\delta(-\infty, \infty)$, and $\gamma(0, \infty)$.

The parameter α, known as the characteristic exponent, defines the "fatness of the tails." For a variety of mathematical reasons, the case $\alpha = 1$ is usually studied separately from the $\alpha \neq 1$ cases. Skewness is governed by β, the symmetric case corresponding to $\beta = 0$. The location and scale of the distributions are denoted by δ and γ.

The two best-known stable distributions are the Normal ($\alpha = 2$) and the Cauchy ($\alpha = 1, \beta = 0$), Sable distributions are usually as in the previous chapter.

Characteristic functions are essentially Fourier transformations of distribution functions, which provide a general and powerful tool to analyze probability distributions.

Definition 4.1 The function $\varphi_X(t) = E\exp(it^\tau X)$ is called the characteristic function (cf) of X [12].

Every distribution on $\mathbb{R}^p$ has a cf regardless of whether moments exist. Recall from complex analysis that $\exp(iu) = \cos(u) + i\sin(u)$. So, we see that $exp(it^\tau x)$ is indeed bounded as a function of x for each t.

Example 4.1 (Normal distribution). Let $f_X(x) = \frac{\exp\frac{-x^2}{2}}{\sqrt{2\pi}}$ be the density of X [12]. Then

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}}\int \exp\left(itx - \frac{x^2}{2}\right)dx$$

$$= \frac{1}{\sqrt{2\pi}}\int \exp\left(-\frac{1}{2}[x-it]^2 - \frac{t^2}{2}\right)dx$$

$$= \exp\left(\frac{-t^2}{2}\right).$$

Example 4.2 (Uniform distribution). Let $f(x) = \frac{1}{2}$ for $-1 < x < 1$ [12]. Then

$$\varphi(t) = \frac{1}{2}\int_{-1}^{1}\exp(itx)dx = \frac{\exp(it) - \exp(-it)}{2it} = \frac{\sin(t)}{t}.$$

Example 4.3 (Cauchy distribution). Let $f_X(x) = [\pi(1+x^2)]^{-1}$ [12]. Then

$$\varphi_X(t) = \exp(-\mid t\mid).$$

(Basic properties of cf). All cf's have the following properties:

1. $\varphi(0) = 1, \mid \varphi(t) \mid \leq 1$,

2. $\varphi(-t) = \overline{\varphi(t)}$ (complex conjugate),

3. $\mid \varphi(t+h) - \varphi(t) \mid \leq E \mid e^{ihX}1 \mid$ (uniform continuity),

4. $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.

The next result gives a sufficient condition for $\varphi(t)$ to be a cf.

Theorem 4.1 (Polya's Criterion). Let φ be continuous, real, nonnegative, symmetric, decreasing and convex on $[0, \infty)$, such that $\varphi(0) = 1$, $\lim_{t\to\infty} \varphi(t) = 0$, then φ is a characteristic function [12].

Proposition 4.1 If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$ [12].

The remaining theorems about convergence in distribution are

• the inversion/uniqueness theorem that says that each cf corresponds to a unique distribution,

• the continuity theorem that says that $X_n \xrightarrow{D} X$ if and only if $\varphi_{Xn}(t)\varphi_X(t)$ for all t(the "only if" direction being trivial), and

• the central limit theorem that says that certain normalized sums of independent (not necessarily identically distributed) random variables with finite variance converge in distribution to a standard normal distribution.

Theorem 4.2 (Inversion and uniqueness). Let φ be the cf for the probability P on $(\mathrm{I\!R^p}, \mathrm{B^p})$ [12]. Let A be a rectangular region of the form

$$A = \{(x_1, ..., x_p) : a_j \leq x_j \leq b_j \ all \ j\},$$

where $a_j < b_j$ for all j and $P(\partial A) = 0$. For each $T > 0$, let

$$B_T = (t_1, ..., t_p) : -T \leq t_j \leq T \ for \ all \ j.$$

Then

$$P(A) = \lim_{T\to\infty} \frac{1}{(2\pi)^p} \int_T \prod_{j=1}^{p} \left[\frac{\exp(-it_j a_j) - \exp(-it_j - b_j)}{it_j} \right] \varphi(t) dt_1 ... dt_p.$$

Distinct probability measures have distinct cf's.

The characteristic function $\varphi(.)$ of a stable distribution is given by equation (3.8) as in the previous chapter.

where $i = \sqrt{-1}$ and the sign (.) function is given by (3.7) as in the previous chapter.

It is important to point out that if $\alpha < 1$, the variance is infinite and the mean of the stable distribution does not exist. Although this class of distributions is a good alternative for data modeling in different areas, we usually have difficulties to obtain estimates under a classical inference approach due to the lack of closed form expressions for their probability density functions. One possibility in applications is to get the probability density function from the inversion formula,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt \tag{4.1}$$

where $\varphi(t)$ is the characteristic function. In applications, we need use numerical methods to solve the integral in (4.1), usually taking a great computational time. An alternative is the use of Bayesian methods. However, the computational cost can be further high to get the posterior summaries of interest. A good alternative is to use latent or artificial variables that could improve the simulation computation of samples of the joint posterior distributions of interest. In this way, a Bayesian analysis of stable distributions was using Markov Chain Monte Carlo (MCMC) methods and latent variables. The use of Bayesian methods with MCMC simulation can have great flexibility by considering latent variables where samples of latent variables are simulated in each step of the Gibbs or Metropolis-Hastings algorithms.

The Bayesian paradigm is to update prior parameter knowledge, in the form of a density function $\pi(\theta)$, using observed data x, through the parametric model density function $f(x \mid \theta)$. The result is a posterior density given by $\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta)$. If the model density for x is not available in closed form, but the model density,$f(x, y \mid \theta)$, jointly with some extra random variables y is available, then the posterior is formally obtainable through Bayes's theorem by integrating out the unwanted variables, so that $\pi(\theta \mid x) \propto \int f(x, y \mid \theta)\pi(\theta)dy$. Such a representation is available for x modeled by stable distributions.

Theorem 4.3 Let the bivariate probability density function of Z and Y, conditional on α and β,$f : (-\infty, 0) \times (-\frac{1}{2}, l_{\alpha,\beta}) \cup (0, \infty) \times (l_{\alpha,\beta}, \frac{1}{2}) \to (0, \infty)$ [6],be given by

$$f(z, y \mid \alpha, \beta) = \frac{\alpha}{\mid \alpha - 1 \mid} \exp\left\{ - \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \right\} \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \frac{1}{\mid z \mid}, \tag{4.2}$$

Where $\theta = \frac{\alpha}{\alpha - 1}$

$$t_{\alpha,\beta}(y) = \left(\frac{\sin[\pi\alpha y + b_{\alpha,\beta}]}{\cos \pi y} \right) \left(\frac{\cos \pi y}{\cos[\pi(\alpha - 1)y + b_{\alpha,\beta}]} \right)^{\frac{1}{\theta}} \tag{4.3}$$

and $\alpha \in (0, 2], \beta \in [-1, 1], \delta \in (-\infty, \infty)$ and $\gamma \in (0, \infty),)$, with $b_{\alpha,\beta} = \beta \min(\alpha, 2 - \alpha)\frac{\pi}{2}$ and $l_{\alpha,\beta} = -b_{\alpha,\beta}/\pi\alpha$. Then f is a proper bivariate probability density for the distribution of (Z, Y), and the marginal distribution of Z is $S_\alpha(\beta, 0, 1)$.

proof: $t_{\alpha,\beta}$ is defined, continuous, and strictly monotonic on (-0.5,0.5); $t_{\alpha,\beta} = (\pm 0.5) = \pm\infty$; $t_{\alpha,\beta}(y) = 0 \Longleftrightarrow y = l_{\alpha,\beta}$; $t_{\alpha,\beta}(y) = -t_{\alpha,-\beta}(-y)$; and the transformation $y \to t_{\alpha,\beta}(y)$ is 1-1.

We begin with random variables W and Y distributed independently as exponential(1) and uniform $(-0.5, 0.5)$. The joint probability density function of the distribution of (W, Y) is given as $f(w, y) = \exp\{-w\}$. Now the transformation

$T : (0, \infty) \times (-0.5, 0.5) \to (-\infty, 0) \times (-0.5, l_{\alpha,\beta}) \cup (0, \infty) \times (l_{\alpha,\beta}, 0.5)$,

where $T(W, Y) = (t_{\alpha,\beta}(Y)W^{\frac{\alpha-1}{\alpha}}, Y)$.

Note that the transformation is 1-1 and that the inverse transformation is given by

$$T^{-1}(Z, Y) = \left(\left| \frac{Z}{t_{\alpha,\beta}(Y)} \right|^{\frac{\alpha-1}{\alpha}}, Y \right).$$

By calculating the Jacobian of this transformation, we have that the joint probability density function of the distribution of (Z, Y) is given as

$$f(z, y \mid \alpha, \beta) = \frac{\alpha}{\mid \alpha - 1 \mid} \exp\left\{ - \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \right\} \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \frac{1}{\mid z \mid}, \tag{4.4}$$

and so our function is indeed a proper bivariate density function. Moreover, the marginal of the distribution of Z has density function

$$\begin{aligned} f(z \mid \alpha, \beta) &= \int_{l_{\alpha,\beta}}^{0.5} f(z, y \mid \alpha, \beta) dy \quad if \quad z > 0, \\ &= \int_{-0.5}^{l_{\alpha,\beta}} f(z, y \mid \alpha, \beta) dy \quad if \quad z < 0, \end{aligned}$$

which, using the properties of $t_{\alpha,\beta}$, can be written as

$$\begin{aligned} f(z \mid \alpha, \beta) &= \int_{l_{\alpha,\beta}}^{0.5} f(z, y \mid \alpha, \beta) dy \quad if \quad z > 0, \\ &= \int_{l_{\alpha,-\beta}}^{0.5} f(z, y \mid \alpha, -\beta) dy \quad if \quad z < 0. \end{aligned}$$

[21] proved that the cumulative distribution function of a stable distribution can be represented as

$$\begin{aligned} F(z \mid \alpha, \beta) &= \frac{1}{2} + l_{\alpha,\beta} + \int_{l_{\alpha,\beta}}^{0.5} \exp\left\{ - \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \right\} dy \quad if \quad \alpha < 1, \\ &= 1 - \int_{l_{\alpha,\beta}}^{0.5} \exp\left\{ - \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\theta} \right\} dy \quad if \quad \alpha > 1, \end{aligned}$$

where $\theta = \frac{\alpha}{\alpha - 1}$, $z > 0$. (For $z < 0$, the same representation is used with β replaced by $-\beta$.) By differentiating, we have that the probability density function of a stable distribution can be represented as

$$f(z \mid \alpha, \beta) = \int_{l_{\alpha,\beta}}^{0.5} f(z, y \mid \alpha, \beta) dy \quad if \quad z > 0,$$

$$= \int_{l_{\alpha,-\beta}}^{0.5} f(z, y \mid \alpha, -\beta) dy \quad if \quad z < 0.$$

which is marginal probability density function derived earlier.

4.2 Linear Regression Models Assuming a Stable Distribution

A random variable X related to a controlled variable V given by the linear relationship

$$x_i = d_0 + d_1 v_i + \epsilon_i, \; for \; i = 1, 2, ..., n, \tag{4.5}$$

where

• the random variable X_i represents the response for the ith unit associated with an experimental value of the independent or explanatory variable v, which is assumed to have a fixed value (a common regression model assumption). In this way, x_i it is an observation of X_i;

• the variables $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are considered as components of unknown errors and are unobserved random variables. Assume that these random variables ϵ_i, for $i = 1, 2, ..., n$, are independent and identically distributed with normal distribution $(0, \sigma_\epsilon^2)$;

• the parameters d_0 and d_1 are unknown.

From the above assumptions, we have normality for the responses, that is,

$$X_i \sim N(d_0 + d_1 v_i; \sigma_\epsilon^2). \tag{4.6}$$

In this way X_i has a normal distribution with mean $d_0 + d_1 v_i$ and common variance σ_ϵ^2. Usually we get estimators for the regression parameters using the least squares approach or standard maximum likelihood methods.

Standard generalization for the linear model (4.5) is given in the presence of k independent or explanatory variables, that is, a multiple linear regression model given by

$$x_i = d_0 + d_1 v_{i1} + d_2 v_{i2} + ... + d_k v_{ik} + \epsilon_i \tag{4.7}$$

From the normality assumption for the error ϵ_i in (4.7), the random variable X_i has a normal distribution with mean $d_0 + d_1 v_{i1} + d_2 v_{i2} + ... + d_k v_{ik}$ and variance σ_ϵ^2.

In practical applications, we need to check if the above assumptions are satisfied. As such, we consider graphical approaches to verify if the model residuals satisfy the above assumptions.

In the presence of outliers or discordant observations, we could have a large impact on the estimators obtained for the regression model given by (4.7), which could invalidate the inferences obtained. In this situation, we could use nonparametric regression models or assume more robust probability distributions for the data. One possibility is to assume that the random variable X in (4.7) or (4.5) has a stable distribution $S_\alpha(\beta, \delta, \gamma)$.

4.3 A Bayesian Analysis for Linear Regression Models Assuming a Stable Distribution

let us assume that the response x_i in the linear regression model (4.7) for $i = 1, ..., n$, have a stable distribution $X_i \sim S_\alpha(\beta, \delta, \gamma)$, that is,

$$Z_i = \frac{X_i - \delta}{\gamma} \sim S_\alpha(\beta, 0, 1)$$

and where the location parameter δ of the stable distribution is related to the explanatory variables by a linear relation given by

$$\delta = \beta_0 + \beta_1 v_{1i} + \beta_2 v_{2i} + ... + \beta_k v_{ki} \tag{4.8}$$

Assuming a joint prior distribution for α, β, d and γ, where $d = (d_0, d_1, d_2, ..., d_k)$ given by $\pi_0(\alpha, \beta, d, \gamma)$, [6] shows that the joint posterior distribution for parameters α, β, d and γ, is given by,

$$\pi_0(\alpha, \beta, d, \gamma \mid x) \propto \int \left(\frac{\alpha}{\mid \alpha - 1 \mid \gamma} \right)^n \exp\left\{ -\sum_{i=1}^{n} \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right|^{\theta} \right\} \prod_{i=1}^{n} \left| \frac{z_i}{t_{\alpha,\beta}(y_i)} \right|^{\theta} \frac{1}{\mid z_i \mid} \times \pi_0(\alpha, \beta, \delta, \gamma) dy, \tag{4.9}$$

where $\theta = \frac{\alpha}{\alpha-1}$, $z_i = \frac{x_i - \delta}{\gamma}$, for $i = 1, ..., n$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\delta \in (-\infty, \infty)$, $\gamma \in (0, \infty)$, $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ are respectively, the observed and non-observed data vectors. Observe that the bivariate distribution in expression (4.9) is given in terms of x_i and the latent variables y_i, and not in terms of z_i and y_i (there is the Jacobian σ^{-1} multiplied by the right-hand- side of expression (4.2)).

Observe that when $\alpha = 2$ we have $\theta = 2$ and $b_{\alpha,\beta} = 0$. In this case we have a Gaussian distribution with mean equals to γ and variance equals to $2\sigma^2$. For a Bayesian analysis of the proposed model, we assume uniform $U(a, b)$ priors for $,\alpha, \beta$ and γ where the hyperparameters a and b are assumed to be known in each application following the restrictions $\alpha \in (0, 2]$, $\beta \in [-1, 1]$ and $\gamma \in (0, \infty)$. We also assume Normal $N(a, b^2)$ prior distributions for the regression parameters $d_0, d_1, ..., d_k$ considering known hyperparameter values a and b^2. We further assume independence among all parameters.

In the simulation algorithm to obtain a Gibbs sample for the random quantities α, β, d and γ, having the joint posterior distribution (4.9), we assume a uniform $U(-0.5, 0.5)$ prior distribution for the latent random quantities Y_i for $i = 1, ..., n$. Observe that, in this case, we are assuming $a_{\alpha,\beta} = 0(b_{\alpha,\beta})$ With this choice of priors, we use standard available software packages like OpenBugs.

From expression (4.9), the joint posterior probability distribution for α, β, d, γ and $y = (y_1, ..., y_n)$ is given by

$$\pi(\alpha, \beta, d, \gamma, y \mid x) \propto \int \left(\frac{\alpha}{\mid \alpha - 1 \mid \gamma}\right)^n \exp\left\{\sum_{i=1}^{n} \left|\frac{Z_i}{t_{\alpha,\beta}(y_i)}\right|\right\} \prod_{i=1}^{n} \left|\frac{Z_i}{t_{\alpha,\beta}(y_i)}\right|^{\theta} \frac{1}{\mid Z_i \mid} \prod_{i=1}^{n} h(y_i)\pi_0(\alpha, \beta, d, \tag{4.10}$$

where θ and $t_{\alpha,\beta}(.)$ are respectively defined in (4.2) and (4.3) and $h(y_i)$ is a $U(-0.5, 0.5)$ density function, for $i = 1, ..., n$.

4.3.1 The Gibbs Sampler

The Gibbs sampler is a Markovian updating scheme orig- inally developed by Geman and Geman (1984) for use in image processing and publicized as a powerful tool in general Bayesian statistics by Gelfand and Smith (1990) . The algorithm is as follows. Suppose that we have a model driven by the parameter vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$, and that we have observed $x = (x_l, x_2, ..., x_n)$. By taking a set of starting values $\theta^{(0)}$ we can generate $\theta_1^{(1)}$ from $\pi(\theta_1 \mid \theta_2^{(0)}, ..., \theta_k^{(0)}, x), \theta_2^{(1)}$ from $\pi(\theta_2 \mid \theta_1^{(1)}\theta_3^{(0)}, ..., \theta_k^{(0)}, x)$, and so on up to $\theta_k^{(1)}$ from $\pi(\theta_k \mid \theta_1^{(1)}, ..., \theta_{k-1}^{(1)}, x)$ there by performing one iteration producing the sample $\theta^{(1)} = \theta_1^{(1)}, \theta_2^{(1)}, ..., \theta_k^{(1)}$. Iterations of this scheme produces a sequence $\theta^{(0)}, \theta^{(1)}, ..., \theta^{(t)}, ...$ Under mild regularity conditions, it can be shown that the sample $\theta^{(t)}$ produced after t such iterations tends, in distribution, to a sample from the joint posterior $\pi(\theta \mid x)$ as t tends to infinity, and that ergodic averages of suitable functions of the Markov chain realizations provide consistent estimates of features of $\pi(\theta \mid x)$ of interest. But there is no reason why θ need be restricted just to parameters. It is often necessary (or a conceptual simplification) to include auxiliary variables, and such unobservables can simply be added to the parameter vector and the MCMC run for the augmented vector.

Once we have generated the entire n vector y , we generate from

$\pi(\alpha \mid \beta, \delta, \gamma, x, y), \quad \pi(\beta \mid \alpha, \delta, \gamma, x, y),$

$\pi(\delta \mid \alpha, \beta, \gamma, x, y), \quad \pi(\gamma \mid \alpha, \beta, \delta, x, y).$

A long run, iterating this procedure, enables us to estimate and summarize features of $\pi(\alpha, \beta, \delta, \gamma \mid x)$ as required. The crucial feature of this method of analysis is that we have circumvented the problem of being unable to specify the stable likelihood in closed form.

4.3.2 Random Variate Generation From $\pi(\alpha \mid \beta, \delta, \gamma, x, y)$

The characteristic exponent, α, is the most difficult parameter to sample from. The way in which it enters into terms of the likelihood function renders a study of the posterior density of a impossible, even with a uniform prior. Assuming that the prior density for α is $\pi(\alpha)$, and recalling that $z_i = \frac{x_i - \delta}{\gamma}$, the posterior for α is

$$\pi(\alpha \mid \beta, \delta, \gamma, x, y) \propto \left(\frac{\alpha}{|\alpha-1|}\right)^n \times \exp\left\{-\sum_{i=1}^n \left|\frac{Z_i}{t_{\alpha,\beta}(y_i)}\right|^\theta\right\} \prod_{i=1}^n \left|\frac{Z_i}{t_{\alpha,\beta}(y)}\right|^\theta \pi(\alpha)$$

Detailed study of plots of this form makes it clear that $\pi(\alpha \mid \beta, \delta, \gamma, x, y)$ can be quite undulating and rather concentrated. We thus decided that a reparameterization would greatly facilitate any sampling strategy.

By transforming from y to $v = t_{\alpha,\beta}(y)$, the new density becomes, on the whole, unimodal, supporting the parameter range much more evenly:

$$\pi(\alpha \mid \beta, \delta, \gamma, x, v) \propto \left(\frac{\alpha}{|\alpha-1|}\right)^n \times \exp\left\{-\sum_{i=1}^n \left|\frac{Z_i}{v_i}\right|^\theta\right\} \prod_{i=1}^n \left|\frac{Z_i}{v_i}\right|^\theta \left|\frac{dt_{\alpha,\beta}}{dy}\right|^{-1}_{t_{\alpha,\beta}(y)=v_i} \pi(\alpha)$$

Unfortunately, we must now solve $t_{\alpha,\beta}(y_i) = v_i$ for $i = 1, ..., n$ to compute the likelihood; but this can be done quickly using the safeguarded Newton method and can be accelerated by ordering the vi's prior to computation. The fact that we have no information about the shape of the likelihood suggests using the Hastings (1970) generalization of the Metropolis sampling algorithm, which runs as follows. Assume that we are currently performing the ith iteration of the sampler; then:

1. Generate α_* from a distribution with density g.

2. Generate u From a Uniform $(0, 1)$.

3. If $u < \pi(\alpha_* \mid \beta, \delta, \gamma, x, v) g(\alpha_i \mid \alpha_*) / \pi(\alpha_i \mid \beta, \delta, \gamma, x, v) g(\alpha_* \mid \alpha_i)$, then $\alpha_{i+1} = \alpha_*$; otherwise $\alpha_{i+1} = \alpha_i$.

4.3.3 Random Variate Generation From $\pi(\beta \mid \alpha, \delta, \gamma, x, y)$

The conditional posterior density for β, with prior density $\pi(\beta)$, is given as

$$\pi(\beta \mid \alpha, \delta, \gamma, x, y) \propto \exp\left\{-\sum_{i=1}^{n} \left|\frac{Z_i}{t_{\alpha,\beta}(y_i)}\right|^{\theta}\right\} \prod_{i=1}^{n} \left|\frac{1}{t_{\alpha,\beta}(y_i)}\right|^{\theta} \pi(\beta)$$

where $\theta = \frac{\alpha}{\alpha-1}$, $z_i = \frac{x_i-\delta}{\gamma}$.

We know that $t_{\alpha,\beta}(y) = 0$ if and only if $\pi\alpha y = -\beta\min(\alpha, 2-\alpha)\pi/2$; thus the posterior density $\pi(\beta \mid \alpha, y, z)$ has zeros at $\beta = -2\alpha y_i/\min(\alpha, 2-\alpha)$ for $i = 1, ..., n$. Depending on the size of a, a proportion of these zeros will lie outside $(-1, 1)$, but the implication is that posterior density of β is highly multimodal, becoming ever more so as the number of observations increases. This presents sampling difficulties that are confounded with the fact that there can be a strong correlation between β and y. Fortunately, these problems can be solved by transforming y to $v = t_{\alpha,\beta}(y)$, so that the new density becomes

$$\pi(\beta \mid \alpha, \delta, \gamma, x, v) = \prod_{i=1}^{n} \left|\frac{dt_{\alpha,\beta}}{dy}\right|_{t_{\alpha,\beta}(y)=v_i}^{-1} \pi(\beta)$$

We have little knowledge of the shape of $\pi(\beta \mid \alpha, \delta, \gamma, x, v)$, so once again, we use Hasting's version of the Metropolis algorithm. Assume that we are currently performing the ith iteration of the sampler; then:

1. Generate β_* from a distribution with density h.

2. Generate u from a Uniform $(0, 1)$.

3. If $u < \pi(\beta_* \mid \alpha, \delta, \gamma, x, v)h(\beta_i \mid \beta_*)/\pi(\beta_i \mid \alpha, \delta, \gamma, x, v)h(\beta_* \mid \beta_i)$, then $\beta_{i+1} = \beta_*$ otherwise, $\beta_{i+1} = \beta_i$.

4.3.4 Random Variate Generation From $\pi(\delta \mid \alpha, \beta, \gamma, x, y)$

We have that

$$\pi(\delta \mid \alpha, \beta, \gamma, x, y) \propto \exp\left\{-\sum_{i=1}^{n} \left|\frac{Z_i}{t_{\alpha,\beta}(y_i)}\right|^{\theta}\right\} \times \prod_{i=1}^{n} \left|\frac{Z_i}{t_{\alpha,\beta}(y)}\right|^{\theta} \frac{1}{|x_i-\delta|}\pi(\delta)$$

where $\theta = \frac{\alpha}{\alpha-1}$, $z_i = \frac{x_i-\delta}{\gamma}$ and $\pi(\delta)$ is the prior density for δ. Obvious problems in sampling from this density $\pi(\delta \mid \alpha, \beta, \gamma, x, y)$ begin with the fact that it equals zero whenever $\delta = x_i$ for any $i = 1, ..., n$. Additionally, this multimodality is coupled with a very high dependence between δ and the y_i's, thereby forcing the density to be very sharply spiked. But under the transformation

$$\phi_i = \frac{t_{\alpha,\beta}(y_i)}{x_i - \delta}$$

the conditional density for δ becomes

$\pi(\delta \mid \alpha, \beta, \gamma, x, \phi) \propto \prod_{i=1}^{n} \left| \frac{dt_{\alpha,\beta}}{dy} \right|^{-1}_{t_{\alpha,\beta}(y)=\phi_i(x_i-\delta)} \pi(\delta)$

and the new density is more spread and appears to be uni- modal. Once again, no real information on the shape of the $\pi(\delta \mid \alpha, \beta, \gamma, x, \phi)$ is available, and we must resort to the Metropolis algorithm.

Assume that we are currently performing the *ith* iteration of the sampler; then:

1. Generate δ_* from a distribution with density f.

2. Generate u from a Uniform $(0, 1)$.

3. If $u < \pi(\delta_* \mid \alpha, \beta, \gamma, x, \phi) f(\delta_i \mid \delta_*)/\pi(\delta_i \mid \alpha, \beta, \gamma, x, \phi) f(\delta_* \mid \delta_i)$ then $\delta_{i+1} = \delta_*$; otherwise, $\delta_{i+1} = \delta_i$.

4.3.5 Random Variate Generation From $\pi(\gamma \mid \alpha, \beta, \delta, x, y)$

Noting that the density of γ given a α, β, δ, x, and $v = t_{\alpha,\beta}(y)$ is

$\pi(\gamma \mid \alpha, \beta, \delta, x, v) \propto \left(\frac{1}{\gamma^\theta}\right)^n \exp\left\{-\frac{1}{\gamma^\theta}\sum_{i=1}^{n}\left|\frac{x_i-\delta}{v_i}\right|^\theta\right\}\pi(\gamma),$

we have that if $\pi(\gamma) \propto \Theta^{-(a+1)} \exp -b/\Theta$ where $\Theta = \gamma^\theta$, then

$\pi(\gamma \mid \alpha, \beta, \delta, x, v) \propto \left(\frac{1}{\Theta}\right)^{a+n+1} \exp\left\{-\frac{1}{\Theta}\left(b + \sum_{i=1}^{n}\left|\frac{x_i-\delta}{v_i}\right|^\theta\right)\right\}.$

That is the inverse delta distribution is conjugate for the scale parameter γ^θ and a simple transformation gives us γ.

A sample is easily extracted from the inverse delta distribution, because if $\Theta \sim$

$ID(a, b)$, then $\Theta^{-1} \sim D(a, b)$ and delta generators exist in multitude and have been well documented by both Devroye (1986) and Ripley (1987). If an inverse delta prior is not assumed, then we can use the same algorithm as is used to generate from $\pi(\delta \mid \alpha, \beta, \gamma, x, y)$ as follows. Assume that we are currently performing the *ith* iteration of the sampler; then:

1. Generate γ_* from a transformed $ID(n, \sum \mid (x_i - \delta)/v_i \mid^{\theta})$.

2. Generate u from a Uniform $(0, 1)$.

3. If $u < \pi(\gamma_*)/\pi(\gamma_i)$ then $\gamma_{i+1} = \gamma$; otherwise, $\gamma_{i+1} = \gamma_i$.

Chapter 5

Applications

5.1 Data Analysis

To illustrate the methodology developed in the previous chapters for Simple Linear Regression Model with and without assuming stable distribution, we examine data set described in Table 5.1, contains hypothetical sales estimates, list prices, and physical specifications for various makes and models of vehicles. The list prices and physical specifications were obtained alternately from *edmunds.com* and manufacturer sites. The complete data set were presented in Appendix A.

Table 5.1: A sample data set of car Prices and Sales in thousands.

n	1	10	25	59	70	90	100	120	135	150
Price (V)	16.919	91.561	32.299	12.855	12.698	20.38	42.643	92.364	142.535	16.957
Sales (X)	21.5	21.975	13.96	26.6	37.805	22.51	13.499	21.665	13.108	23.4

From a preliminary data analysis, we see that a linear regression model (2.9) is suitable for this data set, see Table 5.2.

Table 5.2: ANOVA table of linear regression model

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	66902.899	1	66902.899	15.663	.000^b
	Residual	653543.735	153	4271.528		
	Total	720446.634	154			

The estimated regression straight line obtained by least squares estimates and by using the software SPSS (version 23) is given by

$$\hat{x}_i = 93.026 - 1.452 v_i, \tag{5.1}$$

where the regression parameter d_1 is statistically different from zero (p-value < 0.05) see results in Table 5.3. From standard residuals graphs (Figures 5.2 and 5.3) we can verify that the required assumptions (residuals normality and homoscedastic variance are not satisfied), also outliers were presented, even though, the model and the coefficients were statistically significant, this leads for more investigations and other approaches for such case.

Table 5.3: Coefficients of linear regression model

Model Par.	B	Std. Error	t	Sig.
(Constant)	93.026	11.340	8.203	.000
Price in thousands	-1.452	.367	-3.958	.000

Dependent Variable: Sales in thousands

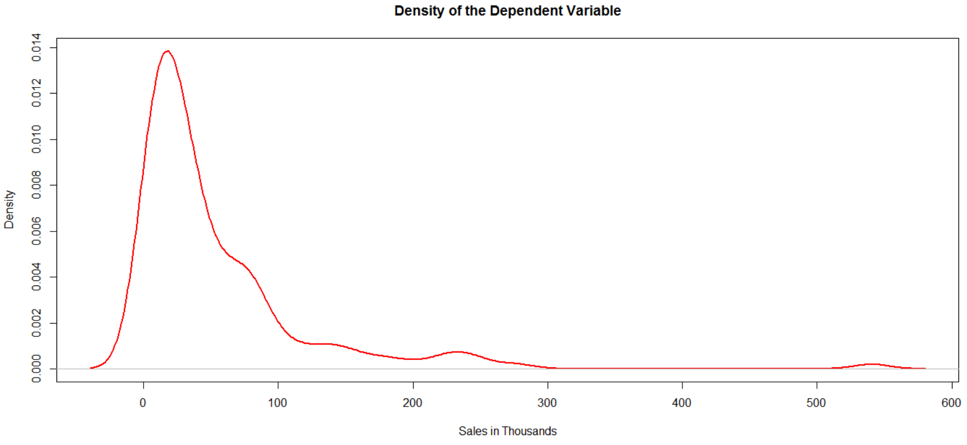

Figure 5.1: Density Plot of the Dependent Variable (Sales).

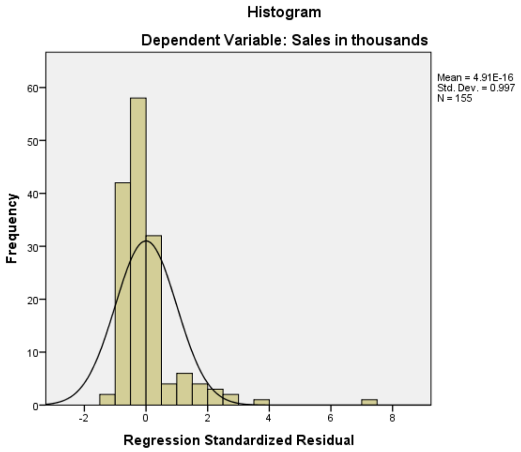

Figure 5.2: Standardized residuals plot

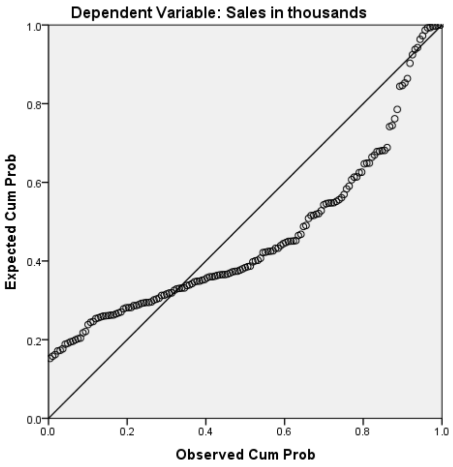

Figure 5.3: Normal P-P Plot of Regression Standardized Residual

Moreover, summary statistics in table 5.4 indicates that the dependent variable (Sales) is not normally distributed it is about right skewed and positively Kurtosis, also a lot of outlier values were presented see Figure 5.1 and 5.4.

Table 5.4: Sales variable descriptive statistics

			Statistic	Std. Error
Sales in thousands	Mean		53.24583	5.493823
	95% Confidence Interval for Mean	Lower Bound	42.39285	
		Upper Bound	64.09881	
	5% Trimmed Mean		43.76296	
	Median		29.45000	
	Variance		4678.225	
	Std. Deviation		68.397550	
	Minimum		.110	
	Maximum		540.561	
	Range		540.451	
	Interquartile Range		54.613	
	Skewness		3.388	.195
	Kurtosis		17.337	.387

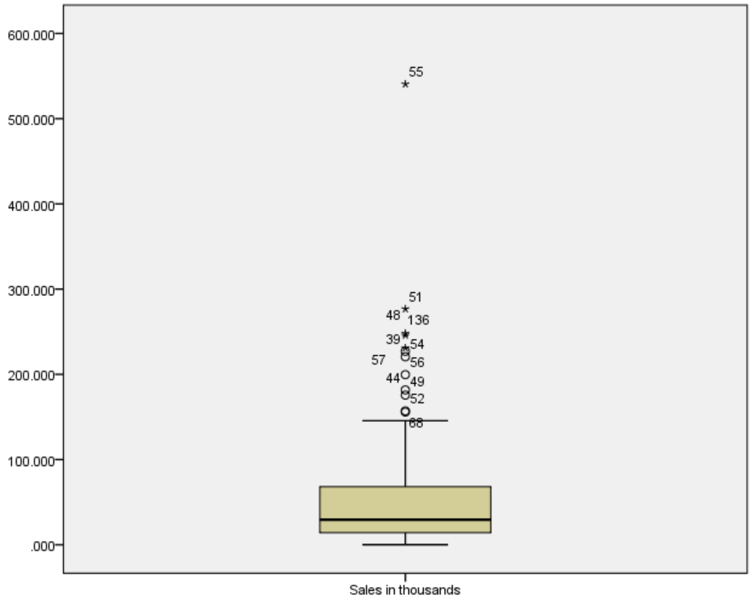

Figure 5.4: Box-and-Whiskers Plot of dependent variable.

5.2 Bayesian Approach

For a comparison purpose, we first fit the linear model (5.2) assuming no stable distribution.

$$X_i = d_0 + d_1 V_i + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{5.2}$$

Let us now turn to the Bayesian version and show that under the reference prior, we will obtain the posterior distributions of d_0, d_1 and σ^2 analogous with the frequentist OLS results. The Bayesian model starts with the same model as the classical frequentist approach:

With the assumption that the errors, ϵ_i, are independent and identically distributed as normal random variables with mean zero and variance σ^2. This assumption is exactly the same as in the classical inference case for testing and constructing confidence intervals for d_0 and d_1. Our goal is to update the distributions of the unknown parameters d_0, d_1, and σ^2 , based on the data $v_1, v_2, \ldots, v_n$, and $x_1, x_2, \ldots, x_n$ where *n* is the number of observations.

Under the assumption that the errors ϵ_i are normally distributed with constant variance σ^2, we have for the random variable of each response X_i, conditioning on the observed data v_i and the parameters $d_0,\ d_1,\ \sigma^2$, is normally distributed:

$$X_i|v_i, d_0,\ d_1, \sigma^2 \sim N(d_0 + d_1 v_i, \epsilon_i), \quad i = 1, 2, \dots, n \tag{5.3}$$

That is, the likelihood of each X_i given $v_i, d_0,\ d_1,$ and σ^2 is given by

$$P(x_i|v_i, d_0,\ d_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{\big(x_i - (d_0 + d_1 v_i)\big)^2}{2\sigma^2}\right) \tag{5.4}$$

The likelihood of $X_1, \dots, X_n$ is the product of each likelihood $P(y_i|v_i, d_0,\ d_1, \sigma^2)$, since we assume each response X_i is independent from each other. Since this likelihood depends on the values of $d_0,\ d_1,$ and σ^2, it is sometimes denoted as a function of $d_0,\ d_1,$ and σ^2: $\mathcal{L}(d_0,\ d_1, \sigma^2)$.

Using the reference prior, we will obtain familiar distributions as the posterior distributions of $d_0,\ d_1,$ and σ^2, which gives the analogue to the frequentist results as in chapter 4. Here we assume the joint prior distribution of $d_0,\ d_1,$ and σ^2 to be proportional to the inverse of σ^2

$$P(d_0,\ d_1, \sigma^2) \propto \frac{1}{\sigma^2} \tag{5.5}$$

The full conditional density of the parameters $\{d_0, d_1, \sigma^2\}$ is a product of it's prior density and some standard distribution density, then, with the choice of conjugate priors:

$$d_0 \sim N\big(\mu_{d_0}, \sigma_{d_0}\big), \quad d_1 \sim N\big(\mu_{d_1}, \sigma_{d_1}\big), \quad \sigma^2 \sim IG(a, b)$$

where, $\mu_{d_0}, \sigma_{d_0}, \mu_{d_1}, \sigma_{d_1}, a,$ and b are defined in chapter 4. For this application the hyperparameters and the initial values for Gibbs sampler were calculated due to the ordinary least squares approach see the *r* software codes in Appendix A.

Thus, drawing random variates from their full conditional distribution is straight forward, therefore, we will use the full conditional density as a proposal density in Gibbs sampler algorithm, and in sampling process each updating step for these parameters, a new draw from the full conditional density is always accepted. With initial values of all parameters set at their MLEs, we perform this algorithm for each parameter 25,000 Gibbs samples after 10000 burn-in. The time series plots of one sequence of Gibbs samples for different number of iterations, posterior density, and the average number of these iterations for these parameters are presented in Figure 5.4. It is clear that these sequences are mix well and converge within 15,000 iterations.

With the initial values of the parameters for which the data are generated considered as the truth values of the parameters, estimate Monte Carlo Summary statistics, Monte Carlo Standard Deviation (MCSD), Mean Squared Error (MSE), 95% Confidence Converge Rate (CCR), and Bias in Percentage Terms (BPT) are presented in Table 5.5 Where, MCE stand for Monte Carlo Error and it can be evaluated as follows : In our simulation study we used 155 data replications, thus the resulting estimates are subject to sampling variation (Monte Carlo Error), this variation for the point estimate can be calculated as $\hat{p} = MCSD/\sqrt{155}$, the MCE then can be found by $MCSD = \sqrt{\frac{\hat{p}(1-\hat{p})}{155}}$.

Results in Table 5.5 assert the convergence of the Markov Chain and the samplers reached the convergence after 15,000 iterations after 10,000 iterations are burn-in. Also, all results are closed to that in ordinary least squares approach in section 5.1.

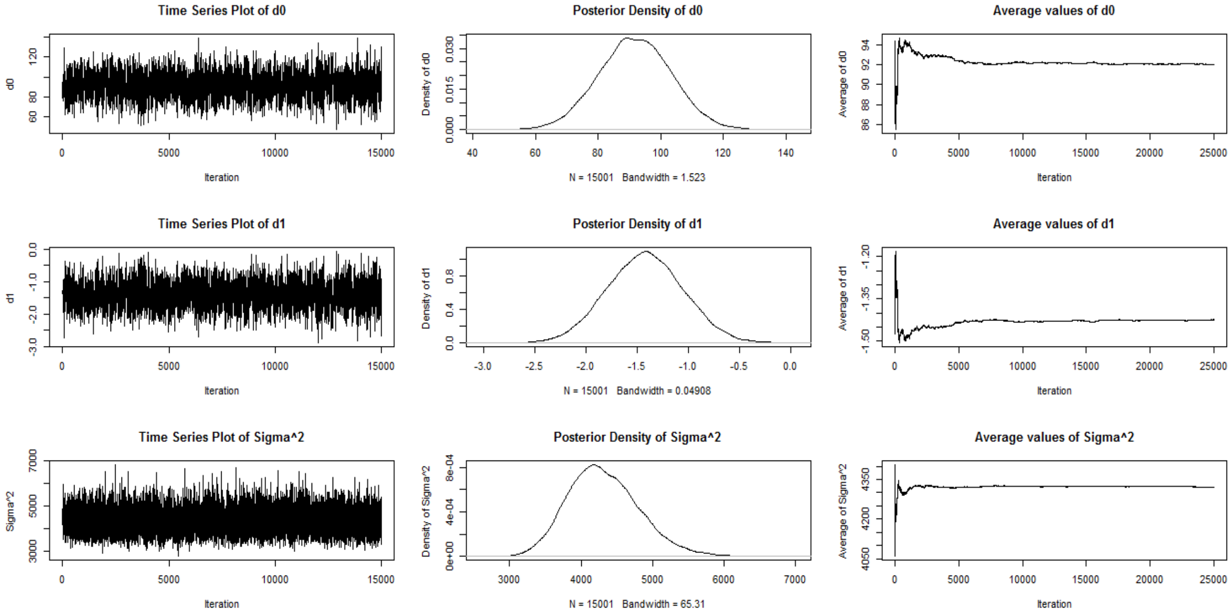

Figure 5.4: Posterior Time Series, Density, Average Values Plot of All Parameters

Table 5.5: Monte Carlo Summary statistics of the parameters estimate

Parameter	Initial Value	Estimated Value	MCSD	MSE	95% CCR	BP	MCE
d_0	98.4285	91.37549	0.016	0.0006	98%	-0.021%	0.005
d_1	-1.6662	-1.402252	0.021	0.0007	98%	0.023%	0.005
σ^2	4184.63	4325.206	0.083	0.0015	96%	0.044%	0.009

5.3 Bayesian Approach Assuming Stable Distribution.

As we see in the above application we need to check if all model assumptions are verified. In this way, we consider graphical approaches to verify if the residuals of the model satisfy the above assumptions.

In presence of outliers or discordant observations we could have great effects on the obtained estimators for the regression model given by (5.2) which could invalidate the obtained inferences. In this way, we could use non-parametric regression models or to assume more robust probability distributions for the data. One possibility is to assume that the random variable X in (5.2) has a *stable distribution* $S_\alpha(\beta, \gamma, \delta)$.

So that, assume model (5.2) have a stable distribution, that is

$$X_i \sim S_\alpha(\beta, \delta_i, \gamma), \qquad \text{with} \qquad Z_i = \frac{X - \delta_i}{\gamma} \sim S_\alpha(\beta, 0, 1), \quad i = 1, 2, \dots, n \tag{5.6}$$

where the location parameter δ_i of the stable distribution is related to the *explanatory* variables by a linear relation given by,

$$\delta_i = \beta_0 + \beta_1 v_{1i} + \beta_2 v_{2i} + \cdots + +\beta_k v_{ki} \tag{5.7}$$

Hence, our unknown parameters in model (5.6) are $\Omega = \{\alpha, \beta, \gamma, d_0, d_1\}$. From the analysis of chapter two, the joint posterior distribution $\pi(\alpha, \beta, \gamma, d_0, d_1 |\, x)$ for these parameters, is given by,

$$\begin{aligned}\pi(\alpha, \beta, \gamma, d_0, d_1 |\, x) \quad &\propto \int \left(\frac{\alpha}{|\alpha - 1|\gamma}\right)^n \exp\left[-\sum\nolimits_{i=1}^{n} \left|\frac{z_i}{t_{\alpha,\beta}(Y_i)}\right|^\theta\right] \\ &\times \prod\nolimits_{i=1}^{n} \left|\frac{z_i}{t_{\alpha,\beta}(y_i)}\right|^\theta . \frac{1}{|z_i|} \pi(\alpha, \beta, \gamma, d_0, d_1)\ dy\end{aligned} \tag{5.8}$$

where,

$$\theta = \frac{\alpha}{\alpha - 1}, \qquad z_i = \frac{x_i - \delta_i}{\gamma}, \qquad i = 1,2, \dots, n, \ \ \alpha \in (0, 2], \qquad \beta \in [-1, 1], \text{and } \ \gamma \in (0, \infty);$$

$x = (x_1, x_1, \dots, x_n)$, and $y = (y_1, y_1, \dots, y_n)$ are the observed and non-observed data vectors respectively.

Note that, the bivariate distribution in expression (5.8) is given in terms of x_i and y_i and not in terms of z_i and y_i. simplification of expression (5.8) gives:

$$\begin{aligned} \pi(\alpha, \beta, \gamma, d_0, d_1 | \, x) \quad & \propto \left(\frac{\alpha}{|\alpha - 1|\gamma}\right)^n \exp\left[\sum\nolimits_{i=1}^{n} \left|\frac{z_i}{t_{\alpha,\beta}(y_i)}\right|\right] \\ & \times \prod\nolimits_{i=1}^{n} \left|\frac{z_i}{t_{\alpha,\beta}(y_i)}\right|^{\theta} \cdot \frac{1}{|z_i|} \prod\nolimits_{i=1}^{n} h(y_i)\pi(\alpha, \beta, \gamma, d_0, d_1) \end{aligned} \tag{5.9}$$

where

$$t_{\alpha,\beta}(y_i) = \left(\frac{\sin\left(\pi\alpha y_i + b_{\alpha,\beta}\right)}{\cos(\pi y_i)}\right)\left(\frac{\cos(\pi y_i)}{\cos\left(\pi(\alpha - 1)y_i + b_{\alpha,\beta}\right)}\right), \quad b_{\alpha,\beta} = \beta \min\{\alpha, 2 - \alpha\}\frac{\pi}{2}.$$

and $h(y_i)$ is a $\mathrm{U}(-0.5, 0.5)$ for $i = 1, \dots, n$.

With the choice of conjugate priors, the full conditional density of the parameters $\{d_0, d_1\}$ are same as in section 5.2. Whereas, the full conditional density of the parameters $\{\alpha, \beta, \gamma\}$ is then $\mathrm{U}(a, b)$, where the hyperparameters a and b are assumed to be known in each application following the restrictions $0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1$ and $0 \leq \gamma \leq \infty$.

In this application using Bayesian approach via (Stable Regression Program stabelreg.exe), with the choice of the following prior distributions:

$$\alpha \sim \mathrm{U}(0, 2)\,, \qquad \beta \sim U(0, \pi/2)\,, \qquad \gamma \sim U(0, 10)$$

All response distributions are significantly different from Gaussian $(\alpha = 2, \ \beta = 0)$.

With the initial values of the parameters for which the data are generated considered as the truth values of the parameters, Posterior Estimates, Standard Deviation (SD), and 95% Credible Interval, are presented in Table 5.6 assuming stable distribution for the response variable.

Table 5.6: Stable Distribution Summary Statistics of the Parameters Estimate

Parameter	Initial Value	Estimated Value	SD	95% Credible Interval
d_0	98.4285	92.3362	8.21	(91.043, 93.629)
d_1	-1.6662	-1.3250	1.38	(-1.617, -1.183)
α	1	1.63	0.18	(1.602, 1.658)
β	0.5	0.34	0.24	(0.302, 0.378)
γ	5	8.71	3.13	(8.217, 9.203)

Results in the above table indicate that all parameter estimates are statistically significant.

In order to make a comparisons between the three approaches above, especially when Normal assumption is not satisfied and in the presence of outliers, we present plots of observed, fitted mean considering three approaches versus samples. Results are presented in Figures 5.6.a – 5.6.c

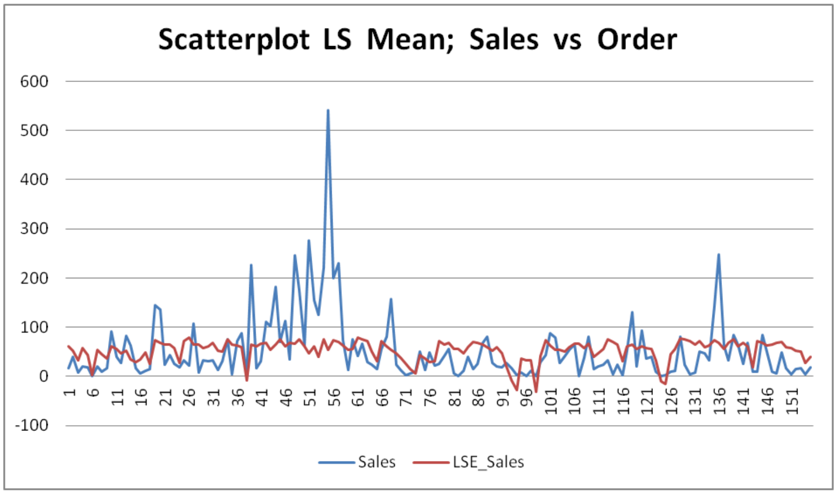

Figure 5.6.a: Plots of Observed, Fitted Means Considering Least Squares Model.

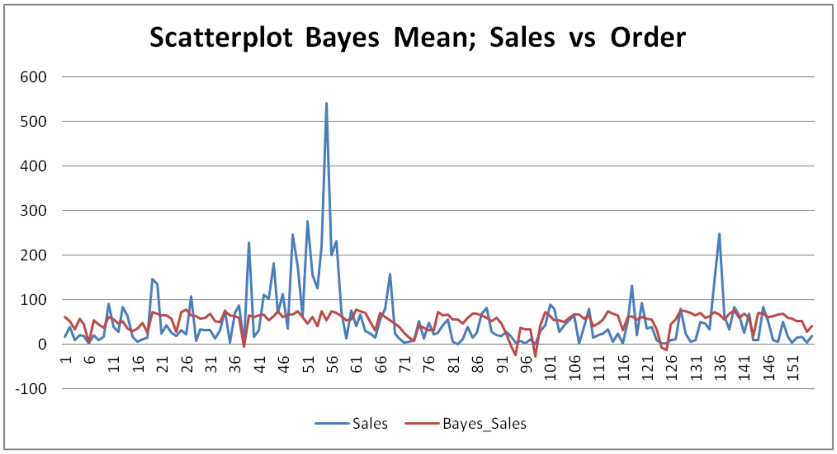

Figure 5.6.b: Plots of Observed, Fitted Means Considering Bayesian Model.

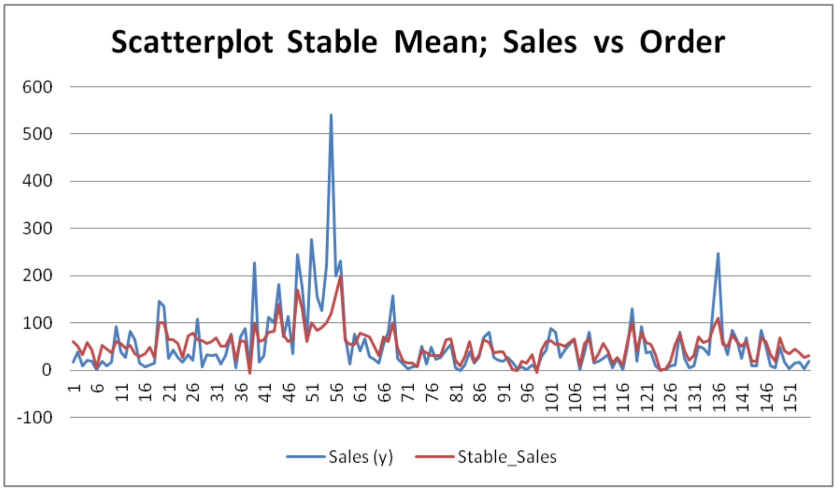

Figure 5.6.c: Plots of Observed, Fitted Means Considering Stable Model.

In Figures 5.6.a, and 5.6.b we have the plots of observed, fitted means considering models with (Least squares and Bayesian approaches) versus samples. We observe similar fit of both models (linear regression model assuming normality). In Figures 5.6.c, we observe that model with a stable distribution is very robust to the presence of the outlier given similar inference results as obtained without the presence of the outlier.

5.4 Maximum Likelihood Approach

Again for comparison purposes, and assuming different heavy-tailed distribution namely (Cauchy distributions) is compared with the normal distribution using maximum likelihood (ML) approach, to see the performance of the above model. In the previous section, the OLS method was implemented, which minimizes the sum of absolute errors (MSAE) by fitting regression through the Bayesian method. The maximum likelihood method is superior to the MSAE method. However, MSAE procedure that do not depend on the distribution of error. Back to likelihood function in (5.4) for the normal distribution, the log likelihood is given by

$$l_N(d_0,\ d_1, \sigma^2) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum\nolimits_{i=1}^{n}\big(x_i - (d_0 + d_1 v_i)\big)^2 \qquad (5.10)$$

In the method of maximum likelihood, we pick the parameter values which maximize the likelihood, or, equivalently, maximize the log-likelihood. Using the advantages of the software r packages, we can use the function *optimize*( ) to find the values of $d_0,\ d_1,$ and σ^2 that maximize (5.10), which are the estimated value of the required parameters, see r codes in Appendix C.

In the case of Cauchy distribution, the probability density function (pdf) is given by,

$$f(x, \delta, \gamma) = \frac{1}{\pi\gamma}\left(\frac{\gamma^2}{(x-\delta)^2 + \gamma^2}\right) \qquad (5.11)$$

where, δ is the location parameter, specifying the location of the peak of the distribution, and γ is the scale parameter which specifies the half-width at half-maximum. The likelihood of each x_i given $v_i, d_0,\ d_1, \delta$ and γ is then given by

$$L_C(d_0,\ d_1, \delta, \gamma) = \prod_{i=1}^{n} \frac{1}{\pi\gamma}\left(\frac{\gamma^2}{\left((d_0 + d_1 v_i) - \delta\right)^2 + \gamma^2}\right) \tag{5.12}$$

Hence, with some simplifications, the log likelihood is given by

$$l_C(d_0,\ d_1, \delta, \gamma) = -n\ln(\pi\gamma) - \sum_{i=1}^{n} \ln\left(1 + \left(\frac{(d_0 + d_1 v_i) - \delta}{\gamma}\right)^2\right) \tag{5.13}$$

Again the function *optimize*( ) was used to estimate all unknown parameters in (5.13).

Tables 5.7 shows the maximum likelihood estimates of model parameters (Est.), standard errors (se), and 95% confidence intervals (95% CI) of the proposed distribution with regression model for the dataset in Table (5.1). For comparison purpose, also Table 5.7 shows the Log-likelihood and Akaike Information Criteria (AIC) associated to the above distributions.

Unsurprisingly, regardless of the parameter estimates, the model fit under the Cauchy distribution is better than the normal distribution. The smaller the AIC, the better the model fit. This is due to the heavy tail distribution (Cauchy), which some authors called it "Supper heavy tailed".

Table 5.7: Model Parameters' Estimates for Different Distributions

Distribution	Parameter	Estimated Value	se.	95% Credible Interval	Log-likelihood	AIC
Normal	d_0	93.0868	7.9668	(77.4717, 108.702)	-1176.129	2358.257
	d_1	-1.4543	0.2578	(-1.9597, -0.9491)		
	σ	64.9364	2.6080	(59.8246, 70.0482)		
Cauchy	d_0	51.7263	296.582	(-5761.284, 5864.736)	-879.979	1767.958
	d_1	-0.6155	0.0732	(-0.7573, -0.4736)		
	δ	8.4261	296.582	(-5804.582, 5821.435)		
	γ	14.9295	1.2860	(12.409, 17.450)		

5.5 Discussion

With the assumption of normality and no outliers, we conclude that we do not need to assume a stable distribution of the data, because the results are very similar to the results obtained from the erroneous normality assumption. In addition, the computational cost of using a stable distribution is very high (see Achcar, Achcar and Martinez 2013).

However, when there are abnormal (outlier) values or inconsistent observations, it is due to multiple measurement errors or violation of the normality assumption. This situation is very common in the application of regression analysis. In the presence of these inconsistent observations, it is usually based on the assumption of normality of error and constant variance, which has a significant impact on the inferences usually obtained in regression parameters or predictions based on least squares or maximum likelihood methods, which may means wrong results. The use of stable distributions could be a good alternative for many applications in the data analysis to have robust inference results, since this distribution has a great flexibility to fit for the data. With the use of Bayesian methods and MCMC simulation algorithms, it is possible to get inferences for the model despite of the nonexistence of an analytical form for the density function as it was showed in this chapter. It is important to point out that the computational work in the sample simulations for the joint posterior distribution of interest can be greatly simplified using standard free software like the r software and Stable regression program (stablereg.exe).

Generally, the appearance of outliers will absolutely affect the regression model under standard normality assumptions. The ideal results not affected by outliers could be obtained using the stable distribution methodology as observed in our application. These results could be of great interest in applications.

It has been noticed that the lack of closed formulas for densities and distribution functions for all but a few stable distributions (Cauchy and Levy) has been a major drawback to the use of stable distributions by practitioners. Also there are multiple parameterization for stable laws and much confused has been caused by these parameterizations. However, there are now reliable computer programs to compute stable densities, distribution functions and quantiles. With these programs, it is possible to use stable models in a variety of practical problems.

References

[1] Jorge A Achcar and Sılvia RC Lopes. “Linear and Non-Linear Regression Models Assuminga Stable Distribution”. In: *Revista Colombiana de Estadıstica* 39.1 (2016), pp. 109–128.

[2] Jorge A Achcar et al. “A bayesian approach for stable distributions: some computational aspects”. In: (2013).

[3] Carl A Bache et al. “Polychlorinated biphenyl residues: Accumulation in Cayuga Lake trout with age”. In: *Science* 177.4055 (1972), pp. 1191–1192.

[4] Szymon Borak, Wolfgang Härdle, and Rafał Weron. “Stable distributions”. In: *Statistical tools for finance and insurance.* Springer, 2005, pp. 21–44.

[5] TM Brown. “Social Choice and Individual Values. By Kenneth J. Arrow. New York: John Wiley & Sons, Inc.; London: Chapman and Hall, Limited [Toronto: University of Toronto Press]. 1951. Pp. xi, 99. 3.00.”. In: *Canadian Journal of Economics and Political Science/Revue canadienne de economiques et science politique* 18.3 (1952), pp. 400–403.

[6] DJ Buckle. “Bayesian inference for stable distributions”. In: *Journal of the American Statistical Association* 90.430 (1995), pp. 605–613.

[7] Paul Damlen, John Wakefield, and Stephen Walker. “Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.2 (1999), pp. 331–344.

[8] Norman R Draper and Harry Smith. “Applied Regression Analysis, John Wiley and Sons”. In: *New York* 407 (1981).

[9] Il'dar Abdullovich Ibragimov and KE Chernin. “On the unimodality of geometric stable laws”. In: *Theory of Probability & Its Applications* 4.4 (1959), pp. 417–419.

[10] Michael Kateregga, Sure Mataramvura, and David Taylor. “Parameter estimation for stable distributions with application to commodity futures log-returns”. In: *Cogent Economics & Finance* 5.1 (2017), p. 1318813.

[11] Ioannis A Koutrouvelis. “Regression-type estimation of the parameters of stable laws”. In: *Journal of the American statistical association* 75.372 (1980), pp. 918–928.

[12] Jing Lei et al. “Distribution-free predictive inference for regression”. In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1094–1111.

[13] Eugene Lukacs. “Characteristic functions. revised and enlarged”. In: *Hafner Publishing Co., New York. Math. Review* 49 (1970), p. 11595.

[14] J Huston McCulloch. “Simple consistent estimators of stable distribution parameters”. In: *Communications in Statistics-Simulation and Computation* 15.4 (1986), pp. 1109–1136.

[15] Stefan Mittnik, Toker Doganoglu, and David Chenyao. “Computing the probability density function of the stable Paretian distribution”. In: *Mathematical and Computer Modelling* 29.10-12 (1999), pp. 235–240.

[16] John Nolan. *Stable distributions: models for heavy-tailed data*. Birkhauser New York, 2003.

[17] Gennady Samoradnitsky. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Routledge, 2017.

[18] GAF Seber and AJ Lee. “Linear Regression Analysis, 549 pp”. In: *Wiley, New York, doi* 10 (2003), p. 9780471722199.

[19] David J Spiegelhalter et al. “Bayesian measures of model complexity and fit”. In: *Journal of the royal statistical society: Series b (statistical methodology)* 64.4 (2002), pp. 583–639.

[20] Martin A Tanner and Wing Hung Wong. “The calculation of posterior distributions by data augmentation”. In: *Journal of the American statistical Association* 82.398 (1987), pp. 528–540.

[21] Vladimir M Zolotarev. *One-dimensional stable distributions.* Vol. 65. American Mathematical Soc., 1986.

Appendix *A*

Data Set:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	16.919	39.384	8.588	20.397	18.78	1.38	19.747	9.231	17.527	91.561	39.35	27.851	83.257	63.729	15.943
V	21.5	28.4	42	23.99	33.95	62	26.99	33.4	38.9	21.975	25.3	31.965	27.885	39.895	44.475
n	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
X	6.536	11.185	14.785	145.519	135.126	24.629	42.593	26.402	17.947	32.299	21.855	107.995	7.854	32.775	31.148
V	39.665	31.01	46.225	13.26	16.535	18.89	19.39	24.34	45.705	13.96	9.235	18.89	19.84	24.495	22.245
n	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
X	32.306	13.462	30.696	76.034	4.734	71.186	88.028	0.916	227.061	16.767	31.038	111.313	101.323	181.749	70.227
V	16.48	28.34	29.185	12.64	19.045	20.23	22.505	69.725	19.46	21.315	18.575	16.98	26.31	19.565	12.07
n	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
X	113.369	35.068	245.815	175.67	63.403	276.747	155.787	125.338	220.65	540.561	199.685	230.902	73.203	12.855	76.029
V	21.56	17.035	17.885	12.315	22.195	31.93	21.41	36.135	12.05	26.935	12.885	15.35	20.55	26.6	26
n	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
X	41.184	66.692	29.45	23.713	15.467	55.557	80.556	157.04	24.072	12.698	3.334	6.375	9.126	51.238	13.798
V	9.699	11.799	14.999	29.465	42.8	14.46	21.62	26.895	31.505	37.805	46.305	54.005	60.105	34.605	39.08
n	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
X	48.911	22.925	26.232	42.541	55.616	5.711	0.11	11.337	39.348	14.351	26.529	67.956	81.174	27.609	20.38
V	43.33	42.66	13.987	19.047	17.357	24.997	25.45	31.807	22.527	16.24	16.54	19.035	22.605	27.56	22.51
n	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
X	18.392	27.602	16.774	3.311	7.998	1.526	11.592	0.954	28.976	42.643	88.094	79.853	27.308	42.574	54.158
V	31.75	49.9	69.7	82.6	38.9	41	41.6	85.5	35.3	13.499	20.39	26.249	26.399	29.299	22.799
n	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
X	65.005	1.112	38.554	80.255	14.69	20.017	24.361	32.734	5.24	24.155	1.872	51.645	131.097	19.911	92.364
V	17.89	18.145	24.15	18.27	36.229	31.598	25.345	12.64	16.08	18.85	43	21.61	19.72	25.31	21.665
n	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
X	35.945	39.572	8.982	1.28	1.866	9.191	12.115	80.62	24.546	5.223	8.472	49.989	47.107	33.028	142.535
V	23.755	25.635	41.43	71.02	74.97	33.12	26.1	10.685	12.535	14.29	18.835	15.01	22.695	20.095	13.108
n	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
X	247.994	63.849	33.269	84.087	65.119	25.106	68.411	9.835	9.761	83.721	51.102	9.569	5.596	49.463	16.957
V	17.518	25.545	16.875	11.528	22.368	16.888	22.288	51.728	14.9	16.7	21.2	19.99	17.5	15.9	23.4

Appendix *B*

r codes of Gibbs Sampler Process:

```
##########################################################
#function returns samples from the model:
# Y is the vector of outcomes
# X is the vector of predictors
##########################################################
Bayes.slm<-function(y,X,
                    mu=rep(0,2),tau=rep(100,2),
                    a=0.01,b=0.01,
                    n.samples=Niteration){
   n <- length(y)
   #intial values:
   ols      <- lm(y~X)
   sigma2   <- var(ols$residuals)
   beta     <- ols$coef
   beta1_av <- 0; beta2_av <- 0;  sigma2_av <- 0
   s1 <- 0; s2 <- 0; s3 <- 0;
   #Initialize matrix to store the results:
   samples           <- matrix(0,n.samples,6)
   colnames(samples) <-
c("beta1","beta2","sigma^2","beta1_av","beta2_av","sigma^2_av")
   #colnames(av) <- c("beta1_av","beta2_av","sigma^2_av")
   #Start the MCMC sampler:
   for(i in 1:n.samples){

     #update sigma^2:
      SSE    <- sum((y-beta[1]-X*beta[2])^2)
      sigma2 <- 1/rgamma(1,n/2+a,SSE/2+b)
      s1 <- s1+ sigma2
      sigma2_av <- s1/i
     #update beta1:
      VVV     <- n/sigma2 + 1/tau[1]^2
      MMM     <- sum(y-X*beta[2])/sigma2 + mu[1]/tau[1]^2
      beta[1] <- rnorm(1,MMM/VVV,1/sqrt(VVV))
      s2 <- s2+ beta[1]
      beta1_av <- s2/i

     #update beta2:
      VVV     <- sum(X^2)/sigma2 + 1/tau[2]^2
      MMM     <- sum(X*(y-beta[1]))/sigma2 + mu[2]/tau[2]^2
      beta[2] <- rnorm(1,MMM/VVV,1/sqrt(VVV))


      s3 <- s3+ beta[2]
      beta2_av <- s3/i
    #store results:
      samples[i,]  <- c(beta,sigma2,beta1_av,beta2_av,sigma2_av)
    }
#return a list with the posterior samples:
      return(samples)}
#####################################################################
```

```
#Fit the model by Least squares Approach
#####################################################################
library (foreign)
setwd("d:/My courses/Master/Layla Khalid/simulation/")
main_data <- read.spss("Car_Sales_Price.sav",to.data.frame=TRUE)
library(BAS)
data(main_data)
summary(main_data)
sales.lm = lm(Sales ~ Price, data = main_data)
summary(sales.lm)
#####################################################################
#set data values and call the MCMC sampler function
X <- main_data$Sales
V <- main_data$Price
n <- length(X)
sigma <- 1;  beta  <- c(2,1)
Niteration <- 25000;  Nburn <- 10000
post<-Bayes.slm(X,V)
#####################################################################
#Posterior summaries
#####################################################################
summary(post)
#####################################################################
#Posterior Plots
#####################################################################
par(mfrow=c(3,3))
plot(post[Nburn:Niteration ,1],type="l",ylab="d0",xlab="Iteration", main="Time Series
Plot of d0")
d <- density(post[Nburn:Niteration ,1]) # returns the density data
plot(d, main="Posterior Density of d0", ylab="Probability deensity of d0") # plots
the results
plot(post[1:Niteration ,4],type="l",ylab="Average of d0",xlab="Iteration")

plot(post[Nburn:Niteration ,2],type="l",ylab="d1",xlab="Iteration", main="Time Series
Plot of d1")
d <- density(post[Nburn:Niteration ,2]) # returns the density data
plot(d, main="Posterior Density of d1", ylab="Probability deensity of d1") # plots
the results
plot(post[1:Niteration ,5],type="l",ylab="Average of d1",xlab="Iteration")

plot(post[Nburn:Niteration ,3],type="l",ylab="Sigma^2",xlab="Iteration", main="Time
Series Plot of Sigma^2")
d <- density(post[Nburn:Niteration ,3]) # returns the density data
plot(d, main="Posterior Density of Sigma^2", ylab="Probability deensity of Sigma^2")
# plots the results
plot(post[1:Niteration ,6],type="l",ylab="Average of Sigma^2",xlab="Iteration")
```

Appendix *C*

r codes for optimization of LE :

```
library (foreign)
library (rmutil)
setwd("d:/My courses/Master/Layla Khalid/simulation/")
main_data <- read.spss("Car_Sales_Price.sav",to.data.frame=TRUE)
#########Build the Loss Function / Determine log-likelihood, and AIC ################
lm.loss <- function(par) {
  d1.par <- par[1]        # The slope
  d0.par <- par[2]        # The intercept
  scale.par <- par[3]     # The scale parameter
  loc.par <- par[4]       # The location parameter
      if(scale.par < 0) {deviance <- 10000000} # If the scale.par is invalid reject
it
      if(scale.par > 0) {
             likelihoods <- dcauchy(main_data$Sales, location = (main_data$Price *
d1.par + d0.par)- loc.par, scale = scale.par, log = FALSE)
             log.likelihoods <- log(likelihoods)
log_like <<- sum(log.likelihoods)
deviance <- -2 * log_like
AIC_n <<- deviance + 2 * length(par)
}
return(deviance)
}

##################### Call the loss function ##################################
dev.temp <- lm.loss(c(1, 5, 20, 20))
dev.temp # print value
log_like
AIC_n
##############optimization functions to find parameter values########################
parameter.fits <- optim(par = c(1, 5, 20, 20),
      fn = lm.loss, hessian = T
      )
parameter.fits # print value
###################parameter values standard error###################################
hessian <- parameter.fits$hessian
hessian.inv <- solve(hessian)
parameter.se <- sqrt(diag(hessian.inv))
parameter.se # print value
################Getting confidence intervals for parameter values###################
CI.matrix <- as.data.frame(matrix(NA, nrow = 3, ncol = 4))
CI.matrix[1,] <- parameter.fits$par
CI.matrix[2,] <- parameter.fits$par - 1.96 * parameter.se
CI.matrix[3,] <- parameter.fits$par + 1.96 * parameter.se
names(CI.matrix) <- c("d1", "d0", "scale", "location")
rownames(CI.matrix) <- c("ML", "95% Lower bound", "95% Upper bound")
CI.matrix # print value
```

نماذج الانحدار الخطي بفرض توزيع مستقر مع التطبيقات

اعداد : ليلى خالد محمود اللهاليه.

اشراف : الدكتور خالد صلاح.

الملخص

الانحدار الخطي هو شكل من أشكال النماذج الرياضية الذي يعكس النتائج في علاقة خطية بين متغير مستقل V ومتغير تابع X، حيث يعطى نموذج الانحدار البسيط وفق العلاقة التالية:

عادة ϵ تتبع التوزيع الطبيعي بوسط حسابي صفر وتباين σ^2 ، ومن هذا الفرض فان المتغير العشوائي التابع X ياخذ نفس التوزيع بالوراثة.

في التطبيقات العملية، ولكي يصح استخدام نماذج الانحدار الخطي يتطلب استخدام نموذج احصائي يعتمد علي عدة افتراضات من ضمنها واهمها التوزيع المثالي (الطبيعي) للمتغر التابع. ولكن اذا لم يتحقق هذا الشرط او شروط اخرى مثل وجود قيم متطرفة، فقد لا يكون اختبار الانحدار الخطي لاختبار الملائمة هو أقوى اختبار متاح، او قد لا يكون هذا يعني وجود علاقة خطية بين المتغر التابع والمستقل. في هذه الحالة، يمكننا استخدام نماذج الانحدار اللامعلمية أو افتراض توزيعات احتمالية أكثر قوة للبيانات. أحد هذه التوزيعات هو افتراض أن المتغير العشوائي X له توزيع مستقر $X{\sim}S_{\alpha}(\beta,\delta,\gamma)$.

من المعروف أنه، بشكل عام ، لا يوجد شكل معروف لدالة الكثافة الاحتمالية للتوزيعات المستقرة. ومع ذلك، في ظل تطبيق طريقة بيز، فإن استخدام متغير عشوائي كامن أو مساعد يعطي بعض التبسيط للحصول على أي توزيع لاحق (posterior) عندما يتعلق بالتوزيعات المستقرة. لتوضيح فائدة هذه الطريقة وخاصة في العمليات الحسابية، تم استخدام تطبيقين: أحدهما مرتبط بنموذج الانحدار الخطي القياسي بافتراض التوزيع الطبيعي ، والآخر مرتبط بنفس النموذج بافتراض توزيع مستقر. وبتطبيق ما سبق باستخدام طريقة (Markov Chain Monte Carlo) وبرنامج r ، تم الحصول على posterior summaries وكانت النتائج مرضية ومقبولة.