

NATIONAL AND KAPODISTRIAN UNIVERSITY OF
ATHENS



MASTER THESIS

Classical and Bayesian Inference for Threshold Regression Models

Author:
Maria BALTOUKA

Supervisor:
Dr. Loukia MELIGKOTSIDOU

*A thesis submitted in fulfillment of the requirements for the degree of M.Sc in
Statistics and Operational Research
in the*

Department of Mathematics

June 20, 2018

Abstract

This dissertation is concerned with Classical and Bayesian theory for the threshold regression model with one or two threshold variables. Threshold regression models have a wide variety of applications, mainly in the field of econometrics, and belong to the family of regression models with structural breaks that were introduced by Quandt (1960). In the literature most of the interest is focused on the discontinuous Threshold Regression Models, because of the non-standard asymptotic distribution of the statistical functions of the threshold parameters. Among others, we will examine this particular issue not only in theoretical but also in a more applied level.

Estimating the model parameters and obtaining asymptotic distributions of the respective estimators concentrates most of the interest in the area of Regression Analysis. The main goal is the construction of confidence intervals and hypotheses testing regarding the significance of each parameter. From the scope of Bayesian analysis, it is of great importance to take advantage of all the available information in order to define the prior distribution and finally get the posterior. The computation of the posterior distribution is a procedure that becomes more complex as the number of parameters increases, since it demands the calculation of composite integrals and the utilization of simulation techniques when the former is not applicable. Such methods are the Markov Chain Monte Carlo (MCMC) algorithms and a special case of them, the Gibb's sampler. All these methods are presented extensively in this dissertation and cover a wide variety of regression models.

Although the estimation of a model's parameters is the primary objective for a statistician, the selection of the most appropriate model for a given dataset comes first. Therefore, model comparison is the first step that one needs to do for precise and complete inference results. In essence, this is a hypotheses test that concerns the kind of relationship between the dependent variable and the explanatories, namely the type of model. Such tests are accomplished from the scope of Classical theory by using appropriate statistical tools, such as the LR statistic, and from the scope of Bayesian theory with the computation of each model's posterior probability. Having selected, either way, the most appropriate model, shall one proceed to statistical inference regarding its parameters.

Περίληψη

Η παρούσα διπλωματική εργασία πραγματεύεται την Κλασσική και Μπεϋζιανή θεωρία για το Μοντέλο Παλινδρόμησης με μια ή δυο μεταβλητές τύπου κατωφλίου (**threshold**). Τα **Threshold** Μοντέλα Παλινδρόμησης χρησιμοποιούνται σε μια πληθώρα εφαρμογών, κυρίως στον κλάδο της Οικονομετρίας, και ανήκουν στην ευρύτερη οικογένεια μοντέλων παλινδρόμησης με δομικές αλλαγές που εισήγαγε ο Quandt (1960). Στη βιβλιογραφία το ενδιαφέρον εστιάζεται συνήθως στα "μη συνεχή" **Threshold** Μοντέλα Παλινδρόμησης, λόγω της μη κανονικής ασυμπτωτικής κατανομής των στατιστικών συναρτήσεων που εμπεριέχουν τις παραμέτρους. Εδώ θα ασχοληθούμε, μεταξύ άλλων, και με την περίπτωση αυτή τόσο σε θεωρητικό αλλά και σε πρακτικό επίπεδο.

Στα Μοντέλα Παλινδρόμησης το ενδιαφέρον επικεντρώνεται στην εκτίμηση των παραμέτρων, καθώς και στην ασυμπτωτική κατανομή στατιστικών συναρτήσεων που τις εμπεριέχουν. Ανώτερος σκοπός είναι η κατασκευή διαστημάτων εμπιστοσύνης και ο έλεγχος υποθέσεων σχετικά με τη στατιστική σημαντικότητα κάθε παραμέτρου. Από την πλευρά της Μπεϋζιανής θεωρίας, κυρίαρχο ρόλο παίζει ο ορισμός των εκ των προτέρων (**prior**) κατανομών, δεδομένης της δοθείσας πληροφορίας, με σκοπό την εύρεση των εκ των υστέρων (**posterior**) κατανομών. Η εύρεση των **posterior** κατανομών είναι ένα σύνθετο πρόβλημα καθώς αυξάνει το πλήθος των παραμέτρων του μοντέλου και απαιτεί τον υπολογισμό σύνθετων ολοκληρωμάτων ή τη χρήση αλγορίθμων προσομοίωσης όταν αυτό δεν είναι αμφιγνό. Τέτοιοι αλγόριθμοι είναι οι **Markov Chain Monte Carlo (MCMC)** αλγόριθμοι και μια ειδική περίπτωση αυτών, ο δειγματολήπτης **Gibbs (Gibbs sampler)**. Όλες αυτές οι μέθοδοι παρουσιάζονται αναλυτικά στην παρούσα διπλωματική εργασία και καλύπτουν ένα ευρύ φάσμα μοντέλων παλινδρόμησης.

Παρ' ότι η εκτίμηση ενός μοντέλου είναι ο βασικός στόχος ενός στατιστικού, αυτό που προηγείται είναι η εύρεση του κατάλληλου μοντέλου για ένα δοθέν δείγμα. Η σύγκριση μοντέλων είναι λοιπόν το πρώτο βήμα που πρέπει να ακολουθήσει κανείς προκειμένου η συμπερασματολογία να είναι ολοκληρωμένη και ακριβής. Πρόκειται ουσιαστικά για έναν έλεγχο υποθέσεων ο οποίος υποδεικνύει τη σχέση που περιγράφει καλύτερα τη σύνδεση της εξαρτημένης μεταβλητής με τις επεξηγηματικές, δηλαδή το είδος του μοντέλου. Τέτοιοι έλεγχοι πραγματοποιούνται από την πλευρά της Κλασσικής θεωρίας με τη χρήση κατάλληλων στατιστικών συναρτήσεων, όπως είναι το **LR** στατιστικό μέτρο, και από την πλευρά της Μπεϋζιανής θεωρίας με τον υπολογισμό της εκ των υστέρων πιθανότητας για κάθε μοντέλο. Έχοντας καταλήξει λοιπόν, με τον ένα ή τον άλλο τρόπο, στο πιο κατάλληλο μοντέλο, τότε συνεχίζει κανείς με την ανάλυση και τη συμπερασματολογία για τις παραμέτρους του.

Acknowledgements

I would first like to thank my thesis advisor Dr. Loukia Meigkotsidou of the Department of Mathematics at the National and Kapodistrian University of Athens. The door to her office was always open whenever I ran into a trouble spot or had a question. She consistently allowed this dissertation to be my own work, but steered me in the right direction whenever it was necessary.

I would also like to thank the other two committee members, Dr. Apostolos Burnetas and Dr. Samis Trevezas, such as the rest of the academic professors that introduced me to Statistics and Operations Research. Moreover, I wish to acknowledge all the professors of the Department of Mathematics that inspired me and shared their knowledge with me during these two years of the M.Sc in Statistics and Operations Research.

Finally, I have to express my profound gratitude to all my beloved people and my family for providing me with unfailing support and continuous encouragement throughout my years of study and the process of researching and writing this thesis. This dissertation would not have been possible without them. Thank you.

Author,
Maria Baltouka.

Contents

Abstract	iv
Περίληψη	vii
Acknowledgements	vii
1 Introduction	1
2 Classical Inference	3
2.1 The multiple linear regression model	3
2.1.1 Estimation	3
2.1.2 Inference	4
2.2 The Threshold Regression model with one threshold variable	5
2.2.1 Estimation	7
2.2.2 Distributions of the Estimators	8
2.3 The Threshold Regression model with two threshold variables	12
2.3.1 Estimation	13
2.3.2 Distributions of Estimators	14
2.4 Model comparison	16
3 Bayesian Inference	19
3.1 The multiple linear regression model	19
3.2 The Threshold Regression model with one threshold variable	22
3.3 The Threshold Regression model with two threshold variables	25
3.4 Bayesian Model Comparison	29
4 Simulation Experiments	31
4.1 Model M_1 as DGP	32
4.1.1 Classical Inference	33
4.1.2 Bayesian Inference	34
4.2 Model M_3 as DGP	36
4.2.1 Classical Inference	37
4.2.2 Bayesian Inference	40
4.3 Model M_4 as DGP	49
4.3.1 Bayesian Inference	51
4.4 Comparing Classical and Bayesian results	57
5 Application	59
5.1 Model	59
5.2 Classical inference	61
5.3 Bayesian inference	62
5.4 Comparing Classical and Bayesian results	69
6 Conclusion	71

A	The Bootstrap Technique	73
B	The Gibbs Sampler	75

List of Figures

4.1	Marginal prior and posterior distribution of ϕ , for the linear regression model.	34
4.2	Joint Prior and Posterior distribution of b_1, b_2 , for the linear regression model.	35
4.3	Marginal prior and posterior distribution of b_1 , for the linear regression model.	35
4.4	Marginal prior and posterior distribution of b_2 , for the linear regression model.	36
4.5	The LR statistic for the one-threshold regression model.	38
4.6	Consecutive realizations of the marginal distribution of ϕ	41
4.7	Consecutive realizations of the marginal distribution of $b_0^{(2)}, b_1^{(2)}, \delta_{n,0}, \delta_{n,1}$	42
4.8	Consecutive sample means of ϕ	43
4.9	Consecutive sample means of $b_0^{(2)}, b_1^{(2)}, \delta_{n,0}, \delta_{n,1}$	44
4.10	Marginal prior and posterior distribution of ϕ , for the one-threshold regression model.	45
4.11	Marginal prior and posterior distribution of $b_0^{(2)}$, for the one-threshold regression model.	45
4.12	Marginal prior and posterior distribution of $b_1^{(2)}$, for the one-threshold regression model.	46
4.13	Marginal prior and posterior distribution of $\delta_{n,0}$, for the one-threshold regression model.	46
4.14	Marginal prior and posterior distribution of $\delta_{n,1}$, for the one-threshold regression model.	47
4.15	Marginal prior and posterior distribution of ϕ , for the linear regression model.	48
4.16	Joint prior and posterior distribution of b_1, b_2 , for the linear regression model.	48
4.17	Marginal prior and posterior distribution of b_1 , for the linear regression model.	49
4.18	Marginal prior and posterior distribution of b_2 , for the linear regression model.	49
4.19	The LR statistic for the two-threshold regression model.	50
4.20	The LR statistic as a function of γ_1 , for the two-threshold regression model.	50
4.21	The LR statistic as a function of γ_2 , for the two-threshold regression model.	51
4.22	Consecutive realizations of the marginal distribution of ϕ	52
4.23	Consecutive realizations of the marginal distribution of $b_0^{(1)}, b_0^{(2)}, b_0^{(3)}, b_0^{(4)}, b_1^{(1)}, b_1^{(2)}, b_1^{(3)}, b_1^{(4)}$	53
4.24	Consecutive sample means of ϕ	54

4.25	Plot of ergodic averages of $b_0^{(1)}, b_0^{(2)}, b_0^{(3)}, b_0^{(4)}, b_1^{(1)}, b_1^{(2)}, b_1^{(3)}, b_1^{(4)}$	55
4.26	Marginal prior and posterior distribution of ϕ , for the two-threshold regression model.	56
4.27	Marginal prior and posterior distribution of $b_0^{(1)}, b_0^{(2)}, b_0^{(3)}, b_0^{(4)}$, for the two-threshold regression model.	56
4.28	Marginal prior and posterior distribution of $b_1^{(1)}, b_1^{(2)}, b_1^{(3)}, b_1^{(4)}$, for the two-threshold regression model.	57
5.1	Annual U.S GDP real growth rate	60
5.2	GDP/debt ratio 1791-2009	60
5.3	Scatterplot of real GDP growth and debt/GDP	61
5.4	Consecutive realizations of the marginal distribution of ϕ	65
5.5	Consecutive realizations of the marginal distribution of $b_3^{(0)}, b_3^{(1)}, b_1, b_2$	65
5.6	Consecutive sample means of ϕ	66
5.7	Consecutive sample means of $b_3^{(0)}, b_3^{(1)}, b_1, b_2$	66
5.8	Marginal prior and posterior distribution of ϕ , for the kink model.	67
5.9	Prior and Posterior marginal distribution of $b_3^{(0)}$, for the kink model.	67
5.10	Marginal prior and posterior distribution of $b_3^{(1)}$, for the kink model.	68
5.11	Marginal prior and posterior distribution of b_1 , for the kink model.	68
5.12	Marginal prior and posterior distribution of b_2 , for the kink model.	69

List of Tables

2.1	Asymptotic Critical Values	10
2.2	Asymptotic Critical Values	15
4.1	Linear regression model. Estimation and Confidence Intervals	33
4.2	One-threshold regression model. Estimation and Confidence Intervals.	38
4.3	Linear regression model. Estimation and Confidence Intervals	39
4.4	Two-threshold regression model. Estimation and Confidence Intervals.	51

Chapter 1

Introduction

Regression analysis is a set of statistical procedures for estimating the relationship between a variable of interest (dependent variable) and an available set of explanatory variables (independent variables). The estimation of this relationship not only is useful for future predictions of the dependent variable given the values of the independents, but also helps us to understand the impact of a change of one of the explanatory variables on the value of the dependent variable.

The earliest form of regression was the method of least squares, which was published by Legendre (1805) and by Gauss (1809). Although the term "regression", which was coined by Francis Galton in the 19th century for the explanation of a biological phenomenon, had at first only a biological meaning, it was later extended to a more general statistical context.

The most simple and commonly used type of regression is linear regression. In linear regression analysis, a linear relationship between the dependent and the independent variables is assumed and it is of high interest to estimate the straight line that best describes this kind of relationship. Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways such as likelihood based methods.

An interesting issue in the linear regression analysis is to examine whether the regression coefficients remain stable when the model is estimated on proper subsamples. The selection of these subsamples is done on the basis of a variable that could be either categorical, such as the gender, or continuous, such as the size of a firm or time. In the latter case, we are interested to examine at which value of this threshold variable we shall split the sample. Of course, there may exist more than one threshold variables and according to their values the sample is splitted in more than two subsamples.

Detecting structural breaks or instability in regression models has attracted a vast amount of attention since the work of Quandt (1960). Since then, threshold regression models have become popular mainly in the field of econometrics and non-linear time series. Tong (1983, 1990) introduced the regression discontinuity models, such as the Threshold Autoregressive model (TAR), the Smooth Transition Autoregressive model (STAR) and the Self Exciting Autoregressive model (SETAR). There is a large literature on discontinuous threshold regression models, including interesting theoretical results and a plenty of applications. For the issue of testing for a threshold effect, see Chan (1990,1991), Chan and Tong (1990), Hansen (1996) and Lee,Seo and Shin (2011). For inference on the model's parameters relevant contributions include Chan (1993), Hansen (2000) and Seo and Linton (2007). Panel data methods have been developed by Hansen (1999) and Ramirez-Rondan (2013).

On the other hand, Chan and Tsay (1998) introduced the continuous threshold

model, which is identical to a regression kink model with piecewise linear regression segments. Economic applications of the continuous threshold regression models include those of Cox, Hansen and Jimenez (2004) and Hansen (2017).

This dissertation is concerned with classical and Bayesian theory for the multiple linear regression model, the discontinuous and kink regression models with one threshold variable, the threshold regression model with two threshold variables, and is organised as follows. Chapter 2 describes in detail methods to estimate each model's parameters, construct confidence intervals and test hypotheses from the scope of classical theory. Chapter 3 presents extensively the Bayesian theory for the linear regression model and the threshold regression models with one and two threshold variables and concludes with the computation of the evidence for the three models in order to use them for model selection. In this chapter, the Gibb's sampler is utilized for inference for the models under consideration. Chapter 4 contains simulation experiments designed to assess the adequacy of the model selection techniques and the compatibility of the ordinary and Bayesian results. Chapter 5 reports the results from an application of the kink model with one threshold variable (examined also by Hansen (2017)), using the data of Reinhart and Rogoff (2010). In this chapter, inference is made from the scope of both the classical and the Bayesian approach to inference and the respective results are compared.

Chapter 2

Classical Inference

2.1 The multiple linear regression model

Linear regression analysis is one of the most used techniques to describe the relationship between a variable of interest and a set of related explanatory variables. Given a set of data (y_i, x_i) for $i = 1, 2, \dots, n$, where $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$, the multiple linear regression attempts to model this relationship by fitting a linear equation to the observed data.

The straight line connecting these variables has the form $y = b_0 + b_1x_1 + \dots + b_px_p = x_i^*b$, where $x_i^* = (1, x_{i1}, \dots, x_{ip})'$ and $b = (b_0, b_1, \dots, b_p)'$, but since the observed values vary about this line (having the same standard deviation σ) a term denoting these deviations is added to the model. This error term, denoted by e_i , is a random variable that accounts for the model to fit the data precisely. Formally, the multiple linear regression model is:

$$y_i = x_i^*b + e_i, \quad (2.1)$$

or in matrix form:

$$Y = Xb + e, \quad (2.2)$$

where $Y = (y_1, y_2, \dots, y_n)'$ is an $n \times 1$ vector containing the real valued y_i , $b = (b_0, \dots, b_p)'$ is the vector of unknown coefficients, $X = (x_1^*, x_2^*, \dots, x_n^*)'$ is an $n \times (p + 1)$ matrix, called the design matrix, and $e = (e_1, \dots, e_n)'$ is the vector of error terms.

We assume that the mean and variance of e_i is 0 and σ^2 respectively, and that the errors are uncorrelated. Then, it is easy enough to see that $E(y_i) = x_i^*b$ and $V(y_i) = \sigma^2$. Under the more restrictive hypothesis that e_i are normally distributed, y_i are also normally distributed with the above mean and variance.

2.1.1 Estimation

After assuming a model for the observed data, we aim to estimate its unknown parameters. The most common technique to get the best fitting line in linear regression is the Ordinary Least Squares method (OLS). This method utilizes the least-squares equation

$$S(b) = \sum_{i=1}^n e_i^2 = (Y - Xb)'(Y - Xb)$$

and the obtained estimator \hat{b} is the vector minimizing $S(b)$.

Assumptions A

A1. $E(e_i) = 0, \forall i = 1, 2, \dots, n$

$$A2. V(e_i) = \sigma^2, \forall i = 1, 2, \dots, n$$

$$A3. Cov(e_i, e_j) = 0, \forall i \neq j$$

Under Assumptions A we get the unbiased and efficient estimator for b ,

$$\hat{b} = (X'X)^{-1}X'Y,$$

provided that the matrix $(X'X)^{-1}$ exists, which means that any column of the design matrix can not be expressed as a linear combination of any other columns. The next step of the linear model analysis concerns σ^2 , which should be also estimated. An unbiased estimator of σ^2 is the mean squared error (MSE), i.e

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{n-p-1}, \text{ where } \hat{e}_i = y_i - \hat{y}_i = y_i - x_i^* \hat{b}.$$

2.1.2 Inference

Under the assumption that the error terms are normally distributed with mean 0 and common variance σ^2 we obtain the distribution of the estimator,

$$\hat{b} \sim N_{p+1}(b, \sigma^2(X'X)^{-1}).$$

That is,

$$\hat{b}_i \sim N(b_i, \sigma^2(X'X)_{ii}^{-1}),$$

where $(X'X)_{ii}^{-1}$ denotes the ii -th item of the matrix $(X'X)^{-1}$. It has been also proven that

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{n-p-1} \sim X_{(n-p-1)}^2,$$

and thus,

$$t = \frac{\hat{b}_i - b_i}{\sqrt{\hat{\sigma}^2(X'X)_{ii}^{-1}}} \sim t\text{-student}_{(n-p-1)}.$$

Having obtained the distribution of \hat{b}_i , the $100(1-\alpha)\%$ confidence interval for b_i , is

$$\left[\hat{b}_i - \sqrt{\hat{V}(\hat{b}_i)} t_{(n-p-1)}^{\alpha/2}, \hat{b}_i + \sqrt{\hat{V}(\hat{b}_i)} t_{(n-p-1)}^{\alpha/2} \right].$$

Here, we have denoted the estimator of b_i 's variance, namely $\hat{\sigma}^2(X'X)_{ii}^{-1}$, as $\hat{V}(\hat{b}_i)$.

Once we have estimated the unknown parameters of the model and have obtained the parameters' distributions, we are faced with two questions:

- does the specific linear model fit well to the data?
- which of the explanatory variables seem important?

Those questions can be answered under the normality assumption for the error terms.

The test for the significance of the regression is an overall test, used to determine if there is a linear connection between y and any of the covariates x_1, \dots, x_p . This test's competing hypotheses are:

$$\begin{aligned} H_0 : b_1 = b_2 = \dots = b_p = 0 \\ H_1 : b_j \neq 0 \text{ for at least one } j = 1, 2, \dots, p. \end{aligned}$$

Rejection of the null hypothesis means that at least one of the regressors impacts linearly on y . Under H_0 , it has been proven that

$$\frac{SSR}{\sigma^2} \sim X_{(p)}^2 \text{ and } \frac{SSE}{\sigma^2} \sim X_{(n-p-1)}^2,$$

where $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ and $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. Hence,

$$F = \frac{\frac{SSR}{p}}{\frac{SSE}{n-p-1}} \sim F_{(p, n-p-1)}.$$

Large values of F lead to the rejection of the null hypothesis.

Once we have been led to the decision that at least one of the regressors is significant, the next question is which one(s). The main goal for a statistician is to explain the variability of y_i with as few explanatory variables as possible. Adding variables to a model increases the information, but adding unimportant covariates may negatively affect the statistical modeling in terms of parsimony. The set of hypothesis tests consists of tests of the form:

$$\begin{aligned} H_0 : b_i = 0 \text{ for some } i = 1, 2, \dots, p, \\ H_1 : b_i \neq 0. \end{aligned}$$

H_0 is rejected if the observed data provide evidence that the specific regressor is important in predicting y and has to be included in the model.

2.2 The Threshold Regression model with one threshold variable

The threshold regression model with one threshold variable differs from the linear regression model in the way that it models the relationship between the variable of interest y_i and a set of explanatory variables $(x_{i1}, x_{i2}, \dots, x_{ip})$. In contrast to the linear regression model, the threshold regression model assumes that there is a linear relationship between the observed data (y_i, x_i) , where $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, but there are two instead of one linear equations modeling this relationship depending on the value of another variable called threshold variable. If the value of the threshold variable is less or equal to the value of a threshold parameter, that can be known or not, then the linear equation connecting y_i and x_i is $y_i = b_0^{(1)} + b_1^{(1)}x_{i1} + b_2^{(1)}x_{i2} + \dots + b_{m_1}^{(1)}x_{ip_1} + e_i$. Otherwise, there is a different linear equation that expresses the relationship between y_i and x_i and this is $y_i = b_0^{(2)} + b_1^{(2)}x_{i1} + b_2^{(2)}x_{i2} + \dots + b_{m_2}^{(2)}x_{ip_2} + e_i$. Therefore, the threshold regression model in general has the form:

$$y_i = \begin{cases} b_0^{(1)} + b_1^{(1)}x_{i1} + b_2^{(1)}x_{i2} + \dots + b_{m_1}^{(1)}x_{ip_1} + e_i, & \text{if } q_i \leq \gamma, \\ b_0^{(2)} + b_1^{(2)}x_{i1} + b_2^{(2)}x_{i2} + \dots + b_{m_2}^{(2)}x_{ip_2} + e_i, & \text{otherwise,} \end{cases} \quad (2.3)$$

where q_i is the threshold variable, classifying the observations y_i into two regimes, depending on whether the threshold variable q_i exceeds or not the threshold parameter γ . We assume that γ lies in $\Gamma = [\underline{\gamma}, \bar{\gamma}]$, which is a strict subset of the support of q_i . The two regimes are distinguished by differing slope parameters $b^{(j)} = (b_0^{(j)}, b_1^{(j)}, \dots, b_{p_j}^{(j)})'$, $j = 1, 2$, where p_j , $j = 1, 2$, is the number of regressors in each regime. Regarding the error terms e_i , they are assumed to be independent and identically distributed with mean equal to zero and common variance (at least in each regime). We will examine the case of common variance σ^2 for both regimes.

Without loss of generality, we set $p = \max(p_1, p_2)$ and $b_i^{(j)} = 0$ when $i > p_j$ and then, given the sample (y_i, x_i, q_i) for $i = 1, \dots, n$ and $x_i^* = (1 \ x_{i1} \dots \ x_{ip})'$, we aim to estimate the threshold parameter γ and the structural parameters $b^{(j)}$.

Regarding the threshold variable q_i , it can be exogenous or a function of x_i and this is what discriminates the threshold regression model with a jump from the threshold regression model with a kink. In the former model, which is identical to the continuous threshold regression model introduced by Chan and Chay (1998), the regression function is continuous but the slope has a discontinuity at a threshold point named "kink". In the latter model, which belongs to the discontinuity threshold regression models introduced by Tong (1983, 1990), the regression model is split in two (or more) regimes according to a threshold indicator. Although, as we will see in more detail, the estimation does not depend in essence on whether the threshold regression model is continuous or not, the distribution theory does. More specifically, if we decide that the model has a jump, one has to follow Hansen (2000), while if we decide for a kink, then one shall employ the asymptotic normal inference as in Feder (1975a) and others.

The simplest and most normal way for someone to write the threshold regression model with one threshold parameter, is this of equation 2.3. However, this equation can be written in more convenient forms, depending on whether we are in the case of a model with a jump or with a kink, which is useful not only for reasons of comprehension but also because in this way estimation and inference techniques are more applicable.

The threshold regression model with a jump. Denoting by $d_i(\gamma) = 1_{(q_i \leq \gamma)}$ and $x_i^*(\gamma) = x_i^* 1_{(q_i \leq \gamma)}$ model (2.3) becomes:

$$y_i = x_i^{*'} b + x_i^*(\gamma)' \delta_n + e_i, \quad (2.4)$$

where $b = b_2$ and $\delta_n = b_1 - b_2$ or in matrix form :

$$Y = Xb + X_\gamma \delta_n + e = X_\gamma \underline{b} + e. \quad (2.5)$$

X_γ is an $n \times (p + 1)$ matrix, where every item of its i -th row is the respective item of the i -th row of the matrix X multiplied with 1 if $q_i \leq \gamma$ or with 0 otherwise. Y , X and e are as defined under equation (2), $X_\gamma^* = [X \ X_\gamma]$ and $\underline{b} = (b' \ \delta_n)'$.

The threshold regression model with a kink. Suppose that q_i is a coordinate of the vector x_i , namely one of the available explanatory variables and z_i is the vector of the explanatory variables (x_i^*) of the i -th observation, having excluded q_i . Then the regression kink model takes the form:

$$y_i = d_1(q_i - \gamma)^- + d_2(q_i - \gamma)^+ + d_3' z_i + e_i, \quad (2.6)$$

where we use $(a)^+ = \max\{0, a\}$ and $(a)^- = \min\{0, a\}$ to denote the positive and respectively the negative part of a number a . In this model, the slope with respect to

variable q_i is d_1 for values of q_i less than γ and d_2 for values of q_i greater than γ . This means that the regression function is continuous in the variables q and z , but the slope has a kink (discontinuity) at $q = \gamma$. Model (2.6) could also be written as:

$$y_i = x_i^*(\gamma)' \underline{b} + e_i, \quad (2.7)$$

where $x_i^*(\gamma) = \begin{pmatrix} (q_i - \gamma)^- \\ (q_i - \gamma)^+ \\ z_i \end{pmatrix}$ and $\underline{b} = (d_1 \ d_2 \ d_3)'$. In matrix form, model (2.7) can now be written as follows.

$$Y = X_\gamma^* \underline{b} + e, \quad (2.8)$$

where $X_\gamma^* = \begin{bmatrix} (q - \gamma)^- \\ (q - \gamma)^+ \\ Z \end{bmatrix}$.

2.2.1 Estimation

The first step of classical inference is point estimation of the model's parameters. For the threshold regression model with a jump the parameters are (b, δ_n, γ) , whereas for the threshold regression model with a kink they are (d_1, d_2, d_3, γ) . Thus, denoting by \underline{b} each model's slope parameters, the threshold regression model's parameters are (\underline{b}, γ) . Let

$$s_n(\underline{b}, \gamma) = (Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) \quad (2.9)$$

be the sum of squared errors. Then, the least squares (LS) estimator for the model's parameters denoted by $(\hat{\underline{b}}, \hat{\gamma})$ minimizes the quantity in equation (2.9) and the estimation method does not depend on whether the model has a jump or a kink. The LS estimators are obtained in three steps, based on the assumption that γ lies in the bounded set Γ .

Step 1: Equation (6), conditional on γ is linear on \underline{b} , so for each value of $\gamma \in \Gamma$ the OLS estimators $\hat{\underline{b}}(\gamma)$ are obtained via the regression of Y on X_γ^* .

Step 2: The concentrated sum of squared errors is that of equation (2.9), given the estimators obtained from Step 1. Thus,

$$s_n(\gamma) = s_n(\hat{\underline{b}}(\gamma), \gamma) = Y'Y - Y'X_\gamma^*(X_\gamma^{*'}X_\gamma^*)^{-1}X_\gamma^{*'}Y$$

and

$$\hat{\gamma} = \underset{\gamma \in \Gamma_n}{\operatorname{argmin}} s_n(\gamma), \text{ where } \Gamma_n = \bar{\Gamma} \cap (q_1, \dots, q_n).$$

Step 3: Finally, given $\hat{\gamma}$ we obtain the slope parameters' estimators which are $\hat{\underline{b}} = \hat{\underline{b}}(\hat{\gamma})$.

The above method requires at least n evaluations to get $\hat{\gamma}$, but in the case of n being very large, we can, for some $N < n$, let $q_{(j)}$ denote the $\frac{j}{N}$ -th quantile of the sample (q_1, \dots, q_n) and $\Gamma_N = \bar{\Gamma} \cap (q_{(1)}, q_{(2)}, \dots, q_{(N)})$. Then $\hat{\gamma}_N = \underset{\gamma \in \Gamma_N}{\operatorname{argmin}} s_n(\gamma)$ is a good approximation to $\hat{\gamma}$, which requires only N evaluations.

2.2.2 Distributions of the Estimators

The threshold regression model with a jump

According to Hansen (2000), in order to obtain the sampling distributions of the estimators, we must first define the moment functionals:

1. $M(\gamma) = E[x_i^* x_i^{*'} 1_{q_i \leq \gamma}]$
2. $D(\gamma) = E[x_i^* x_i^{*'} | q_i = \gamma]$
3. $V(\gamma) = E[x_i^* x_i^{*'} e_i^2 | q_i = \gamma]$

We also let $f(q)$ be the density function of q_i , γ_0 the true value of γ , $D = D(\gamma_0)$, $V = V(\gamma_0)$, $f = f(\gamma_0)$ and $M = E[x_i^* x_i^{*'}]$.

Assumptions B

- B1. (x_i^*, q_i, e_i) for $i=1, \dots, n$ is strictly stationary, ergodic and ρ -mixing with ρ -mixing coefficients satisfying $\sum_{m=1}^{+\infty} \rho_m^{1/2} < +\infty$
- B2. $E[e_i | F_{i-1}] = 0$
- B3. $E[|x_i^*|^4] < +\infty$ and $E[|x_i^* e_i|^4] < +\infty$
- B4. $E[|x_i^*|^4 | q_i = \gamma] \leq C$ and $E[|x_i^*|^4 e_i^4 | q_i = \gamma] \leq C$ for some $C \leq +\infty$ and $f(\gamma) < \bar{f}$ for every $\gamma \in \Gamma$
- B5. $f(\gamma)$, $V(\gamma)$ and $D(\gamma)$ are continuous at $\gamma = \gamma_0$
- B6. $\delta_n = cn^{-a}$ with $c \neq 0$ and $0 < a < 1/2$
- B7. $c'Dc > 0$, $c'Vc > 0$ and $f > 0$
- B8. $M > M(\gamma) > 0 \forall \gamma \in \Gamma$

Assumption B1 implies that all the regressors are stationary and ergodic, and is automatically satisfied for independent observations. Stationarity excludes time trends, while the condition of stationary ergodic y_i allows us to apply the law of large numbers. The ρ -mixing assumption controls the degree of time series dependence. B2 and B3 require that the e_i , $i = 1, 2, \dots, n$, is a martingale sequence, which means that the model is correctly specified. Assumptions B3 and B4 refer to the conditional and unconditional fourth order bounds, while B5 implies that γ is continuous with a positive density function. Moreover, the condition of continuous variance at γ_0 excludes regime-dependent heteroskedasticity. B6 means that the slope difference decreases and converges to zero as the sample size increases. Under this assumption we are able to obtain a simple limiting distribution of $\hat{\gamma}$, free of nuisance parameters. Assumption B7 is used in order to have a non-degenerating distribution for the threshold estimator and to exclude the case of the continuous threshold model (the restriction $c'Dc > 0$ ensures that property). The continuous threshold model is (4)-(5), with $x_i^* = (1 \ q_i)'$, and $\delta_n \gamma_0^* = 0$, where $\gamma_0^* = (1 \ \gamma_0)'$. Finally, B8 excludes multicollinearity and restricts Γ on a proper subset for q_i .

Asymptotic distribution of the threshold estimate

Theorem 1 Under assumptions B and according to Hansen (2000)

$$n^{1-2a}(\hat{\gamma} - \gamma_0) \rightarrow \omega T \quad (2.10)$$

where $\omega = \frac{c'Vc}{(c'Dc)^2 \bar{f}}$ and $T = \underset{r \in \mathbb{R}}{\operatorname{argmax}}[-\frac{1}{2}|r| + W(r)]$, for $W(r)$ being a two-sided Brownian motion on the real line.

At this point we define as a two-sided Brownian motion:

$$W(r) = \begin{cases} W_1(-r) & , \text{if } r < 0 \\ 0 & , \text{if } r=0 \\ W_2(r) & , \text{otherwise} \end{cases}$$

where $W_1(r), W_2(r)$ are two independent standard Brownian motions on $[0, +\infty)$.

The distribution function for T is given in Bhattacharya and Brockwell (1976) as follows:

$$P(T \leq x) = \begin{cases} 1 + \sqrt{\frac{x}{2\pi}} \exp(-\frac{x}{8}) + \frac{3}{2} \exp(x) \Phi(-\frac{3\sqrt{x}}{2}) - \frac{x+5}{2} \Phi(-\frac{\sqrt{x}}{2}) & , \text{if } x \geq 0 \\ 1 - P(T \leq -x) & , \text{otherwise} \end{cases}$$

with Φ being the cumulative standard normal distribution function.

Confidence interval of the threshold parameter

In order to obtain a confidence interval for γ we utilize the quantities T and ω from Theorem 1. Although T is free of parameters, ω is a function of δ_n and also depends, through $D(\gamma_0)$, from γ_0 . For that reason, the most common way to obtain a confidence interval, that is the inversion of Wald statistic, can have really poor sample behavior. Another proposed way to get confidence intervals is the use of the likelihood ratio statistic $LR_n(\gamma)$.

To test the hypothesis $H_0 : \gamma = \gamma_0$, the likelihood ratio test rejects the null hypothesis for large values of the statistic $LR_n(\gamma_0)$, where

$$LR_n(\gamma) = n \frac{s_n(\gamma) - s_n(\hat{\gamma})}{s_n(\hat{\gamma})}. \quad (2.11)$$

Theorem 2 *It holds that under Assumptions B,*

$$LR_n(\gamma_0) \rightarrow \eta^2 \xi, \quad (2.12)$$

where $\xi = \max_{s \in R} [2W(s) - |s|]$, for $\eta^2 = \frac{c'Vc}{\sigma^2 c'Dc}$.

The distribution function of ξ is $P(\xi \leq x) = (1 - e^{-x/2})^2$. Unless error homoscedasticity given q_i holds, which means that $E(e_i^2|q_i) = \sigma^2$ (it is then $\eta^2 = 1$), $LR_n(\gamma_0)$ is not free of nuisance parameters and η^2 has to be estimated. However, we are only concerned for the case of homoscedastic errors. Thus, although the asymptotic distribution of the likelihood ratio test is non standard, it is though free of nuisance parameters and since the distribution function of ξ is given in a simple form, we can get asymptotic p-values for observed test statistics. Under H_0 , the p-value is determined as:

$$p_n = P(\xi > LR_n(\gamma_0)) = 1 - \left(1 - e^{-\frac{LR_n(\gamma_0)}{2}}\right)^2.$$

The null hypothesis is rejected with significance level of α for values of $LR_n(\gamma_0)$ greater than the critical value $c_\xi(\alpha)$. Note here that the critical value $c_\xi(\alpha)$ is determined such as $P(\xi > c_\xi(\alpha)) = \alpha$. The calculation of $c_\xi(\alpha)$, which from now on we will write for simplicity reasons $c(\alpha)$, is accomplished through inversion of the distribution function of ξ and therefore, $c(\alpha) = -2\log(1 - \sqrt{1 - \alpha})$. The next table contains

some of these critical values.

Table 2.1: Asymptotic Critical Values

	0.80	0.85	0.90	0.925	0.95	0.975	0.99
$P(\xi \leq x)$	4.50	5.10	5.94	6.53	7.35	8.75	10.59

At this point, we are ready to construct the $(1 - \alpha)$ confidence interval for γ that is

$$\hat{\Gamma} = \{\gamma : LR_n(\gamma) \leq c(\alpha)\}.$$

A proposed graphical way to obtain the confidence interval, is by plotting $LR_n(\gamma)$ against γ and drawing a flat line at $c(\alpha)$. Then, the confidence interval of γ is the set of the values for which the graph of the $LR_n(\gamma)$ lies beneath the horizontal line $c(\alpha)$.

Asymptotic distribution of the slope parameters

Although $\hat{b} = \hat{b}(\hat{\gamma})$ depends on the estimator $\hat{\gamma}$ and inference on \hat{b} would be difficult, Chan (1993) and Hansen (1997) showed that inference on \hat{b} can be based on $\hat{\gamma}$ as if it was the true value of γ . Then,

$$\sqrt{n}(\hat{b} - \underline{b}) \xrightarrow{d} N(0, V(\underline{b})), \quad (2.13)$$

where $V(\underline{b})$ is the standard asymptotic covariance matrix, which can be estimated by

$$\widehat{V(\underline{b})} = (X_{\hat{\gamma}}^{*'} X_{\hat{\gamma}}^*)^{-1} \hat{\sigma}^2.$$

While the confidence interval for γ was constructed under the assumption that the error terms are iid, this assumption is not necessary for the slope parameters' confidence intervals. In this case,

$$\widehat{V(\underline{b})} = (X_{\hat{\gamma}}^{*'} X_{\hat{\gamma}}^*)^{-1} (X_{\hat{\gamma}}^{*'} X_{\hat{\gamma}}^* \Omega) (X_{\hat{\gamma}}^{*'} X_{\hat{\gamma}}^*)^{-1},$$

where $\Omega = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$.

As stated by Hansen (2000), since $\hat{\gamma}$ will not be equal to the real value of $\gamma = \gamma_0$ at each sample, if someone would like to incorporate that type of uncertainty in the confidence intervals, a Bonferroni-type bound should be used. If we let $\hat{B}(\gamma)$ denote the α -level confidence region for b , then from (11), $P(b \in \hat{B}(\gamma)) \rightarrow \alpha$ as $n \rightarrow +\infty$ and γ is known. The Bonferroni-type bound, is a construction of confidence regions of b , for any γ lying in a ρ -level confidence region for γ . Let for any $\rho < 1$, $\hat{\Gamma}(\rho)$ denote the ρ -level confidence interval for γ . For each $\gamma \in \hat{\Gamma}(\rho)$, construct the pointwise confidence region $\hat{B}(\gamma)$ and then set $\hat{B}_\rho = \bigcup_{\gamma \in \hat{\Gamma}(\rho)} \hat{B}(\gamma)$. Given that $\hat{B}_\rho \supset \hat{B}(\hat{\gamma})$, $P(b \in \hat{B}_\rho) \geq P(b \in \hat{B}(\hat{\gamma})) \rightarrow \alpha$ as $n \rightarrow +\infty$.

The Threshold Regression model with a kink

The following assumptions are necessary for obtaining asymptotic distributions, as they are presented in Hansen (2017).

Assumptions C. For some $r > 1$,

- C1. (y_i, x_i^*) for $i = 1, 2, \dots, n$ is strictly stationary, ergodic and absolutely regular with mixing coefficients $\eta(m) = O(m^{-A})$ for some $A > \frac{r}{r-1}$.
- C2. $E[|y_i|^{4r}] < +\infty, E[|x_i^*|^{4r}] < +\infty$.
- C3. $\inf_{\gamma \in \Gamma} \det Q(\gamma) > 0$, where $Q(\gamma) = E[x_i^*(\gamma)' x_i^*(\gamma)]$.
- C4. x_i^* has a density function $f(x)$ satisfying $f(x) \leq \bar{f} < +\infty$.
- C5. $\gamma \in \Gamma$, where Γ is a compact set.

Assumptions C1 and C2 are weak dependence conditions that allow the application of the central limit theorem. For independent observations we can set r arbitrarily close to one, while in general the choice of r depends on the allowable degree of serial dependence and the number of finite moments. Assumption C3 requires that the projection coefficients are well defined for all values of γ in the parameter space, while C4 requires that the threshold variable, as in the jump case, has a bounded density function.

Asymptotic distribution of the regression coefficients

Chan and Chay (1998) showed that the least square estimates (including the threshold coefficient), as they were described in subsection 2.2.1, in the continuous threshold regression model are jointly normally distributed. Hansen (2017) in his recent paper, extends this distribution theory for the regression kink model. Let $\theta = (\underline{b}, \gamma)$ denote the model's parameters, $\hat{\theta} = (\hat{\underline{b}}, \hat{\gamma})$ and $\theta_0 = (\underline{b}_0, \gamma_0)$. Set also,

$$\begin{aligned} H_i(\theta) &= -\frac{\partial}{\partial \theta} (y - \underline{b}' x_i(\gamma)) \\ &= \begin{pmatrix} x_i(\gamma) \\ -d_1 \mathbf{1}_{q_i \leq \gamma} - d_2 \mathbf{1}_{q_i > \gamma} \end{pmatrix} \end{aligned}$$

and $H_i = H_i(\theta_0)$.

Theorem 3 *Under Assumptions C, it holds that*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V(\theta)) \quad (2.14)$$

where,

$$V(\theta) = Q^{-1} S Q^{-1}, S = \sum_{j=-\infty}^{+\infty} E[H_i H_{i+j}' e_i e_{i+j}]$$

and

$$Q = E[H_i H_i'] + E \begin{bmatrix} 0 & 0 & 0 & e_i \mathbf{1}_{q_i \leq \gamma_0} \\ 0 & 0 & 0 & e_i \mathbf{1}_{q_i > \gamma_0} \\ 0 & 0 & 0 & 0 \\ e_i \mathbf{1}_{q_i \leq \gamma_0} & e_i \mathbf{1}_{q_i > \gamma_0} & 0 & 0 \end{bmatrix}.$$

Notice that if the threshold model is correctly specified, hence $E[e_i | x_i^*] = 0$, then the second term of Q equals 0. However, it may be nonzero is the case of model misspecification. From Theorem 3, one can conclude that the slope and threshold estimates are jointly asymptotically normal and they have a non zero asymptotic covariance. On the contrary, in the discontinuous threshold regression model the slope and threshold estimates are not only independent, but γ has a nonstandard distribution

and this is because in the kink model the regression function is continuous. Hansen (2017), suggests the following estimate for the covariance matrix $V(\theta)$,

$$\widehat{V}(\theta) = \widehat{Q}^{-1} \widehat{S} \widehat{Q}^{-1},$$

where,

$$\widehat{S} = \frac{1}{n-p-2} \sum_{i=1}^n \widehat{H}_i \widehat{H}_i' \widehat{e}_i^2,$$

$$\widehat{H}_i = \begin{pmatrix} x_i(\widehat{\gamma}) \\ -\widehat{d}_1 1_{q_i \leq \widehat{\gamma}} - \widehat{d}_2 1_{q_i > \widehat{\gamma}} \end{pmatrix} \text{ and}$$

$$\widehat{Q} = \frac{1}{n} \sum_{i=1}^n \left(\widehat{H}_i \widehat{H}_i' + \begin{bmatrix} 0 & 0 & 0 & \widehat{e}_i 1_{q_i \leq \widehat{\gamma}} \\ 0 & 0 & 0 & \widehat{e}_i 1_{q_i > \widehat{\gamma}} \\ 0 & 0 & 0 & 0 \\ \widehat{e}_i 1_{q_i \leq \widehat{\gamma}} & \widehat{e}_i 1_{q_i > \widehat{\gamma}} & 0 & 0 \end{bmatrix} \right).$$

Under the normal asymptotic distribution, the construction of confidence intervals for the parameters could be done using the conventional rule. For example the 95% confidence interval of d_2 is $\widehat{d}_2 \pm z_{0.025} s(\widehat{d}_2)$.

Although the normal distribution is convenient for the construction of the confidence interval of γ , for small samples it may have poor coverage and this happens because the least square criterion is nonquadratic with respect to γ . For this reason, we shall use the *LR* statistic for the construction of the confidence interval of γ . Suppose that we aim to test the hypotheses:

$$\begin{aligned} H_0 : \gamma &= \gamma_0 \\ \text{vs } H_1 : \gamma &\neq \gamma_0. \end{aligned}$$

Then, the null hypothesis is rejected for large values of the statistic $LR_n(\gamma_0)$, as it is defined in equation (2.11). In comparison to the model with a jump, this test has an asymptotic X_1^2 distribution under H_0 , due to the asymptotic normality of Theorem 3. Thus, the $(1 - \alpha)$ confidence interval of γ is:

$$\widehat{\Gamma} = \{\gamma : LR_n(\gamma) \leq c(\alpha)\},$$

where $c(\alpha)$ is the critical value from the X_1^2 distribution, such that $P(X > c(\alpha)) = \alpha$ for a random variable $X \sim X_1^2$.

2.3 The Threshold Regression model with two threshold variables

The threshold regression model with two threshold variables assumes linear relationship among the given data (y_i, x_i) which is expressed by the respective linear equation depending on the values of two threshold variables. The classification of the observations depends on the threshold parameters. If we choose two threshold variables and one threshold parameter, then the data y_i are classified into three regimes, while if there are two threshold variables and two threshold parameters the data y_i are classified into four regimes. We will examine the latter case, hence the following model with two threshold variables is the one classifying the observations y_i into four regimes.

$$y_i = \begin{cases} b_0^{(1)} + b_1^{(1)}x_{i1} + \cdots + b_{p_1}^{(1)}x_{ip_1} + e_i, & \text{if } q_{i1} \leq \gamma_1, q_{i2} \leq \gamma_2 \\ b_0^{(2)} + b_1^{(2)}x_{i1} + \cdots + b_{p_2}^{(2)}x_{ip_2} + e_i, & \text{if } q_{i1} \leq \gamma_1, q_{i2} > \gamma_2 \\ b_0^{(3)} + b_1^{(3)}x_{i1} + \cdots + b_{p_3}^{(3)}x_{ip_3} + e_i, & \text{if } q_{i1} > \gamma_1, q_{i2} \leq \gamma_2 \\ b_0^{(4)} + b_1^{(4)}x_{i1} + \cdots + b_{p_4}^{(4)}x_{ip_4} + e_i, & \text{otherwise} \end{cases} \quad (2.15)$$

where,

$q_i = (q_{i1}, q_{i2})$ are the threshold variables,

$\gamma = (\gamma_1, \gamma_2) \in \Gamma, \Gamma = [\underline{\gamma}_1, \overline{\gamma}_1] \times [\underline{\gamma}_2, \overline{\gamma}_2]$ is a bounded subset of the support of q_i ,

p_j , for $j=1,2,3,4$ is the order in each regime,

and $b^{(j)} = (b_0^{(j)}, b_1^{(j)}, \dots, b_{p_j}^{(j)})$ are the structural parameters.

As in the previous section, where there was only one threshold variable, we set for simplicity $p = \max(p_1, p_2, p_3, p_4)$ and $b_i^{(j)} = 0$ for $i > p_j$, for each $j=1,2,3,4$. Given the sample (y_i, x_i, q_i) for $i = 1, \dots, n$ our aim is to estimate the threshold variables γ_1, γ_2 and the slope parameters for each regime. The model can be written in a unique equation as:

$$y_i = \sum_{j=1}^4 d_i^{(j)}(\gamma) [b_0^{(j)} + \sum_{k=1}^p b_k^{(j)} x_{ik} + e_i], \text{ for } i = 1, 2, \dots, n, \quad (2.16)$$

where $d_i^{(j)}(\gamma)$ is an indicator function denoting one of the four regimes. More specifically, we have:

$$d_i^{(1)}(\gamma) = 1_{(q_{i1} \leq \gamma_1, q_{i2} \leq \gamma_2)}$$

$$d_i^{(2)}(\gamma) = 1_{(q_{i1} \leq \gamma_1, q_{i2} > \gamma_2)}$$

$$d_i^{(3)}(\gamma) = 1_{(q_{i1} > \gamma_1, q_{i2} \leq \gamma_2)}$$

$$d_i^{(4)}(\gamma) = 1_{(q_{i1} > \gamma_1, q_{i2} > \gamma_2)}.$$

It is also convenient to rewrite model (2.12) in matrix form, as follows.

$$Y = \sum_{j=1}^4 D^{(j)}(\gamma) X b^{(j)} + e, \quad (2.17)$$

where $X = (x_1^*, x_2^*, \dots, x_n^*)'$ is the $n \times (p+1)$ design matrix, $x_i^* = (1, x_{i1}, \dots, x_{ip})'$ for $i = 1, 2, \dots, n$, $D^{(j)}(\gamma) = \text{diag}(d_1^{(j)}(\gamma), \dots, d_n^{(j)}(\gamma))$, for $j=1,2,3,4$, $Y = (y_1, y_2, \dots, y_n)'$ and $e = (e_1, e_2, \dots, e_n)'$.

For simplicity, we set $D^{(j)}(\gamma)X = X_\gamma^{(j)}$.

2.3.1 Estimation

Given $\gamma = (\gamma_1, \gamma_2)$, model (2.13) is linear and the conditional OLS estimator for $b^{(j)}$ is:

$$\hat{b}^{(j)}(\gamma) = (X_\gamma^{(j)'} X_\gamma^{(j)})^{-1} X_\gamma^{(j)'} Y. \quad (2.18)$$

Now, the sum of squared errors becomes

$$\begin{aligned} s_n(\gamma) &= s_n(\widehat{b}^{(1)}(\gamma), \widehat{b}^{(2)}(\gamma), \widehat{b}^{(3)}(\gamma), \widehat{b}^{(4)}(\gamma), \gamma) \\ &= (Y - \sum_{j=1}^4 X_\gamma^{(j)} \widehat{b}^{(j)}(\gamma))' (Y - \sum_{j=1}^4 X_\gamma^{(j)} \widehat{b}^{(j)}(\gamma)) \end{aligned} \quad (2.19)$$

and $\widehat{\gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2)$ is the value that minimizes (2.15).

That is,

$$\widehat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} s_n(\gamma).$$

Finally, the slope parameters' estimators are $\widehat{b}^{(j)} = \widehat{b}^{(j)}(\widehat{\gamma})$. The estimators obtained through this procedure, are unbiased and consistent for the model's parameters.

2.3.2 Distributions of Estimators

In order to derive the asymptotic distribution of $\widehat{\gamma}$, Assumptions B from subsection 2.2.2 have to be generalized for the two-threshold variable model case. Assumptions (B1), (B2) and (B3) remain unchanged, with F denoting the joint distribution function of γ . (B4) requires a bounded joint density function of γ and moreover we assume that $0 < f_i(\gamma) \leq \bar{f}_i$, for $i=1,2$ and $f_i(\gamma) = \frac{\partial F(\gamma)}{\partial \gamma_i}$. (B5) is directly generalized for $\gamma = (\gamma_1, \gamma_2)$ and $\gamma_0 = (\gamma_1^0, \gamma_2^0)$. Assumption (B6) in the case of one threshold variable implies that as the sample size increases, the difference of the slope parameters increases. In our case this means that $\delta_n = (\delta^{(2)'}, \delta^{(3)'}, \delta^{(4)'}) = cn^{-\alpha} = (c'_2, c'_3, c'_4)n^{-\alpha}$, where $0 < \alpha < \frac{1}{2}$, c is a $3p$ -dimensional vector and $\delta^{(j)} = b^{(j)} - b^{(1)}$. For (B7) and (B8), we set $M_j(\gamma) = E[x_i x_i' d_i^{(j)}(\gamma)]$ and $d_1 = (c'_2 - c'_4, c'_3)$, $d_2 = (c'_2, c'_3 - c'_4)$. We assume that $M > M_j(\gamma) > 0$ and $d_1' D d_1 > 0$, $d_2' D d_2 > 0$.

Asymptotic distribution of the threshold estimate

Theorem 4 *Under the previous generalized assumptions, it is proven that:*

$$n^{1-2\alpha}((\widehat{\gamma}_1 - \gamma_1^0), (\widehat{\gamma}_2 - \gamma_2^0)) \xrightarrow{d} \omega T \quad (2.20)$$

where,

$$\omega = \left(\frac{d_1' V d_1}{d_1' D d_1 f_1}, \frac{d_2' V d_2}{d_2' D d_2 f_2} \right),$$

$$T = \operatorname{argmax}_{r_1 \in R, r_2 \in R} \left[-\frac{1}{2} + W_1(r_1) - \frac{1}{2} + W_2(r_2) \right],$$

$W_i(r_i)$ is a two-sided Brownian motion on the real line.

Note here, that in case that the homoskedasticity of the error terms holds, $d_i' V d_i = \sigma^2$, respectively to the one threshold case.

Confidence interval of the threshold estimate

The construction of the confidence interval for γ is accomplished utilizing the likelihood ratio statistic $LR_n(\gamma)$, in accordance with the one threshold regression case. Suppose that we aim to test the hypothesis:

$$H_0 : \gamma = \gamma_0 = (\gamma_1^{(0)}, \gamma_2^{(0)}).$$

Under the assumption of i.i.d $e_i \sim N(0, \sigma^2)$, the LR statistic is:

$$LR_n(\gamma) = n \frac{s_n(\gamma) - s_n(\hat{\gamma})}{s_n(\hat{\gamma})}.$$

The null hypothesis is rejected for large values of $LR_n(\gamma_0)$.

Theorem 5 Under the generalized assumptions B,

$$LR_n(\gamma_0) \rightarrow \eta^2 \zeta \quad (2.21)$$

where $\zeta = \zeta_1 + \zeta_2$ and $\zeta_i = \max_{s_i \in R} [2W(s_i) - |s_i|]$, $i = 1, 2$, for $\eta^2 = \frac{c'Vc}{\sigma^2 c'Dc}$.

Since the error terms are supposed to be homoskedastic, $\eta^2 = 1$ and the distribution of $LR_n(\gamma)$ depends on the distribution of ζ . The distribution of ζ_i , $i = 1, 2$ is $P(\zeta_i \leq x) = (1 - e^{-x/2})^2$ and $f_{\zeta_i}(x) = (1 - e^{-x/2}) e^{-x/2}$. Thus,

$$\begin{aligned} P(\zeta \leq x) &= P(\zeta_1 + \zeta_2 \leq x) \\ &= \int_0^x P(\zeta_1 \leq x - y) f_{\zeta_2}(y) dy \\ &= 1 - (x + 5)e^{-x} - 2(x - 2)e^{-x/2}. \end{aligned}$$

Given the distribution of ζ , we can get the asymptotic p-value for observed test statistics. The test's p-value under H_0 is:

$$p_n = P(\zeta > LR_n(\gamma_0)) = 1 - \left[1 - (LR_n(\gamma_0) + 5)e^{-LR_n(\gamma_0)} - 2(LR_n(\gamma_0) - 2)e^{-LR_n(\gamma_0)/2} \right],$$

and the null hypothesis is rejected at a level of significance α for $p_n > c_{\zeta}(\alpha)$, where $c_{\zeta}(\alpha)$, $c_1(\alpha)$ from now on, is the α -critical value of the distribution of ζ , namely $P(\zeta > c_1(\alpha)) = \alpha$. In contrast to the one threshold regression model, where the critical values are calculated through inversion, in this case we solve the critical values by simulations and the results are summarized in the following table.

Table 2.2: Asymptotic Critical Values

	0.80	0.85	0.90	0.925	0.95	0.975	0.99
$P(\zeta \leq x)$	8.33	9.13	10.21	10.96	11.98	13.68	15.85

Thus, the confidence interval of γ is $\hat{\Gamma} = \{\gamma \in \Gamma : LR_n(\gamma) \leq c(\alpha)\}$.

Asymptotic distribution of the slope parameters

Regarding the slope parameters, in order to derive the asymptotic distribution we write model (14) as:

$$Y = Xb^{(1)} + \sum_{j=2}^4 X_{\gamma}^{(j)} \delta^{(j)} + e, \quad (2.22)$$

where $\delta^{(j)} = b^{(j)} - b^{(1)}$. Now the analysis for the asymptotic distribution and confidence intervals for $b^{(j)}$ is the same as the one used in subsection 2.2.2.

2.4 Model comparison

Apart from statistical inference, another interesting problem in statistics is model comparison. This problem involves comparing a number of competing models in order to decide which one is the most appropriate for a specific dataset. In the context of threshold regression models it is of interest to determine the number of threshold variables. Firstly, we consider the null hypothesis of no threshold effect against the alternative that there is one threshold variable:

$$H_0 : m = 0$$

$$H_1 : m = 1.$$

Under the null hypothesis, there is one regime only, which means that we are in the case of the multiple linear model described in subsection 2.1, while under H_1 we are in the case of the threshold regression model with one threshold variable that can be either γ_1 or γ_2 . In order to test H_0 we define a likelihood ratio statistic as:

$$LR_n^i = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2(\hat{\gamma}_i)}{\hat{\sigma}^2(\hat{\gamma}_i)} \quad i = 1, 2, \quad (2.23)$$

where $\tilde{\sigma}^2$ is the estimator of the error variance σ^2 under H_0 and $\hat{\sigma}^2(\hat{\gamma}_i)$ is the estimator of error variance from the regression with γ_i being the only threshold variable. Rejecting the null hypothesis, we have evidence for the existence of more than one regimes. Since under H_0 γ_i is not identified, the asymptotic distribution of LR_n^i is not a standard chi-square density. The solution to that problem was given from Hansen (1996), who suggested a bootstrap method to approximate the LR_n^i asymptotic distribution. Next follows his recommended algorithm, which tests for the existence of the threshold variable q_{i1} :

Algorithm 1.

1. Generate iid $u_i \sim N(0, 1)$ for $i = 1, 2, \dots, n$.
2. Set $y_i^* = \tilde{e}_i u_i$, where \tilde{e}_i are the ordinary least-squares (OLS) residuals from the multiple linear regression model.
3. Using the sample (y_i^*, x_i, q_{1i}) estimate the linear regression model (2.2) and the one-threshold regression model (2.5) (with γ_1 being the threshold variable) and get $\tilde{\sigma}^{*2}$ and $\hat{\sigma}^{*2}(\hat{\gamma}_1)$.
4. Compute $LR_n^{1*} = n \frac{\tilde{\sigma}^{*2} - \hat{\sigma}^{*2}(\hat{\gamma}_1)}{\hat{\sigma}^{*2}(\hat{\gamma}_1)}$.
5. Repeat this procedure N times and get a sample of $(LR_n^{1*}(1), \dots, LR_n^{1*}(N))$.
6. The percentage of draws that LR_n^{1*} is greater than the real value LR_n^1 is the bootstrap p-value p_n^1 .
7. Reject H_0 in favor of H_1 at significance level α if $p_n^1 < \alpha$.

We can then compute the bootstrap p-value p_n^2 in order to compare the linear model with the threshold regression model with one threshold variable, q_{i2} .

If the null hypothesis can not be rejected for both γ_1 and γ_2 there is evidence for the existence of just one regime. On the contrary, rejecting the null hypothesis for at least one of the threshold variables, the presence of threshold effects is implied. The next question is what is the number of threshold variables. To answer this question we have now to test the hypotheses:

$$H_0 : m = 1$$

$$H_1 : m = 2$$

In essence, the above set of hypotheses corresponds to tests comparing the model with one threshold variable, that it could be either q_{i1} or q_{i2} , with the two threshold regression model. For each one of them we will use an LR statistic, namely:

$$LR_{ni} = n \frac{\tilde{\sigma}^2(\hat{\gamma}_i) - \hat{\sigma}_1^2(\hat{\gamma}_1, \hat{\gamma}_2)}{\hat{\sigma}^2(\hat{\gamma}_1, \hat{\gamma}_2)} \quad i = 1, 2 \quad (2.24)$$

where $\tilde{\sigma}^2(\hat{\gamma}_i)$ is the error variance estimator under the γ_i -threshold regression model and $\hat{\sigma}^2(\hat{\gamma}_1, \hat{\gamma}_2)$ is the estimator of the error variance under model (2.11). The p-value of the test is obtained using bootstrap as proposed by Hansen (1996). As presented by ... (2012) and ... (2014) the bootstrap algorithm is as follows, where for simplicity reasons we present the algorithm for the comparison between the γ_1 -threshold regression model and the model with both threshold variables.

Algorithm 2.

1. Estimate the two-threshold regression model, take its residuals and draw the bootstrap e_i^* residuals from them.
2. Set $y_i^* = x_i^{*'} \hat{b} + x_i^{*'} (\hat{\gamma}_1)' \hat{\delta}_n + e_i^*$ where \hat{b} , $\hat{\delta}_n$ and $\hat{\gamma}_1$ are the parameter estimators from γ_1 -threshold regression model.
3. Using the sample $(y_i^*, x_i, q_{1i}, q_{2i})$ estimate γ_1 -threshold regression model and model (2.11) and get $\tilde{\sigma}^{*2}(\hat{\gamma}_1)$ and $\hat{\sigma}^{*2}(\hat{\gamma}_1, \hat{\gamma}_2)$.
4. Compute $LR_{n1}^* = n \frac{\tilde{\sigma}^{*2}(\hat{\gamma}_1) - \hat{\sigma}^{*2}(\hat{\gamma}_1, \hat{\gamma}_2)}{\hat{\sigma}^{*2}(\hat{\gamma}_1, \hat{\gamma}_2)}$.
5. Repeat N times and get a sample $(LR_{n1}^*(1), \dots, LR_{n1}^*(N))$.
6. The percentage of draws that LR_{n1}^* is greater than the real value LR_{n1} , is the bootstrap p-value.

Rejection of both null hypotheses leads to the conclusion that there are two threshold variables. Regarding γ_1 , if we reject the null in the first step and not in the second, then γ_1 is the only threshold variable of the model, while regarding γ_2 a similar argument is applied. Note here that we may face the problem of rejecting the null hypothesis in the first step but accepting it in the second for both γ_1 and γ_2 . Although it should not occur in large samples, in case of smaller samples when such a problem appears we choose the threshold variable that better fits on the data. This could be done for example using the value of R_{adj}^2 or some other test of goodness of fit.

Further details on the Bootstrap can be found in Appendix A.

Chapter 3

Bayesian Inference

The fundamental difference between classical and Bayesian inference lies on how the unknown parameters are treated. In classical inference, unknown parameters are considered as constants, while in Bayesian inference as random variables. Let θ denote the unknown parameter, which we wish to make inferences about, and $f(x|\theta)$ is the likelihood model of our data x . After specifying a prior distribution for θ , which reflects our beliefs about it, using Bayes' theorem we can calculate the posterior density function $f(\theta|x)$. This posterior distribution of θ is in essence the inference. However, sometimes it is desirable to summarize our conclusions about θ using a point estimate, or a credibility area. Before analyzing the three models presented in the previous section from the scope of Bayesian inference, we remind Bayes' theorem.

Theorem 6 *Assuming that we have the likelihood model $f(x|\theta)$, with unknown parameters denoted by θ and a prior distribution for them, $f(\theta)$, Bayes' theorem takes the form:*

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} \quad (3.1)$$
$$\propto f(x|\theta)f(\theta).$$

3.1 The multiple linear regression model

As described in subsection 2.1 the multiple linear regression model has the form:

$$y_i = x_i^{*'}b + e_i, \quad (3.2)$$

where (y_i, x_i) for $i = 1, 2, \dots, n$ are the observed data, $x_i^* = (1, x_{i1}, x_{i2}, \dots, x_{ip})'$ and b is the $(p+1) \times 1$ vector of coefficients. In matrix form, model (3.2) can be equivalently written as:

$$Y = Xb + e. \quad (3.3)$$

Assuming that e_i are normally distributed with mean 0 and variance $\frac{1}{\phi}$, i.e

$$e_i \sim N\left(0, \frac{1}{\phi}\right),$$

then y_i are also normally distributed with mean $x_i^{*'}b$ and common variance $\frac{1}{\phi}$. Thus,

$$\underline{y} \sim N_n\left(Xb, \frac{1}{\phi}I_{n \times n}\right)$$

and the model's likelihood function is:

$$f(\underline{y}|b, \phi) = (2\pi)^{-n/2} \phi^{n/2} \exp\left\{-\frac{\phi}{2}(\underline{Y} - Xb)'(\underline{Y} - Xb)\right\}.$$

Regarding the prior distribution of the unknown parameters, it is a subjective matter which one to choose and is based on the prior information one may have about them. Different choices of priors lead to different posteriors, and therefore different conclusions. This is the disadvantage of Bayesian theory with respect to classical inference, according to the latter's advocates. However, as more and more data are obtained, posterior inference is mainly based on the likelihood model rather than the prior distribution. For the linear regression model, as prior distributions for the parameters we choose the following.

$$b|\phi \sim N_{p+1}\left(\mu, \frac{1}{\phi}C\right) \text{ and } \phi \sim \text{Gamma}(s, q).$$

Thus,

$$\begin{aligned} f(b, \phi) &= f(b|\phi)f(\phi) \\ &= \left(\frac{\phi}{2\pi}\right)^{\frac{p+1}{2}} |C|^{-1/2} \exp\left\{-\frac{\phi}{2}(b - \mu)'C^{-1}(b - \mu)\right\} \frac{\phi^{s-1} \exp\{-q\phi\} q^s}{\Gamma(s)}. \end{aligned}$$

In order to get the posterior density function up to a proportionality constant, we apply Bayes' theorem as it is presented in equation (3.1) and so we get:

$$\begin{aligned} f(b, \phi|\underline{y}) &\propto f(\underline{y}|b, \phi)f(b, \phi) \\ &= c \phi^{\frac{n+p+1}{2} + s - 1} \exp\left\{-\frac{\phi}{2}(\underline{Y} - Xb)'(\underline{Y} - Xb) - \frac{\phi}{2}(b - \mu)'C^{-1}(b - \mu) - q\phi\right\}, \end{aligned}$$

$$\text{where } c = \frac{|C|^{-1/2} q^s}{(2\pi)^{\frac{n+p+1}{2}} \Gamma(s)}.$$

Now that the joint posterior density function is available, we can also write down, up to a proportionality constant, the conditional density functions of b and ϕ . To do this, we consider the joint posterior as a function of the one parameter only, with the other being fixed. Hence,

$$\begin{aligned} f(b|\phi, \underline{y}) &\propto f(b, \phi|\underline{y}) \propto \exp\left\{-\frac{\phi}{2}(\underline{Y} - Xb)'(\underline{Y} - Xb) - \frac{\phi}{2}(b - \mu)'C^{-1}(b - \mu)\right\} \\ &\propto \exp\left\{-\frac{\phi}{2}[b'C_1^{-1}b - 2b'C_1^{-1}\mu_1]\right\} \\ &\equiv N_{p+1}\left(\mu_1, \frac{1}{\phi}C_1\right), \end{aligned}$$

where,

$$C_1 = (C^{-1} + X'X)^{-1} \text{ and } \mu_1 = C_1(C^{-1}\mu + X'\underline{Y}).$$

$$f(\phi|b, \underline{y}) \propto f(b, \phi|\underline{y})$$

$$\begin{aligned} &\propto \phi^{\frac{n+p+1}{2} + s - 1} \exp\left\{-\phi \left[\frac{(\underline{Y} - Xb)'(\underline{Y} - Xb) + (b - \mu)'C^{-1}(b - \mu)}{2} + q \right]\right\} \\ &\equiv \text{Gamma}\left(\frac{n+p+1}{2} + s, \frac{(\underline{Y} - Xb)'(\underline{Y} - Xb) + (b - \mu)'C^{-1}(b - \mu)}{2} + q\right). \end{aligned}$$

Although the conditional posterior densities belong to known distribution families, each of them depends on the other parameter and this prevents us from making inference separately. Exact Bayesian inference about each parameter can be made only if we have in hand the marginal posterior density of each parameter and this could be done by integrating the joint posterior distribution over the other parameter. The resulting function, i.e the marginal posterior of the parameters, can be used among others, either for point estimates or for the construction of credibility intervals. The marginal density function carries all the available information provided by the data for the specific parameter.

$$\begin{aligned}
f(\phi|\underline{y}) &= \int_{-\infty}^{+\infty} f(b, \phi|\underline{y}) db \\
&\propto c \phi^{\frac{n+p+1}{2}+s-1} \exp\left\{-q\phi - \frac{\phi}{2} Y'Y - \frac{\phi}{2} \mu' C^{-1} \mu\right\} \\
&\quad \times \int_{-\infty}^{+\infty} \exp\left\{-\frac{\phi}{2} [b' C_1^{-1} b - 2b' C_1^{-1} \mu_1]\right\} db \\
&= c \phi^{\frac{n+p+1}{2}+s-1} \exp\left\{-\left(q + \frac{Y'Y + \mu' C^{-1} \mu - \mu_1' C_1^{-1} \mu_1}{2}\right) \phi\right\} \\
&\quad \times \int_{-\infty}^{+\infty} \exp\left\{-\frac{\phi}{2} [b' C_1^{-1} b - 2b' C_1^{-1} \mu_1 + \mu_1' C_1^{-1} \mu_1]\right\} db \\
&= c \phi^{\frac{n+p+1}{2}+s-1} \exp\left\{-\left(q + \frac{Y'Y + \mu' C^{-1} \mu - \mu_1' C_1^{-1} \mu_1}{2}\right) \phi\right\} \\
&\quad \times (2\pi)^{\frac{p+1}{2}} \phi^{-\frac{p+1}{2}} |C_1|^{1/2} \\
&\propto \phi^{\frac{n}{2}+s-1} \exp\left\{-\left(q + \frac{Y'Y + \mu' C^{-1} \mu - \mu_1' C_1^{-1} \mu_1}{2}\right) \phi\right\} \\
&\equiv \text{Gamma}\left(\frac{n}{2} + s, q + \frac{Y'Y + \mu' C^{-1} \mu - \mu_1' C_1^{-1} \mu_1}{2}\right) \\
&\equiv \text{Gamma}(s^*, q^*)
\end{aligned}$$

$$\begin{aligned}
f(b|\underline{y}) &= \int_0^{+\infty} f(b, \phi|\underline{y})d\phi \\
&= \alpha c \int_0^{+\infty} \phi^{\frac{n+p+1}{2}+s-1} \exp\left\{-\phi \left[q + \frac{(Y-Xb)'(Y-Xb) + (b-\mu)'C^{-1}(b-\mu)}{2} \right]\right\} d\phi \\
&= c \frac{\Gamma\left(\frac{n+p+1}{2} + s\right)}{\left(q + \frac{(Y-Xb)'(Y-Xb) + (b-\mu)'C^{-1}(b-\mu)}{2}\right)^{\frac{n+p+1}{2}+s}} \\
&= \alpha \left(q + \frac{(Y-Xb)'(Y-Xb) + (b-\mu)'C^{-1}(b-\mu)}{2}\right)^{-\left(\frac{n+p+1}{2}+s\right)} \\
&= \left(q + \frac{(Y-Xb)'(Y-Xb) + (b-\mu)'C^{-1}(b-\mu)}{2}\right)^{-\frac{\nu+p+1}{2}} \\
&= \left(q + \frac{Y'Y + \mu'C^{-1}\mu - \mu_1'C_1^{-1}\mu_1}{2} + \frac{(b-\mu_1)'C_1^{-1}(b-\mu_1)}{2}\right)^{-\frac{\nu+p+1}{2}} \\
&= \left(q^* + \frac{(b-\mu_1)'C_1^{-1}(b-\mu_1)}{2}\right)^{-\frac{\nu+p+1}{2}} \alpha \left(1 + \frac{(b-\mu_1)'C_1^{-1}(b-\mu_1)}{2q^*}\right)^{-\frac{\nu+p+1}{2}} \\
&= \left(1 + \frac{\nu}{2q^*} \frac{(b-\mu_1)'C_1^{-1}(b-\mu_1)}{\nu}\right)^{-\frac{\nu+p+1}{2}} \\
&= \left(1 + \frac{(b-\mu_1)'\Sigma^{-1}(b-\mu_1)}{\nu}\right)^{-\frac{\nu+p+1}{2}}
\end{aligned}$$

which is equivalent to a multivariate t-student distribution and more specifically,

$$b|\underline{y} \sim t_\nu(\mu_1, \Sigma),$$

where $\nu = n + 2s = 2s^*$ and $\Sigma = \frac{2q^*}{\nu} C_1 = \frac{q^*}{s^*} C_1$.

3.2 The Threshold Regression model with one threshold variable

We remind that the threshold regression model with one threshold variable is of the form:

$$y_i = \begin{cases} b_0^{(1)} + b_1^{(1)}x_{i1} + b_2^{(1)}x_{i2} + \dots + b_{m_1}^{(1)}x_{ip_1} + e_i, & \text{if } q_i \leq \gamma \\ b_0^{(2)} + b_1^{(2)}x_{i1} + b_2^{(2)}x_{i2} + \dots + b_{m_2}^{(2)}x_{ip_2} + e_i, & \text{otherwise} \end{cases}$$

or in matrix form:

$$Y = Xb + X_\gamma \delta_n + e = X_\gamma^* \underline{b} + e, \quad (3.4)$$

where, $X_\gamma^* = [XX_\gamma]$ and $\underline{b} = [b' \delta_n']'$. We will also keep the notation for that model the same as in subsection (2.2). Under this agreement and the fact that the error terms

distribution is multivariate normal with mean $E(e) = 0_n$ and covariance matrix $\Omega = \frac{1}{\phi} I_{n \times n}$, i.e

$$e \sim N_n(0_n, \frac{1}{\phi} I_{n \times n}),$$

we can write the likelihood function of the model:

$$f(\underline{y}|\underline{b}, \phi, \gamma) = \left(\frac{\phi}{2\pi}\right)^{n/2} \exp\left\{-\frac{\phi}{2}(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b})\right\}.$$

As noted in subsection 2.2 the prior specification reflects our knowledge about the model parameters without having the sample in hand. Our model's parameters are \underline{b} , ϕ and γ and so we have to determine their joint prior density function. Regarding γ , the choice of its prior distribution depends on our choice of q_i , which means that if q_i is discrete the prior density of γ is also discrete, while if q_i is continuous so is the prior of γ . We assume a non-informative prior for γ , to reflect our ignorance about the threshold parameter, namely we choose the discrete uniform in the subset $\{q_1, q_2, \dots, q_{n-1}\}$, i.e

$$\gamma \sim DU(q_1, q_2, \dots, q_{n-1}).$$

We also assume that the prior distribution of \underline{b} and ϕ is independent of the prior of γ and choose the conjugate prior for (\underline{b}, ϕ) . That is,

$$\underline{b}|\phi \sim N_{2(p+1)}\left(a, \frac{1}{\phi}V\right) \text{ and } \phi \sim \text{Gamma}(s, q).$$

The joint prior density function is:

$$f(\underline{b}, \phi, \gamma) = \frac{1}{n-1} \left(\frac{\phi}{2\pi}\right)^{p+1} |V|^{-1/2} \exp\left\{-\frac{\phi}{2}(\underline{b} - a)'V^{-1}(\underline{b} - a)\right\} \frac{\phi^{s-1} \exp\{-q\phi\} q^s}{\Gamma(s)}.$$

We shall note at this point that in Bayesian inference there is no need for a discrimination between the case of a jump or a kink. The joint posterior density is :

$$\begin{aligned} f(\underline{b}, \phi, \gamma|\underline{y}) &\propto f(\underline{y}|\underline{b}, \phi, \gamma) f(\underline{b}, \phi, \gamma) \\ &= c_1 \phi^{\frac{n}{2} + p + s} \exp\left\{-\frac{\phi}{2}(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) - q\phi - \frac{\phi}{2}(\underline{b} - a)'V^{-1}(\underline{b} - a)\right\}, \end{aligned}$$

$$\text{where } c_1 = \frac{q^s |V|^{-1/2}}{(n-1)(2\pi)^{\frac{n}{2} + p + 1} \Gamma(s)}.$$

The conditional density function of each parameter can be found by considering the joint posterior as a function of the specific parameter only. Thus, we get for \underline{b} , ϕ and γ :

$$\begin{aligned} f(\phi|\underline{y}, \underline{b}, \gamma) &\propto f(\underline{b}, \phi, \gamma|\underline{y}) \\ &\propto \phi^{\frac{n}{2} + p + s} \exp\left\{-\left[q + \frac{(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) + (\underline{b} - a)'V^{-1}(\underline{b} - a)}{2}\right]\phi\right\}, \end{aligned}$$

which is proportional to a Gamma distribution with parameters $\Lambda = \frac{n}{2} + p + s + 1$ and $M = q + \frac{(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) + (\underline{b} - a)'V^{-1}(\underline{b} - a)}{2}$.

$$\begin{aligned}
f(\underline{b}|\underline{y}, \phi, \gamma) &\propto f(\underline{b}, \phi, \gamma|\underline{y}) \\
&\propto \exp\left\{-\frac{\phi}{2}[(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) + (\underline{b} - a)'V^{-1}(\underline{b} - a)]\right\} \\
&\propto \exp\left\{-\frac{\phi}{2}\exp\{\underline{b}'(V^{-1} + X_\gamma^{*'}X_\gamma^*)\underline{b} - 2\underline{b}'(V^{-1}a + X_\gamma^{*'}Y)\}\right\} \\
&= \exp\left\{-\frac{\phi}{2}\exp\{\underline{b}'V_1^{-1}\underline{b} - 2\underline{b}'V_1^{-1}V_1(V^{-1}a + X_\gamma^{*'}Y)\}\right\} \\
&\equiv N(a_1, \frac{1}{\phi}V_1),
\end{aligned}$$

where $V_1 = (V^{-1} + X_\gamma^{*'}X_\gamma^*)^{-1}$ and $a_1 = V_1(V^{-1}a + X_\gamma^{*'}Y)$.

$$\begin{aligned}
f(\gamma|\underline{y}, \phi, \underline{b}) &\propto f(\underline{b}, \phi, \gamma|\underline{y}) \\
&\propto \exp\left\{-\frac{\phi}{2}(\underline{b}'X_\gamma^{*'}X_\gamma^*\underline{b} - 2\underline{b}'X_\gamma^{*'}Y)\right\}.
\end{aligned}$$

The exact conditional density function of γ can be found by normalising the above expression, i.e

$$f(\gamma|\underline{y}, \phi, \underline{b}) = \frac{\exp\left\{-\frac{\phi}{2}(\underline{b}'X_\gamma^{*'}X_\gamma^*\underline{b} - 2\underline{b}'X_\gamma^{*'}Y)\right\}}{\sum_{\gamma \in \{q_1, q_2, \dots, q_{n-1}\}} \exp\left\{-\frac{\phi}{2}(\underline{b}'X_\gamma^{*'}X_\gamma^*\underline{b} - 2\underline{b}'X_\gamma^{*'}Y)\right\}}.$$

The computation of the marginal densities for \underline{b} and ϕ will be done in two steps. First we will find the posterior density for both \underline{b} and ϕ conditional on γ and the data and then we will integrate out γ .

$$\begin{aligned}
f(\phi|\underline{y}, \gamma) &= \int_{-\infty}^{+\infty} f(\underline{b}, \phi, \gamma|\underline{y}) d\underline{b} \\
&\propto c_1 \phi^{\frac{n}{2} + p + s} \exp\left\{-q\phi - \frac{\phi}{2}(Y'Y + a'V^{-1}a)\right\} \\
&\quad \times \int_{-\infty}^{+\infty} \exp\left\{-\frac{\phi}{2}(\underline{b}'V_1^{-1}\underline{b} - 2\underline{b}'V_1^{-1}a_1)\right\} d\underline{b} \\
&= c_1 \phi^{\frac{n}{2} + p + s} \exp\left\{-q\phi - \frac{\phi}{2}(Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1)\right\} \\
&\quad \times (2\pi)^{p+1} \phi^{-(p+1)} |V_1|^{1/2},
\end{aligned}$$

which is proportional to the density function of a $Gamma(s_1^*, q_1^*)$ distribution, where $s_1^* = \frac{n}{2} + s$ and $q_1^* = q + \frac{Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1}{2}$.

$$\begin{aligned}
f(\underline{b}|\underline{y}, \gamma) &= \int_0^{+\infty} f(\underline{b}, \phi, \gamma|\underline{y})d\phi \\
&\propto c_1 \int_0^{+\infty} \phi^{\Lambda-1} \exp\{-M\phi\}d\phi \\
&= c_1 \frac{\Gamma(\Lambda)}{M^\Lambda} \alpha \left[1 + \frac{n+2s}{2q_1^*} \frac{(\underline{b}-a_1)'V_1^{-1}(\underline{b}-a_1)}{n+2s} \right]^{-\frac{n+2s+2(p+1)}{2}} \\
&= \left[1 + \frac{(\underline{b}-a_1)'\Sigma_1^{-1}(\underline{b}-a_1)}{\nu} \right]^{-\frac{\nu+2(p+1)}{2}},
\end{aligned}$$

which is equivalent to the density function of a multivariate student-t distribution with $\nu = 2(\frac{n}{2} + s)$ degrees of freedom, mode a_1 and scale $\Sigma_1 = \frac{V_1 q_1^*}{s_1^*}$.

For the posterior of γ we will integrate the joint posterior density with respect to \underline{b} and ϕ . Thus,

$$\begin{aligned}
f(\gamma|\underline{y}) &= \int_0^{+\infty} \int_{-\infty}^{+\infty} f(\underline{b}, \phi, \gamma|\underline{y})d\underline{b}d\phi \\
&\propto c_1 \int_0^{+\infty} \phi^{\frac{n}{2}+p+s} \exp\{-q\phi\} \int_{-\infty}^{+\infty} \exp\left\{-\frac{\phi}{2} \left[(\underline{b}-a)'V^{-1}(\underline{b}-a) + (Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) \right]\right\} d\underline{b}d\phi \\
&= c_1 \int_0^{+\infty} \phi^{\frac{n}{2}+p+s} \exp\left\{-\left(q + \frac{Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1}{2}\right)\phi\right\} (2\pi)^{(p+1)} \phi^{-(p+1)} |V_1|^{1/2} d\phi \\
&= c_1 (2\pi)^{(p+1)} |V_1|^{1/2} \int_0^{+\infty} \phi^{s_1^*-1} \exp\{-q_1^* \phi\} d\phi = c_1^* |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^* s_1^*}.
\end{aligned}$$

In order to get the exact postrior density function of γ we need to find the normalising constant so that the posterior sums to one. Hence,

$$f(\gamma|\underline{y}) = \frac{|V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^* s_1^*}}{\sum_{\gamma \in \{q_1, q_2, \dots, q_{n-1}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^* s_1^*}}.$$

Finally, we get the marginal posterior densities of \underline{b} and ϕ as,

$$\begin{aligned}
f(\underline{b}|\underline{y}) &= \sum_{\gamma \in \{q_1, q_2, \dots, q_{n-1}\}} f(\underline{b}|\underline{y}, \gamma) f(\gamma|\underline{y}), \\
f(\phi|\underline{y}) &= \sum_{\gamma \in \{q_1, q_2, \dots, q_{n-1}\}} f(\phi|\underline{y}, \gamma) f(\gamma|\underline{y}).
\end{aligned}$$

3.3 The Threshold Regression model with two threshold variables

As presented in subsection 2.3 the threshold regression model with two threshold variables is

$$y_i = \begin{cases} b_0^{(1)} + b_1^{(1)} x_{i1} + \cdots + b_{p_1}^{(1)} x_{ip_1} + e_i, & \text{if } q_{i1} \leq \gamma_1, q_{i2} \leq \gamma_2, \\ b_0^{(2)} + b_1^{(2)} x_{i1} + \cdots + b_{p_2}^{(2)} x_{ip_2} + e_i, & \text{if } q_{i1} \leq \gamma_1, q_{i2} > \gamma_2, \\ b_0^{(3)} + b_1^{(3)} x_{i1} + \cdots + b_{p_3}^{(3)} x_{ip_3} + e_i, & \text{if } q_{i1} > \gamma_1, q_{i2} \leq \gamma_2, \\ b_0^{(4)} + b_1^{(4)} x_{i1} + \cdots + b_{p_4}^{(4)} x_{ip_4} + e_i, & \text{otherwise.} \end{cases} \quad (3.5)$$

Writing the above model in matrix form, the Bayesian analysis is similar to the case of one threshold variable, but in this case $\gamma = (\gamma_1, \gamma_2)$. Thus, in matrix form model (3.5) can be written as:

$$Y = \sum_{j=1}^4 D^{(j)}(\gamma) X b^{(j)} + e = \sum_{j=1}^4 X_{\gamma}^{(j)} b^{(j)} + e \quad (3.6)$$

where,

$X = (x_1^*, x_2^*, \dots, x_n^*)'$ is the $n \times (p+1)$ design matrix

$x_i^* = (1, x_{i1}, \dots, x_{ip})'$ for $i = 1, 2, \dots, n$

$D^{(j)}(\gamma) = \text{diag}(d_1^{(j)}(\gamma), \dots, d_n^{(j)}(\gamma))$, for $j=1,2,3,4$ and $\gamma = (\gamma_1, \gamma_2)$

$Y = (y_1, y_2, \dots, y_n)'$

and $e = (e_1, e_2, \dots, e_n)'$.

In order to simplify the analysis and keep up with the notation of the previous section, equation (3.6) can be also written as

$$Y = X_{\gamma}^* \underline{b} + e \quad (3.7)$$

where $X_{\gamma}^* = [X_{\gamma}^{(1)} \ X_{\gamma}^{(2)} \ X_{\gamma}^{(3)} \ X_{\gamma}^{(4)}]$ and $\underline{b} = [b^{(1)} \ b^{(2)} \ b^{(3)} \ b^{(4)}]$.

Despite the fact that there are two threshold variables instead of one, the specification of the likelihood function and the prior distributions is the same as in section 3.2. More specifically, assuming that the error terms are uncorrelated, having as distribution the multivariate normal with mean 0 and common variance $\frac{1}{\phi}$, namely

$$e \sim N_n(0_n, \frac{1}{\phi} I_{n \times n}),$$

the likelihood function of the model is

$$f(\underline{y} | \underline{b}, \phi, \gamma) = \left(\frac{\phi}{2\pi} \right)^{n/2} \exp\left\{ -\frac{\phi}{2} (Y - X_{\gamma}^* \underline{b})' (Y - X_{\gamma}^* \underline{b}) \right\}.$$

As priors we decide to choose the conditional conjugate ones for reasons of convenience. Regarding the threshold parameters we assume that they are a priori independent, each one having the discrete uniform distribution in the bounded subsets $\{q_{11}, q_{21}, \dots, q_{n-1,1}\}$ and $\{q_{12}, q_{22}, \dots, q_{n-1,2}\}$, i.e

$$P(\gamma_j = k_j) = \frac{1}{n-1} \text{ for } k_j \in \{q_{1j}, q_{2j}, \dots, q_{n-1,j}\} \text{ and } j = 1, 2.$$

We also assume that the prior distribution of \underline{b} and ϕ is independent of the prior of γ and

$$\underline{b} | \phi \sim N_{4(p+1)}(a, \frac{1}{\phi} V),$$

$$\phi \sim \text{Gamma}(s, q).$$

Therefore, the joint prior density function is:

$$f(\underline{b}, \phi, \gamma) = \frac{1}{(n-1)^2} \left(\frac{\phi}{2\pi}\right)^{2(p+1)} |V|^{-1/2} \exp\left\{-\frac{\phi}{2}(\underline{b}-a)'V^{-1}(\underline{b}-a)\right\} \frac{\phi^{s-1} \exp\{-q\phi\} q^s}{\Gamma(s)}.$$

The joint posterior density function is:

$$\begin{aligned} f(\underline{b}, \phi, \gamma | \underline{y}) &\propto f(\underline{y} | \underline{b}, \phi, \gamma) f(\underline{b}, \phi, \gamma) \\ &= c_1 \phi^{\frac{n}{2}+2p+s+1} \exp\left\{-\frac{\phi}{2}(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) - q\phi - \frac{\phi}{2}(\underline{b}-a)'V^{-1}(\underline{b}-a)\right\}, \end{aligned}$$

$$\text{where } c_1 = \frac{q^s |V|^{-1/2}}{(n-1)^2 (2\pi)^{\frac{n}{2}+2(p+1)} \Gamma(s)}.$$

The conditional density function of each parameter can be found by considering the joint posterior as a function of that specific parameter only. Thus, we have for \underline{b} , ϕ and γ :

$$\begin{aligned} f(\phi | \underline{y}, \underline{b}, \gamma) &\propto f(\underline{b}, \phi, \gamma | \underline{y}) \\ &\propto \phi^{\frac{n}{2}+2p+s+1} \exp\left\{-\left[q + \frac{(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) + (\underline{b}-a)'V^{-1}(\underline{b}-a)}{2}\right] \phi\right\} \\ &\equiv \text{Gamma}\left(\frac{n}{2} + 2p + s + 2, q + \frac{(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) + (\underline{b}-a)'V^{-1}(\underline{b}-a)}{2}\right) \\ &\equiv \text{Gamma}(\Lambda, M). \end{aligned}$$

$$\begin{aligned} f(\underline{b} | \underline{y}, \phi, \gamma) &\propto f(\underline{b}, \phi, \gamma | \underline{y}) \\ &\propto \exp\left\{-\frac{\phi}{2}[(Y - X_\gamma^* \underline{b})'(Y - X_\gamma^* \underline{b}) + (\underline{b}-a)'V^{-1}(\underline{b}-a)]\right\} \\ &\propto \exp\left\{-\frac{\phi}{2} \exp\{\underline{b}'(V^{-1} + X_\gamma^{*'} X_\gamma^*) \underline{b} - 2\underline{b}'(V^{-1}a + X_\gamma^{*'} Y)\}\right\} \\ &= \exp\left\{-\frac{\phi}{2} \exp\{\underline{b}' V_1^{-1} \underline{b} - 2\underline{b}' V_1^{-1} V_1 (V^{-1}a + X_\gamma^{*'} Y)\}\right\} \\ &\equiv N(a_1, \frac{1}{\phi} V_1) \end{aligned}$$

where $V_1 = (V^{-1} + X_\gamma^{*'} X_\gamma^*)^{-1}$ and $a_1 = V_1 (V^{-1}a + X_\gamma^{*'} Y)$.

$$\begin{aligned} f(\gamma | \underline{y}, \phi, \underline{b}) &\propto f(\underline{b}, \phi, \gamma | \underline{y}) \\ &\propto \exp\left\{-\frac{\phi}{2}(\underline{b}' X_\gamma^{*'} X_\gamma^* \underline{b} - 2\underline{b}' X_\gamma^{*'} Y)\right\}. \end{aligned}$$

Thus,

$$f(\gamma|\underline{y}, \phi, \underline{b}) = \frac{\exp\{-\frac{\phi}{2}(\underline{b}'X_\gamma^*X_\gamma^*\underline{b}-2\underline{b}'X_\gamma^*Y)\}}{\sum_{\gamma \in \{q_1, q_2, \dots, q_{n-1}\}} \exp\{-\frac{\phi}{2}(\underline{b}'X_\gamma^*X_\gamma^*\underline{b}-2\underline{b}'X_\gamma^*Y)\}}.$$

The computation of the marginal densities for \underline{b} and ϕ as in section 3.2 will be done in two steps. First we will find the posterior density for both \underline{b} and ϕ conditional on γ and the data and then we will integrate out γ .

$$\begin{aligned} f(\phi|\underline{y}, \gamma) &= \int_{-\infty}^{+\infty} f(\underline{b}, \phi, \gamma|\underline{y})d\underline{b} \\ &\propto c_1 \phi^{\frac{n}{2}+2p+s+1} \exp\{-q\phi - \frac{\phi}{2}(Y'Y + a'V^{-1}a)\} \\ &\quad \times \int_{-\infty}^{+\infty} \exp\{-\frac{\phi}{2}(\underline{b}'V_1^{-1}\underline{b} - 2\underline{b}'V_1^{-1}a_1)\}d\underline{b} \\ &= c_1 \phi^{\frac{n}{2}+2p+s+1} \exp\{-q\phi - \frac{\phi}{2}(Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1)\} \\ &\quad \times (2\pi)^{2(p+1)} \phi^{-2(p+1)} |V_1|^{1/2} \\ &\propto \text{Gamma}\left(\frac{n}{2} + s, q + \frac{Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1}{2}\right) \\ &\equiv \text{Gamma}(s_1^*, q_1^*). \end{aligned}$$

$$\begin{aligned} f(\underline{b}|\underline{y}, \gamma) &= \int_0^{+\infty} f(\underline{b}, \phi, \gamma|\underline{y})d\phi \\ &\propto c_1 \int_0^{+\infty} \phi^{\Lambda-1} \exp\{-M\phi\}d\phi \\ &= c_1 \frac{\Gamma(\Lambda)}{M^\Lambda} \propto \left[1 + \frac{n+2s}{2q_1^*} \frac{(\underline{b}-a_1)'V_1^{-1}(\underline{b}-a_1)}{n+2s}\right]^{-\frac{n+2s+4(p+1)}{2}} \\ &= \left[1 + \frac{(\underline{b}-a_1)'\Sigma_1^{-1}(\underline{b}-a_1)}{\nu}\right]^{-\frac{\nu+4(p+1)}{2}}, \end{aligned}$$

which is the density function of a multivariate student-t distribution with $\nu = 2(\frac{n}{2} + s)$ degrees of freedom, mode a_1 and scale $\Sigma_1 = \frac{V_1 q_1^*}{s_1^*}$.

For the posterior of γ we will integrate the joint posterior density with respect to \underline{b} and ϕ . Thus,

$$\begin{aligned} f(\gamma|\underline{y}) &= \int_0^{+\infty} \int_{-\infty}^{+\infty} f(\underline{b}, \phi, \gamma|\underline{y})d\underline{b}d\phi \\ &\propto c_1 \int_0^{+\infty} \phi^{\frac{n}{2}+2p+s+1} \exp\{-q\phi\} \exp\left\{-\left(\frac{Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1}{2}\right)\phi\right\} (2\pi)^{2(p+1)} \phi^{-2(p+1)} |V_1|^{1/2} \\ &= c_1 (2\pi)^{2(p+1)} |V_1|^{1/2} \int_0^{+\infty} \phi^{s_1^*-1} \exp\{-q_1^*\phi\}d\phi = c_1^* |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^* s_1^*}. \end{aligned}$$

Therefore, the exact posterior density function of γ_1 and γ_2 is

$$f(\gamma_1|\underline{y}) = \frac{\sum_{\gamma_2 \in \{q_{12}, q_{22}, \dots, q_{n-1,2}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}{\sum_{\gamma_1 \in \{q_{11}, q_{21}, \dots, q_{n-1,1}\}} \sum_{\gamma_2 \in \{q_{12}, q_{22}, \dots, q_{n-1,2}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}},$$

$$f(\gamma_2|\underline{y}) = \frac{\sum_{\gamma_1 \in \{q_{11}, q_{21}, \dots, q_{n-1,1}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}{\sum_{\gamma_1 \in \{q_{11}, q_{21}, \dots, q_{n-1,1}\}} \sum_{\gamma_2 \in \{q_{12}, q_{22}, \dots, q_{n-1,2}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}.$$

The marginal posterior densities of \underline{b} and ϕ are respectively,

$$f(\underline{b}|\underline{y}) = \sum_{\gamma_1 \in \{q_{11}, q_{21}, \dots, q_{n-1,1}\}} \sum_{\gamma_2 \in \{q_{12}, q_{22}, \dots, q_{n-1,2}\}} f(\underline{b}|\underline{y}, \gamma) f(\gamma|\underline{y}),$$

and

$$f(\phi|\underline{y}) = \sum_{\gamma_1 \in \{q_{11}, q_{21}, \dots, q_{n-1,1}\}} \sum_{\gamma_2 \in \{q_{12}, q_{22}, \dots, q_{n-1,2}\}} f(\phi|\underline{y}, \gamma) f(\gamma|\underline{y}).$$

3.4 Bayesian Model Comparison

After summarizing Bayesian inference for the multiple linear regression model and the threshold regression models with one and two threshold variables, we may be interested in which one of them best fits the data. This means that the three models have to be compared in order to see which one is the most probable, or in terms of Bayesian inference which one is the most possible to have generated the observed data. Note that, under the Bayesian perspective, all competing models are compared simultaneously. The competing models in our case are:

$$\begin{aligned} M_1 : y_i &= x_i^{*'} b + e_i \quad \forall i = 1, 2, \dots, n, \\ M_2 : y_i &= x_{i\gamma}^{*'} b + e_i \quad \forall i = 1, 2, \dots, n, \\ M_3 : y_i &= \sum_{j=1}^4 x_{i\gamma}^{*(j)'} b^{(j)} + e_i \quad \forall i = 1, 2, \dots, n. \end{aligned}$$

Firstly, we need to specify prior model probabilities and then obtain the posterior probability of each model. The model with the highest posterior probability, is the most preferable. Since we have no prior information about the three competing models we consider them to be equally probable a-priori, that is we set $P(M_1) = P(M_2) = P(M_3) = \frac{1}{3}$. The posterior probability of model M_j is calculated using Bayes theorem as follows:

$$P(M_j|\underline{y}) = \frac{P(M_j)f(\underline{y}|M_j)}{P(M_1)f(\underline{y}|M_1) + P(M_2)f(\underline{y}|M_2) + P(M_3)f(\underline{y}|M_3)}, \quad j = 1, 2, 3,$$

where $P(M_1|\underline{y}) + P(M_2|\underline{y}) + P(M_3|\underline{y}) = 1$ and $f(\underline{y}|M_j)$ is the marginal likelihood of model M_j . The marginal likelihood or evidence of a given model is obtained by integrating the product of the likelihood times the model parameters' prior distribution

over the models' parameters. Therefore for our three competing models we have:

Evidence of M_1

$$\begin{aligned} f(\underline{y}) &= \int_0^{+\infty} \int_{-\infty}^{+\infty} f(b, \phi) f(\underline{y}|b, \phi) db d\phi \\ &= c \int_0^{+\infty} \int_{-\infty}^{+\infty} \phi^{\frac{n+p+1}{2}+s-1} \exp\left\{-\left[\frac{(Y-Xb)'(Y-Xb) + (b-\mu)'C^{-1}(b-\mu)}{2} + q\right] \phi\right\} db d\phi. \end{aligned}$$

Since we have already integrated the joint posterior density function (which is proportional to the product of the prior times the likelihood function) with respect to b , we have that:

$$\int_{-\infty}^{+\infty} f(b, \phi) f(\underline{y}|b, \phi) db = c^* \phi^{\frac{n}{2}+s-1} \exp\left\{-\left(q + \frac{Y'Y + \mu'C^{-1}\mu - \mu_1'C_1^{-1}\mu_1}{2}\right) \phi\right\},$$

where $c^* = c(2\pi)^{\frac{p+1}{2}} |C_1|^{1/2}$. Thus,

$$\begin{aligned} f(\underline{y}|M_1) &= c^* \int_0^{+\infty} \phi^{\frac{n}{2}+s-1} \exp\left\{-\left(q + \frac{Y'Y + \mu'C^{-1}\mu - \mu_1'C_1^{-1}\mu_1}{2}\right) \phi\right\} d\phi \\ &= c^* \int_0^{+\infty} \phi^{s^*-1} \exp\{-q^* \phi\} d\phi \\ &= c^* \frac{\Gamma(s^*)}{q^{*s^*}}. \end{aligned}$$

Evidence of M_2

$$\begin{aligned} f(\underline{y}|M_2) &= \sum_{\gamma \in \{q_1, q_2, \dots, q_{n-1}\}} \int_0^{+\infty} \int_{-\infty}^{+\infty} f(b, \phi, \gamma|\underline{y}) db d\phi \\ &= \sum_{\gamma \in \{q_1, q_2, \dots, q_{n-1}\}} c_1^* |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}, \end{aligned}$$

where

$$\begin{aligned} s_1^* &= \frac{n}{2} + s \text{ and } q_1^* = q + \frac{Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1}{2}, \\ V_1 &= (V^{-1} + X_\gamma^{*'} X_\gamma^*)^{-1}, a_1 = V_1(V^{-1}a + X_\gamma^{*'} Y), \\ c_1^* &= c_1(2\pi)^{(p+1)}, c_1 = \frac{q^s |V|^{-1/2}}{(2\pi)^{\frac{n+2(p+1)}{2}} \Gamma(s)n}. \end{aligned}$$

Evidence of M_3

$$\begin{aligned} f(\underline{y}|M_3) &= \sum_{\gamma_1 \in \{q_{11}, q_{21}, \dots, q_{n-1,1}\}} \sum_{\gamma \in \{q_{12}, q_{22}, \dots, q_{n-1,2}\}} \int_0^{+\infty} \int_{-\infty}^{+\infty} f(b, \phi, \gamma|\underline{y}) db d\phi \\ &= \sum_{\gamma_1 \in \{q_{11}, q_{21}, \dots, q_{n-1,1}\}} \sum_{\gamma \in \{q_{12}, q_{22}, \dots, q_{n-1,2}\}} c_1^* |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}, \end{aligned}$$

where

$$\begin{aligned} s_1^* &= \frac{n}{2} + s \text{ and } q_1^* = q + \frac{Y'Y + a'V^{-1}a - a_1'V_1^{-1}a_1}{2}, \\ V_1 &= (V^{-1} + X_\gamma^{*'} X_\gamma^*)^{-1}, a_1 = V_1(V^{-1}a + X_\gamma^{*'} Y), \\ c_1^* &= c_1(2\pi)^{2(p+1)}, c_1 = \frac{q^s |V|^{-1/2}}{(2\pi)^{\frac{n+4(p+1)}{2}} \Gamma(s)n}. \end{aligned}$$

Chapter 4

Simulation Experiments

In this section, we design a number of simulation experiments with the aim to assess the performance of both the classical and the Bayesian approach to identify the correct data generating process (DGP) among the competing models. Specifically, we consider four competing models:

$$\begin{aligned}M_1 : y_i &= x_i^{*'} b + e_i \\M_2 : y_i &= x_i^{*'} b_2 + x_i^* (\gamma_1)' \delta_n + e_i, \\M_3 : y_i &= x_i^{*'} b_2 + x_i^* (\gamma_2)' \delta_n + e_i, \\M_4 : y_i &= \sum_{j=1}^4 d_i^{(j)} x_i^{*'} b^{(j)} + e_i,\end{aligned}$$

where,

$$x_i^* (\gamma_1) = x_i^* \mathbf{1}_{q_{i1} \leq \gamma_1}, \quad x_i^* (\gamma_2) = x_i^* \mathbf{1}_{q_{i2} \leq \gamma_2} \quad \text{and} \quad \delta_n = b_1 - b_2$$

$$d_i^{(1)} = \mathbf{1}_{(q_{i1} \leq \gamma_1, q_{i2} \leq \gamma_2)},$$

$$d_i^{(2)} = \mathbf{1}_{(q_{i1} \leq \gamma_1, q_{i2} > \gamma_2)},$$

$$d_i^{(3)} = \mathbf{1}_{(q_{i1} > \gamma_1, q_{i2} \leq \gamma_2)},$$

$$d_i^{(4)} = \mathbf{1}_{(q_{i1} > \gamma_1, q_{i2} > \gamma_2)}.$$

Each of the above competing models is considered in turn as the true DGP from which a data set is simulated. Because models M_2 and M_3 belong to the same category of the regression models with one threshold variable, we will consider only M_3 as a DGP. In each case, we first compare the four models from the scope of classical theory in order to decide which one fits the data best. For this model, we then proceed to the estimation of its parameters. Then, for the same simulated data we perform Bayesian model selection and Bayesian inference for the model with the highest posterior probability. Finally, we comment on the comparison between classical and Bayesian results. Note that in this chapter we focus on the threshold regression model with a jump.

In the following three subsections we will consider the same prior distributions for each model's parameters and thus, it is more convenient to define them once. We also consider the four competing models a priori equally in order to express the lack of information, i.e $P(M_1) = P(M_2) = P(M_3) = P(M_4) = \frac{1}{4}$.

Prior distributions for the parameters of model M_1

$$\begin{aligned}\phi &\sim \text{Gamma}(1, 1) \\ b|\phi &\sim N\left(0_{2 \times 1}, \frac{1}{\phi} I_{2 \times 2}\right)\end{aligned}$$

Prior distributions for the parameters of model M_2

$$\begin{aligned}\phi &\sim \text{Gamma}(1, 1) \\ \underline{b}|\phi &\sim N\left(0_{4 \times 1}, \frac{1}{\phi} I_{4 \times 4}\right) \\ \gamma_1 &\sim \text{DU}(q_{(10,1)}, q_{(15,1)}, \dots, q_{(90,1)})\end{aligned}$$

Prior distributions for the parameters of model M_3

$$\begin{aligned}\phi &\sim \text{Gamma}(1, 1) \\ \underline{b}|\phi &\sim N\left(0_{4 \times 1}, \frac{1}{\phi} I_{4 \times 4}\right) \\ \gamma_2 &\sim \text{DU}(q_{(10,2)}, q_{(15,2)}, \dots, q_{(90,2)})\end{aligned}$$

Prior distributions for the parameters of model M_4

$$\begin{aligned}\phi &\sim \text{Gamma}(1, 1) \\ \underline{b}|\phi &\sim N\left(0_{8 \times 1}, \frac{1}{\phi} I_{8 \times 8}\right) \\ \gamma_1 &\sim \text{DU}(q_{(10,1)}, q_{(15,1)}, \dots, q_{(90,1)}) \\ \gamma_2 &\sim \text{DU}(q_{(10,2)}, q_{(15,2)}, \dots, q_{(90,2)})\end{aligned}$$

Note that in each case we have a priori assumed the coefficients to be uncorrelated.

4.1 Model M_1 as DGP

Considering M_1 as the data generating process, we simulate $n=200$ i.i.d observations from the univariate normal distribution $N(0, 1)$, in order to create the design matrix $X = (x_1^* \ x_2^* \ \dots \ x_n^*)'$, where $x_i^* = (1 \ x_i)'$ and $x_i \sim N(0, 1)$ for $i = 1, 2, \dots, n$. We also set the vector of the coefficients $b = (1 \ 2)'$ and $e_i \sim N(0, 1)$. Therefore, we compute the values of the dependent variable y_i according to M_1 as follows:

$$y_i = x_i^{*'} b + e_i.$$

Regarding the threshold variables we simulate a sample of $n = 200$ observations for each one of them, assuming that $q_{i1} \sim N(0, 1)$, $q_{i2} \sim N(2, 1)$. We also assume that $(\gamma_1, \gamma_2) \in \bar{\Gamma}_1 \times \bar{\Gamma}_2$, where $\bar{\Gamma}_j = \{q_{j(10)}, q_{j(15)}, q_{j(20)}, \dots, q_{j(90)}\}$ and $q_{j(k)}$ is the k -th percentile of q_{ij} for $i = 1, 2, \dots, n, j = 1, 2$.

4.1.1 Classical Inference

Under the previous assumptions and given the simulated data we will start by comparing the four models, that is testing for the existence of one or more threshold parameters. In terms of hypotheses testing, we start by testing the null hypothesis of the existence of one threshold variable that could be γ_1 or γ_2 against the alternative of two threshold parameters. Hence,

$$\begin{aligned} H_0 : m &= 1 (\gamma_1) \\ H_1 : m &= 2 (\gamma_1), (\gamma_2) \end{aligned}$$

and

$$\begin{aligned} H_0 : m &= 1 (\gamma_2) \\ H_1 : m &= 2 (\gamma_1), (\gamma_2) \end{aligned}$$

The observed values of the LR statistic are in this case, $LR_1^* = 12.3958$ and $LR_2^* = 13.1979$. Applying the first bootstrap technique as it was described in Algorithm 2 of subsection 2.4, with 5000 replications we get the tests' p-values: $p - value_1 = 0.1420$ and $p - value_2 = 0.1148$ which are both greater than the significance level 0.05 and thus we cannot reject neither of the null hypotheses. This leads us to the next two hypotheses tests,

$$\begin{aligned} H_0 : m &= 0 \\ H_1 : m &= 1 (\gamma_1) \end{aligned}$$

and

$$\begin{aligned} H_0 : m &= 0 \\ H_1 : m &= 1 (\gamma_2) \end{aligned}$$

The above tests' observed values of the LR statistic are , $LR_1^* = 5.8965$ and $LR_2^* = 5.1220$. The bootstrap technique of Algorithm 1 having been applied with 5000 replications returns the test p-values $p - value_1 = 0.270$ and $p - value_2 = 0.4302$. Both of them are greater than the specified level of significane, therefore we do not reject neither of the two null hypotheses. Hence, we come to the conclusion that among the four models M_1 is finally the one that better fits on the given data. This was the desired conclusion, since the process that initially generated the data is M_1 .

For the model selected after the hypotheses testing, namely M_1 in this case, we proceed to the estimation of its parameters and the construction of the respetive confidence intervals. Table summarizes our findings, where for the construction of the respective confidence interval we have set as level of significance $\alpha = 95\%$.

Table 4.1: Linear regression model. Estimation and Confidence Intervals

Parameter	Estimate	Confidence Interval
b	1.0022	[0.8872, 1.1173]
	1.8862	[1.7700, 2.0024]
σ^2	0.9604	

4.1.2 Bayesian Inference

The posterior model probabilities of the competing models, computed as discussed in subsection 3.4, are calculated for our simulated data as $P(M_1|\underline{y}) = 0.8528$, $P(M_2|\underline{y}) = 0.0717$, $P(M_3|\underline{y}) = 0.0755$ and $P(M_4|\underline{y}) = 8.0606e \wedge -06$. This means that model M_1 is a posteriori the most possible to have generated this specific data set. For this model, that is the most probable, we shall continue with the Bayesian inference for its parameters. The marginal posterior densities are available in close form and, more specifically are given by

$$\phi|\underline{y} \sim \text{Gamma} \left(\frac{n}{2} + s, q + \frac{Y'Y + \mu'C^{-1}\mu - \mu_1'C_1^{-1}\mu_1}{2} \right) = \text{Gamma}(s^*, q^*),$$

where,

$$C_1 = (C^{-1} + X'X)^{-1} \text{ and } \mu_1 = C_1(C^{-1}\mu + X'Y).$$

$$b|\underline{y} \sim t_v(\mu_1, \Sigma),$$

where,

$$v = n + 2s = 2s^* \text{ and } \Sigma = \frac{2q^*}{v} C_1 = \frac{q^*}{s^*} C_1.$$

We shall start with a visualization of the prior and the posterior probability density functions of the parameters in order to get an idea of how close are our prior beliefs to the posterior results, after observing the data.

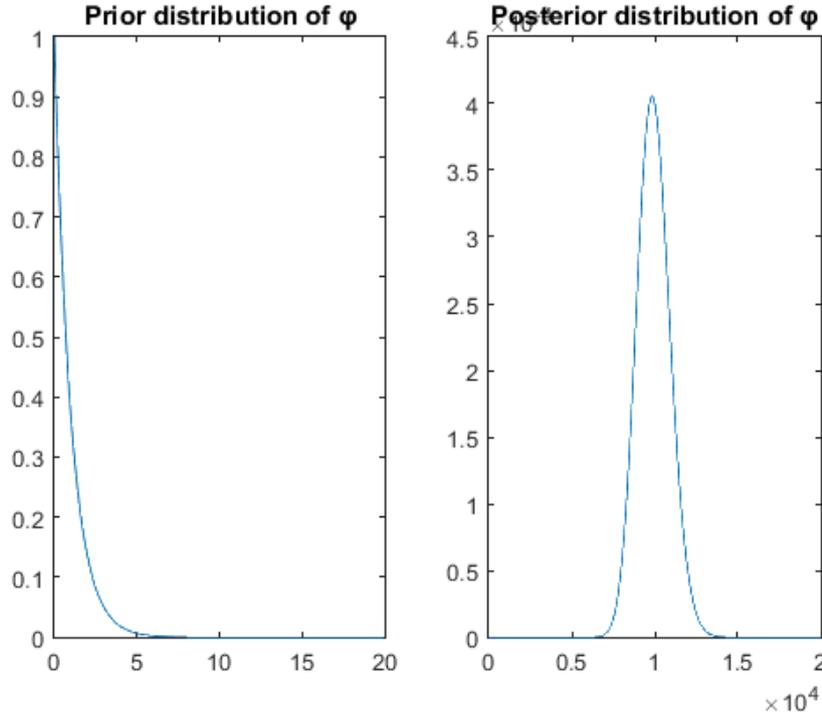


Figure 4.1: Marginal prior and posterior distribution of ϕ , for the linear regression model.

Note that the posterior distribution of ϕ is Gamma, but the parameter values are much greater than the ones assumed a priori. More specifically, $s^* = 101$ and $q^* = 98.3509$, which justifies the difference between the two distribution plots. However, we have to notice that the posterior mean of ϕ equals to $\frac{101}{98.5176} = 0.9754$ which is close to 1, that is the prior mean of ϕ .

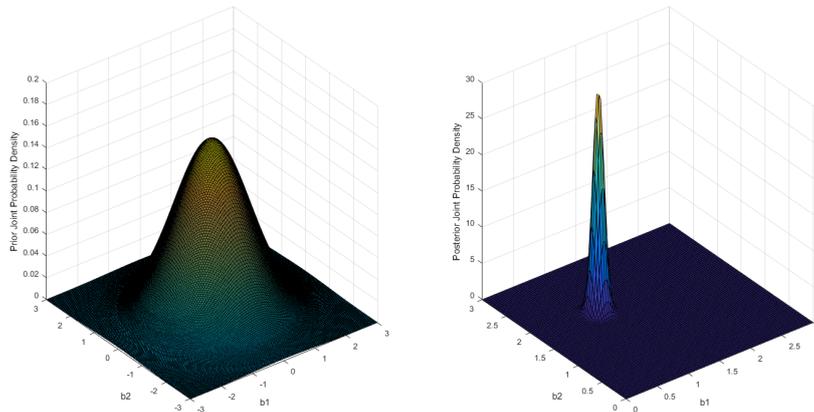


Figure 4.2: Joint Prior and Posterior distribution of b_1, b_2 , for the linear regression model.

Regarding the joint posterior distribution of b_1 and b_2 we shall focus our interest in three points in the above graph; the form, the means and the variances. The form of the posterior density function is that of a normal distribution function graph, the means seem different in comparison to those assumed a priori, and the variances seem also less than those assumed a priori for the two parameters. Turning to the two graphs and considering the probability density function as a function of just one parameter at a time, we can make inferences about each parameter separately.

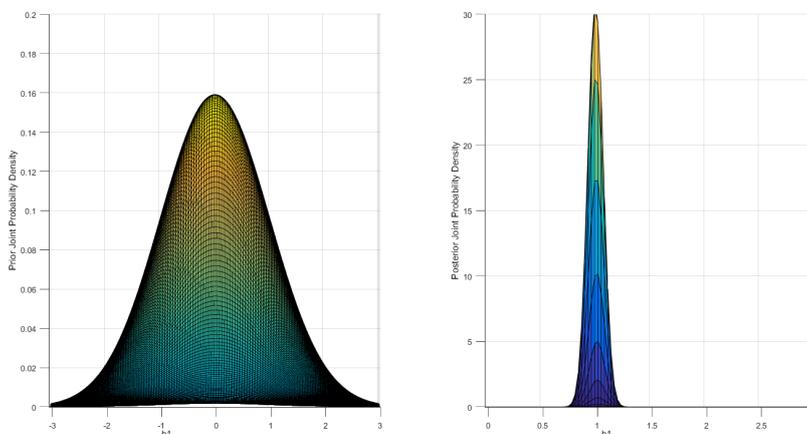


Figure 4.3: Marginal prior and posterior distribution of b_1 , for the linear regression model.

Starting with b_1 the posterior distribution seems to be normal but with mean 1 rather than 0 and variance less than 1, though not stable. It depends on b_2 which is the variance of b_1 and this is a result of the existence of correlation between b_1 and b_2 a posteriori.

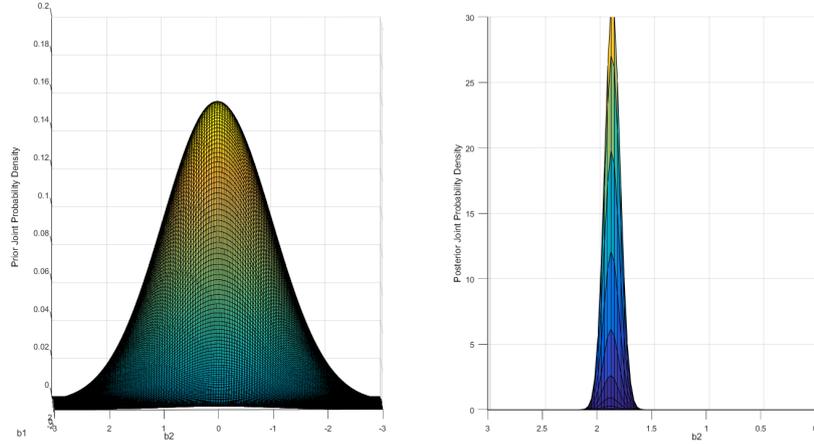


Figure 4.4: Marginal prior and posterior distribution of b_2 , for the linear regression model.

The posterior distribution of b_2 is also normal, but with mean close to 2 instead of 0 and variance less than 1. Because of the correlation between the two parameters, we can see in the right graph that the exact form and variance of the distribution of b_2 is not fixed, but depends on b_1 .

More specifically the posterior vector of means for $b = (b_1 \ b_2)'$ is $\mu_1 = (0.9981 \ 1.8770)'$ and the covariance matrix is $C_1 = \begin{pmatrix} 0.0050 & -0.0005 \\ -0.0005 & 0.0051 \end{pmatrix}$.

4.2 Model M_3 as DGP

In this subsection we consider M_3 as the data generating process, simulating $n = 200$ i.i.d observations from the univariate $N(0,1)$ distribution for x_i and from the univariate $N(2,1)$ for q_{i2} . Then, $x_i^* = (1 \ x_i)'$, where $x_i \sim N(0,1)$, $e_i \sim N(0,1)$ for $i = 1, 2, \dots, n$ and $b_2 = (1 \ 2)'$, $\delta_n = (0.5 - 1)$, $\gamma = 2$ are the parameter values. Thus, the values of y_i are generated according to M_3 as:

$$y_i = x_i^{*'} b_2 + x_i^* (\gamma)' \delta_n + e_i.$$

For the estimation of the parameters of the model that will be selected as the most appropriate, we also need to specify the values that the second threshold variable can take on. We assume, as in subsection 4.1 that $q_{i1} \sim N(0,1)$ and that $(\gamma_1, \gamma_2) \in \bar{\Gamma}_1 \times \bar{\Gamma}_2$, where $\bar{\Gamma}_j = \{q_{j(10)}, q_{j(15)}, q_{j(20)}, \dots, q_{j(90)}\}$ and $q_{j(k)}$ is the k -th percentile of q_{ij} for $i = 1, 2, \dots, n, j = 1, 2$.

Given the simulated data and under the previous assumptions, we proceed to the comparison of the three models in a trial to identify which one of them has generated the data. Of course, we already know where have the data come from and thus, we only need to verify it.

4.2.1 Classical Inference

As we did in subsection 4.1 we will start by testing the null hypothesis of the existence of one threshold variable that could be either γ_1 or γ_2 against the alternative that the linear regression model fits better on the data. In terms of hypotheses testing, the described tests are as follows.

$$\begin{aligned} H_0 : m = 1 (\gamma_1) \\ H_1 : m = 2 (\gamma_1), (\gamma_2) \end{aligned}$$

and

$$\begin{aligned} H_0 : m = 1 (\gamma_2) \\ H_1 : m = 2 (\gamma_1), (\gamma_2) \end{aligned}$$

For the first test we get $LR^* = 65.3059$ and respective $p - value_1 = 0$, while for the second test the observed value of the LR statistic is $LR^* = 5.3755$ and $p - value_2 = 0.8724$. Therefore, we reject the first null hypothesis but not the second. This leads us to the next hypothesis testing, that is deciding over the linear regression model and the model with γ_2 being the threshold parameter.

$$\begin{aligned} H_0 : m = 0 \\ H_1 : m = 2 (\gamma_2) \end{aligned}$$

The above test's observed value of the LR statistic is $LR^* = 67.3956$ and the respective bootstrap $p - value = 0$. This means, that the model fitting better the data is the threshold regression model with γ_2 being the threshold parameter.

Regarding the case that the data have occurred from a threshold regression model with one threshold variable, before estimating its parameters we shall take a look at the graph of the LR function and the values of γ_2 that lie beneath the critical value $c(\alpha)$.

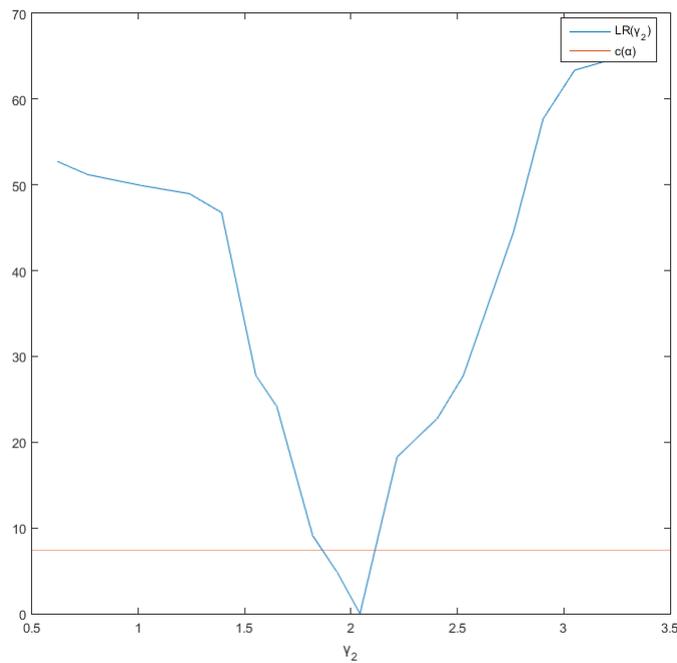


Figure 4.5: The LR statistic for the one-threshold regression model.

Figure 4.5 shows a narrow confidence interval for γ_2 which includes the estimator of the threshold parameter and maybe some other of the possible values of γ_2 ; remember that γ_2 is not a continuous variable but can take distinct values. This happens probably because we have generated the data set used in this subsection from a threshold regression model where the real values of δ_n are not close to 0. For that reason, when trying to estimate the value of the threshold parameter and provide a confidence interval for it, our inference is very accurate. Nevertheless, we construct the respective confidence interval for γ_2 following the procedure below:

1. Starting from the smallest value of γ_2 and moving across the x-axis, find the largest γ_2 for which $LR \leq c(\alpha)$. This is the lower bound of the confidence interval.
2. Starting from the opposite direction and moving across the x-axis, find the smallest value of γ_2 for which $LR \leq c(\alpha)$. This is the upper bound of the confidence interval.

Therefore, using the Bonferroni-type bound for the confidence intervals of the slope parameters we sum up the results for the one-threshold regression model in the following table.

Table 4.2: One-threshold regression model. Estimation and Confidence Intervals.

Parameter	Estimate	Confidence Interval
b_2	0.9338	[0.7392, 1.1482]
	2.2047	[2.0100, 2.3995]
δ_n	0.5989	[0.3192, 0.8612]
	-0.9449	[-1.2161, -0.6737]
γ_2	2.0415	[1.9350, 2.0415]
σ^2	1.2548	

The identification of the threshold regression model is based among others to the real values of the coefficients. If these values are close for the two regimes it is more difficult to decide the presence of a threshold variable than in the case they are not so close. In the above examples, the real values of the coefficients have been chosen to be far from each other and thus, the correct model was easily identified. It is of great interest to examine what will happen if we decide to simulate a new data set from model M_3 but choose as the parameters' real values $b_2 = (1 \ 2)'$, $\delta_n = (0.15 - 0.2)$. Regarding the distributions of x_i , q_{i1} , e_i , the real value of γ_2 and the subset in which γ_1 and γ_2 lie, we assume the same as in the first part of this subsection.

The model selection procedure begins with the following two hypotheses tests.

$$\begin{aligned} H_0 : m &= 1 (\gamma_1) \\ H_1 : m &= 2 (\gamma_1), (\gamma_2) \end{aligned}$$

and

$$\begin{aligned} H_0 : m &= 1 (\gamma_2) \\ H_1 : m &= 2 (\gamma_1), (\gamma_2) \end{aligned}$$

The observed values of the LR -statistic for the above tests are $LR_1^* = 13$ and $LR_2^* = 8.7771$ while the respective p -values are $p - value_1 = 0.1090$ and $p - value_2 = 0.4676$. This means that we cannot reject neither of the null hypotheses, at a level of significance 95% and thus, we proceed to the next two tests

$$\begin{aligned} H_0 : m &= 0 \\ H_1 : m &= 1 (\gamma_1) \end{aligned}$$

and

$$\begin{aligned} H_0 : m &= 0 \\ H_1 : m &= 1 (\gamma_2) \end{aligned}$$

The observed values of the LR statistic for these two tests are $LR_1^* = 3.6433$ and $LR_2^* = 8.2337$ and the utilization of Algorithm 1 gives the bootstrap p -values, $p - value_1 = 0.5988$ and $p - value_2 = 0.0662$. therefore, neither of the null hypotheses is rejected and the standard linear regression model is the one selected by the LR statistic method. For this model, table 4.3 summarizes its parameters' estimates and their respective confidence intervals.

Table 4.3: Linear regression model. Estimation and Confidence Intervals

Parameter	Estimate	Confidence Interval
b	1.1447	[1.0225, 1.2668]
	2.0020	[1.8780, 2.1260]
σ^2	1.0921	

4.2.2 Bayesian Inference

Here we consider Bayesian model comparison to compare simultaneously the four competing models for the first simulated dataset from model M_3 . Comparing the four competing models, which are a priori equally probable to occur, we get the posterior probabilities $P(M_1|\underline{y}) = 1.0544e \wedge -09$, $P(M_2|\underline{y}) = 2.3599e \wedge -10$, $P(M_3|\underline{y}) = 1$ (numerically equal to 1) and $P(M_4|\underline{y}) = 3.1606e \wedge -05$. It is obvious that the data come from the threshold regression model with γ_2 being the threshold parameter and this happens with a posterior model probability that numerically equals 1. For the selected model, in order to compare the prior and posterior distributions of the parameters, we shall first utilize the Gibb's sampler algorithm. To achieve this, we should first remind the posterior probability function of γ_2 and the conditional density functions of ϕ and \underline{b} . Given that the prior distributions of the parameters are

$$\begin{aligned}\gamma_2 &\sim DU(q_{1,2}, q_{2,2}, \dots, q_{n-1,2}) \equiv DU(q_{(10,2)}, q_{(15,2)}, \dots, q_{(90,2)}), \\ \phi &\sim \text{Gamma}(s, q) \equiv \text{Gamma}(1, 1), \\ \underline{b}|\phi &\sim N_{2(p+1)}(a, \frac{1}{\phi}V) \equiv N\left(0_{4 \times 1}, \frac{1}{\phi}I_{4 \times 4}\right),\end{aligned}$$

then, as it was proved in subsection 3.2, the conditional distributions of ϕ and \underline{b} are given by

$$\begin{aligned}\phi|\underline{y}, \underline{b}, \gamma_2 &\sim \text{Gamma}\left(\frac{n}{2} + p + s + 1, q + \frac{(Y - X_{\gamma_2}^* \underline{b})'(Y - X_{\gamma_2}^* \underline{b}) + (\underline{b} - a)'V^{-1}(\underline{b} - a)}{2}\right) \\ &\equiv \text{Gamma}(\Lambda, M)\end{aligned}$$

$$\begin{aligned}\underline{b}|\underline{y}, \phi, \gamma_2 &\sim N_{2(p+1)}\left(V_1(V^{-1}a + X_{\gamma_2}^* 'Y),, \frac{1}{\phi}\left(V^{-1} + X_{\gamma_2}^* 'X_{\gamma_2}^*\right)^{-1}\right) \\ &\equiv N_{2(p+1)}\left(a_1, \frac{1}{\phi}V_1\right)\end{aligned}$$

and

$$f(\gamma_2|\underline{y}) = \frac{|V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}{\sum_{\gamma_2 \in \{q_{1,2}, q_{2,2}, \dots, q_{n-1,2}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}.$$

Therefore, the Gibb's algorithm is as follows:

Algorithm 3.

1. Begin the chain with some values $(\gamma_2^{(0)}, \phi^{(0)}, \underline{b}^{(0)})$. As initial values we will use draws from the prior distributions.
2. Simulate $\gamma_2^{(1)}$ from the distribution with probability function

$$f(\gamma|\underline{y}) = \frac{|V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}{\sum_{\gamma_2 \in \{q_{1,2}, q_{2,2}, \dots, q_{n-1,2}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}.$$

3. Simulate $\phi^{(1)}$ from the distribution

$$\text{Gamma}\left(\frac{n}{2} + p + s + 1, q + \frac{(Y - X_{\gamma_2^{(1)}}^* \underline{b}^{(0)})'(Y - X_{\gamma_2^{(1)}}^* \underline{b}^{(0)}) + (\underline{b}^{(0)} - a)'V^{-1}(\underline{b}^{(0)} - a)}{2}\right).$$

4. Simulate $\underline{b}^{(1)}$ from the distribution

$$N_{2(p+1)}\left(V_1(V^{-1}a + X_{\gamma_2^{(1)}}^*{}'Y),, \frac{1}{\phi^{(1)}}\left(V^{-1} + X_{\gamma_2^{(1)}}^*{}'X_{\gamma_2^{(1)}}^*\right)^{-1}\right).$$

5. Iterate this procedure.

Iterating this procedure, it is proven that the Gibb's sampler finally returns draws from the the posterior joint distribution. In our case, we decided to simulate 30000 samples, which have been stored in a 30000×6 matrix. the first column contains the draws of γ_2 , the second column contains the draws of ϕ and the rest 4 the draws of \underline{b} . After a so called burn-in period (that is the number of iterations until the algorithm converges and for which the draws are discarded), each row of this matrix consists a realization of the joint distribution, while each item of a column is a realization of the marginal posterior distribution of the respective parameter. For further details on Gibb's algorithm refer to Appendix B.

To visualize the draws of each marginal distribution and also get an idea of the convergence of the sampler one could plot, for example, 3000 equally spaced draws, out of the 30000 for each parameter.

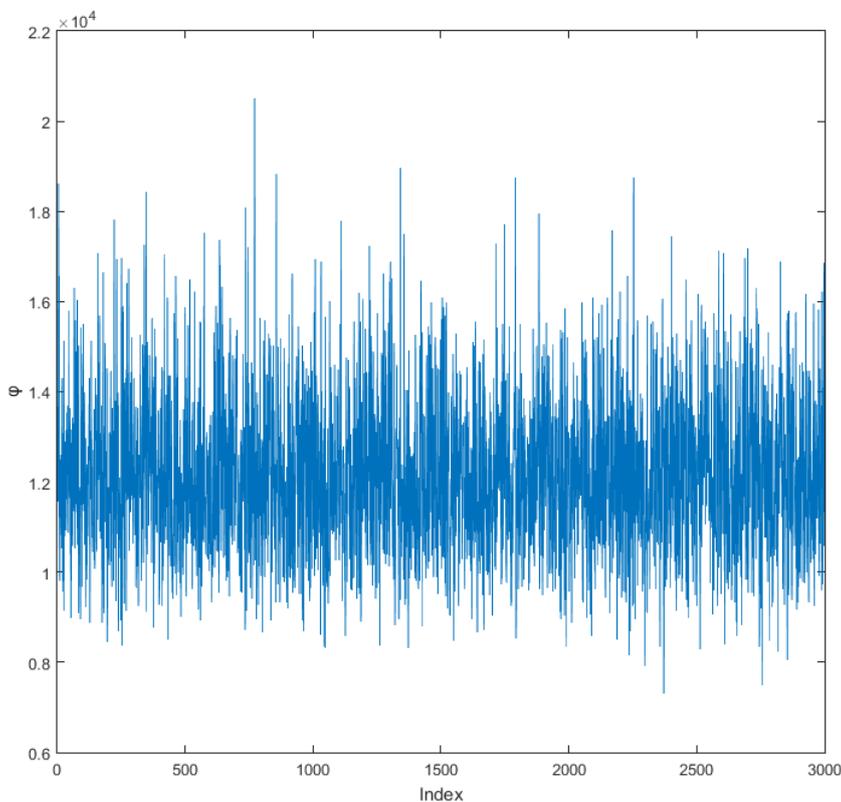


Figure 4.6: Consecutive realizations of the marginal distribution of ϕ .

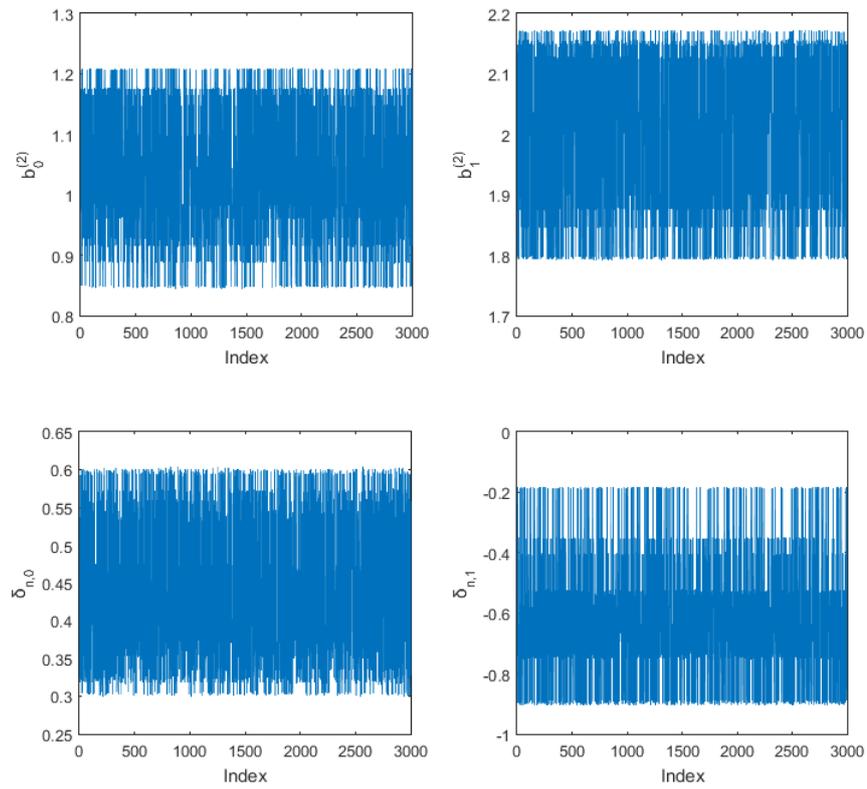


Figure 4.7: Consecutive realizations of the marginal distribution of $b_0^{(2)}$, $b_1^{(2)}$, $\delta_{n,0}$, $\delta_{n,1}$.

Another informal and maybe more convenient way to check the sampler's convergence is to plot the posterior mean of the parameters against the number of iterations of the sampler. The mean (which should be provided that exists) should settle around a certain value if convergence is achieved.

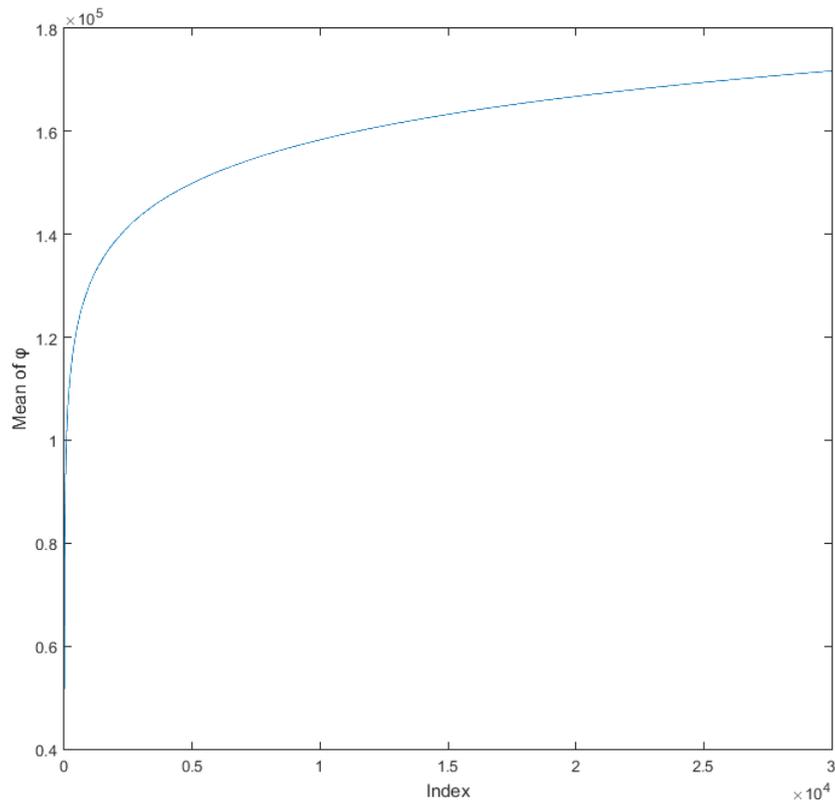


Figure 4.8: Consecutive sample means of ϕ .

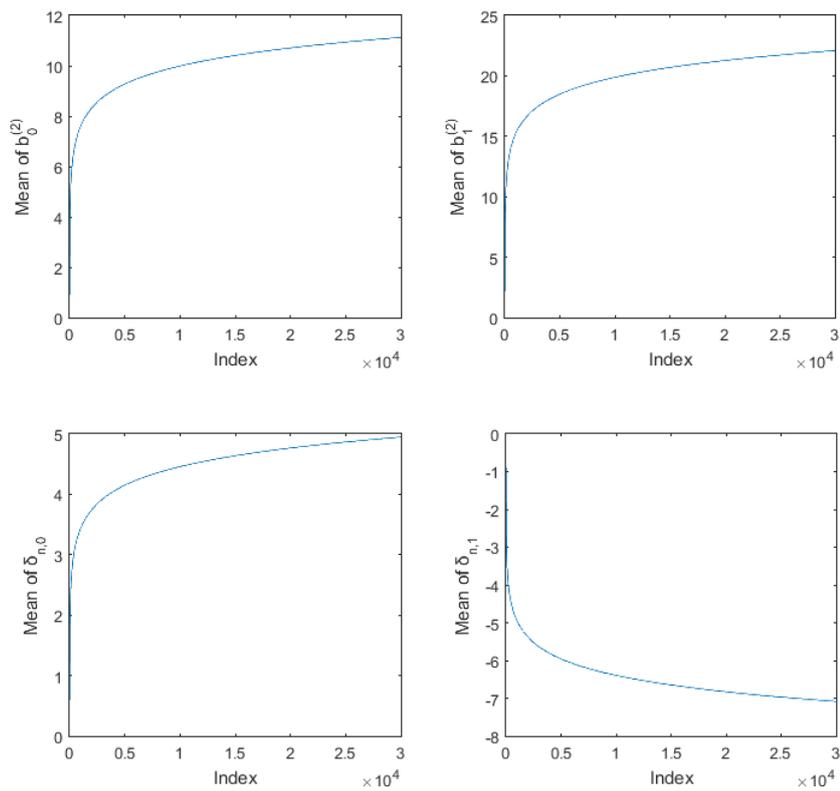


Figure 4.9: Consecutive sample means of $b_0^{(2)}$, $b_1^{(2)}$, $\delta_{n,0}$, $\delta_{n,1}$.

All the above plots suggest that convergence has been achieved after about 20000 iterations and thus, we remove the first 20000 draws (this is the burn in period) and keep the remaining 10000 to conduct inference.

The next plots display the prior and posterior distributions of the parameters.

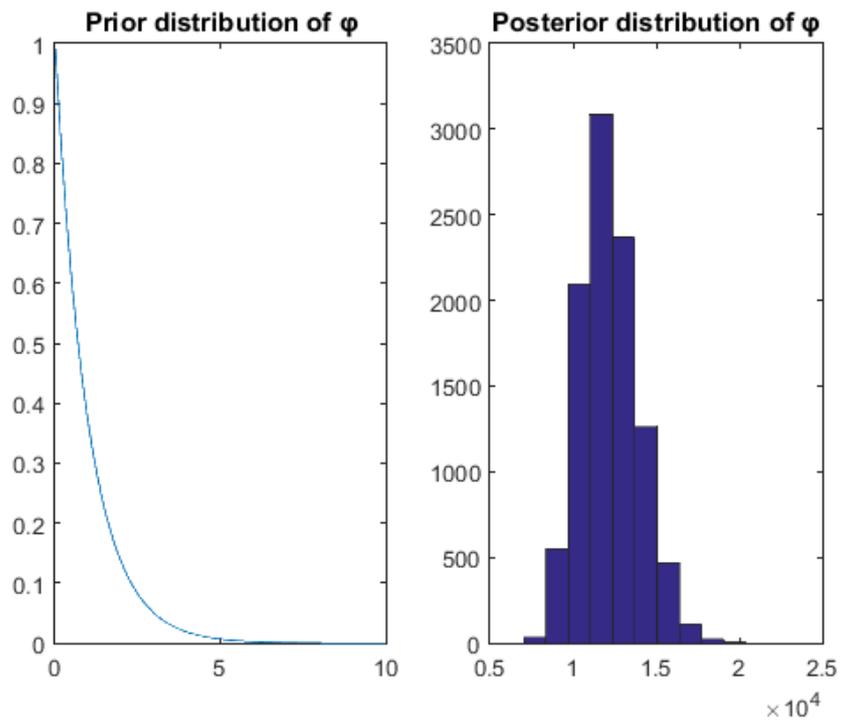


Figure 4.10: Marginal prior and posterior distribution of ϕ , for the one-threshold regression model.

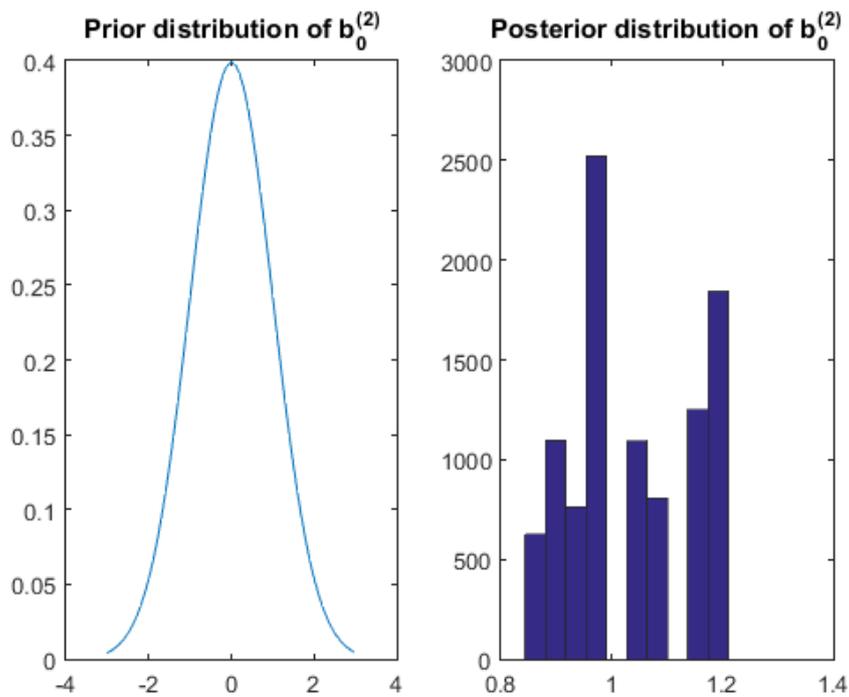


Figure 4.11: Marginal prior and posterior distribution of $b_0^{(2)}$, for the one-threshold regression model.

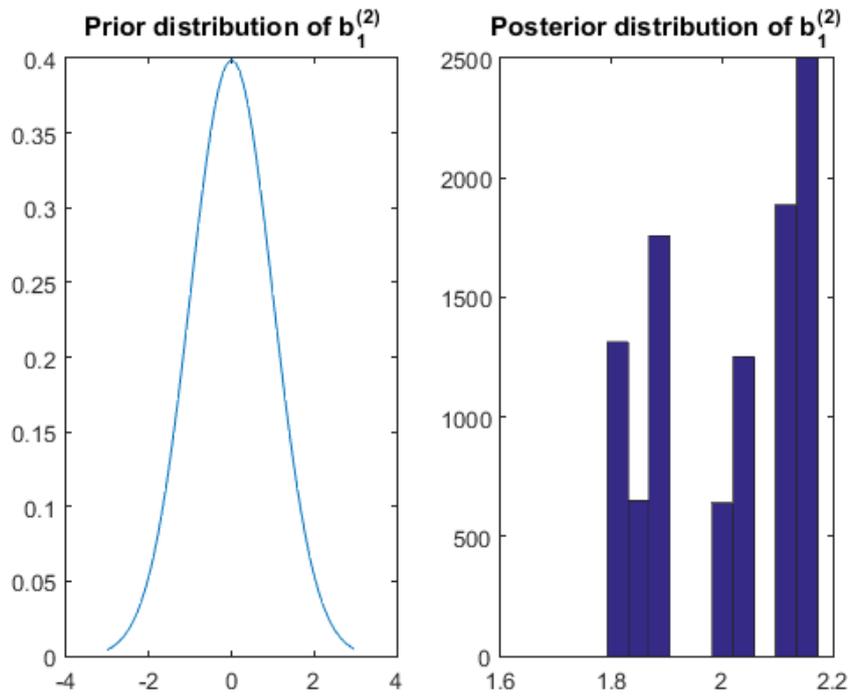


Figure 4.12: Marginal prior and posterior distribution of $b_1^{(2)}$, for the one-threshold regression model.

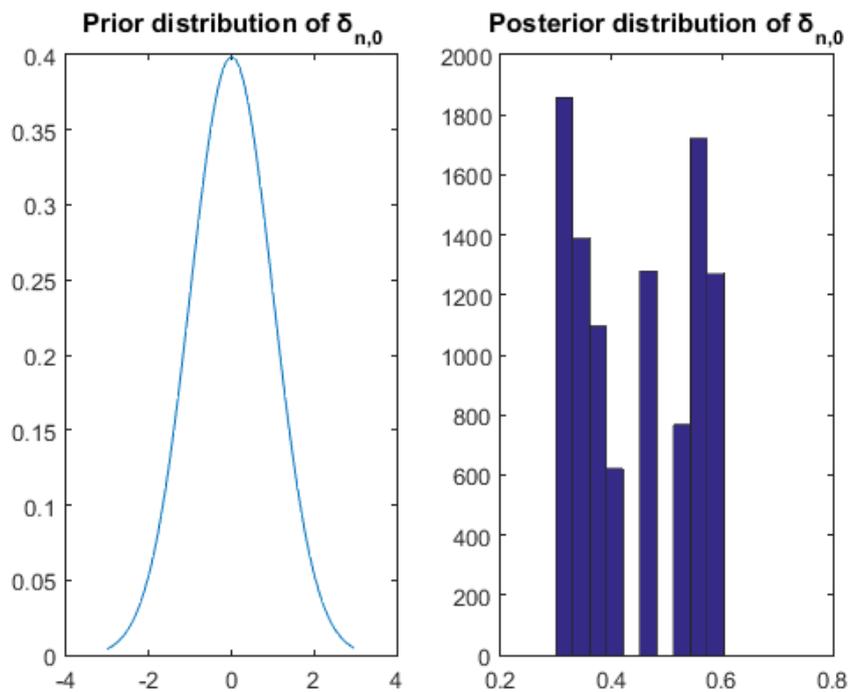


Figure 4.13: Marginal prior and posterior distribution of $\delta_{n,0}$, for the one-threshold regression model.

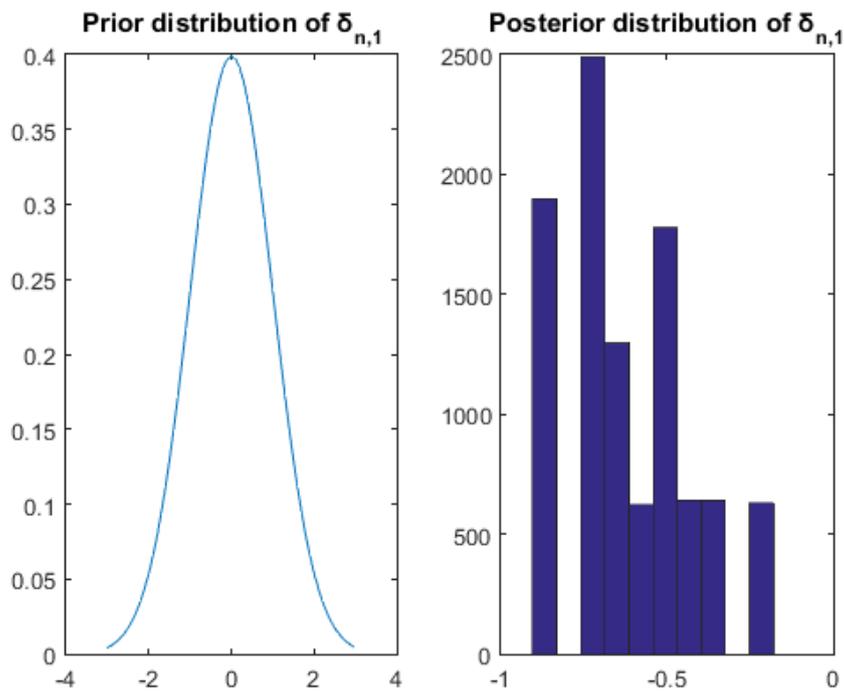


Figure 4.14: Marginal prior and posterior distribution of $\delta_{n,1}$, for the one-threshold regression model.

At this point is interesting to examine what happens to the posterior model probabilities for the second simulated data set of subsection 4.2.1, where the real values of b_2 and δ_n are closer. Recall that when using classical theory for the model comparison, the model that had generated the data was not identified.

Using the Bayesian approach to model selection, the posterior model probabilities are calculated as $P(M_1|y) = 0.7715$, $P(M_2|y) = 0.0342$, $P(M_3|y) = 0.1943$, $P(M_4|y) = 1.7390e \wedge -06$ and therefore, the most probable model is M_1 which is compatible to the conclusion of 4.2.1.

For the selected model, the next figures display the prior and the posterior distributions of the parameters.

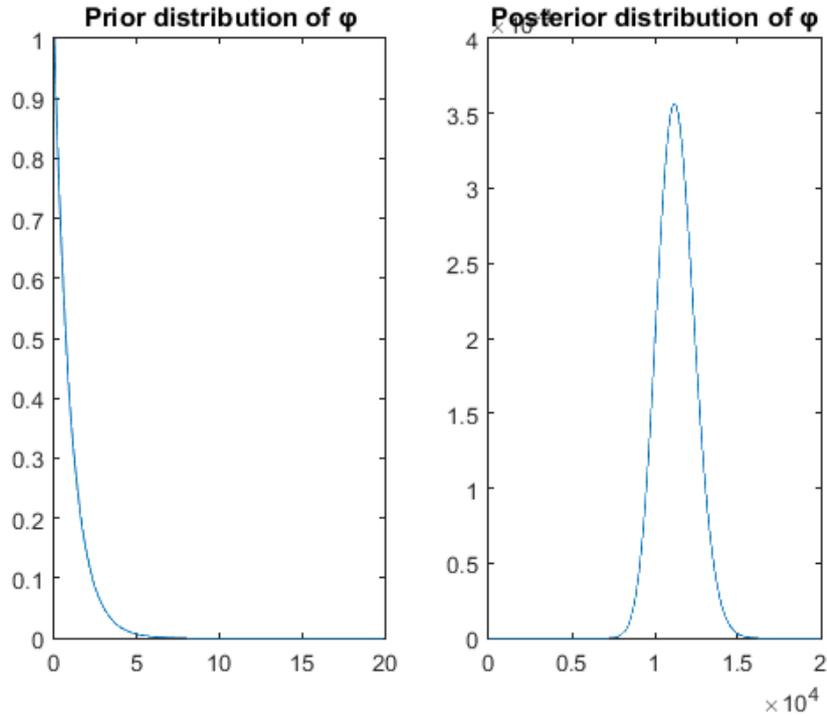


Figure 4.15: Marginal prior and posterior distribution of ϕ , for the linear regression model.

Note that the posterior distribution of ϕ is also Gamma but with different hyperparameter values $s^* = 101$ and $q^* = 111.7685$.

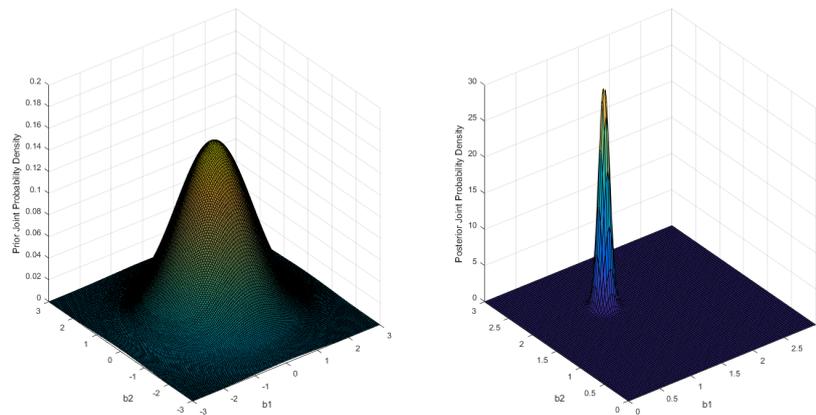


Figure 4.16: Joint prior and posterior distribution of b_1 , b_2 , for the linear regression model.

The above figure displays the joint prior and posterior distribution of b_1 and b_2 . The posterior distribution is also normal but with different hyperparameters and non-zero correlation between b_1 and b_2 . More specifically we have calculated that the vector of the posterior means of b_1 and b_2 is $\mu_1 = (1.1390 \ 1.9918)'$ and the covariance matrix is $C_1 = \begin{pmatrix} 0.0051 & -0.0000 \\ -0.0000 & 0.0069 \end{pmatrix}$. The plots of the prior and posterior distributions of each parameter follow and confirm the calculated results.

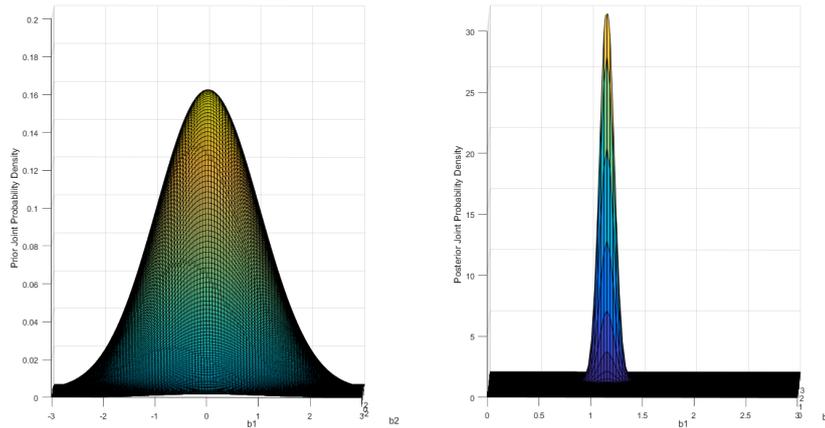


Figure 4.17: Marginal prior and posterior distribution of b_1 , for the linear regression model.

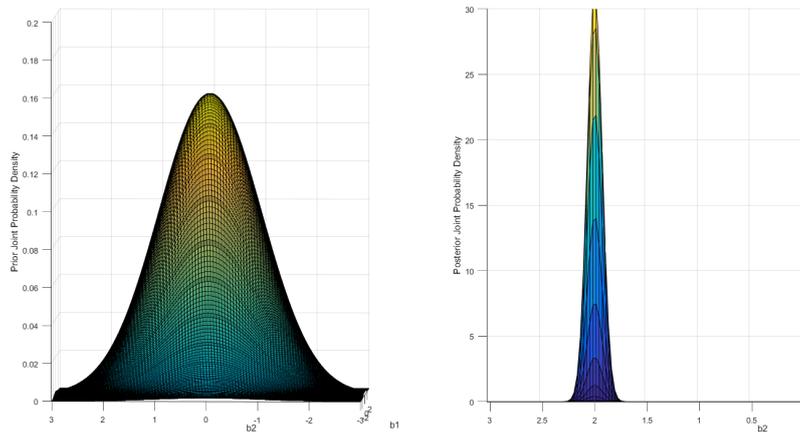


Figure 4.18: Marginal prior and posterior distribution of b_2 , for the linear regression model.

4.3 Model M_4 as DGP

In this subsection we consider M_4 as the DGP. Assume that $x_i^* = (1 \ x_i)'$, where $x_i \sim N(0, 1)$, $e_i \sim N(0, 1)$, $q_{i1} \sim N(0, 1)$ and $q_{i2} \sim N(2, 1)$ for $i = 1, 2, \dots, n$, where $n = 200$. Regarding the real values of the slope parameters we set $b^{(1)} = (1 \ 2.5)'$, $b^{(2)} = (1.5 \ 3.5)'$, $b^{(3)} = (2.5 \ 3)'$, $b^{(4)} = (0 \ 1)'$ while for the threshold parameters, $\gamma_1 = 1$ and $\gamma_2 = 2$. Therefore, we generate the data set of the dependent variable according to the model

$$y_i = \sum_{j=1}^4 x_i^{*'} b^{(j)} + e_i.$$

Having the data set $(y_i, x_i^*, e_i, q_{i1}, q_{i2})$ in hand we shall proceed to the comparison of the four regression models. As we did in the previous subsection we set $(\gamma_1, \gamma_2) \in \bar{\Gamma}_1 \times \bar{\Gamma}_2$, where $\bar{\Gamma}_j = \{q_{j(10)}, q_{j(15)}, q_{j(20)}, \dots, q_{j(90)}\}$ for $j = 1, 2$.

We shall start by testing the hypothesis of the existence of one threshold variable in the model, with respective parameters γ_1 or γ_2 against the alternative that both the

threshold variables enter the model. That is,

$$\begin{aligned} H_0 &: m = 1 (\gamma_1) \\ H_1 &: m = 2 (\gamma_1), (\gamma_2) \end{aligned}$$

$$\begin{aligned} H_0 &: m = 1 (\gamma_2) \\ H_1 &: m = 2 (\gamma_1), (\gamma_2) \end{aligned}$$

The observed values of the LR statistic and the respective bootstrap p -values having evaluated 5000 replications are, $LR_1^* = 137.8685$, $LR_2^* = 129.9317$, $p - value_1 = 0$ and $p - value_2 = 0$ and thus we reject, as expected, both the null hypotheses and come to the conclusion that the model fittingbest on the data is the two-threshold regression model.

Proceeding to the analysis of the two-threshold regression model, it is first examined the graph of the LR function and then one gets the table of the estimates of its parameters with their respective confidence intervals.

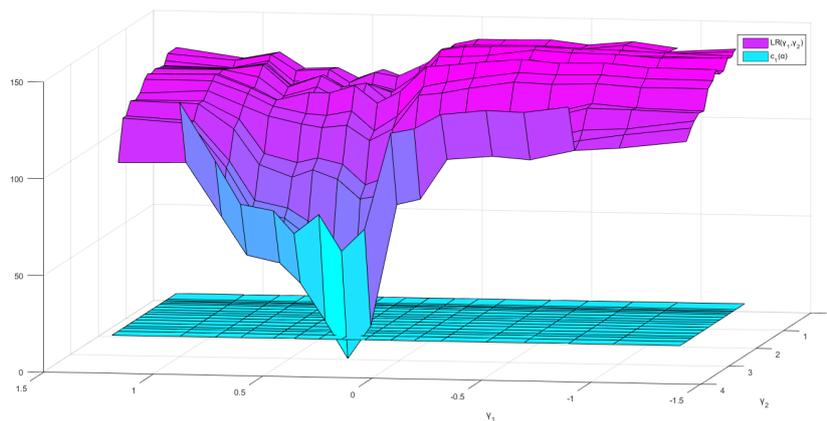


Figure 4.19: The LR statistic for the two-threshold regression model.

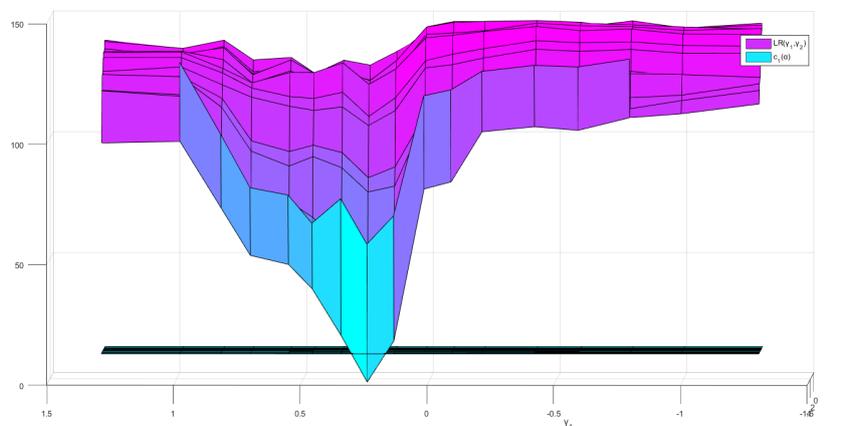


Figure 4.20: The LR statistic as a function of γ_1 , for the two-threshold regression model.

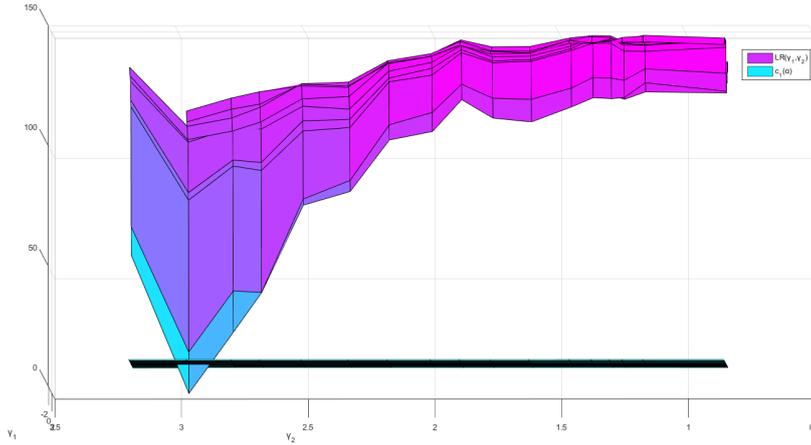


Figure 4.21: The LR statistic as a function of γ_2 , for the two-threshold regression model.

Table 4.4: Two-threshold regression model. Estimation and Confidence Intervals.

Parameter	Estimate	Confidence Interval
$b^{(1)}$	1.1201	[0.9551, 1.2851]
	2.4216	[2.2710, 2.5722]
$b^{(2)}$	1.6125	[1.4290, 1.7960]
	3.4389	[3.2287, 3.6491]
$b^{(3)}$	2.5205	[2.1174, 2.9236]
	3.0874	[2.6010, 3.5737]
$b^{(4)}$	-0.2542	[-0.7042, 0.1958]
	0.3324	[-0.1955, 0.8604]
γ_1	0.9814	[0.1370, 0.3457]
γ_2	1.9986	[2.96, 3.1876]
σ^2	0.9461	

4.3.1 Bayesian Inference

Completing the simulation chapter, we consider the data set that was generated from the fourth model. For this dataset the posterior probabilities of the four competing models are $P(M_1|\underline{y}) = 1.2431e^{-11}$, $P(M_2|\underline{y}) = 7.5935e^{-12}$, $P(M_3|\underline{y}) = 6.0882e^{-12}$ and $P(M_4|\underline{y}) = 1$ which means that model M_4 , almost with probability 1, is the one that has generated the data. For the posterior density and probability function of the parameters we utilize the Gibb's algorithm as follows.

Algorithm 4.

1. Begin the chain with some values $(\gamma_1^{(0)}, \gamma_2^{(0)}, \phi^{(0)}, \underline{b}^{(0)})$. We will use as initial values drops from the prior distributions.
2. Simulate $\gamma_1^{(1)}$ from the distribution with probability function

$$f(\gamma_1|\underline{y}) = \frac{\sum_{\gamma_2 \in \{q_{(10,2)}, q_{(11,2)}, \dots, q_{(90,2)}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{s_1^*}}}{\sum_{\gamma_1 \in \{q_{(10,1)}, q_{(11,1)}, \dots, q_{(90,1)}\}} \sum_{\gamma_2 \in \{q_{(10,2)}, q_{(11,2)}, \dots, q_{(90,2)}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{s_1^*}}}.$$

3. Simulate $\gamma_2^{(1)}$ from the distribution with probability function

$$f(\gamma_2|\underline{y}) = \frac{\sum_{\gamma_1 \in \{q_{(10,1)}, q_{(11,1)}, \dots, q_{(90,1)}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}{\sum_{\gamma_1 \in \{q_{(10,1)}, q_{(11,1)}, \dots, q_{(90,1)}\}} \sum_{\gamma_2 \in \{q_{(10,2)}, q_{(11,2)}, \dots, q_{(90,2)}\}} |V_1|^{1/2} \frac{\Gamma(s_1^*)}{q_1^{*s_1^*}}}.$$

4. Simulate $\phi^{(1)}$ from the distribution

$$\text{Gamma}\left(\frac{n}{2} + p + s + 1, q + \frac{(Y - X_{\gamma(1)}^* \underline{b}^{(0)})'(Y - X_{\gamma(1)}^* \underline{b}^{(0)}) + (\underline{b}^{(0)} - a)'V^{-1}(\underline{b}^{(0)} - a)}{2}\right).$$

5. Simulate $\underline{b}^{(1)}$ from the distribution

$$N_{4(p+1)}\left(V_1(V^{-1}a + X_{\gamma(1)}^* 'Y),, \frac{1}{\phi^{(1)}}\left(V^{-1} + X_{\gamma(1)}^* 'X_{\gamma(1)}^*\right)^{-1}\right).$$

6. Iterate this procedure.

The Gibbs algorithm, after a burn-in period, returns samples from the joint posterior distribution of the parameters. As in subsection 4.2.2 we simulate 30000 samples and store them in an 3000×11 matrix, the first two columns of which are the draws of γ_1 and γ_2 respectively, the third columns contains the draws of ϕ , while columns 4 through 11 contain the draws of $b_0^{(1)}, b_1^{(1)}, b_0^{(2)}, b_1^{(2)}, b_0^{(3)}, b_1^{(3)}, b_0^{(4)}, b_1^{(4)}$. In order to check the algorithm's convergence we shall take a look at the next figures containing consecutive draws and consecutive means for each parameter.

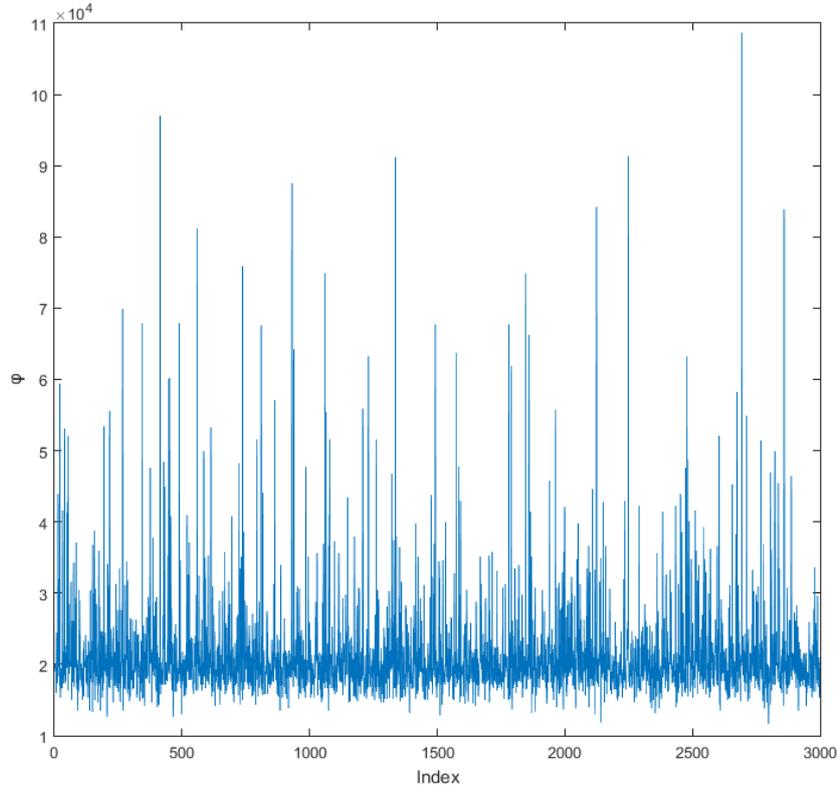


Figure 4.22: Consecutive realizations of the marginal distribution of ϕ .

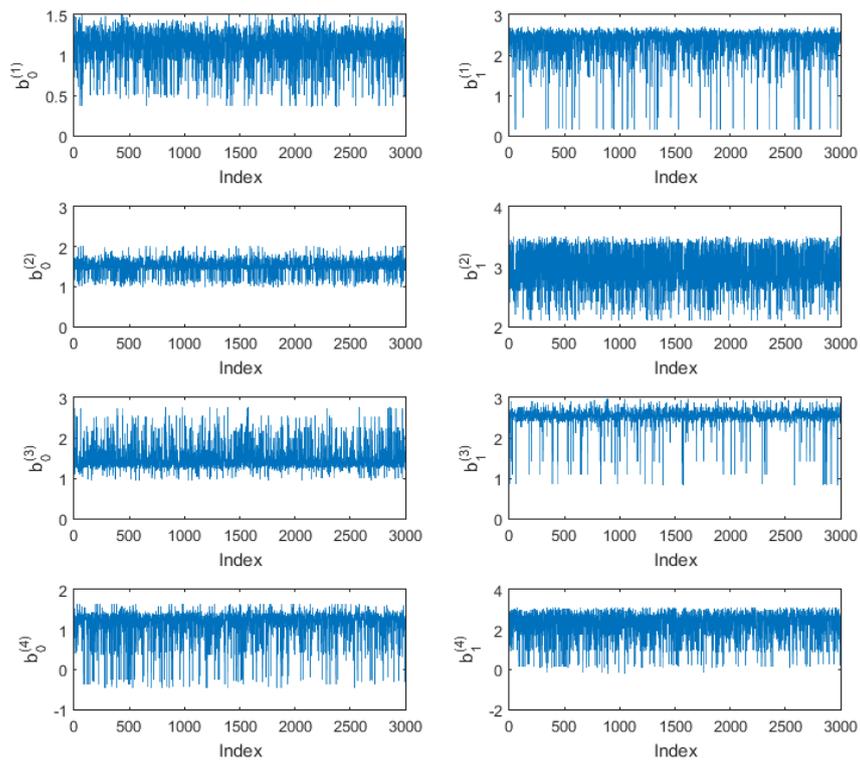


Figure 4.23: Consecutive realizations of the marginal distribution of $b_0^{(1)}, b_0^{(2)}, b_0^{(3)}, b_0^{(4)}, b_1^{(1)}, b_1^{(2)}, b_1^{(3)}, b_1^{(4)}$.

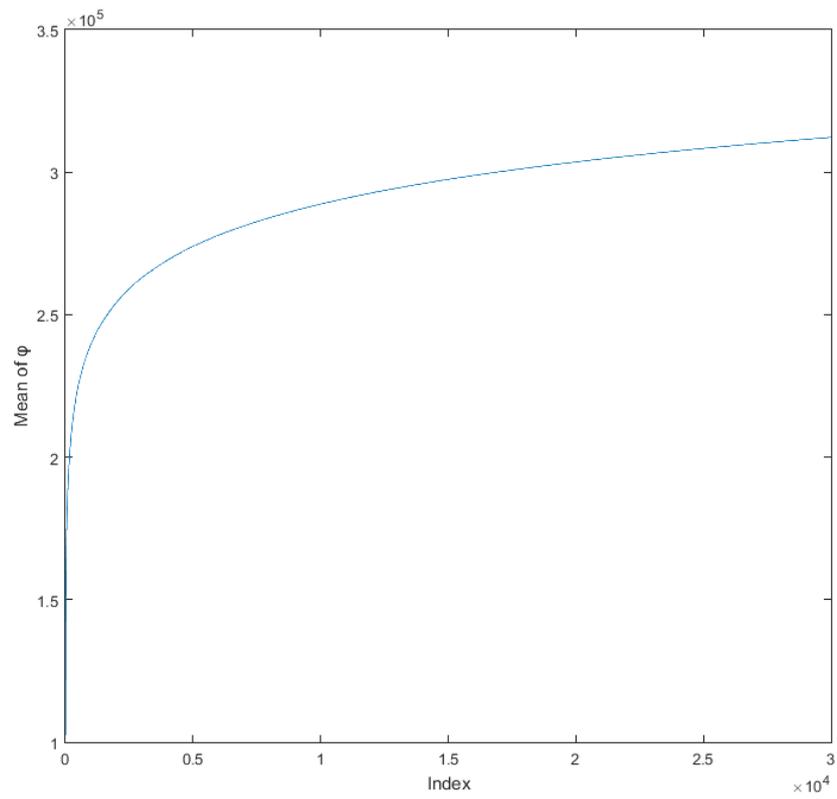


Figure 4.24: Consecutive sample means of ϕ .

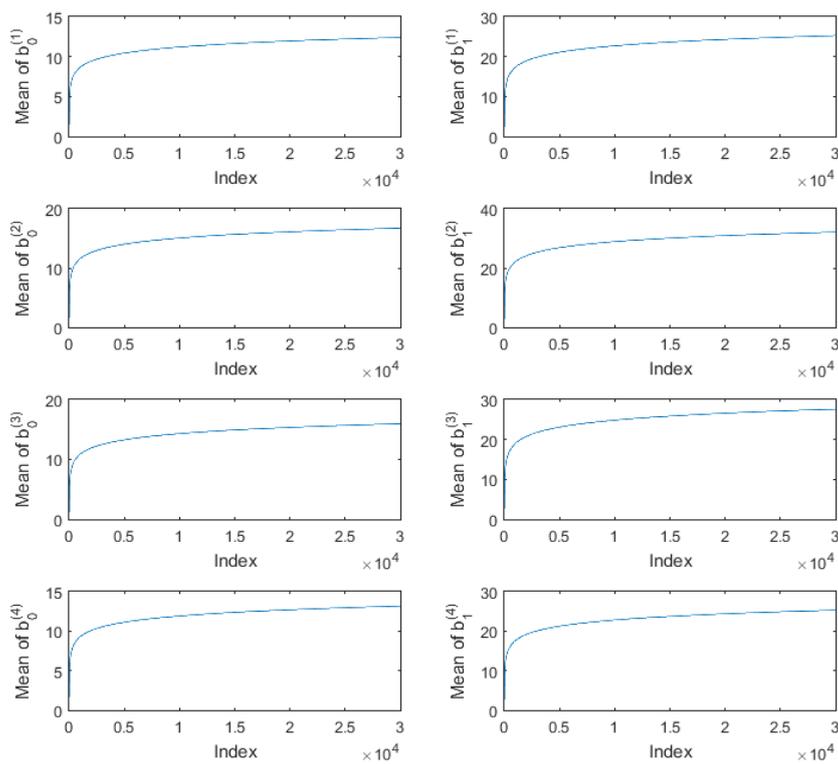


Figure 4.25: Plot of ergodic averages of $b_0^{(1)}, b_0^{(2)}, b_0^{(3)}, b_0^{(4)}, b_1^{(1)}, b_1^{(2)}, b_1^{(3)}, b_1^{(4)}$.

We consider the first 20000 draws as the burn-in period and store in a new matrix the remaining 10000. Each row of this new matrix consists of a sample of draws from the joint posterior distribution, while each column is a sample of the respective parameter's marginal distribution. For each one of the models parameters, the next figures illustrates the prior and posterior density functions.

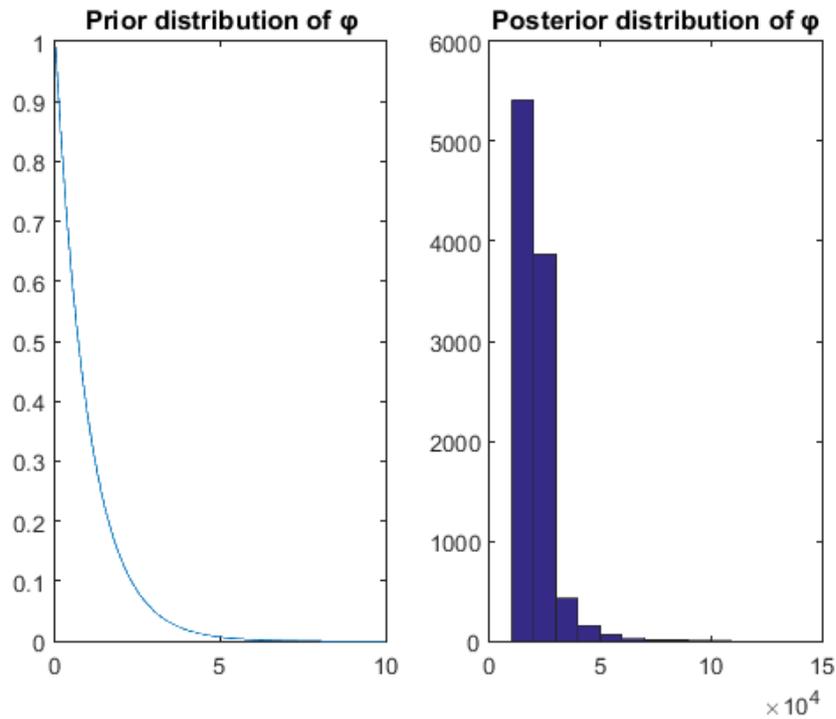


Figure 4.26: Marginal prior and posterior distribution of ϕ , for the two-threshold regression model.

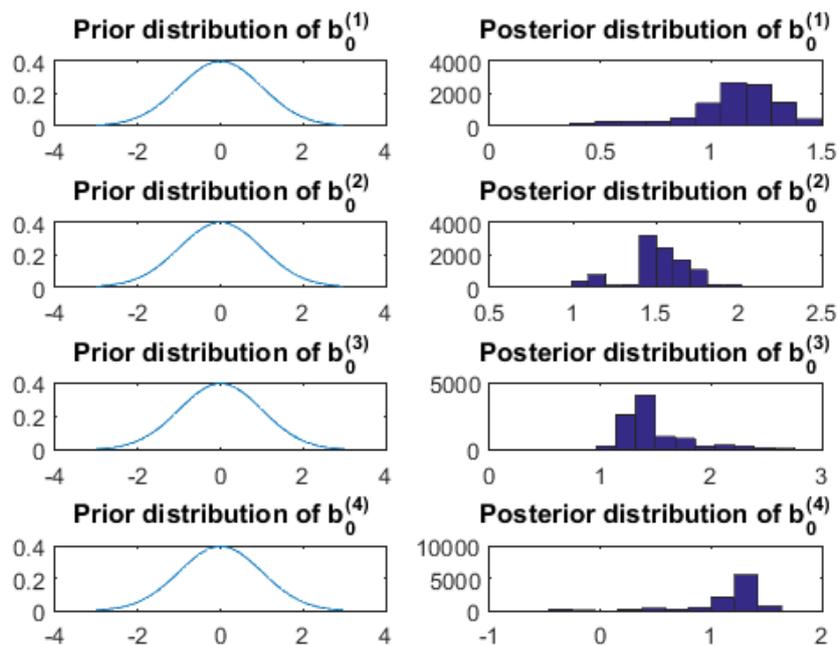


Figure 4.27: Marginal prior and posterior distribution of $b_0^{(1)}$, $b_0^{(2)}$, $b_0^{(3)}$, $b_0^{(4)}$, for the two-threshold regression model.

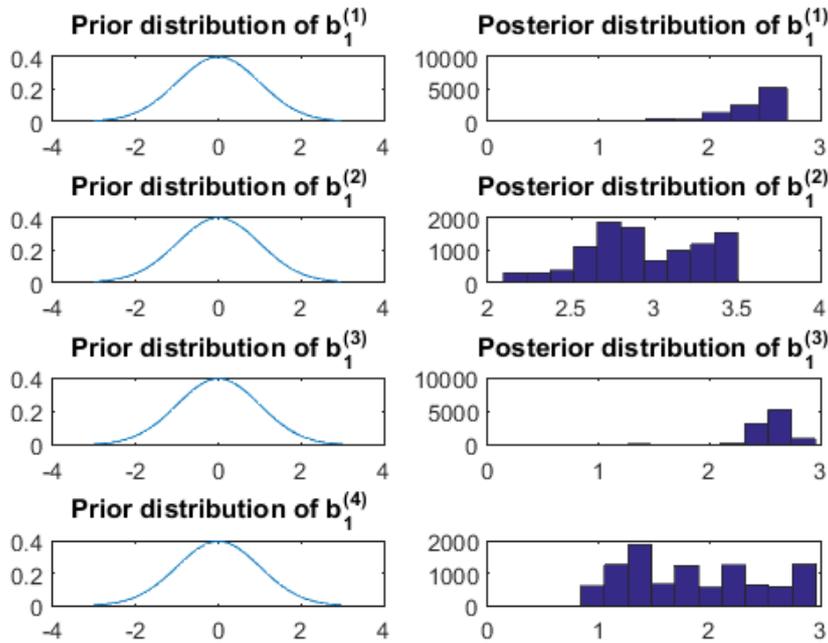


Figure 4.28: Marginal prior and posterior distribution of $b_1^{(1)}$, $b_1^{(2)}$, $b_1^{(3)}$, $b_1^{(4)}$, for the two-threshold regression model.

4.4 Comparing Classical and Bayesian results

Before ending this chapter, it is worthwhile to summarize and compare the results gained by applying both classical and bayesian theory. Regarding the model selection, it seems that there is a total accordance between the two methods since the correct models where identified except the one case of the threshold regression model but with real parameter values that were very close to each other. For each one of the selected models, from the scope of classical theory the estimators of the parameters where close enough to the real values, which were also included in the respective confidence intervals. On the other hand, from the scope of bayesian theory, we manage not only to get point estimates of the models' parameters but an overall representation of their distributions, via the posterior distributions.

Chapter 5

Application: Regression kink model with an unknown threshold

Reinhart and Rogoff in their paper *Growth in a Time of Debt* (2010), use an extensive dataset incorporating data on 44 countries for over about 200 years in order to search for the existence of a relationship between public debt levels, growth and inflation. The data used in their research cover a wide range of political systems and historic circumstances, such as war or economic crises periods, data for both rich countries and emerging markets.

Their conclusion of a nonlinear effect of debt on growth arises from the fact that there is a nonlinear response of market interest rates as countries reach debt tolerance limits. Regarding inflation, it is connected with debt in the sense that high levels of inflation can reduce the real cost of servicing the debt. Of course, the genesis of debt buildups can be important since for example debts accumulated during a war period are not as problematic for future growth and inflation as those accumulated in peacetime. However, the way in which debt builds up is not of the interest of the authors, who only try to determine how debt connects to growth and inflation outcomes.

The authors' main result is that as the level of government debt to GDP exceeds a threshold, economic growth tends to slow. Regarding advanced economies and emerging markets, it is remarkable that the relationship between public debt and growth is similar. And this is not because of the inflation since there is no systematic connection between high debt levels and inflation for the advanced economies, whereas in emerging markets the higher the debt levels are, the higher is the inflation.

Taking advantage of this paper and the available data, Hansen in his paper *Regression kink model with an unknown threshold* (2017) suggests that the relationship between debt and growth as presented by Reinhart and Rogoff (2010) can be formalized as a regression kink model. In this model, GDP growth is the dependent variable, while debt to GDP is the threshold variable.

5.1 Model

This application is concerned with the data referring to the United States for the years 1792-2009, as they were gathered by Reinhart and Rogoff. There are $n = 218$ observations and we set y_i the real GDP growth in year i , q_i the debt to GDP percentage from the year $i - 1$ and $x_i = (1 \ y_{i-1})$, for $i = 2, 3, \dots, n$. The next two figures, display time series plots for y_i and q_i respectively, while the third is a scatterplot of real GDP Growth and Debt/GDP.

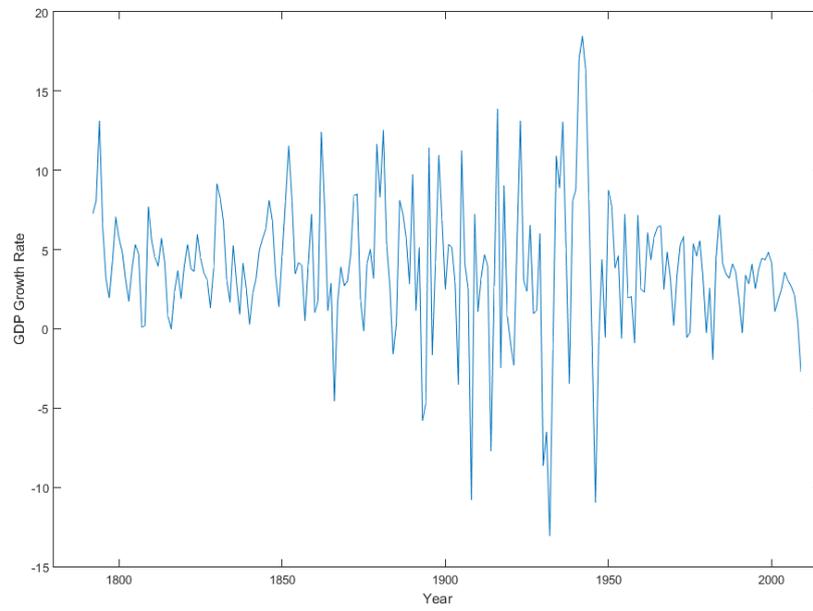


Figure 5.1: Annual U.S GDP real growth rate

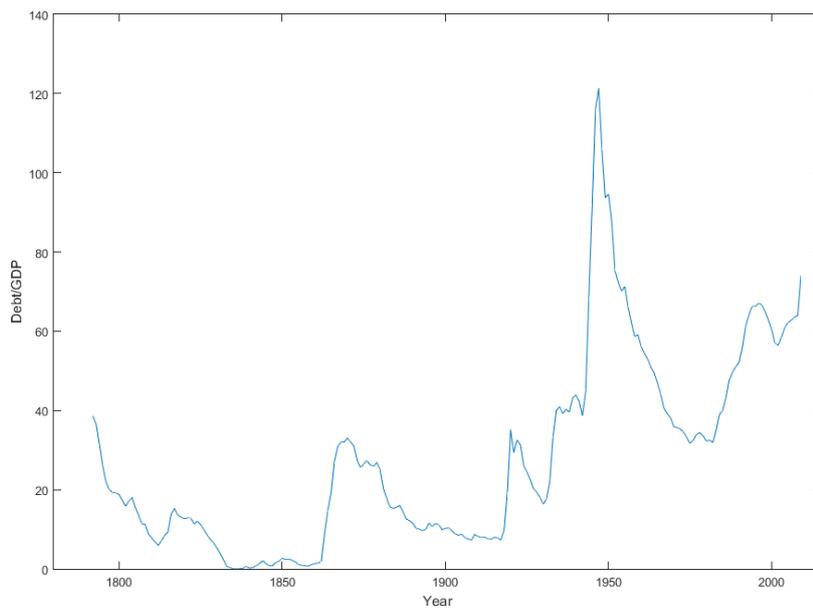


Figure 5.2: GDP/debt ratio 1791-2009

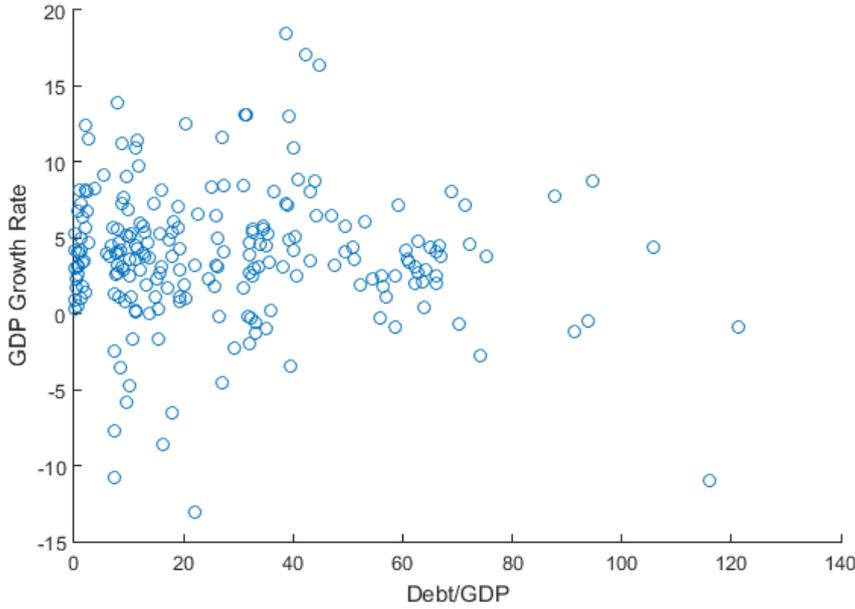


Figure 5.3: Scatterplot of real GDP growth and debt/GDP

Figure 5.3 presents a scatterplot of (y_i, q_i) in which one can see the existence of a threshold variable. The slope of the fitted line of the linear regression changes from positive to negative for debt to GDP ratios that are above a threshold variable. Therefore, the threshold regression model assumed for these data is as follows:

$$y_i = b_1(q_i - \gamma)^- + b_2(q_i - \gamma)^+ + b_3'x_i + e_i, \quad i = 2, 3, \dots, n, \quad (5.1)$$

where b_3 is a 2×1 vector. Regarding the support of γ , according to Hansen (2017) it is $\Gamma = [10, 70]$, so that at least 5% of the sample and 10% of the support of the threshold variable are below the lower bound and above the upper bound. As always, we assume that the error terms have 0 mean, are uncorrelated and homoskedastic, with common variance σ^2 . Thus, the model's parameters are $\theta = (b, \gamma)$ and σ^2 , where we have set $b = (b_1 \ b_2 \ b_3)'$.

After specifying the model's parameters, one comes across with the question *Is the threshold regression model statistically different from the linear regression model?* We aim to answer this question, using both ordinary and Bayesian theory, as it was extensively described in the subsections 2.1, 2.2 and 3.1, 3.2. It is also interesting to compare the results coming from the two approaches, point similarities and differences and come to a more solid conclusion.

5.2 Classical inference

From the scope of classical inference, in order to answer the question of whether the threshold regression model is statistically different from the linear regression model or not, we are going to use the *LR* statistic and a bootstrap technique in order to get the test's p-value. Namely, we are interested in testing the hypotheses

$$H_0 : m = 0$$

$$H_1 : m = 1$$

Recall that for the computation of the observed value of the LR statistic, that is

$$LR_n^* = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2(\gamma)}{\hat{\sigma}^2(\gamma)},$$

we have first to estimate the variance of the error terms for both the linear and the threshold regression model.

The respective linear regression model is nested in model (5.1), as the next equation holds for the case that $b_1 = b_2$

$$y_i = b_1 q_i + b_3' x_i + e_i \quad (5.2)$$

$$= b_1 q_i + b_{31} + b_{32} y_{i-1} + e_i, \quad i = 2, 3, \dots, n \quad (5.3)$$

and $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2$. The OLS estimators of the linear model (5.2) are:

$$\hat{b}_1 = -0.008, \quad \hat{b}_2 = (2.89 \ 0.3030)' \text{ and } \tilde{\sigma}^2 = 17.59.$$

On the other hand, the estimates of the parameters of model (5.1) are:

$$\hat{b}_1 = 0.033, \quad \hat{b}_2 = 0.067, \quad \hat{b}_3 = (3.78, 0.28)', \quad \hat{\gamma} = 43.8 \text{ and } \hat{\sigma}^2(\gamma) = 17.2.$$

Therefore, the real value of the LR statistic is $LR_n^* = 5.6583$ and next comes the bootstrap technique for the computation of the test's p-value. Following Algorithm 1, as it was described in subsection 2.3, and setting $J=10000$, that is applying 10000 bootstrap replications the obtained p-value at a 95% level of statistical significance is $p_n = 0$. Thus, we do reject the null hypothesis and confirm our initial point of view about the existence of a threshold variable.

Our results are in total agreement with those of Hansen (2017).

5.3 Bayesian inference

In order to utilize Bayesian theory we shall first rewrite models (5.1) and (5.2) in matrix form, which is more convenient not only for the calculations but also for the syntax of the code that we will use.

Model 1

$$y_i = b_3' x_i + b_1 q_i + e_i = b_{31} + b_{32} y_{i-1} + b_1 q_i + e_i, \quad i = 2, 3, \dots, n \quad (5.4)$$

Model M_1 in matrix form is expressed as follows.

$$Y = Xb + e, \quad (5.5)$$

$$\text{where, } X = (1_n \ Y_{-1} \ q) = \begin{pmatrix} 1 & y_1 & q_2 \\ 1 & y_2 & q_3 \\ \vdots & \vdots & \vdots \\ 1 & y_{n-1} & q_n \end{pmatrix}$$

and

$$b = (b_3^{(0)} \ b_3^{(1)} \ b_1)'$$

Model 2

Model (5.1) can equivalently be written as

$$y_i = b_3^{(0)} + b_1\gamma\mathbf{1}_{q_i \leq \gamma} + b_2\gamma\mathbf{1}_{q_i \geq \gamma} + b_3^{(1)}y_{i-1} + b_1q_i\mathbf{1}_{q_i \leq \gamma} + b_2q_i\mathbf{1}_{q_i \geq \gamma} + e_i, \quad i = 2, 3, \dots, n, \quad (5.6)$$

or in matrix form

$$Y = X_\gamma^* b + e, \quad (5.7)$$

where,

$$X_\gamma^* = (1_n \ Y_{-1} \ q\mathbf{1}_{q \leq \gamma} \ q\mathbf{1}_{q \geq \gamma}) = \begin{pmatrix} 1 & y_1 & q_2\mathbf{1}_{q_2 \leq \gamma} & q_2\mathbf{1}_{q_2 \geq \gamma} \\ 1 & y_2 & q_3\mathbf{1}_{q_3 \leq \gamma} & q_3\mathbf{1}_{q_3 \geq \gamma} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_{n-1} & q_n\mathbf{1}_{q_n \leq \gamma} & q_n\mathbf{1}_{q_n \geq \gamma} \end{pmatrix}$$

and

$$b = (b_0 \ b_3^{(1)} \ b_1 \ b_2)'$$

Note that the constant term of model (5.6) depends on γ , and it is $b_3^{(0)} + b_1\gamma$ when $q_i \leq \gamma$ and $b_3^{(0)} + b_2\gamma$ otherwise. However, it does not affect the bayesian analysis and thus we denote as b_0 the sum $b_0 + b_1\gamma\mathbf{1}_{q_i \leq \gamma} + b_2\gamma\mathbf{1}_{q_i \geq \gamma}$.

Thus, the two competing models are

$$M_1 : Y = Xb + e$$

$$M_2 : Y = X_\gamma^* b + e$$

Now that the two models are written in an appropriate matrix form, the likelihood function of them can be specified under the assumption of normally distributed error terms. More specifically, assuming that the error terms joint ditribution is normal, i.e

$$e \sim N_n \left(0_{n \times 1}, \frac{1}{\phi} 1_{n \times n} \right)$$

the models' likelihood functions are

$$M_1 : f(\underline{y}|b, \phi) = \left(\frac{\phi}{2\pi} \right)^{\frac{n}{2}} (Y - Xb)'(Y - Xb)$$

$$M_2 : f(\underline{y}|b, \phi, \gamma) = \left(\frac{\phi}{2\pi} \right)^{\frac{n}{2}} (Y - X_\gamma^* b)'(Y - X_\gamma^* b).$$

The specification of the prior distributions of the parameters is critical and one shall consider if there is any prior information about them. Regarding ϕ and b we decide to use the conditional conjugate families, while for γ we decide to follow the proposal of Hansen (2017), as we did for the classical inference.

Model 1

$$\phi \sim \text{Gamma}(1, 1)$$

$$b|\phi \sim N_3 \left(0_{3 \times 1}, \frac{1}{\phi} 1_{3 \times 3} \right)$$

Model 2

$$\phi \sim \text{Gamma}(1, 1)$$

$$b|\phi \sim N_4 \left(0_{4 \times 1}, \frac{1}{\phi} 1_{4 \times 4} \right)$$

$$\gamma \sim \text{DU}(10, 10.1, 10.2, \dots, 70).$$

Although the scatterplot indicates the existence of a threshold variable, and one could set a larger prior probability to this model, we decide to ignore it and consider the two models a priori equally probable, i.e. $P(M_1) = P(M_2) = \frac{1}{2}$. Following the respective theory of subsection 3.4, we compute the evidence of each model and thus we get the two models' posterior probabilities. More specifically, it is $P(M_1) = 0$ and $P(M_2) = 1$, which is a strong evidence for the existence of a threshold variable. For the threshold regression model, that seems to have generated the data, we proceed to its parameters inference.

The utilization of Gibb's sampler is required in order to get a visualization of the parameters' posterior distributions. Following Algorithm 3 of subsection 4.2.2, we get the subsamples of the joint posterior distribution of the parameters, as well as of each parameter's marginal distribution using 30000 replications.

To investigate the convergence of the algorithm we shall take a look at the following figures.

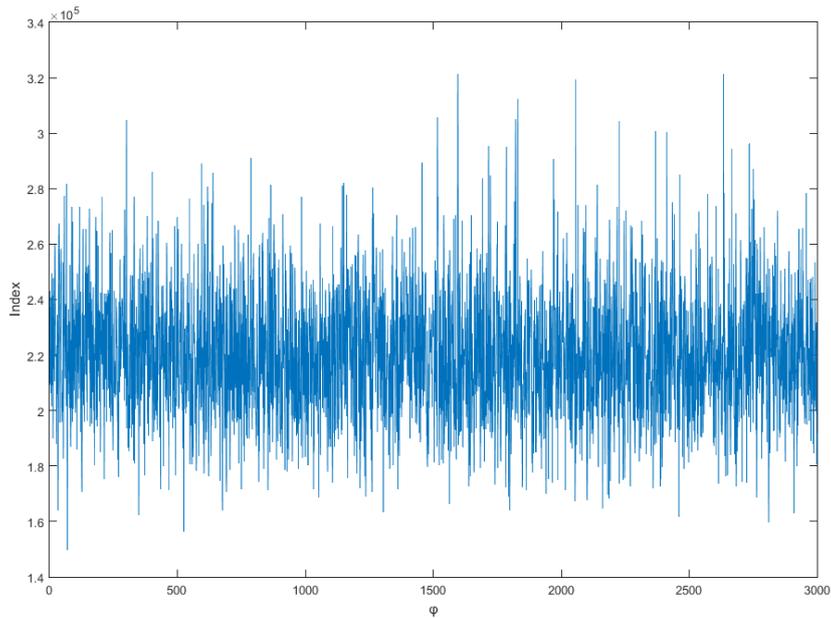


Figure 5.4: Consecutive realizations of the marginal distribution of ϕ

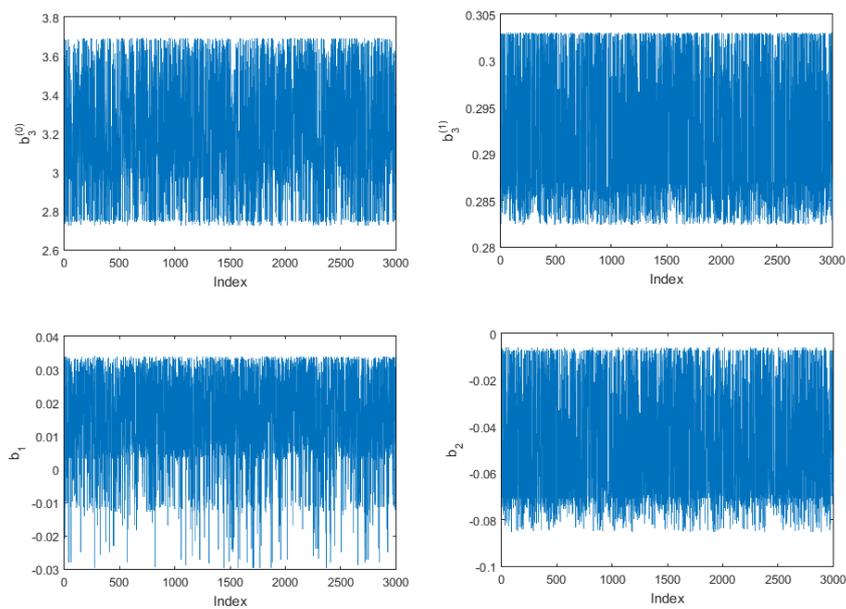
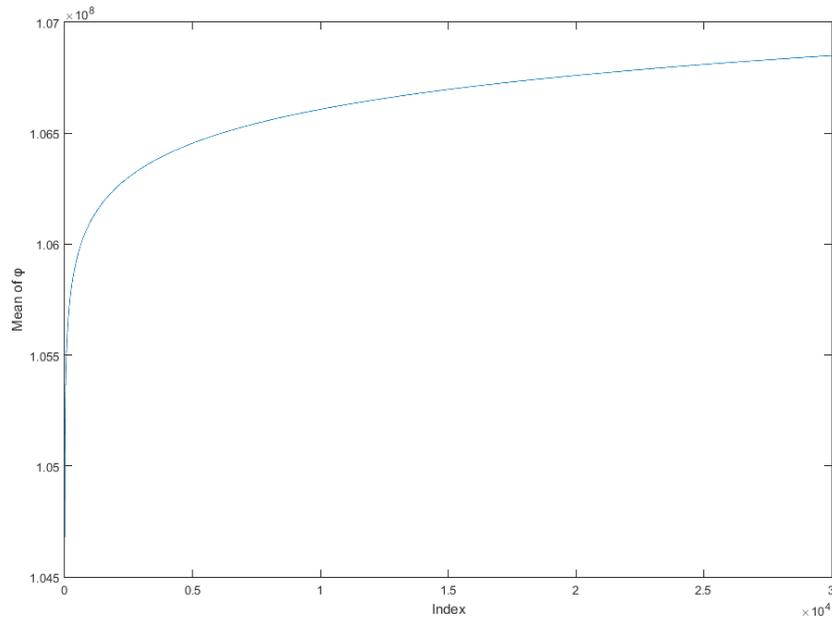
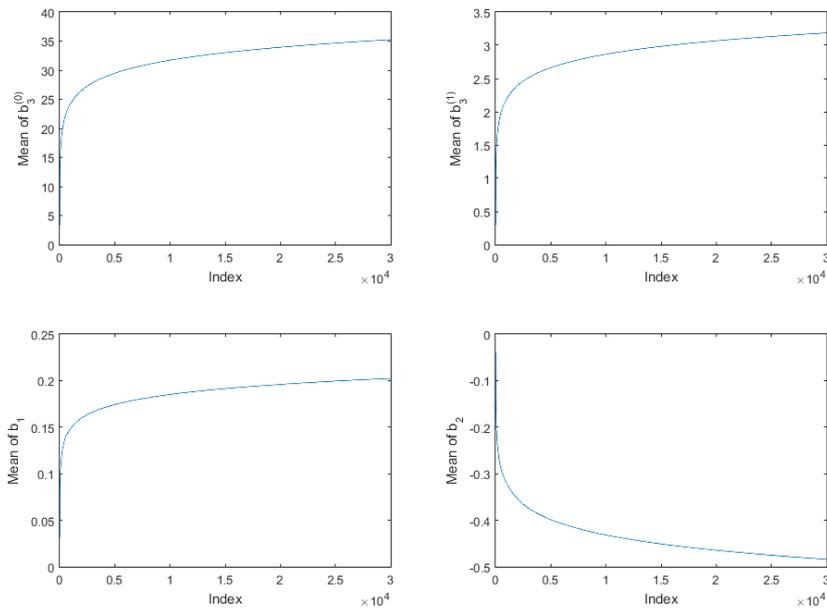


Figure 5.5: Consecutive realizations of the marginal distribution of $b_3^{(0)}$, $b_3^{(1)}$, b_1 , b_2

Figure 5.6: Consecutive sample means of ϕ Figure 5.7: Consecutive sample means of $b_3^{(0)}$, $b_3^{(1)}$, b_1 , b_2

Inspecting the above figures we decide to reject the first 20000 samples and keep the remaining 10000 in a 10000×6 matrix. Each column of this matrix is a sample of the respective parameter's marginal distribution and by plotting the respective histogram next to the prior distribution we manage to compare our prior beliefs to the posterior evidence.

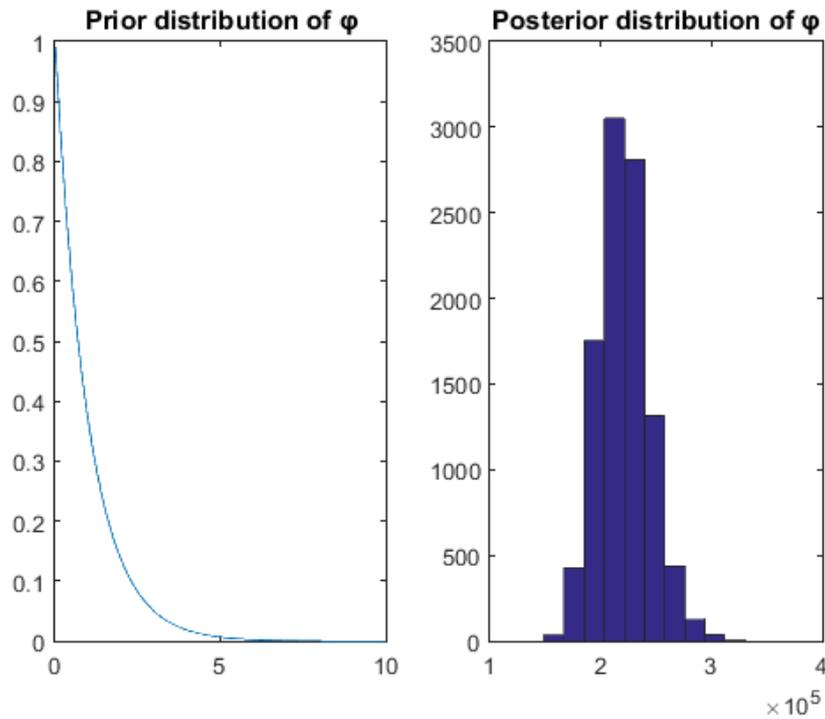


Figure 5.8: Marginal prior and posterior distribution of ϕ , for the kink model.

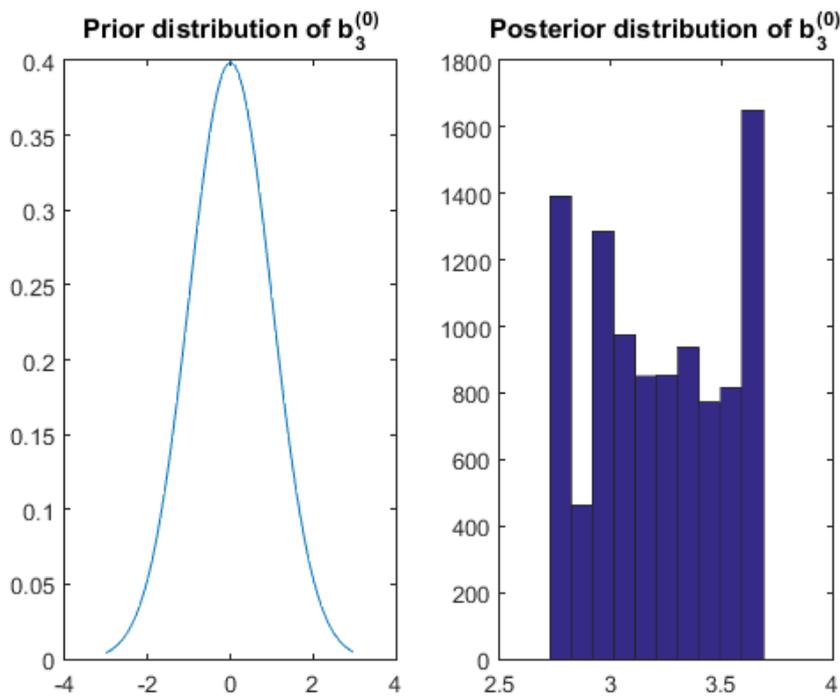


Figure 5.9: Prior and Posterior marginal distribution of $b_3^{(0)}$, for the kink model.

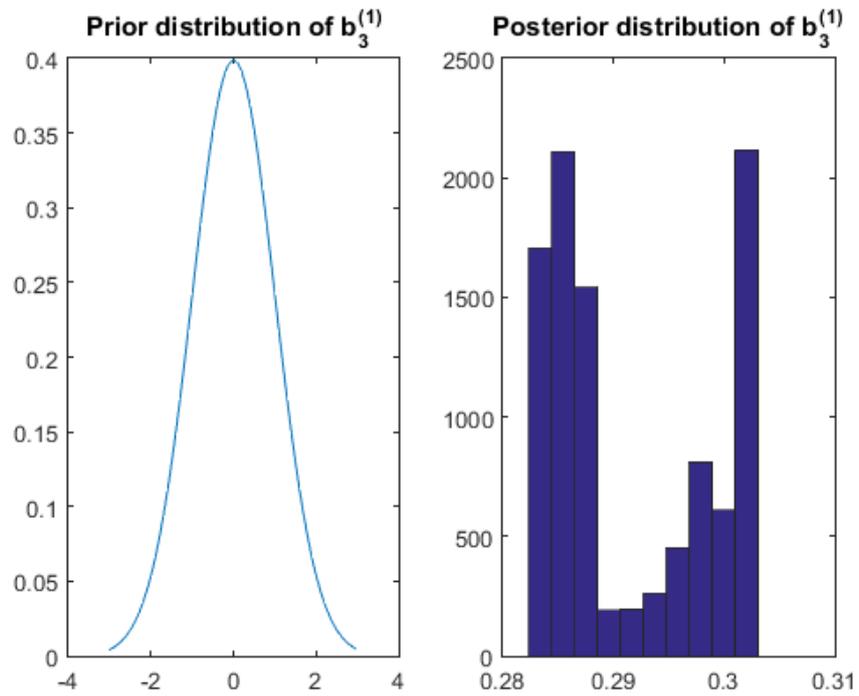


Figure 5.10: Marginal prior and posterior distribution of $b_3^{(1)}$, for the kink model.

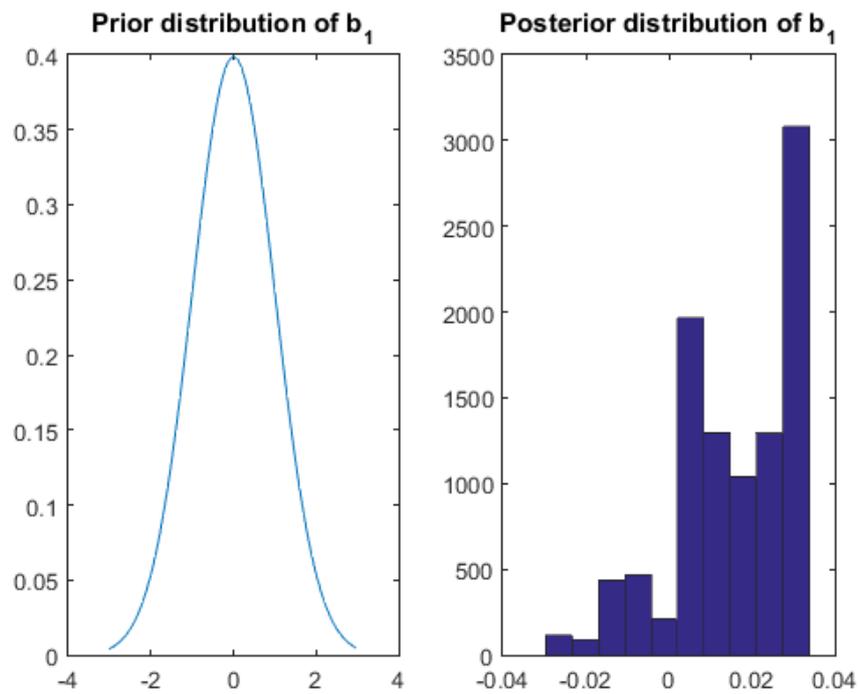


Figure 5.11: Marginal prior and posterior distribution of b_1 , for the kink model.

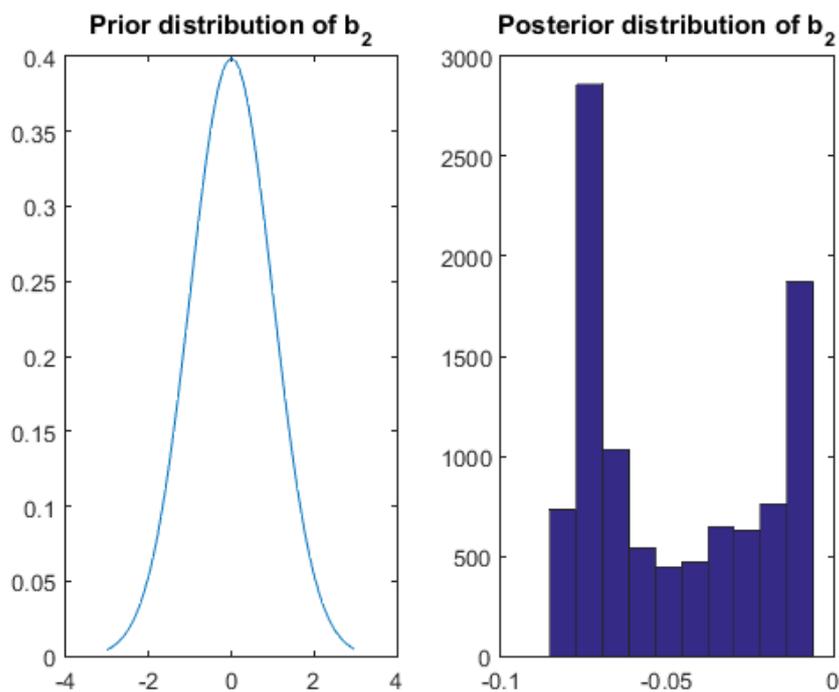


Figure 5.12: Marginal prior and posterior distribution of b_2 , for the kink model.

5.4 Comparing Classical and Bayesian results

One of the main purposes of this thesis is the verification (or not) of the classical results of this application, from the scope of bayesian theory.

According to Hansen (2017) and as we presented in subsection 5.2 the linear relationship connecting the real GDP growth of year i , with the previous year DGP growth and the rate of debt to GDP in the previous year, depends on the latter's rate value. Namely, the underlying model is a threshold regression model, where debt/GDP is the threshold variable. From the scope of classical theory it is estimated that the change of the linear equation describing the data among the dependent and the independent variables occurs when the threshold variable equals to 43.8. The respective estimators of the coefficients are also calculated in subsection 5.2.

Bayesian techniques come then to verify the above result, indicating that not only the threshold regression model describes the relationship of the regressors with the dependent variable, but this happens with probability that is numerically equal to one. Using as prior information the indication of Hansen (2017) about the threshold variable and the conditional conjugate distributions for the other parameters, we managed to get, utilizing the Gibbs sampler, a brief idea of the parameters' posterior distributions.

Chapter 6

Conclusion

This dissertation develops classical and Bayesian methods for the linear regression model, the threshold regression model with one threshold variable and the threshold regression model with two threshold variables. For the case of one threshold variable, the cases of the "jump" and "kink" models are examined separately from the scope of classical theory, since Bayesian inference is not affected from this discrimination.

Classical hypotheses testing is accomplished using two Bootstrap algorithms in the case of the discontinuous threshold regression model, since the sampling distribution of the likelihood ratio statistic appears to be non standard. For the hypotheses testing in the case of the continuous threshold regression model, one does not need to apply neither of the two Bootstrap algorithms, since the distribution of the likelihood ratio statistic is a chi-square. From the scope of Bayesian theory, a lot of attention is paid in the selection of the parameters' prior distributions and the conditional conjugacy. Also, Gibb's sampler is utilized in order to simulate draws from the marginal posterior distributions of each model's parameters.

The models and methods studied in this dissertation were applied to simulated and real data, relevant comparisons were made and useful conclusions were drawn.

Appendix A

The Bootstrap Technique

Initiated by Bradley Efron in 1979, the bootstrap method lies on the fact that an observed sample of data carries all the available information about the underlying unknown sampling distribution. Thus, resampling the original sample is the best way to get an approximation of the unknown sampling distribution.

Let $X = (X_1, X_2, \dots, X_n)$ denote a random sample of size n observed under a totally unknown probability distribution F , and $x = (x_1, x_2, \dots, x_n)$ be its observed realization. Given a statistical function $\hat{\theta} = f(X, F)$ (which can be a function both of the sample and the unknown distribution F , but not of the unknown parameter θ) that estimates a parameter θ , we aim to estimate the sampling distribution of $\hat{\theta}$. Note that if the distribution F were known, we could get an estimation for the sampling distribution of $\hat{\theta}$, using Monte Carlo methods as follows:

1. Draw a sample $X^{(j)} = (X_1^{(j)}, X_2^{(j)}, \dots, X_n^{(j)})$ from F .
2. Compute $\hat{\theta}^{(j)} = f(X^{(j)}, F)$.
3. Repeat this procedure J times and get the samples $X^{(1)}, X^{(2)}, \dots, X^{(J)}$.

Then $P(\hat{\theta} \in A) = P^*(\hat{\theta} \in A)$, where $P^*(\hat{\theta} \in A) = \frac{1}{J} \sum_{j=1}^J 1_A(\hat{\theta}^{(j)})$.

However, the sampling distribution is unknown and thus the idea is to resample the original sample in order to get the so-called *empirical distribution* F_e . This empirical distribution function $F_e(x)$, namely the probability that a datum value is less than or equal to x is defined as the proportion of the n observed data values that are less than or equal to x . Hence,

$$F_e(x) = \frac{\sum_{i=1}^n 1_{x_i \leq x}}{n}$$

From the above definition comes out that F_e is the distribution function of a random variable X_e , which takes on any of the n values (x_1, x_2, \dots, x_n) with probability $\frac{1}{n}$. (If the values of x_i are not all distinct then X_e equals to x_i with probability equal to the number of times that the value of x_i occurs in the sample divided by n .)

According to the strong law of large numbers, $F_e(x)$ converges to $F(x)$ with probability 1 as $n \rightarrow +\infty$ and moreover, this convergence will be uniform in x , with probability 1 due to the Glivenko-Cantelli theorem. Thus, assuming that θ is, in some sense, a continuous function of the distribution, $\theta(F)$ will be close to $\theta(F_e)$.

Sampling from the empirical distribution. Generating a random variable X from the empirical distribution is quite easy, since such a variable as referred above can take on equally likely (or with probability multiple of $1/n$ for values that occur more than once in the original sample) one of the values (x_1, x_2, \dots, x_n) . Therefore, if we want to

draw n i.i.d realizations from F_e , we follow the below procedure:

1. Generate n random numbers U_1, U_2, \dots, U_n and set $I_j = \text{Int}(nU_j) + 1$ for $j = 1, 2, \dots, n$, where $\text{Int}(k)$ is the integer part of k (i.e the largest integer less than or equal to k).

2. Set $x_j^* = x_{I_j}$ and $x^* = (x_1^*, x_2^*, \dots, x_n^*)$.

Then, x^* is an i.i.d sample consisting of n realizations of the random variable X having cumulative distribution function the empirical distribution F_e .

(It is easy to see that this is in essence sampling with replacement from the original sample and will work even when x_i are not all distinct.)

The bootstrap technique. Suppose that we have in hand the realization $x = (x_1, x_2, \dots, x_n)$ of a random sample $X = (X_1, X_2, \dots, X_n)$ from an unknown distribution F , θ is a parameter of the distribution and $\hat{\theta} = f(X, F)$ estimates θ . We wish, for purposes of statistical inference, to estimate the sampling distribution of θ . The bootstrap method is extremely simple, at least in principle:

1. Construct the empirical probability distribution F_e , as defined above.

2. Draw a random sample of size n from F_e , say $X^* = x^*$, $X_i^* \sim F_e$. This is the bootstrap sample $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ and its realization $x^* = (x_1^*, x_2^*, \dots, x_n^*)$.

3. Approximate the sampling distribution of $\hat{\theta} = f(X, F)$ by the bootstrap distribution of $\hat{\theta}^* = f(X^*, F_e)$.

For the calculation of the bootstrap distribution, i.e the distribution of $\hat{\theta}^*$, three methods are possible:

1. Direct theoretical calculation.

2. Monte Carlo approximation

3. Taylor series expansion methods, which can give an approximation of mean and variance of the bootstrap distribution of $\hat{\theta}^*$.

Monte Carlo approximation method is the one used more often and also at the present thesis, so we will only analyze this method.

As mentioned in step 3 of the bootstrap method, we aim to approximate the sampling distribution of $\hat{\theta} = f(X, F)$ by the bootstrap distribution of $\hat{\theta}^* = f(X^*, F_e)$ but now the distribution F_e is considered known. Thus, Monte Carlo method is used as referred at the begining of this appendix. More specifically, we follow the next steps:

1. Draw J independent samples $X^{*(1)}, X^{*(2)}, \dots, X^{*(J)}$ from the empirical sample distribution F_e , where

$$X^{*(j)} = (X_1^{*(j)}, X_2^{*(j)}, \dots, X_n^{*(j)}), X_i^{*(j)} \sim F_e \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, J$$

2. Evaluate bootstrap replications and get

$$\hat{\theta}^{*(j)} = f(X^{*(j)}, F_e), \quad j = 1, 2, \dots, J$$

3. Estimate the sampling distribution of $\hat{\theta}$ by the empirical distribution of the bootstrap replications as:

$$P^*(\hat{\theta} \in A) = \frac{1}{J} \sum_{j=1}^J 1_A(\hat{\theta}^{*(j)})$$

Appendix B

The Gibbs Sampler

Markov Chain Monte Carlo (MCMC) methods

Markov Chain Monte Carlo (MCMC) methods, largely developed in the nineties, are techniques for sampling from probability distributions using Markov chains. These algorithms, that are mostly used in data modeling for bayesian inference and numerical integration, allow us to handle problems of extreme difficulty that were almost impossible to handle before. Suppose that we aim to simulate the value of a random vector X , whose components are dependent. The idea of simulating such a complex system using a Markov chain was proposed in Metropolis et al. (1953) and has been used extensively in statistical physics. Then it was generalized by Hastings (1970) and thus we have the Metropolis-Hastings algorithm which is the base of the MCMC techniques, since all other MCMC methods are considered as special cases of it.

Markov Chains

Consider the sequence of random variables $\{X_0, X_1, \dots\}$, where X_i is interpreted as the state of the system at time i . Suppose that this system has large but finite number of states, i.e the possible values of X_i is the set $1, 2, \dots, N$, and p_{ij} , $i, j = 1, 2, \dots, N$ is the probability of the process moving from state i to state j independently of the past states. Namely, $P(X_{n+1} = j | X_0, X_1, \dots, X_n = i) = P(X_{n+1} = j | X_n = i) = p_{ij}$ for $i, j = 1, 2, \dots, N$. Then, the sequence $\{X_n, n \neq 0\}$ constitutes a *Markov chain*, with transition probabilities p_{ij} , $i, j = 1, 2, \dots, N$. Obviously, the process after leaving state i must enter some other state the transition probabilities satisfy

$$\sum_{j=1}^N p_{ij} = 1, \quad \forall i = 1, 2, \dots, N$$

A state j of a Markov chain is said to be accessible from state i ($i \rightarrow j$) if the process starting from state i will ever enter state j , while a state i is said to communicate with state j ($i \leftrightarrow j$) if both $i \rightarrow j$ and $j \rightarrow i$. Hence, a Markov chain is said to be irreducible if all of the states communicate, namely if it is possible to get from any state to any state. For an irreducible Markov chain, let π_j denote the proportion of time that the process is in state j . Vector $\underline{\pi} = (\pi_1, \pi_2, \dots, \pi_N)$ is called a stationary distribution of the Markov chain if $\forall j$ satisfies

$$\pi_j = \sum_{i=1}^N \pi_i p_{ij}, \quad j = 1, 2, \dots, N$$
$$\sum_{j=1}^N \pi_j = 1$$

A Markov chain is called aperiodic if all of its states are aperiodic, namely if for some state j

$$P(X_n = j | X_0 = j) > 0 \text{ and } P(X_{n+1} = j | X_0 = j) > 0$$

Regarding an irreducible Markov chain, if one of its states is proved to be aperiodic then it implies that all of the other states are too. For an aperiodic and irreducible Markov chain π_j can be interpreted as the limiting probability that the chain is in state j .

Supposing that we want to simulate from the probability mass function $P(X = j) = p_j, j = 1, 2, \dots, N$, the idea is to generate an aperiodic and irreducible Markov chain with limiting probabilities $p_j, j = 1, 2, \dots, N$. Then, if we run the chain for n steps, and large n , we will be able to generate the value of a random variable X , having probability mass function $p_j, j = 1, 2, \dots, N$. We will next see that the Metropolis-Hastings algorithm accomplishes this task.

The Metropolis-Hastings algorithm

Suppose that $\pi_j, j = 1, 2, \dots, N$ is a probability mass function and that we wish to generate a random variable from it. The desired probability function from which we want to simulate is called target distribution. The Metropolis-Hastings algorithm accomplishes to simulate random variables from the target distribution by simulating a time-reversible Markov chain which has limiting probabilities π_j .

Let Q denote the transition probability matrix of an irreducible Markov chain $\{X_n, n \neq 0\}$. Now suppose that the system's state in time n is $X_n = i$ and a new variable X having probability mass function $q_{ij} = P(X = j), j = 1, 2, \dots, N$ is generated. The value of this random variable X is now the candidate next state of the Markov chain. More specifically, if $X = j$ then with some known probability $\alpha(i, j)$ the system moves to the candidate state j , i.e. $X_{n+1} = j$, and remains to state i , i.e. $X_{n+1} = i$, with probability $1 - \alpha(i, j)$. Iterating this procedure, we manage to construct an irreducible and aperiodic Markov chain $\{X_n, n \neq 0\}$ with transition probabilities

$$p_{ij} = \begin{cases} q_{ij}\alpha(i, j) & j \neq i \\ q_{ii} + \sum_{k \neq i} q_{ik}(1 - \alpha(i, k)) & j = i \end{cases}$$

This Markov chain is time-reversible and has stationary probabilities π_j if the following equalities are satisfied

$$\begin{aligned} \pi_i p_{ij} &= \pi_j p_{ji} \text{ for } j \neq i \\ \iff \pi_i q_{ij} \alpha(i, j) &= \pi_j q_{ji} \alpha(j, i) \end{aligned}$$

One can easily see that the last equation is satisfied if we set

$$\alpha(i, j) = \min \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right)$$

Thus, the Metropolis-Hastings algorithm for generating a time-reversible Markov chain with limiting probabilities $\pi_j, j = 1, 2, \dots, N$ is:

1. Choose an irreducible Markov chain with transition matrix Q and transition probabilities $q_{ij}, i, j = 1, 2, \dots, N$.
2. Set $n = 0$ and $X_0 = k$, an arbitrarily chosen initial state.
3. Generate a random variable X , where $P(X = j) = q_{X_n j}$, and a random number U .

X is the candidate next state of the Markov chain.

4. Set $NS = X$ if $U < \frac{\pi_X q_{X X_n}}{\pi_{X_n} q_{X_n X}}$ (the candidate state is accepted), else $NS = X_n$ (the candidate state is not accepted).
6. $n = n + 1$, $X_{n+1} = NS$.
7. Go to step 3.

The Gibbs Sampler

A special case of the Metropolis-Hastings algorithm is the Gibbs sampler which simplifies a complex high-dimension problem breaking it down into simple, low-dimension problems. The basic idea of the Gibbs sampler is that if we wish to generate a random vector $X = (X_1, X_2, \dots, X_n)$ from a high-dimensional joint distribution, we shall break it down into a series of samples of simpler conditional distributions. Let π_x denote the probability mass function of X and $P(X = x) = P(X_i = x | X_j = x_j, j \neq i)$ is a known mass function from which we can generate a random variable X .

The Gibbs sampler utilizes the Metropolis-Hastings algorithm as follows:

1. Create a Markov chain with states $\mathbf{x} = (x_1, x_2, \dots, x_n)$.
2. Define the transition probabilities as follows: if the chain's present state is \mathbf{x} , choose equally likely one of the $1, 2, \dots, n$ coordinates. If coordinate i is chosen, generate a random variable X having probability mass function $P(X = x) = P(X_i = x | X_j = x_j, j \neq i)$.
3. If $X = x$, the candidate next state is $\mathbf{y} = (x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$ and $q_{\mathbf{y}\mathbf{x}} = \frac{1}{n} P(X_i = x | X_j = x_j, j \neq i) = \frac{\pi_{\mathbf{y}}}{nP(X_j = x_j, j \neq i)}$
4. The candidate next state is accepted with probability

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \left(1, \frac{\pi_{\mathbf{y}} q_{\mathbf{y}, \mathbf{x}}}{\pi_{\mathbf{x}} q_{\mathbf{x}, \mathbf{y}}} \right) = \min \left(1, \frac{\pi_{\mathbf{y}} \pi_{\mathbf{x}}}{\pi_{\mathbf{x}} \pi_{\mathbf{y}}} \right) = 1$$

In essence, Gibbs sampler is a special case of the Metropolis-Hastings algorithm, where the proposed next state of the Markov chain is always accepted.

The Gibbs Sampler for Bayesian inference

As referred in previous chapters, in Bayesian inference the posterior distribution carries all the available information about the parameters, or as it is usually said "the posterior distribution is the inference". However, in multiparameter problems it is difficult to calculate the marginal posterior distributions in order to examine each parameter separately. For that reason, we wish to obtain through simulation samples from the marginal distribution and Gibbs sampler accomplishes this task under conditional conjugacy. *Conditional conjugacy* is called the phenomenon where the conditional posterior distributions belong to the same families as the priors. Under this assumption of conjugate conditional posteriors, the algorithm for sampling from the multivariate distribution $\pi(\theta_1, \theta_2, \dots, \theta_d)$ is as follows:

1. Select initial values for the parameters, i.e from priors, $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)})$.
2. Simulate $\theta_1^{(1)}$ from the conditional distribution $\pi(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_d^{(0)})$.
3. Simulate $\theta_2^{(1)}$ from the conditional distribution $\pi(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_d^{(0)})$.
4. ...
5. Simulate $\theta_d^{(1)}$ from the conditional distribution $\pi(\theta_d | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{d-1}^{(1)})$.
6. Iterate this procedure.

After a burn-in period, it is proven that the above Markov chain converges to the desired stationary distribution $\pi(\theta_1, \theta_2, \dots, \theta_d)$. Thus, having discarded the samples generated in the burn-in period, the subsequent draws $(\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_d^{(1)}), \dots, (\theta_1^{(J)}, \theta_2^{(J)}, \dots, \theta_d^{(J)})$ can be regarded as realizations from the joint posterior distribution. Moreover, with Gibbs sampler we manage to get samples from the marginal posterior distributions, since the i -th component of each of the draws $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(J)}$ constitutes a sample from $\pi(\theta_i | \underline{y})$.

Bibliography

- [1] Thomas Augustin and Gero Walter. Bayesian linear regression- Different conjugate models and their (in)sensitivity to prior-data conflict. Department of Statistics, University of Munich, 2009.
- [2] George E.P. Box and George C. Tiao. *Bayesian inference in Statistical analysis*. Addison-Wesley publishing Company, Apr. 2012.
- [3] Mehmet Caner and Bruce E. Hansen. Threshold autoregression with a unit root. *Econometrica*, 69(6):1555–1596, Nov. 2001.
- [4] Husni A. Charif and Mohamed A. Ismail. Bayesian inference for threshold moving average models. *METRON- International Journal of Statistics*, LXI(1):119–132, Nov. 2002.
- [5] Cathy W.S Chen and Jack C. Lee. Bayesian inference of threshold autoregressive models. *Journal of time series analysis*, 16(5), Nov. 1993.
- [6] Cathy W.S Chen and Jack C. Lee. Bayesian inference of threshold autoregressive models. *Journal of time series analysis*, 16(5), Nov. 1993.
- [7] Terence Tai-Leung Chong and Isabel Kit-Ming Yan. Estimating and testing threshold regression models with multiple threshold variables. The Chinese University of Hong Kong, Mar. 2014.
- [8] Elizabeth A. Peck Douglas C. Montgomery and G. Geoffrey Vining. *Introduction to Linear Regression Analysis, 5th Edition*. WILEY, Apr. 2012.
- [9] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, Vol. 7(No. 1):1–26, Jan. 1979.
- [10] B. Efron. *The Jackknife, the Bootstrap and other resampling plans*. CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia: Society for Industrial and Applied Mathematics (SIAM), Jan. 1982.
- [11] Pedro E. Ferreira. A Bayesian analysis of switching regression model: known number of regimes. *Journal of the American Statistical Association*, 70(350):370–374, Jun. 1975.
- [12] Michael Goldstein and David Wooff. *Bayes linear Statistics, Theory and methods*. Wiley series in Probability and Statistics, 2007.
- [13] Jesus Gonzalo and Jean-Yves Pitarakis. Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, Vol. 7(No. 1):319–352, Oct. 2002.
- [14] Bruce E. Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, Vol. 64(No. 2):413–430, Mar. 1996.

- [15] Bruce E. Hansen. Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics*, Vol. 2(No. 1), Apr. 1997.
- [16] Bruce E. Hansen. Threshold effects in non-dynamic panels: Estimation, testing and inference. Boston College, May 1997.
- [17] Bruce E. Hansen. Sample splitting and threshold estimation. *Econometrica*, Vol. 68(No. 3):575–603, May 2000.
- [18] Bruce E. Hansen. Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, Vol. 68(No. 3):228–240, Mar. 2017.
- [19] Jungyoon Lee Javier Hidalgo and Myung Hwan Seo. Robust inference and testing of continuity in threshold regression models. Feb. 2017.
- [20] Haiqiang Chen Jushan Bai and Terence Tai-Leung Chong. Theory and applications of TAR model with two threshold variables. *Econometric Reviews*, 31(2):142–170, 2012.
- [21] Oleg Korenok. Bayesian methods in non-linear time series. Sep. 2007.
- [22] K. S. Lim and H. Tong. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B*, Vol. 42(No. 3):245–292, 1980.
- [23] Pi-Erh Lin. Some characterizations of the multivariate t distribution. *Journal of Multivariate Analysis*, pages 339–344, 1972.
- [24] D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society*, 34(1):1–41, 1972.
- [25] O. Lintonz and M. Seoy. A smoothed least squares estimator for threshold regression models. London School of Economics and Political Science, Oct. 2005.
- [26] Loukia Meligkotsidou. Bayesian inference. 2008.
- [27] Loukia Meligkotsidou. Bayesian inference II. 2014.
- [28] Martin Sola Michael J. Dueker, Zacharias Psaradakis and Fabio Spagnolo. Multivariate contemporaneous-threshold autoregressive models. Mar. 2010.
- [29] Carmen M. Reinhart and Kenneth S. Rogoff. Growth in a time of debt. *American Economic Review*, 100(2):339–344, May 2010.
- [30] Sheldon M. Ross. *Simulation: 4th (fourth) Edition*. Elsevier Academic Press, Aug. 2006.
- [31] Howell Tong. Threshold models in time series analyss- 30 years on. The University of Hong Kong, Department of Statistics and Actuarial Science, Apr. 2010.
- [32] Ruey S. Tsay. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, Vol. 84(No. 405):231–240, Mar. 1989.