THE TURKISH CATASTROPHE INSURANCE POOL CLAIMS: MODELING 2000-2008 DATA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÖZDE SARIBEKİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

JANUARY 2013

Approval of the thesis:

THE TURKISH CATASTROPHE INSURANCE POOL CLAIMS MODELING 2000-2008 DATA

submitted by **GÖZDE SARIBEKİR** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department**, **Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** ____________

Prof. Dr. İnci BATMAZ
Head of Department, **Statistics** ____________

Assist. Prof. Dr. Berna Burçak Başbuğ Erkan
Supervisor, **Statistics Dept., METU** ____________

Examining Committee Members:

Assoc. Prof. Dr. A. Sevtap Kestel
Actuarial Sciences Dept., METU ____________

Assist. Prof. Dr. Berna Burçak Başbuğ Erkan
Statistics Dept., METU ____________

Prof. Dr. İnci Batmaz
Statistics Dept., METU ____________

Assoc. Prof. Dr. Özlem İlk
Statistics Dept., METU ____________

Assoc. Prof. Dr. Vilda Purutçuoğlu Gazi
Statistics Dept., METU ____________

Date: 18.01.2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Gözde SARIBEKİR

Signature:

ABSTRACT

THE TURKISH CATASTROPHE INSURANCE POOL CLAIMS: MODELING 2000-2008 DATA

Sarıbekir, Gözde
M.Sc., Department of Statistics
Supervisor: Assist. Prof. Dr. Berna Burçak Başbuğ Erkan

January 2013, 93 pages

After the 1999 Marmara Earthquake, social, economic and engineering studies on earthquakes became more intensive. The Turkish Catastrophe Insurance Pool (TCIP) was established after the Marmara Earthquake to share the deficit in the budget of the Government. The TCIP has become a data source for researchers, consisting of variables such as number of claims, claim amount and magnitude. In this thesis, the TCIP earthquake claims, collected between 2000 and 2008, are studied. The number of claims and claim payments (aggregate claim amount) are modeled by using Generalized Linear Models (GLM). Observed sudden jumps in claim data are represented by using the exponential kernel function. Model parameters are estimated by using the Maximum Likelihood Estimation (MLE). The results can be used as recommendation in the computation of expected value of the aggregate claim amounts and the premiums of the TCIP.

Keywords: Earthquake, Claims, Generalized Linear Models, Exponential Kernel Function

ÖZ

2000-2008 YILLARI ARASINDAKİ DOĞAL AFET SİGORTALARI KURUMU TAZMİNAT VERİ MODELLEMESİ

Sarıbekir, Gözde
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Berna Burçak Başbuğ Erkan

Ocak 2013, 93 Sayfa

1999'da yaşanan Marmara Depremi'nden sonra, depremler üzerine yapılan sosyal, ekonomik ve mühendislik çalışmaları daha da önem kazanmıştır. Doğal Afet Sigortaları Kurumu (DASK), Marmara Depremi'nden sonra devletin bütçe açığını paylaşmak için kurulmuştur. DASK, araştırmacılar için hasar sayısı, tazminat tutarı ve deprem büyüklüğü (magnitüd) gibi değişkenlerle bir veri kaynağı haline gelmiştir. Bu tezde, 2000 ve 2008 yılları arasında DASK'tan elde edilen deprem hasar verileri üzerinde çalışılmıştır. Genelleştirilmiş Doğrusal Modeller kullanılarak hasar sayıları ve tazminat ödemelerinin modellemesi yapılmıştır. Hasar verilerinde görülen ani sıçramaları temsil etmesi için üstel çekirdek fonksiyonu kullanılmıştır. Model parametreleri Maksimum Olabilirlik Tahmini kullanılarak tahmin edilmektedir. Sonuçların, DASK'ın hasar tazminatlarının beklenen değerini ve primlerini hesaplamada kullanılması önerilebilir.

Anahtar Kelimeler: Deprem, Hasar, Genelleştirilmiş Doğrusal Modeller, Üstel Çekirdek Fonksiyonu

To my family,

for

their unconditional love

and support…

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor, Assist. Prof. Dr. Berna Burçak Başbuğ Erkan, who accepted me as her M.Sc. student for her guidance, encouragement and support. She always answers my questions with patient and shares her knowledge with me throughout the research.

I would also like to thank to my examining committee, Assoc. Prof. Dr. Sevtap Kestel, Prof. Dr. İnci Batmaz, Assoc. Prof. Dr. Özlem İlk and Assoc. Prof. Dr. Vilda Purutçuoğlu Gazi for their valuable discussions, comments and suggestions. Their help have improved the quality of my thesis.

I am also thankful to Dr. İbrahim Erkan for his relevant discussions, suggestions and interest on my thesis progress.

I also want to express my thanks to my manager Ahmet Hızarcıoğlu and my colleagues; Aybike Var Yazıcı, Rıza Tüzün, Cihat Temel and Harun Arat at the Turkish Employment Agency for their support, understanding and assistance during my thesis writing.

I owe special thanks to Ceyda Yazıcı, M. Tuğba Erdem and Gül İnan for their suggestions, guidance, everlasting patience and kindness during my thesis. They were always very helpful and answered my questions during this study.

I also owe a special thanks to my dear friend Başak Bulut Karageyik for her friendship and kindness. She helped me with her computational knowledge.

But most of all, I would like to thank my parents. I greatly appreciate their support, encouragement, and everlasting patience, unconditional and endless love. They always listened to my complaints and gave me advice, not to mention their tolerance during stressful times.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Natural disasters such as floods, landslides, rock falls, avalanches and earthquakes can cause significant damage to social, economical, environmental and cultural system in many countries with the impact of globalization. Over the past few decades, Turkey has suffered from natural disasters, mainly from earthquakes. Approximately 98% of Turkey's population lives on earthquake risk areas (Pampal et al., 2009). Social scientists conducted research regarding the impact of an earthquake in the community. Construction areas on fault lines, choices of materials used in construction have become a more important research area for civil and geological engineers. From a statistical point of view, earthquakes are high severity and low frequency events.

The Turkish Catastrophe Insurance Pool (TCIP) collects Compulsory Earthquake Insurance (CEI) claims data since 2000. This data can be modeled statistically and in terms of actuarial analysis. Research estimates will bring new perspectives to mitigate and to reduce the possible future losses caused by earthquakes. The aim of this thesis is to construct models for aggregate claim amount and number of claims. The data belongs to the period of 2000-2008.

Pampal et al. (2009) is a significant source to understand earthquake mechanism. The development of the TCIP System and probabilities are expressed by Akın (2008) in Chapter 2 of this thesis. The TCIP premium calculations are studied by Yücemen (2005). He checks the validity of current CEI tariff rates. Yücemen et al. (2009) extend the previous study of Yücemen (2005) by concentrating on premium rates for reinforced concrete and masonry buildings. In both studies, they find higher premium rates than the ones, which are in use.

The aggregate claim amount and the number of claims have a wide use in statistical and actuarial context. Başbuğ (2006 and 2007) had also studied the TCIP claims data modeling for the period 2000-2003 and given important aspects for premium calculations, reserving and disaster management in Turkey. Hekman et al. (1983) studied excess pure premiums and the aggregate claim distributions for collective risk models. In actuarial science, Generalized Linear Modeling (GLM) has a wide use for modeling claims data. In this study, the aggregate claim amount and the number of claims are also modeled by GLM. The Poisson regression is used to model the number of claims by using the log-link function and the aggregate claim amount, which is assumed to be Lognormal and modeled as Gaussian with identity link function. Jong et al. (2008) and Boland (2006) are great sources for GLM insurance data and other statistical methods. Diggle et al. (2002) also examined GLM and Poisson regression in Chapter 14. Boucher et al. (2008) studied on insurance time dependent claim counts data that is panel data and obtained a better fit with random effect models for premium calculations. Moreover, the claim number process and the accumulated claim process mentioned in Bülhmann (2005) (Chapter 2), Rotar (2007) (Chapter 4), Klugman et al. (2008) (Chapter 6) are sources to understand risk process. Bowers et al. (1997) in Chapter 12 and Daykin et al. (1995) in Chapters 2 and 3, study the distributions of the aggregate claims and the number of claims. Achieng (2010) discusses the distribution for claim amounts and tests the distributions such as Gamma, Weibull, Lognormal and Exponential by using Akaike's Information Criterion (AIC) and the Quantile-Quantile Plot. Claim process is an important subject for an insurance company and reveals the risk of the insurers. Mosley (2008) expressed the significance of the predictive modeling, which includes different variables such as loss, time and geography for claim process.

The organization of this thesis is as follows: Chapter 2 provides a background of the study: earthquakes, hazard profile of Turkey and the TCIP is described in detail. In Chapter 3, statistical and actuarial methodologies including individual and collective risk models, the Poisson process, the aggregate claim amount process and the distribution of aggregate claim amount are provided. Chapter 4 explains the mathematics of the study: the GLM, likelihoods of observations, the use of the exponential kernel function and the likelihood of the exponential kernel function with computations of parameter estimations of the number of claims and the aggregate claim amount models. Chapter 5 introduces explanatory data analysis and main models for the variables of interest: number of claims and aggregate claim amount. Finally, conclusion and suggestions take place in Chapter 6.

CHAPTER 2

BACKGROUND OF THE STUDY

2.1 Natural Disaster: “Earthquake”

Since prehistoric times, heavy and unbearable consequences of natural events have turned into natural disasters due to urbanization on the fault lines, bad construction and lack of disaster awareness. As a consequence, many lives have been lost due to natural disasters within the past few decades. Natural disasters can occur in many forms such as earthquakes, landslides, storm, floods and avalanche. The frequency and severity of each event differ. Earthquakes are low frequency and high severity events.

Throughout centuries, movable plates of Earth's crust caused accumulation of energy. This accumulated energy cleared and began to span from the weak areas at the breaking point of rocks, which is called “*faults.*” The movement as a result of the spread is called an “*earthquake.*” After large earthquakes (magnitude 5.0 and greater) smaller earthquakes occur and this it is called an “*aftershock.”* The number of aftershocks is large after a big earthquake and can continue for days, weeks, months or even years but in time the number declines. Smaller earthquakes in the same zone prior to main earthquake are called “*pre-shock.”* S*eismology* is the science, which analyses the occurrence of earthquakes, measurement devices (*seismograph)* and methods and earthquake wave behaviors (Pampal et al., 2009).

Intensity and magnitude are two main terms to explain an earthquake. The first intensity scale was prepared by Rossi- Forel that including 10 degrees in 1883 and has been modified since then. Today, one of the commonly used intensity scales, the Modified Mercalli (MM) Intensity Scale, is in use in The United States (see Table 2.1). It has 12 different degrees, which is represented by Roman numbers from I to XII and is used to measure the earthquake’s intensity. Also, the Medvedev-Sponheuer-Karnik (MSK) intensity scale is used in Europe. These scales are prepared according to many earthquake experiences throughout many years. The effects of an earthquake on people, goods and the environment are determined with these scales. The MSK scale is very similar to the MM scale.

Magnitude is the measurement of “the size of the earthquake.” It was first defined by Charles F. Richter in 1935 and named after him; the *Richter Scale*. It is a logarithmic scale that is based on the amplitude of the seismic waves recorded on a seismograph. There are several types of magnitudes such as *the surface-wave magnitude (Ms)*, *the body-wave magnitude (Mb)*, *the moment magnitude (Mw)* and *the local magnitude (ML)* used to calculate the earthquake’s magnitude. According to logarithmic scale one unit of changes in a magnitude causes 10 times greater ground motion and also triggers 32 times stronger energy. For example, an earthquake with a magnitude 7.6 causes 63 times greater ground motion than an earthquake with a magnitude 5.8 ($10^{7.6} \div 10^{5.8} = 10^{7.6-5.8} = 10^{1.8} = 63$). The magnitude 7.6 causes a 63 times greater devastating effect than a magnitude 5.8.

Table 2.1 The Modified Mercalli Intensity Scale Degree and Definition [41]

Modified Mercalli Intensity Scale Degree	Definition
I	Not felt.
II	Rarely felt by fewer people who are resident on upper floors.
III	Felt indoors by people.
IV	Felt indoors by many people.
V	Severity felt by majority.
VI	Felt by all resulting in panic. Also, heavy furniture movement.
VII	Causes breaks and breaches on rock walls.
VIII	Buildings partially collapse; e.g. columns, walls.
IX	Devastating damages occur and some buildings collapse. Animals start to run around randomly and make noises.
X	Rails bend, most of the buildings are destroyed.
XI	A few structures remain. Bridges, dams and rails are greatly damaged.
XII	An entire destruction has occurred. Objects are thrown into the air.

Around the world, every day approximately 50 and in a year 20,000 earthquakes are detected. In fact, there are millions of earthquakes that are not located because of lack of seismographs (USGS, 2012). It is estimated that 1,300,000 earthquakes occur annually with a magnitude between 2-2.9 and a magnitude 8.0 (Table 2.2).

Table 2.2 Number of Earthquakes According to Their Magnitudes [40]

Magnitude	Annual Average
≥8.0	1
7-7.9	15
6-6.9	134
5-5.9	1,319
4-4.9	13,000 (estimated)
3-3.9	130,000 (estimated)
2-2.9	1,300,000 (estimated)

On December 26, 2004, in Indonesia/Sumatra Island an earthquake with a magnitude of 9.3 occurred. As a result, more than 240,000 people died. Later in 2009, an earthquake with a magnitude of 7.6 occurred in the same location and this time more than 1,100 people died. There are two other earthquakes that occurred with magnitudes of 8.2 and 8.6 off the west coast of Northern Sumatra on April 11, 2012. On May 22, 1960, in Chile the biggest earthquake in the twentieth-century occurred with a magnitude 9.5. In 2011, 332 natural disasters occurred with 336.1 billion USD economic loss that killed 30,773 people and caused 244.7 million victims (the total number of killed and affected people) (CRED, 2012).

Faults are usually named according to aspects of movement. There are three kinds of faults. *Normal Faults*, where the block on fault plane moves down. The *Reverse Fault,* which is the opposite of a normal fault, where the block moves up. Third, in *Strike-Slip Fault,* where the blocks horizontally move in the opposite direction (Pampal et al., 2009) (Figure 2.1). On the other hand, some faults do not move and do not cause earthquakes for years and are considered as ‘inactive.’ However, there is no guarantee for a sudden release of accumulated energy; that is why earthquakes are unpredictable.

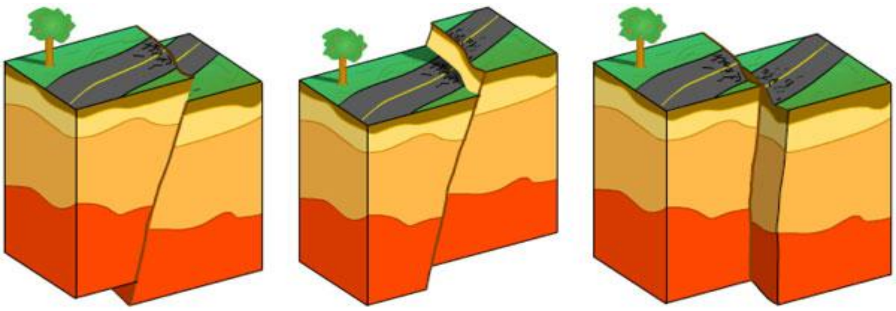


Figure 2.1 Types of Faults [42]

2.1.1 Earthquake Waves

There are two different types of seismic waves: body waves and surface waves.

Body waves: P-waves and S-waves are types of body waves. *P-Waves (Primary waves)* are the initial and fastest waves recorded in a seismograph. The vibration motion has the same direction with the expansion direction of the wave. These waves have the ability to travel through gas, liquid and solid materials with a speed of 6-13 kilometers per second. The speed changes according to features of rigidity, intensity and elasticity of ground. *S-waves (Secondary waves)* are slower than P-waves and cannot move through liquid materials but only through solid objects. The vibration motion is vertical with the expansion direction of the wave (Pampal et al., 2009).

Surface waves: There are two kinds of Surface Waves: The Rayleigh wave and the Love wave. *Rayleigh waves* travel as waves on the surface of water. *Love waves* are faster and have horizontal movement. Surface waves are slower than the body waves and travel the Earth’s surface with a speed of 1.5 kilometers per second. S-waves and surface waves cause destruction. Figure 2.2 shows the movements of the earthquake waves (Pampal et al., 2009).

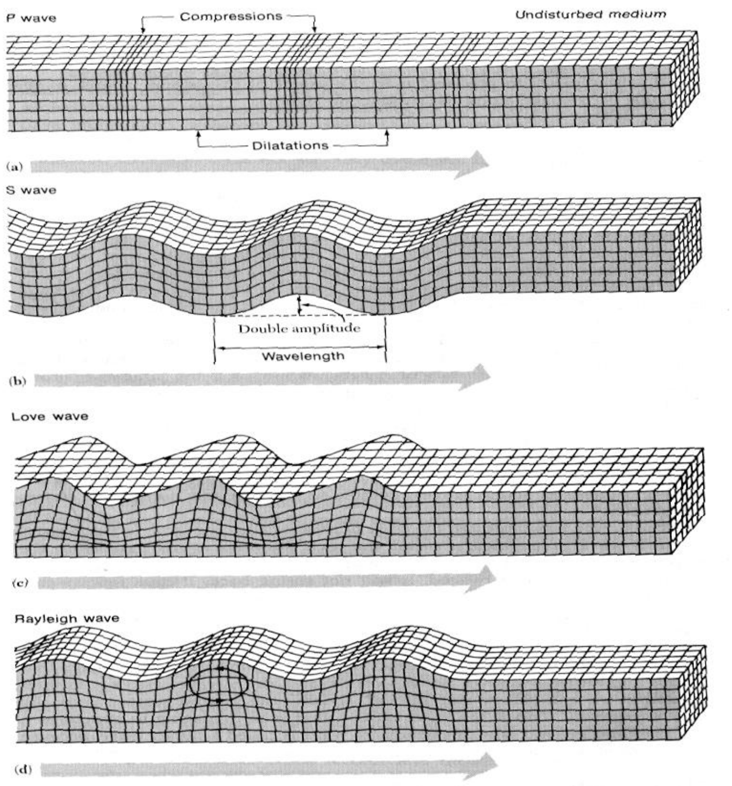


Figure 2.2 Types of Seismic Waves [13]

2.1.2 Types of Earthquakes

The types of earthquake depend on the geological area, where it happens. *Tectonic earthquakes* happen as a result of movements of tectonic plates, which compose the surface of the Earth and move against each other. Most of the earthquakes belong in this group. *Volcanic earthquakes* occur with volcanic activity. *Collapse earthquakes* are observed with the collapse of underground mines, caverns and that of melting of salty spaces of the ceiling blocks. *Explosion earthquakes* are the result of nuclear explosions (Figure 2.3).

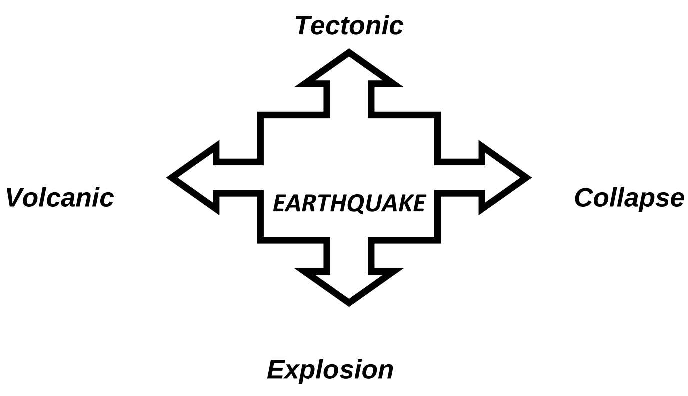


Figure 2.3 Types of Earthquakes

2.1.3 Natural Disasters in Turkey

Geological structure, tectonic formation and the meteorological features of Turkey take an important role in natural disasters. In addition to these, climate change caused by global warming prompt an increase of natural disasters. Thus, Turkey has suffered from various disasters throughout its history. Earthquakes, floods, landslides and erosion are the main natural disasters that have occurred in the twentieth century. When social and physical characteristics of a country are considered, the results become challenging. According to the Emergency Events Database (EM-DAT) between 1900, and 2012 a total of 155 events caused 26 billion USD in damage and 92,086 casualties in Turkey (Table 2.3). Table 2.4 shows the damages of each disaster and it can be seen that earthquake cause the most harmful damage. Flooding is the second most common natural disaster, which causes economic losses. However, it is possible to control flooding to minimize losses. Possible flood areas should not be chosen for settlement. Higher elevations are appropriate to prevent flooding in settlements. On July, 2012 storms in Samsun (Salıpazarı and Ayvacık) resulted in flooding causing 12 deaths. Bridges collapsed and building floors filled with water, which were near the river.

Table 2.3 Natural Disasters in Turkey from 1900-2012 [20]

		Number of Events	Killed	Total Affected	Damage (000 USD)
Earthquake (seismic activity)	Earthquake	76	89,236	6,924,005	24,685,400
Epidemic	Bacterial Infectious Diseases	1	11	150	NA
	Parasitic Infectious Diseases	2	NA	100,000	NA
	Viral Infectious Diseases	5	602	104,705	NA
Extreme temperature	Cold wave	3	69	NA	NA
	Extreme winter conditions	2	17	8,150	NA
	Heat wave	2	14	300	1,000
Flood	Unspecified	11	897	372,617	65,000
	Flash flood	10	243	1,341,382	1,892,000
	General flood	17	189	64,521	238,500
Mass movement dry	Avalanche	1	261	1,069	NA
Mass movement wet	Avalanche	2	146	6	NA
	Landslide	9	286	13,481	26,000
Storm	Unspecified	4	49	3	NA
	Local storm	5	51	13,636	2,200
Wildfire	Forest fire	5	15	1,150	NA
Total		**155**	**92,086**	**8,945,175**	**26,910,100**

Table 2.4 Damage of Buildings Caused by Natural Disasters [33]

Type of Natural Disasters	Number of Destroyed Buildings	Percentage (%)
Earthquake	495,000	76
Landslide	63,000	10
Flood	61,000	9
Rock avalanche	26,500	4
Avalanche	5,154	1
Total	**650,654**	**100**

Turkey is located on one of the most critical fault of the Alpine-Himalayan Seismic Belt, which lies between the Azure Island and Southeast Asia, where the African and Arabian plates and Eurasian plates meet. The North Anatolian Fault (NAF), The East Anatolian Fault, the Thrust belt of the Southeastern Anatolia and the Aegean Graben Systems are the plates that surround the Anatolian plate. Most of the earthquakes that occur are observed in these zones. Tectonic features are one of the main causes of devastating earthquakes in Turkey. These features caused earthquakes in the past that led to the collapse of civilizations. The archaeological ruins are the most important evidence of lost civilizations in West, Central West, North West Anatolia and the Eastern Mediterranean (Pampal et al., 2009).

The most important factor of an earthquake turning into a disaster is the preferences of human beings. Houses built on alluvial plains that are on active faults, inadequate and inappropriate materials used in the construction of houses and deficiencies in the controls are influential factors that cause great loss of life and property. In Turkey, for instance, Erzincan is a province established on an active fault. 18 devastating earthquakes occurred in its 1000 years of history. In addition, Adapazarı, Izmit, Adana, Osmaniye, Hatay and many other cities are established on alluvial plains, which are on active faults (Pampal et al., 2009).

Approximately 96% of various regions in different degrees are located on earthquake belts and 98% of the country's population lives in these risk areas. Natural disasters also cause loss of life, as well as economic losses. Statistics show that every year natural disasters cause economic losses equivalent to 1% of Turkey’s Gross National Product (Pampal et al., 2009).

Table 2.5 Important Earthquakes in Turkey [33]

Event	Date	Killed	Affected Population
Erzincan Earthquake	13/03/1992	653	250,000
Dinar Earthquake	01/10/1995	94	120,000
Çorum Earthquake	14/08/1996	0	17,000
Adana-Ceyhan Earthquake	27/06/1998	145	1,500,000
Gulf of İzmit Earthquake	17/08/1999	17,480	15,000,000
Düzce Earthquake	12/11/1999	763	600,000
Afyon-Sultandağı Earthquake	03/02/2002	42	222,000
Bingöl Earthquake	01/05/2003	177	245,000

Table 2.5 illustrates significant earthquakes in Turkey since 1992. The August 17, 1999 Marmara (Gulf of Izmit) earthquake is one of the biggest earthquakes in Turkey's recent history as well as in the world with a magnitude of 7.4 and damages of 20 billion USD (see Table 2.6). Approximately 15 million people were affected from this catastrophe. Then, an earthquake on November 12, 1999 in Düzce with a magnitude 7.2 occurred. In addition, with earthquakes on July 2, 2004 in Ağrı/Doğubayazıt with a magnitude of 5.1, on January 25, 2005 in Hakkari with a magnitude of 5.5 and on March 8, 2010 in Elazığ with a magnitude of 6.0, most of the 51 people died because of poor quality of construction. In 2011 Kütahya/Simav on May 19 with a magnitude of 5.9, on October 23, Van/ Erciş with a magnitude of 7.2 and on November 9, with a magnitude of 5.6 earthquakes occurred.

Table 2.6 Top 10 Most Important Earthquake Disasters for the Period 1900 to 2012 [20]

Country	**Date**	**Damage (000 USD)**
Japan, Tsunami	11/03/2011	210,000,000
Japan, Earthquake	17/01/1995	100,000,000
China P Rep, Earthquake	12/05/2008	85,000,000
United States, Earthquake	17/01/1994	30,000,000
Chile, Earthquake	27/02/2010	30,000,000
Japan, Earthquake	23/10/2004	28,000,000
Italy, Earthquake	23/11/1980	20,000,000
Turkey, Earthquake	**17/08/1999**	**20,000,000**
New Zealand, Earthquake	22/02/2011	15,000,000
Taiwan (China), Earthquake	21/09/1999	14,100,000

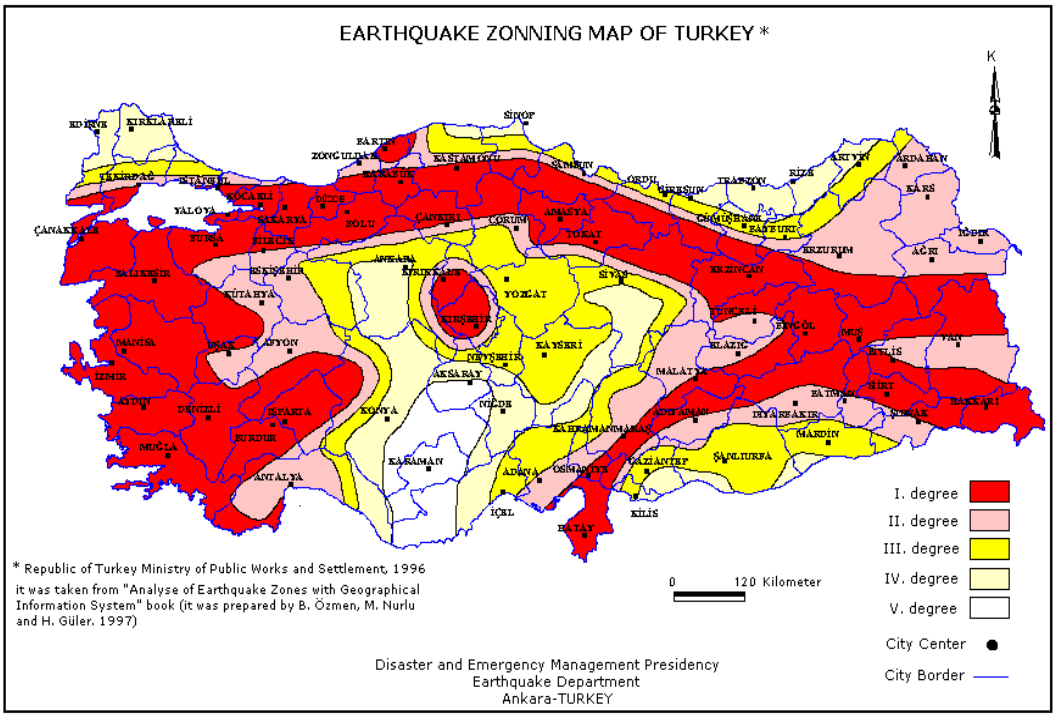


Figure 2.4 The Earthquake Hazard Zone Map [39]

Since 1945 The Earthquake Hazard Zone Map of Turkey (Figure 2.4) was updated six times (1945, 1947, 1948, 1963, 1972 and 1996). The latest map was prepared by the Earthquake Working Group of the Earthquake Research Department of the General Directorate of Disaster Affairs by changing the 1972 earthquake map and is used by the Ministry of Public Works and Settlement since 1996. Furthermore, the General Director of Mineral Research and Exploration (MTA) revised the Earthquake Hazard Zone Map (1996) in 2012. They worked on the new fault line map for 6 years. They detected new fault lines and changes the length of some faults. According to the latest map (1996), Turkey is divided into 5 zones. Expected accelerations that are more than 0.4g are in the first degree earthquake zone, between 0.3g - 0.4g in the second degree zone, between 0.2g - 0.3g in the third degree zone, between 0.2g - 0.1g in the fourth degree zone and less than 0.1g in the fifth degree earthquake zone (g: gravity (981 cm/s*s). The map shows the distribution by risk areas of Turkey. It can be seen that most of the land lies in the first and second degree earthquake zones. Constructions in high hazard zone areas should be earthquake resistant by using the classification in the map (Table 2.7).

Table 2.7 Earthquake Hazard Zones due to Acceleration (g: gravity (981 cm/s*s))

Acceleration	Earthquake Hazard Zone
> 0.4 g	1st degree
0.3g - 0.4g	2nd degree
0.2g - 0.3g	3rd degree
0.2g - 0.1g	4th degree
< 0.1g	5th degree

2.2 The Turkish Catastrophe Insurance Pool (TCIP)

The most devastating earthquake in Turkey occurred on August 17, 1999. The Marmara earthquake (Kocaeli/Izmit Earthquake) claimed many lives and resulted in significant socio-economic damage. The earthquake affected the whole country, since the Marmara region is the most developed industrial region in Turkey. The current Earthquake Hazard Zone Map of Turkey (Figure 2.4) shows that 98% of the territories of the country are under risk in different degrees. Also, 96% of the population lives in these critical areas (Özmen, 1999). In order to minimize earthquake losses many measures have been incorporated by government. One of the most important of these measures is the CEI arrangement. As a result of the disaster, there was a large amount of spending accounted by the government. Disasters like earthquakes that cause great economic damages need a large amount of guarantee with a great resource. In this situation, insurance is an important financial system to cope with this problem. Turkey received assistance from the World Bank with the Marmara Earthquake Emergency Reconstruction Project (MEER). It assisted in creating an insurance system to manage with its own catastrophic risks of Turkey and reconstruct destroyed residential buildings. The project received approval on November 16, 1999 and cost 737.1 million USD. The General Directorate of Insurance received technical and financial help to establish the TCIP.

The TCIP was established with Decree Law No. 587 for the CEI as an ”insurance pool” and is non-profitable. It is the first public-private insurance system in the world. It began to offer guarantees to houses, which are under its coverage on September 27, 2000. Since the establishment of the TCIP, it guaranteed 75 million USD for 19,270 number of claims as of August 17, 2012. The Compulsory Insurance was revised with the Disaster Insurance Law No. 6305 on May 5, 2012 and Decree Law No. 587 annulled. It became effective as of August 18, 2012. During services of water and electricity for homeowners, the CEI is compulsory. With this application, the penetration rate of the TCIP is expected to increase. CEI policies are valid for 1 year. It aims to compensate losses in one month.

The CEI was taken as a model application and system from the New Zealand Earthquake Insurance Commission (NZEQC) and the California Earthquake Authority (CEA) and implemented according to Turkey’s requirements. NZEQC with its new name the Earthquake Commission (EQC) started as a fund in 1941 and then in 1993 as a public institution, began to cover natural disasters such as earthquakes only for houses. The EQC provides mandatory earthquake coverage protection up to a certain limit like in our country. In 1996, the CEA was also formed after the 1994 Northridge Earthquake as a guarantee for its members. As a public institution, the CEA is financed by special funds except taxes.

There were many aims to contrive the TCIP, for instance, ensuring buildings, which are under coverage with a premium, spreading risk over international reinsurance companies and capital markets and decreasing the financial obligation on the government while building constructions. Moreover, the formation of insurance awareness in the community is prompted.

Technical tasks of the TCIP are operated by an insurance or reinsurance company determined by The Undersecretary of Treasury for five years. Fund of the TCIP is managed by an operational manager, who complies with the decisions of board of directors. The TCIP is exempted from tax deductions. Annual accounting, operations and tasks of the TCIP are controlled by the Undersecretary of Treasury.

The TCIP is directed by a board of directors, which consists of seven delegates, four of them are public officials from various areas (the Treasury, the Prime Ministry, the Ministry of Environment and Urban Planning and the Capital Markets Board of Turkey), two private sector delegates (Association of the Insurance and Reinsurance Companies of Turkey and an insurance company leader) and one representative from academia.

Earthquake claims vary due to the features and risks of the disaster profile of regions. Table 2.8 supports this information with realized claims. It can be observed that until 2013, the largest amount of claims belongs to the Eastern Anatolia Region with 70,055,853 USD and 10,118 number of claims. The East Anatolian and the North Anatolian Faults are in this region and as a result of Erzincan, Bingöl and Van earthquakes have affected the total claims. The lowest payments are made in Black Sea Region with 17,225 USD and 32 claims (Table 2.8).

Table 2.8 Claims According to Regions (04/03/2013) [18]

Region	Number of Earthquakes	Number of Claims	Claim Amount (TL)	Claim Amount (USD)
Other	52	582	3,254,001	1,828,090
Mediterranean	35	419	242,395	136,177
Eastern Anatolia	130	10,118	124,699,418	70,055,853
Aegean	97	8,666	23,584,342	13,249,630
Southeast Anatolia	9	47	149,712	84,108
Central Anatolia	38	242	630,824	354,396
Black Sea	11	32	30,661	17,225
Marmara	35	320	394,329	221,533
Total	**407**	**20,426**	**152,985,683**	**85,947,013**

According to the TCIP statistics in Table 2.9, during the past 13 years the highest claim amount was in 2011 (67,427,785 USD). The October 23, 2011 Van/ Erciş earthquake with a magnitude of 7.2 contributed to an important increase in claims and a total of 61,356,576 USD in claim amounts was paid to the policyholders. The second largest claim amount was observed in 2005 due to the October 21, 2005 Izmir/Seferihisar earthquake with a magnitude of 5.9 and 1,836,600 USD in claims.

Table 2.9 Claims According to the Year (until 04/03/2013) [18]

Year	Number of Earthquakes	Number of Claims	Claim Amount (TL)	Claim Amount (USD)
2000	1	6	23.022	19,185
2001	17	336	126.052	87,173
2002	21	1,558	2.284.835	1,394,042
2003	20	2,504	5.203.990	3,735,815
2004	31	587	768.927	575,415
2005	41	3,488	8.119.871	6,051,476
2006	23	500	1.303.673	927,485
2007	42	995	1.381.599	1,191,753
2008	45	481	558.849	367,229
2009	37	267	498.852	335,407
2010	36	454	715.418	465,282
2011	40	7,671	127,364,344	67,427,785
2012	51	1,569	4,624,976	2,556,368
2013	2	10	11,276	6,280
Total	**407**	**20,426**	**152,985,683**	**85,140,696**

The Coverage of the Compulsory Earthquake Insurance (CEI)

The CEI covers damages and losses caused by the earthquake. Losses of fires, explosions, tsunamis and landslides caused as a result of earthquake are also covered. The CEI is only provided for residential buildings within municipality borders. These buildings are private properties with land registry, independent areas within the context of Law No. 634 Condominium Ownership and trading firms, offices in the independent areas of these buildings. Constructions that are built by government or credit use are also within the insurance coverage (TOKI and mortgage).

Constructions that are not under coverage of the CEI (TCIP, 2012):

- Buildings of public corporations and establishments
- Buildings in rural areas
- Trading and industrial firms within or outside the context of Law No. 634 Condominium Ownership such as business centers, trading centers and administrative service buildings
- Uncompleted buildings
- Unlicensed constructions on the areas belonging to the Turkish Treasury
- Neglected and ruined buildings, which are not suitable for residence
- Buildings that are built after December 27, 1999 without a building permit

The CEI does not include all the damages and loss of the earthquake. Some exclusions are listed below (TCIP, 2012):

- Debris removal
- Loss of profit
- Loss of rent revenue
- Movable properties and look alike
- Alternative expenses of business and residence
- Moral claim amount requests
- Death and injuries
- Financial responsibilities
- Damages to constructions that occurred in time because of their own defects

The sum insured is determined according to the structure type of the construction by multiplying the gross square area of the dwelling with the square meter price (see Table 2.10). The gross square area of the dwelling is determined annually by the Turkish Statistical Institute (TurkStat). In an earthquake event without considering the construction structure, the maximum sum insured amount granted by the TCIP is 84,270 USD (150,000 TL). If a property is worth more than this, the property owner can obtain private insurance. A 2% deductible amount of the total insured value is applied for each loss. The amount above the 2% deductible amount of the total insured is paid to the insurer.

Structure types of the building and their explanations are as follows:

a. *Steel, Reinforced Concrete and Frame Structures:* Where the buildings have steel and reinforced concrete frames.

b. *Masonry Stone Structures:* Where the buildings have bearing walls made up of rubble stone, cut stone, brick or concrete blocks with spaces and without spaces, stairs and ceilings with concrete or reinforced concrete structures.

c. *Other Structures,* which are made up of adobe and wood.

Table 2.10 Square Meter Prices for Sum Insured Including Structure Types [18]

Structure Type	Square meter price for sum insured
Steel, Reinforced Concrete Frame	359.55 USD (640 TL)
Masonry Stone Structures	258.43 USD (460 TL)
Other Structures	134.83 USD (240 TL)

Also, the premium of the CEI is determined according to sum insured amount, risk zone and type of building structure. Base premium is obtained by multiplying sum insured amount with tariff rate. For risks in Istanbul a 8.43 USD (15 TL) and for other cities a 5.62 USD (10 TL) fixed premium is added to base premium to find total policy premium. Minimum premium price is 14.04 USD (25 TL) for any type of building and risk zone. There are 15 tariffs consisting of 5 risk zones and 3 types of structures. Table 2.11 provides the tariff rates.

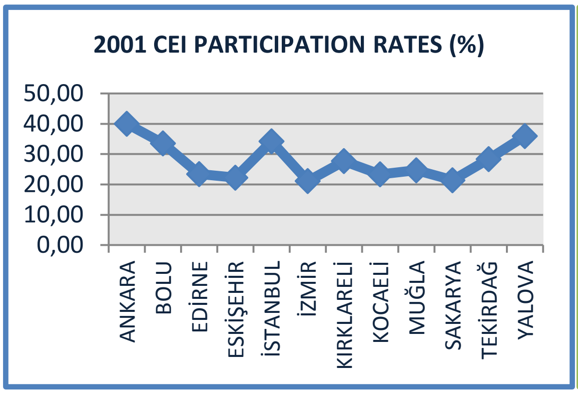
Table 2.11 Tariff Rates [18]

Region Based Rates According to Construction Type	Zone I	Zone II	Zone III	Zone IV	Zone V
A-Steel, Reinforced Concrete Frame Structures	2.20	1.55	0.83	0.55	0.44
B-Masonry Stone Structures	3.85	2.75	1.43	0.60	0.50
C-Other Structures	5.50	3.53	1.76	0.78	0.58

As observed in Table 2.12 in Turkey, Bolu province has the highest CEI participation rate with 44.48%. 95% of the city is located on 1st degree earthquake zone.

Table 2.12 Total House Number and Participation Rates for Years between 2001 and 2007

Year	City	Total Number of Houses	Participation Rate %
2001	ANKARA	902,900	40.02
2002	YALOVA	64,227	36.47
2003	YALOVA	64,227	33.96
2004	BOLU	38,918	40.68
2005	BOLU	38,918	42.80
2006	BOLU	38,918	44.48
2007	BOLU	38,918	42.79



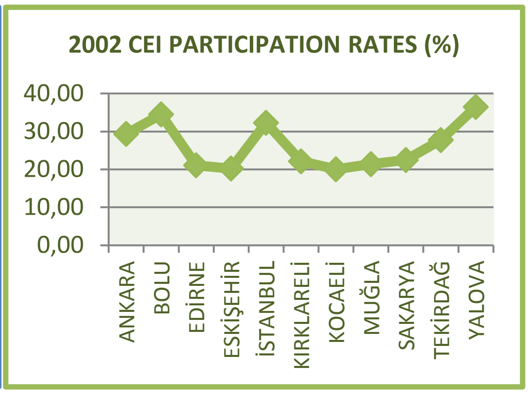
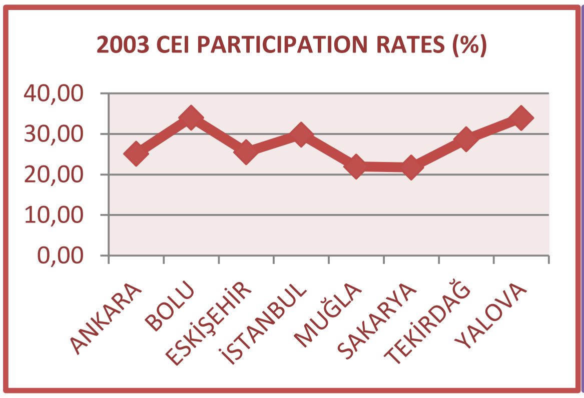


Figure 2.5 2001 CEI Participation Rates

Figure 2.6 2002 CEI Participation Rates



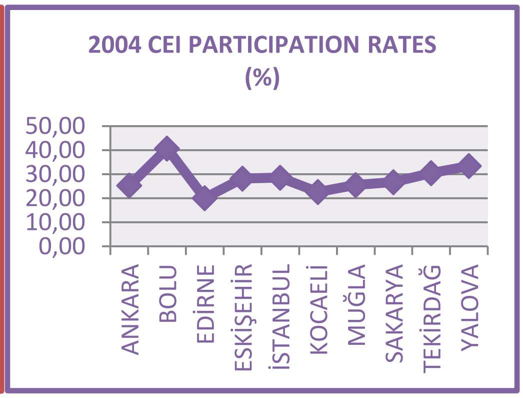
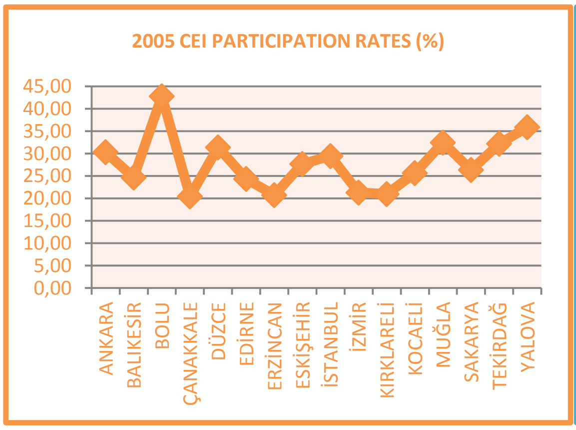


Figure 2.7 2003 CEI Participation Rates

Figure 2.8 2004 CEI Participation Rates



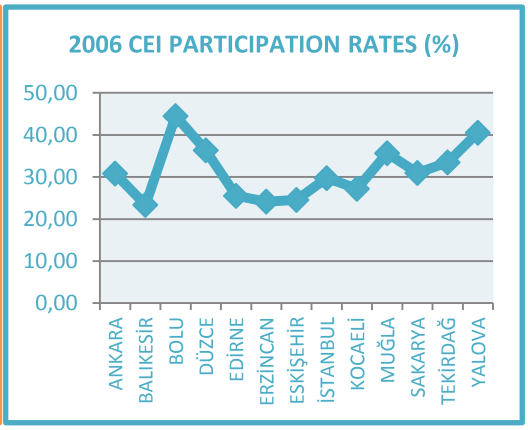


Figure 2.9 2005 CEI Participation Rates

Figure 2.10 2006 CEI Participation Rates

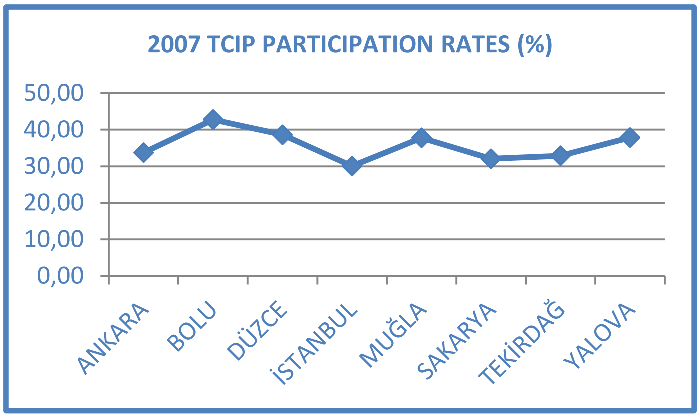


Figure 2.11 2007 CEI Participation Rates

By using the data of the thesis; Figures 2.5, 2.6, 2.7, 2.8, 2.9, 2.10 and 2.11 are generated to show the TCIP participation rates of the provinces between the years 2001 and 2007. As observed, Ankara in 2001, Yalova in 2002 and 2003 and Bolu in 2004, 2005, 2006 and 2007 had the highest rates with 40.02%, 36.47%, 33.96%, 40.68%, 42.80%, 44.48% and 42.79%, respectively. The 1999 Marmara earthquake's significant damages and the potential of future earthquakes have caused people to be more aware of the insurance. This can be observed from the data.

CHAPTER 3

METHODOLOGY

3.1 Risk Models

In actuarial analysis, two different risk models are used to build aggregate claim amount: individual and collective. These are usually used for premium calculation, reserving and reinsurance practices (Klugman, 2008 and Bowers et al., 1997). In these models, it is assumed that S is a random variable that refers to the aggregate claim amount for a portfolio. Moreover, S is the total of policy's claim amount paid by each policyholder in a time period $(0, t]$.

Individual Risk Models

In individual risk models each policy is considered one by one and aggregate claim amount S is calculated by the summation of these individual claim amounts (Bowers et al., 1997 and Kaas et al., 2001),

$$S = X_1 + X_2 + \cdots + X_n, \tag{3.1}$$

where n, number of the policies, is fixed and known during the insurance time period (Kaas et al., 2001). X_i $(i = 1, 2, \ldots, n)$ defines the claim amount of i^{th} policy and X_i's in (3.1) are assumed to be independent. History of the X_i is not needed and it is easier to calculate the aggregate claim amount, S, under these assumptions.

Collective Risk Models

In collective risk models, the aggregate claim amount is also derived with summation of each claim amount (3.2). It is assumed to be a random process and also used in this thesis for modeling the aggregate claim amount S. Then (Kaas et al., 2001 and Bülhmann, 2005),

$$S = \sum_{i=1}^{N} X_i = X_1 + X_2 + \cdots + X_N \,, \tag{3.2}$$

where N is the number of claims in a time period and a random variable. X_i $(i = 1, \ldots, N)$ is the claim amount paid by the insurer for the i^{th} claim. If $N = 0$ then $S = 0$.

Assumptions of the collective risk models are:

1. $X_1, X_2, \ldots, X_N$ are independently and identically distributed,
2. N and $X_1, X_2, \ldots, X_N$ are mutually independent.

Since the claim amounts are independent from each other and identically distributed, also do not have an effect on the claim numbers, both the claim amount and the number of claims are assumed to have their own distribution functions.

3.2 The Poisson Process

The claim number process N_t $(N_t, t \geq 0)$, which is a Poisson counting process, where interarrival times are independent and identically distributed exponentially with parameter λ in a given time interval $(0, t]$. It is a discrete random variable and depends on changes in time. Interarrival times are memoryless and also independent of history of the process. This process has a wide use in insurance and actuarial applications (Klugman et al., 2008, Rolski et al., 2000, Kaas et al., 2001, Bülhmann, 2005 and Rotar, 2007).

There are three properties of the process N_t:

1. $N_0 = 0$,
2. N_t has *independent increments*; for $0 = t_0 < t_1 < t_2 < \cdots < t_n$, the random variables $N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent. Also, it has *stationary increments*; the distribution of $N_t - N_s$ for $t > s$ depends only on the length of the time interval $t - s$; for same length time intervals increments are identically distributed,
3. The number of claims is Poisson distributed with the mean $\lambda(t)$ in time interval $t > 0$.

That is, for all $s, t > 0$ it has,

$$N_{t+s} - N_s \sim Poisson(\lambda(t))$$

$$P(N_{t+s} - N_s = n) = \frac{e^{-\lambda(t)}(\lambda(t))^n}{n!} \quad , \quad n = 0, 1, 2 \ldots \tag{3.3}$$

In Poisson process, N_t is a *renewal counting process* with t_n since it is memoryless, which restarts from any time point in time. t_n is formed by the sum of W_i s, which are independent and identically distributed.

$$t_n = W_1 + W_2 + \cdots + W_n \qquad \text{where} \quad n \geq 1.$$

Also, $t_n = W_{n+1} - W_n$ is the differences of time intervals that is called interarrival time.

Since the times of the events are independent and exponentially distributed, probability of an event can be written as,

$$P(W_n > s | History\ of\ the\ process) = P(N_{t+s} - N_s = 0 > s | History\ of\ the\ process)$$

$$= e^{-\lambda(t)}.$$

Consequently, the probability is derived as independent from the past.

In general, the λ rate of the Poisson process N_t is taken as a constant that is independent of time t. This process is known as the homogeneous Poisson process with stationary and independent increments. However, in many applications, it is appropriate to take the λ as time varying $(\lambda = \lambda(t))$ .When λ depends on both s and $t (s, t > 0)$ in the time interval, the process is called the *nonhomogeneous Poisson process* with mean $\Lambda(t) = \int_s^t \lambda(r) dr = \lambda(t) - \lambda(s)$ (Klugman et al., 2008, Bülhmann, 2005 and Rolski, 2000).

In the nonhomogeneous Poisson process $N_t \sim Poisson(\Lambda(t))$ and the expected value and the variance is equal to rate $\Lambda(t)$.

$$E(N_t) = V(N_t) = \Lambda(t). \tag{3.4}$$

Since the times $W_1, W_2, \ldots, W_n$ are the nonhomogeneous Poisson process, with the condition of W_i, the distribution function of the next event $t_i = W_{i+1} - W_i$ is (see Başbuğ, 2007),

$$F_i(t_i|W_i) = 1 - e^{-\Lambda_i(t)}, \quad \text{where } \Lambda_i(t) = \int_{W_i}^{W_{i+1}} \lambda(r)dr,$$

and when derivative of cumulative distribution function is taken, the probability density function is

$$f_i(t) = \lambda(W_{i+1})e^{-\Lambda_i(t)}. \tag{3.5}$$

In this thesis, the nonhomogeneous Poisson process is used to model the time varying claims. Since we organized data in terms of weeks, the mean of the number of claims changes in time interval and it becomes more relevant to use the nonhomogeneous Poisson process to use in the analysis.

3.3 The Aggregate Claim Amount Process

In actuarial studies the aggregate claim amount process, S_t, has a wide use and is obtained by the summation of the each policies' claim amount X_i's in a given time period. As a collective risk model (Section 3.1) the calculation of S_t depends on both the number of claims N_t arrived by time t and claim amount X_i (Bowers et al., 1997 and Rotar, 2007). It is assumed that N_t and X_i's are independent and also X_i's are identically distributed.

$$S_t = \sum_{i=1}^{N_t} X_i = X_1 + X_2 + \cdots + X_{N_t} \quad (t \geq 0), \tag{3.6}$$

where $S_t = 0$ if $N_t = 0$.

In most studies, researches prefer to use the claim number process N_t as Binomial, Negative Binomial or Poisson. When the claim number process N_t is preferred as a nonhomogeneous Poisson process with mean $\lambda(t)$, the aggregate claim amount process, S_t, $(S_t, t \geq 0)$ is called *Compound Poisson Process.* In other cases, if the claim number process N_t is Binomially or Negative Binomially distributed, S_t has *Compound Binomial or Compound Negative Binomial* distributions (Boland, 2006 and Rotar, 2007). The distribution function of aggregate claim amount process (S_t) is a main case in actuarial studies and it can be denoted by the convolution formula by using the summation of the independent claim amounts, X_is instead of S_t (Rolski et al., 2000, Kaas, 2001 and Rotar, 2007):

$$F_{S_t}(s) = P(S_t \leq s) = P(\sum_{i=1}^{N_t} X_i \leq s),$$

$$= \sum_{n=0}^{\infty} P\big(X_1 + \cdots + X_{N_t} \leq s | N_t = n\big) P(N_t = n), \tag{3.7}$$

For simple notation $F_{S_t}(s) = F(s)$ is used. Therefore,

$$F(s) = \sum_{n=0}^{\infty} P^{*k}(s) P(N_t = n), \quad f(s) = \sum_{n=0}^{\infty} p^{*k}(s) P(N_t = n), \tag{3.8}$$

where these expressions are called convolution formula and $P^{*k}(s)$ is k^{th} convolution of $P(s)$.

If the claim number process N_t is Poisson with rate $\lambda(t)$ the probability is,

$$P(N_t = n) = \frac{e^{\lambda(t)}(\lambda(t))^n}{n!},$$

and hence

$$F(s) = \sum_{n=0}^{\infty} P^{*k}(s)\ \frac{e^{\lambda t}(\lambda(t))^n}{n!}. \tag{3.9}$$

In actuarial applications, the expected value of the aggregate claim $E(S_t)$ is used as net premium when expenses, interest rates and inflation etc. are excluded in net premium calculation. It is calculated as in the following form,

$E(S_t) = \ E\big(E(S_t|N_t)\big),$

when we substitute (3.6) into the equation,

$= \sum_{i=0}^{\infty} E\big(X_1 + X_2 + \cdots + X_{N_t}|N_t = n\big)\, \mathrm{P}(N_t = n),$

with respect to the assumption of $N_0 = 0$, the index of the summation starts from 1,

$= \sum_{i=1}^{\infty} E\big(X_1 + X_2 + \cdots + X_{N_t}|N_t = n\big)\, \mathrm{P}(N_t = n),$

$X_i s$ are identically and independently distributed. Under the assumption of $X_i \sim approximatelly\ Normal(\mu, \sigma^2)$ and $N_t \sim Poisson(\lambda(t))$,

$(E(X_i) = \mu$ and $Var(X_i) = \ \sigma^2)$

$= \sum_{i=1}^{\infty} E(X_i)\, \mathrm{P}(N_t = n) = \sum_{i=1}^{\infty} \mu\, N_t\, \mathrm{P}(N_t = n)$

$= \mu \sum_{i=1}^{\infty} N_t\, \mathrm{P}(N_t = n)$

$(E(N_t) = Var(N_t) = \lambda(t))$

$= \ \mu\, E(N_t) = \mu\lambda(t),$

$$E(S_t) = E(X_i)E(N_t) = \mu\lambda(t). \tag{3.10}$$

Also the variance is,

$Var(S_t) = \ [(S_t - E(S_t))^2]$

$= E[(S_t - \mu\lambda(t))^2]$

$= E[(S_t - \mu N_t + \mu N_t - \mu\lambda(t))^2]$

$$= E[(S_t - \mu N_t)^2] + 2E[\mu(S_t - \mu N_t)(N_t - \lambda(t))] + E[\mu^2(N_t - \lambda(t))^2]. \tag{3.11}$$

If $E[(S_t-\mu N_t)^2]$ is calculated,

$$E[(S_t-\mu N_t)^2]=\sum_{i=0}^{\infty}E((S_t-\mu N_t)^2|N_t=n)\,\mathrm{P}(N_t=n),$$

Use the assumption of $N_0=0$ again,

$$=\sum_{i=1}^{\infty}E\left((X_1+X_2+\cdots+X_{N_t}-\mu N_t)^2\middle|N_t=n\right)\mathrm{P}(N_t=n)$$

$$=\sum_{i=1}^{\infty}N_t Var(X_i)\,\mathrm{P}(N_t=n)$$

$$=\sigma^2\sum_{i=1}^{\infty}N_t\,\mathrm{P}(N_t=n)=\sigma^2\lambda(t). \tag{3.12}$$

Then,

$$E[\mu^2(N_t-\lambda(t))^2]=\mu^2E[(N_t-\lambda(t))^2]=\mu^2\lambda(t)\,. \quad \left(Var(N_t)=E\left[\left(N_t-\lambda(t)\right)^2\right]\right) \tag{3.13}$$

Also,

$$E[\mu(S_t-\mu N_t)(N_t-\lambda(t))]=\mu\sum_{i=0}^{\infty}E((S_t-\mu N_t)(N_t-\lambda(t))|N_t=n)\,\mathrm{P}(N_t=n)$$

$=\ \mu\sum_{i=0}^{\infty}E\left((X_1+X_2+\cdots+X_{N_t}-\mu N_t)(N_t-\lambda(t))\middle|N_t=n\right)\mathrm{P}(N_t=n),$

Since $E(X_1+X_2+\cdots+X_{N_t})=\mu N_t$, $E\left(X_1+X_2+\cdots+X_{N_t}-\mu N_t\right)=0$,

$$E[\mu(S_t-\mu N_t)(N_t-\lambda(t))]=0. \tag{3.14}$$

Finally, when we put (3.12), (3.13) and (3.14) in (3.11),

$$Var(S_t)=\sigma^2\lambda(t)+2(0)+\mu^2\lambda(t)=\lambda(t)(\sigma^2+\mu^2).$$

If we rewrite the equation,

$$Var(S_t)=E(N_t)Var(X_i)+Var(N_t)E(X_i)=\lambda(t)(\sigma^2+\mu^2). \tag{3.15}$$

3.4 The Distribution of Aggregate Claim Amount

In probability theory, the distribution of the aggregate claim amount, in a given time period is connected with the distribution of the number of claims (N) and individual claim amount (X_i). Let N, $X_1, X_2, \ldots, X_N$ be independent and also $X_1, X_2, \ldots, X_N$ be identically distributed. In actuarial sciences, the total of the random variable S is called compound and identifies the aggregate claim amount model in collective risk model. The compound aggregate claim amount as a collective risk model can be obtained by developing the distribution of the individual claim amount and the number of claims separately. There are various methods to approximate the distribution of the aggregate claim amount since it is not easy to determine the exact distribution (Rotar, 2007). Two of these calculation methods are the convolution and moment generating function to model (3.2) and these methods are described in the following sections.

The convolution method:

Distribution function of S can be obtained by using the summation of the independent claim amounts, $X_i s$, (convolution). As it has been stated before the aggregate claim amount is defined in the following form (Section 3.1),

$$S = X_1 + X_2 + \cdots + X_N = \sum_{i=1}^{N} X_i.$$

Let the distribution function of S be $F_s(x)$ and $P(S = 0) = P(N = 0) = 0$. When we obtain the conditional distribution function given $N = n$ (Rotar, 2007 and Kaas, 2001),

$$F_s(x) = P(S \le x) = \sum_{n=0}^{\infty} P(X_1 + \cdots + X_N \le x | N = n)P(N = n)$$

$$= \sum_{n=0}^{\infty} P(S_n \le x | N = n)P(N = n), \tag{3.16}$$

where $S_n = X_1 + X_2 + \cdots + X_n$ and $S_0 = 0$. For $n \ge 1$,

$$P(S_n \le x | N = n) = P(S_n \le x) = F^{*n}(x),$$

So,

$$F_s(x) = \sum_{n=0}^{\infty} F^{*n}(x)P(N = n), \;\; f_s(x) = \sum_{n=1}^{\infty} f^{*n}(x)P(N = n), \tag{3.17}$$

where $F^{*n} = F * F * ... * F$, n^{th} convolution of F and f^{*n} is the n^{th} convolution of density f.

The moment generating function (mgf) method:

The moment generating function (mgf) is one of the most useful methods to derive the moments of a distribution. The definition of the moment generating function of X is,

$$M_x(t) = E(e^{xt}) = \int_{-\infty}^{\infty} e^{xt} f(x)dx. \tag{3.18}$$

The mgf of the aggregate claim amount $(S = \sum_{i=1}^{N_t} X_i)$ can be represented with the moment generating functions of N and X_i to determine the distribution of the aggregate claim amount. Let $X_i s$ be independently and identically distributed and denote i^{th} moment of X as (Bowers, 1997),

$$\mu'_i(x) = E(X^i). \tag{3.19}$$

also, the mgf of the number of claim N is

$$M_{N_t}(t) = E(e^{N_t t}). \tag{3.20}$$

and the mgf of the aggregate claim amount can be denoted by

$$M_S(t) = E(e^{St}). \tag{3.21}$$

The total of the variability of the expected value of the number of claims and the expected value of the individual claim amount compose the variance of the aggregate claim amount (Bowers et al., 1997).

The mgf of S can also be denoted with the conditional expected value:

$$M_S\,(t) = E(e^{St}) = E\;(E(e^{St}|N))$$

$$= \sum_{n=o}^{\infty} E(e^{t(X_1+\cdots+X_N)}|N=n)P(N=n)$$

$$= \sum_{n=o}^{\infty} E\left(e^{t(X_1+\cdots+X_n)}\right)P(N=n)$$

$$= \sum_{n=0}^{\infty} \left(\,M_x\,(t)\right)^n P(N=n)$$

$$= E(\left(e^{logM_x(t)}\right)^N)$$

$$= \;M_N\left(logM_x(t)\right) \tag{3.22}$$

The expected value and the variance of the distributions can be derived by the mgf. The first derivative of the mgf with respect to t, when $t = 0$ gives the first moment, the expected value of the distribution. In addition, the second derivative of the mgf helps to calculate the variance.

The mgf of the Compound Poisson aggregate claim amount, when N_t is Poisson with parameter $\Lambda(t)$ by (3.22), where $M_N(t) = e^{\Lambda(t)(t-1)}$,

$$M_S(t) = E(e^{St}) = M_N\left(logM_x(t)\right) = e^{\Lambda(t)(\log{(e^{M_X(t)})}-1)} = e^{\Lambda(t)(M_X(t)-1)}, \tag{3.23}$$

where X can have any distribution.

If the expected value μ and the variance of Compound Poisson S_t are derived by using (3.23):

$M_S(t) = e^{\Lambda(t)(e^{xt}-1)}$,

$M'_S\,(t) = e^{\Lambda(t)(e^{xt}-1)}\Lambda(t)e^{xt}$,

$E(S_t) = \;M'_{S_t}\,(0) = \Lambda(t)$.

and the second derivative is:

$M''_S(t) = e^{\Lambda(t)(e^{xt}-1)}\Lambda(t)e^{xt} + \Lambda(t)e^{xt}e^{\Lambda(t)(e^{xt}-1)}\Lambda(t)e^{xt}$,

$M''_S\,(0) = E\left({S_t}^2\right) = \Lambda(t) + \Lambda(t)^2$,

therefore the variance is obtained as:

$V(S_t) = \;M''_S\,(0) - \left(\,M'_S(0)\right)^2 = \Lambda(t) + \Lambda(t)^2 - \Lambda(t)^2 = \Lambda(t)$.

The higher moments can be computed by using the moment generating method.

Also, if $X \sim Normal\;(\mu, \sigma^2)$ then the mgf of the aggregate claim S_t is

$$M_S\,(t) = e^{\Lambda(t)\left(\mu t+\frac{1}{2}\sigma^2 t^2\right)}.$$

The first derivative of the mgf with respect to the t gives the expected value

$$M'_S(t) = \Lambda(t)\mu\, e^{\Lambda(t)\mu t} e^{\Lambda(t)\left(\mu t+\frac{1}{2}\sigma^2 t^2\right)} + e^{\Lambda(t)\mu t}\Lambda(t)\sigma^2 t e^{\Lambda(t)\left(\mu t+\frac{1}{2}\sigma^2 t^2\right)}$$

$$= e^{\Lambda(t)\left(\mu t+\frac{1}{2}\sigma^2 t^2\right)}\,\Lambda(t)(\mu + \sigma^2 t)$$

If $t = 0$ is substituted in $M'_S(t)$

$$M'_S(0) = E(S_t) = \Lambda(t)\mu$$

The variance is obtained by following the second derivative

$$M''_S(t) = e^{\Lambda(t)\left(\mu t+\frac{1}{2}\sigma^2 t^2\right)}(\Lambda(t))^2(\mu + \sigma^2 t)^2 + e^{\Lambda(t)\left(\mu t+\frac{1}{2}\sigma^2 t^2\right)}\Lambda(t)\sigma^2$$

and if $t = 0$ is substituted in $M''_S(t)$

$$M''_S(0) = E\left({S_t}^2\right) = (\Lambda(t))^2\,\mu^2 + \Lambda(t)\sigma^2$$

$$V(S) = M''_S(0) - \left(M'_S(0)\right)^2 = \left(\Lambda(t)\right)^2\mu^2 + \Lambda(t)\sigma^2 - (\Lambda(t)\mu)^2 = \Lambda(t)\sigma^2. \qquad (3.24)$$

3.5 Insurance

Insurance is a useful risk transfer system of uncertain losses. Risk is reduced and losses are indemnified by using insurance. The features of the insurance are (Rejda, 2005):

1. *Pooling of Loss*: It is the sharing/spreading of the losses by the few to the entire group.
2. *Payment of causeless losses*: A payment of unexpected loss that occurs by chance.
3. *Risk transfer*: It is the transfer of risk from insured to insurer.
4. *Indemnification:* After a loss, insured restores his/her financial position to initial position.

If there is a high risk, insurer/insurance company will usually want to share that risk with another insurer or insurers. This operation is called *reinsurance*. The insurer, who shares its risk is called *cedant*. The TCIP reinsures its premium risks as well. For instance, if an earthquake occurs with a big magnitude, the number of claims will increase and the financial burden of the TCIP will be challenging. It may not be able to afford all claims. In less time, it might be difficult to overcome a large amount of risk. By reinsuring, the TCIP minimizes this risk. In addition, premium arrangements could increase the claim payment capacity.

Premium

Premium is an adequate price to overcome the risks. Calculation of the premium is important since it should be reasonable for both insurer and insured parts. If the premium is too high, insurance company cannot compete with other insurance companies. Also, if the premium is too low, it will not be easy to face losses (Rolski et al., 1999). The premium is defined with $\Pi(X)$ notation, where X is a random variable that refers to a kind of risk. Some essential properties of premium for risks X and Y are (Rolski et al., 1999):

1. *No unjustified safety loading* if, for all constants $p \geq 0$, $\Pi(p) = p$,
2. *Proportionality* if, for all constants $p \geq 0$, $\Pi(pX) = p\Pi(X)$,
3. *Additivity* if $\Pi(X + Y) = \Pi(X) + \Pi(Y)$(risks are independent),
4. *Subadditivity* if $\Pi(X + Y) \leq \Pi(X) + \Pi(Y)$,

5. *Consistency* if, for all constants $p \geq 0$, $\Pi(X+p) = \Pi(X) + p$,

6. *Preservation of stochastic order* if $X \leq_{st} Y$ implies $\Pi(X) \leq \Pi(Y)$.

If the distribution of the risk X is known, it will be easier to calculate the premium. The simplest premium principle is the 'net (pure) premium principle' $\Pi(X) = E(X)$ and $\Pi(X) - E(X)$ gives the safety loading, which should be positive for the insurance company's reserves (Rolski et al., 1999).

Company takes the whole risk of the portfolio so the premiums should be appropriate and reasonable for company's safety. The basic principles of the premiums are (Rolski et al., 1999)

1. *Expected value principle*: Is the most commonly used principle.

$$\Pi(X) = (1+p)E(X) \text{ where } p \geq 0 \text{ and } E(X) < \infty$$

when $p = 0$ it becomes the net premium principle. The expected value principle has some risks since it does not consider the variability of the risk X. But other principles include safety loading $\Pi(X) - E(X)$ so they can get over from this risk. For $p > 0$,

2. *Variance principle*: $\Pi(X) = E(X) + pV(X)$,

3. *Standard deviation principle*: $\Pi(X) = E(X) + p\sqrt{V(X)}$,

4. *Exponential principle*: $\Pi(X) = p^{-1} log E(e^{pX})$

In classical risk model, the surplus process is (Dickson, 2005)

$$U(t) = u + \Pi(t) - S(t),$$

where u is the initial surplus, $\Pi(t) = ct$ (total premium), c is the loaded premium rate and the aggregate claim amount process $S_t = \sum_{i=1}^{N_t} X_i$.

In insurance, the expected value of the aggregate claim is an important calculation to ensure the risk. It gives an idea about the cost of a disaster to the insurer. For example in this study the net premium is calculated by using the 2000-2008 thesis data,

$$E(S) = \frac{aggregate\ claim\ amount}{total\ number\ of\ claims} = \frac{23{,}668{,}818}{12{,}075} = 1{,}960.151\ TL \approx 1{,}113.72\ USD, \qquad (3.25)$$

which gives the expectation of claim amount payment for the TCIP.

Then the expected payment of the TCIP according to the claims until January 7, 2012 from the beginning of the Pool;

$$E(S) = \frac{aggregate\ claim\ amount}{total\ number\ of\ claims} = \frac{150{,}100{,}857}{20{,}253} = 7{,}411.29\ TL \approx 4{,}210.96\ USD. \qquad (3.26)$$

A big difference is observed between expected aggregate claims of the thesis data and current claims data of the TCIP. This is a result of the claim arrivals from the October 23, 2011 Van/ Erciş earthquake of a magnitude 7.2.

CHAPTER 4

MATHEMATICS OF THE STUDY

The theoretical and mathematical work of the thesis is combined to lead to the computational analysis of the data. The basic ideas used are likelihood of the aggregate claim amount and number of claims, the nonhomogeneous Poisson process and the GLM.

4.1 Generalized Linear Models (GLM)

Modeling the relationships between the variables is one of the main interests in statistical methodology. The aim of the modeling is to find the best model, where the one variable is explained by one or more other variables. The explained variable is called '*response*' and others are '*explanatory*' variables. GLM generalizes the classical linear models by allowing the response variable to be non-normal via link function. In actuarial analysis, the GLM has a wide use to model premium, mortality and reserving (Boland, 2006). It was first introduced by Nelder and Wedderburn in 1972. In GLM, parameter estimations are obtained by maximum likelihood estimation followed by least-square algorithm iterations. If the response variable is continuous, the probability distribution might be Normal or if the response is countable, its distribution might be Poisson, also distribution might be Binomial when the response is binary (e.g. occurred or not occurred).

The GLM have two important features (Jong et al., 2008):

1. Distribution of the response variable is selected from the exponential family (e.g. Binomial, Poisson and Normal).

2. Response variable's mean is transformed and linked with explanatory variables. Some commonly used 'link functions' are; identity for Normal, log for Poisson and logit for Binomial.

The classical linear regression model is the most preferable and first comes to mind to fit a model. The connection between the response and explanatory variables can be shown as,

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{p_i} + \epsilon_i \;\; j = 0,1,\dots,p \; and \; i = 1,2,\dots,n \,,$$

where β_j's are unknown parameters, x_{ji} 's are defined explanatory variables and ϵ_i is the error term. The expected value of the error term $E(\epsilon_i) = 0$ and the variance $V(\epsilon_i)$is constant. There is a linear relationship between response variable and explanatory variables including the error term. The expected value of y, $E(y) = \mu$, and explanatory variables also has a linear relation. But in some cases, the linear connection may not be appropriate with non-linear unknown parameters so non-linear regression models are chosen. Transformations can be used to change a non-linear model into a linear model like log-transformation. The log-transformation is usually used in insurance for claim numbers and claim amount (Boland, 2006).

In insurance, normal assumption is not practical for claim numbers, claim amount and claim occurrences. So, the GLM is preferred for modeling since it is the combination of linear and non-linear models. In the GLM, the response (y) must be from an exponential family and the linear relation between the expected value of the response variable, which is called link and the explanatory variables is generalized as

$$g(E(y) = \mu) = x'\beta = \eta.$$

Observations of the response variables (y_i) and the explanatory variables (x_i) are assumed to be independent. The choice of link is offered by the form of the relationship between the response variable and the explanatory variables. If the response variable is a counting variable, generally the Poisson distribution is used. It is a well-known distribution for which the expected value and the variance are equal. In GLM for count response, Poisson regression is suitable via an appropriate link function. Ideal link functions are identity link $\mu = x'\beta$ and log-link $g(\mu) = log\ \mu = x'\beta$ (Boland, 2006). In this thesis, we model the number of claims N_t, which is a counting process under the assumption that it has a Poisson process and is distributed with rate $\lambda(t)$ that is changing over time t and is dependent on its history. The Poisson regression is preferred with the log-link function since it is generally suggested for insurance claim number modeling. If we consider a simple Poisson regression model, then

$$\log \mu = \beta_0 + \beta_1 x_1,$$

and the expected value of y is $E(y) = \mu = e^{\beta_0+\beta_1 x_1}$.

The probability density function of the Poisson distribution can be written as

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}\ x > 0, \lambda > 0,$$

If the probability density function is rewritten we obtain,

$$f(x) = \frac{e^{-\lambda}e^{xlog\lambda}}{x!} = \frac{e^{xlog\lambda-\lambda}}{x!}, x > 0, \lambda > 0. \quad (4.1)$$

In Poisson regression models, parameters are estimated by the maximum likelihood estimation method. In likelihood since we model counting data N_i in the Poisson process, where the rate λ is replaced with $log\Lambda_{\mathrm{i}}$,

$$L(\theta|x) = \prod_{i=1}^{n} \frac{e^{x_i log\Lambda_{\mathrm{i}}-\Lambda_{\mathrm{i}}}}{x_i!} = \frac{e^{\sum_{i=1}^{n}(x_i log\Lambda_{\mathrm{i}}-\Lambda_{\mathrm{i}})}}{\prod_{i=1}^{n}(x_i!)},$$

where x_i represents the number of claims N_i in the study.

The log-likelihood function of the Poisson distribution is

$$logL(\theta|x) = \sum_{i=1}^{n}(x_i log\Lambda_{\mathrm{i}} - \Lambda_{\mathrm{i}}) - \log\left(\prod_{i=1}^{n} x_i!\right),$$

$$= \sum_{i=1}^{n}(x_i log\Lambda_{\mathrm{i}} - \Lambda_{\mathrm{i}}) - \sum_{\mathrm{i=1}}^{\mathrm{n}} \log(x_i!). \quad (4.2)$$

In the following sections of the study, the estimations of the parameters α_j and β of the exponential kernel function are calculated. In some parts, the study becomes challenging.

4.2 Likelihoods of Observations

The likelihood function has an important role in estimating methods for the parameters. If we assume n random variables (observations) $x_1, x_2 \ldots, x_n$, which are independent and identically distributed (iid) with the probability density function of unknown parameter vector θ $f(x;\theta)$ and the likelihood function $L(\theta|x)$ of the joint probability density function of random variables is

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i;\theta) = f(x_1;\theta) \ldots f(x_n;\theta).$$

In this thesis, we want to find a good estimator for each parameter of the data and use it in modeling. The likelihood of the aggregate claim amount S_i and the time likelihood of Poisson counts are obtained under the assumption of the number of claims N_t and individual claim amounts $X_i$are independent, where the number of claims $N_t = n$ is conditionally given as follows (Başbuğ, 2007)

$$L(\theta|S,t) = \prod_{i=1}^{n} f(S_i|N_t = n)\, f(t_i|N_t = n) f(N_t = n)\,, \tag{4.3}$$

where θ refers to the all distribution parameters.

In chapter 3 (3.2), it is stated that the probability density function of interarrival time is $\lambda(t_i)e^{-\Lambda_i(t_i)}$, where $\Lambda_i(t) = \int_{W_i}^{W_{i+1}} \lambda(r)dr = \Lambda(W_{i+1}) - \Lambda(W_i)$. Thus, the time likelihood function of nonhomogeneous Poisson process is (Başbuğ, 2007),

$$\begin{aligned} L(\theta|t_1, \ldots, t_n) &= \textstyle\prod_{i=1}^{n} \lambda(t_i\,;\theta)\, e^{-\Lambda_i(t_i)}, \\ &= \lambda(t_1)\, e^{-\Lambda(t_1)} \lambda(t_2)\, e^{-(\Lambda(t_2)-\Lambda(t_1))} \ldots \lambda(t_n)\, e^{-(\Lambda(t_n)-\Lambda(t_{n-1}))}\,, \quad (4.4) \\ &= \textstyle\prod_{i=1}^{n} \lambda(t_i;\theta)\, e^{-\Lambda(t_n)}. \end{aligned}$$

Since $N_t = n$ is fixed, the likelihood function takes the form of $L(t_1, \ldots, t_n;\theta) = \prod_{i=1}^{n} f(t_i \mid \theta)$.

4.3 The Use of the Exponential Kernel Function

As mentioned before in Chapter 2 after large earthquakes, many aftershocks can be observed. Observed sudden jumps in claim numbers with the aftershock show a big increase in records. In the study, sudden changes are observed after big earthquakes. This aftershock occurrence is related with two main formulas. One of them is Omori's Law and other is Gutenberg-Richter. The aftershock rate λ, which is the number of earthquakes measured in certain time t, follows Omori's Law. It is an empirical relation for the temporal decay of aftershock rates

$$\lambda = \frac{K}{c+t},$$

where c is the time offset parameter and K is the size of earthquake waves. These two parameters are constant. Omori's Law shows that the aftershock rate decreases as time goes on (Utkucu et al., 2005).

In 1961, Utsu suggested the Modified Omori's Law, used now, and add p constant to equation (Utkucu et al., 2005)

$$\lambda = \frac{K}{(c+t)^p},$$

where the range of p is between 0.7 and 1.5. When $p = 1$ the Modified Omori's Law becomes Omori's Law.

The Gutenberg-Richter Formula gives the relation of the claim numbers and magnitude. As the magnitude of the aftershock increases the claim number decreases (Utkucu et al., 2005 and Gutenberg -Richter, 1944),

$$logN = a - bM,$$

where M is the magnitude, N is the claim number bigger than M and a, b are constants.

In the analysis, the nonhomogeneous Poisson process with the parameter Λ_i for $i = 1, \dots, n$ that is a member of log-linear family is used, where each claim arrival is considered in the given time period. The log-linear Poisson process is used to estimate parameters. Sudden changes after big earthquakes are represented by the exponential kernel function. The exponential kernel function that is used in the claim numbers and the aggregate claim amount are in the following forms respectively when calculating the likelihoods and in modeling sections of the thesis (Başbuğ, 2007):

$$\Lambda_i = e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} \text{ and } log\Lambda_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \quad j = 1,..,k \text{ and } \mathrm{i} = 1, \dots, \mathrm{n}$$

$$\mu_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \quad j = 1,..,k \text{ and } \mathrm{i} = 1, \dots, \mathrm{n},$$

where $s_j, j = 1,..,k$ is the kernel knots, where the earthquake takes place with α_j, effect s_j is empirical kernel knots and t_i is the corresponding kernel knots time, which are chosen to see the jump effect. In notation $(t_i - s_j)|+$, (+) part determines that the difference is always positive. β is an explanatory parameter for the GLM. Moreover, the different features of the each earthquake region of different years are represented by a non-linear β parameter. Additionally, α_j parameter represents the sudden jumps after big earthquakes and α_0 gives ordinary claim arrivals, which occur due to small tremors. Figure 4.1 gives the structure of the exponential kernel function. Jumps in the plot reveal the idea that the kernel function is used in analysis. Also, Figure 4.2 indicates that the exponential kernel function picks the big earthquakes.

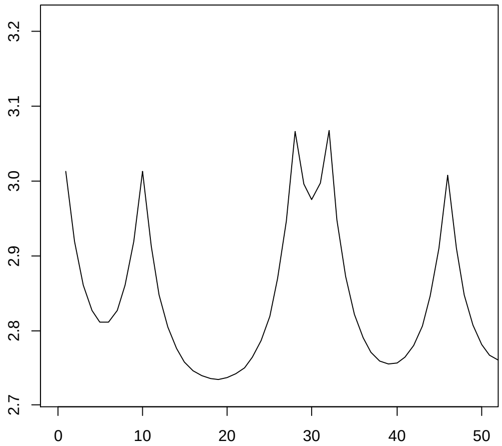


Figure 4.1 The Plot of the Exponential Kernel Function

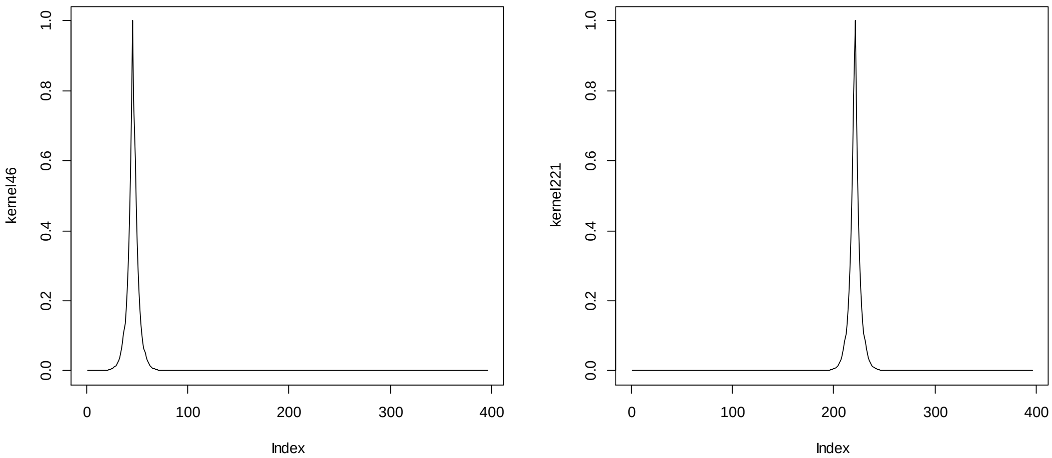


Figure 4.2 The Exponential Kernel Function, which picks the Big Earthquakes in Weeks

4.3.1 Likelihoods of the Exponential Kernel Function

As mentioned before, the exponential kernel functions are used instead of the number of claims. In this section, maximum likelihood estimation of the exponential kernel function is calculated. In the case of $N_i \sim Poisson\ (\Lambda_i)$ the rate has the following form,

$$log\Lambda_i = \begin{cases} \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)} & if\ t_i \geq s_j; \\ 0 & if\ t_i < s_j, \end{cases}$$

where $\Lambda_i = e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)}}$ with $j = 1,..,k$ knots, each ith time t_i, is connected with empirical knots s_js at which the kernels are replaced.

By using (4.2), initial steps of estimation of the non-linear β parameter, which represents the different features of the regions is obtained by the following steps (Başbuğ, 2007),

$$L(t_1, \dots, t_n;\ \theta, \alpha, \beta) = \prod_{i=1}^{n} e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} e^{-\Lambda(t_n)}, \tag{4.5}$$

where $\theta = (\alpha_0, \alpha_1, \dots, \alpha_k; \beta)$ is the vector of unknown parameters. Log-likelihood function is

$$\begin{aligned} logL(t_1, \dots, t_n;\ \theta, \alpha, \beta) &= \log\left(\prod_{i=1}^{n} e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} e^{-\Lambda(t_n)}\right) \\ &= \sum_{i=1}^{n}\left(log\left(e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} e^{-\Lambda(t_n)}\right)\right) \\ &= \sum_{i=1}^{n}\left(\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} - \Lambda(t_n)\right), \end{aligned} \tag{4.6}$$

where $\Lambda(\mathrm{t_n}) = \int_0^t \lambda(r)dr$. The derivative of log-likelihood function of (4.6) regarding non-linear parameter β in the next step,

$$\begin{aligned} \frac{\partial logL}{\partial \beta} &= \frac{\partial}{\partial \beta}\sum_{i=1}^{n}\left(\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} - \Lambda(t_n)\right) \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \beta}\left(\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} - \Lambda(t_n)\right) \\ &= \sum_{i=1}^{n}\left(\sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}(-t_i + s_j)|+ - \frac{\partial}{\partial \beta}\Lambda(t_n)\right). \end{aligned} \tag{4.7}$$

when we find the derivative of $\Lambda(t_n)$,

$$\Lambda(\mathrm{t_n}) = \int_0^{t_n} \lambda(r)dr = \int_0^{t_n} e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} dr$$

$$\begin{aligned} \frac{\partial \Lambda(\mathrm{t_n})}{\partial \beta} &= \int_0^{t_n} e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}(-t_i + s_j)|+ dr \\ &= \sum_{j=1}^{k} \alpha_j \int_0^{t_n} e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} e^{-\beta(t_i - s_j)|+}(-t_i + s_j)|+ dr \end{aligned} \tag{4.8}$$

Then, if (4.8) is replaced in (4.7), the following (4.9) is obtained. By numerical solutions the maximum likelihood estimation of non-linear parameter β can be derived.

$$\frac{\partial logL}{\partial \beta} = \sum_{i=1}^{n} (\sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \left(-t_i + s_j\right)|+$$

$$-\sum_{j=1}^{k} \alpha_j \int_0^{t_n} e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} e^{-\beta(t_i - s_j)|+} \left(-t_i + s_j\right)|+ dr). \qquad (4.9)$$

4.4 Estimation of the Model Parameters for the Number of Claims Model

Estimation of β parameter

In the Poisson regression, log-rate is used to model the number of claims N_i.The estimation of the non-linear parameter β of the exponential kernel function is obtained in the following form. The log-rate is,

$$log\Lambda_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}, \qquad (4.10)$$

so the rate of the process is $\Lambda_i = e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}}$.

Log-likelihood function of Poisson is given as

$$logL(x; \beta, \alpha) = \sum_{i=1}^{n} \left(x_i \left(\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \right) - e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} \right) - \sum_{i=1}^{n} \log(\mathrm{x_i}!) . \quad (4.11)$$

Differentiation of the log-likelihood regarding to non-linear parameter β,

$$\frac{\partial logL}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} (x_i log\Lambda_i - \Lambda_i) - \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \log(x_i!) ,$$

the Fisher's score function $U(\beta)$ is obtained as

$$\frac{\partial logL}{\partial \beta} = U(\beta)$$

$$= \sum_{i=1}^{n} \left(-x_i \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \left(-t_i + s_j\right)|+ \right.$$

$$\left. - \left(e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} \left(\sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \left(-t_i + s_j\right)|+ \right) \right) \right).$$

If the score function, $U(\beta)$, is differentiated regarding to β again, the Hessian, which gives the second derivative of log-likelihood function is obtained as

$$\frac{\partial^2 logL}{\partial\beta^2} = \frac{\partial U(\beta)}{\partial\beta}$$

$$= \sum_{i=1}^{n}\left(-x_i\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+(-t_i+s_j)|+\right.$$
$$-\left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}\ (-t_i+s_j)|+\right)\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\right)\right.$$
$$+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+(-t_i$$
$$\left.\left.+s_j)|+(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}})\right)\right). \tag{4.12}$$

If we rewrite (4.12),

$$\frac{\partial^2 logL}{\partial\beta^2} = \frac{\partial U(\beta)}{\partial\beta} = \sum_{i=1}^{n}\left(-x_i\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}\left((-t_i+s_j)|+\right)^2\right.$$
$$-\left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\right)^2\right.$$
$$\left.\left.+e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}\left((-t_i+s_j)|+\right)^2\right)\right)\right).$$

The variance estimate of the non-linear parameter β can be obtained by the observed information matrix $I(\alpha_0, \alpha_1, \dots, \alpha_k; \beta)$, by using the expectation of Hessian matrix

$$E\big(-H(\alpha_0, \alpha_1, \dots, \alpha_k; \beta)\big) = I(\alpha_0, \alpha_1, \dots, \alpha_k; \beta).$$

Since, sometimes it is hard to calculate the expectation of Hessian matrix, where $-H$ is the observed information matrix I, we can use $-H(\alpha_0, \alpha_1, \dots, \alpha_k; \beta) = I(\alpha_0, \alpha_1, \dots, \alpha_k; \beta)$. The inverse of negative of the Hessian matrix is used to get the confidence interval of the β and α_j parameters.

$$I(\beta) = -H(\beta) \;\; = \sum_{i=1}^{n} \left(-x_i \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \left((-t_i + s_j)|+ \right)^2 \right.$$

$$+ \left(e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} \left(\sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} (-t_i + s_j)|+ \right)^2 \right.$$

$$\left. \left. - e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} \left(\sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \left((-t_i + s_j)|+ \right)^2 \right) \right) \right).$$

The inverse of negative of the Hessian ($[H]^{-1}$) or observed information matrix gives the variance-covariance of the parameters, which are used for the confidence interval of the parameter β that gives the different features of each earthquake region of different years.

$$\hat{\beta} \pm critical\ value * standard\ error(\hat{\beta})$$

Estimation of α Parameter

The linear parameter α_j $(j = 1, .., k)$, which represents the big earthquakes is estimated with the same method used in the non-linear parameter β in log-linear Poisson modeling. The log-likelihood in (4.11) changes in

$$logL(x; \beta, \alpha) = \sum_{i=1}^{n} \left(x_i \left(\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \right) - e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} - \sum_{i=1}^{n} log(x_i!) \right). \tag{4.13}$$

The first derivative of the log-likelihood function (4.13) by α_j,

$$\frac{\partial logL}{\partial \alpha_j} = \sum_{i=1}^{n} \left(x_i \sum_{j=1}^{k} e^{-\beta(t_i - s_j)|+} - e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} \sum_{j=1}^{k} e^{-\beta(t_i - s_j)|+} \right), \tag{4.14}$$

The second derivation (Hessian) of the log-likelihood (4.14) is

$$\frac{\partial^2 logL}{\partial {\alpha_j}^2}$$
$$= \sum_{i=1}^{n} \left(-e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} \sum_{j=1}^{k} e^{-\beta(t_i - s_j)|+} \sum_{j=1}^{k} e^{-\beta(t_i - s_j)|+} \right)$$
$$= -\sum_{i=1}^{n} \left(e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta\left(t_i - s_j\right)|+}} \left(\sum_{j=1}^{k} e^{-\beta(t_i - s_j)|+} \right)^2 \right).$$

Since $I(\alpha_j) = -H(\alpha_j)$, the Fisher Information matrix, which includes $\frac{\partial^2 logL}{\partial\alpha_j\partial\alpha_j}$ is

$$I(\alpha_j) = \sum_{i=1}^{n}\left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}} \left(\sum_{j=1}^{k} e^{-\beta(t_i-s_j)|+}\right)^2\right). \qquad (4.15)$$

Second partial derivative for $j = j' = 1, \ldots, k$ ($\frac{\partial^2 logL}{\partial\alpha_1\partial\alpha_2}, \frac{\partial^2 logL}{\partial\alpha_2\partial\alpha_3}, \ldots$) is

$$\frac{\partial^2 logL}{\partial\alpha_j\partial\alpha_{j'}} = \sum_{i=1}^{n}\left(-e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}} e^{-\beta(t_i-s_j)|+} e^{-\beta(t_i-s_{j'})|+}\right).$$

Also, the first derivative, the Fisher's Score Function, of (4.13) by α_0 is

$$\frac{\partial logL}{\partial\alpha_0} = \sum_{i=1}^{n}\left(x_i - e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\right), \qquad (4.16)$$

So, the Hessian (4.16) (second derivative of α_0) is

$$\frac{\partial^2 logL}{\partial{\alpha_0}^2} = \sum_{i=1}^{n}\left(-e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\right).$$

The Hessian with the parameters α_0 and α_j by (4.16) can be obtained as

$$\frac{\partial^2 logL}{\partial\alpha_0\partial\alpha_j} = \sum_{i=1}^{n}\left(-e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}} \sum_{j=1}^{k} e^{-\beta(t_i-s_j)|+}\right).$$

The confidence interval of α_j parameters can be calculated by using inverse of negative of the Hessian matrix of $\alpha_j\alpha_j$'s for variance estimates of α_j's:

$$\widehat{\alpha_j} \pm critical\ value * standard\ error(\widehat{\alpha_j})$$

The Hessian matrix also contains the entry for $\alpha_j\beta$, which is symmetric with $\beta\alpha_j$ in the following form by using (4.13)

$$\frac{\partial logL}{\partial\beta} = \sum_{i=1}^{n}\left(-x_i\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\right) - \left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}} \left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\right)\right)\right), \qquad (4.17)$$

The second derivative of (4.17) by $\alpha_j$is

$$\frac{\partial^2 logL}{\partial\beta\partial\alpha_j} = \sum_{i=1}^{n}\left(-x_i \sum_{j=1}^{k} e^{-\beta(t_i-s_j)|+}(-t_i + s_j)|+ \; -\left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}} \sum_{j=1}^{k} e^{-\beta(t_i-s_j)|+} \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+ + \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+ e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\right)\right). \tag{4.18}$$

If we take the derivative of log-likelihood respectively by α_j then β

$$\frac{\partial logL}{\partial\,\alpha_j} = \sum_{i=1}^{n}\left(-x_i\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\right) - e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}} \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}\right),$$

Then,

$$\frac{\partial^2 logL}{\partial\alpha_j\partial\beta} = \sum_{i=1}^{n}\left(x_i \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+ \; -\left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}} \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}\right) + \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)\Big|+ \left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\right)\right)\right). \tag{4.19}$$

We obtain the same result with (4.18) because of the symmetry.

By using (4.13) the Fisher Score Function and the Hessian for $\frac{\partial^2 logL}{\partial\beta\partial\alpha_0}$ we obtain,

$$\begin{aligned}\frac{\partial^2 logL}{\partial\beta\partial\alpha_0} &= \sum_{i=1}^{n}\left(-e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\right)\right)\\ &= -\sum_{i=1}^{n}\left(e^{\alpha_0+\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}}\left(\sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+}(-t_i+s_j)|+\right)\right).\end{aligned}$$

the result will be same with $\frac{\partial^2 logL}{\partial\alpha_0\partial\beta}$ because of the symmetry.

The score function U, the first derivative of log-likelihood function, according to β and α_j $(j = 1,..,k)$ is described as in matrix form:

$$U(L(\beta;\alpha_0,\dots\alpha_k)) = \begin{bmatrix} \frac{\partial logL(\beta;\alpha_0,\dots\alpha_k)}{\partial\alpha_0} \\ \frac{\partial logL(\beta;\alpha_0,\dots\alpha_k)}{\partial\alpha_1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial logL(\beta;\alpha_0,\dots\alpha_k)}{\partial\alpha_k} \\ \frac{\partial logL(\beta;\alpha_0,\dots\alpha_k)}{\partial\beta} \end{bmatrix},$$

and the Hessian matrix is obtained by using the corresponding second partial derivatives of log-likelihood function;

$$H(\beta;\alpha_0,\dots,\alpha_k) = \begin{bmatrix} \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_0\partial\alpha_0} & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_0\partial\alpha_1} & \cdots & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_0\partial\alpha_k} & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_0\partial\beta} \\ & & \vdots & & \\ \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_k\partial\alpha_0} & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_k\partial\alpha_1} & \cdots & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_k\partial\alpha_k} & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\alpha_k\partial\alpha_\beta} \\ \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\beta\partial\alpha_0} & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\beta\partial\alpha_1} & \cdots & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\beta\partial\alpha_k} & \frac{\partial^2 logL(\beta;\alpha_0,\dots,\alpha_k)}{\partial\beta\partial\beta} \end{bmatrix}.$$

The variance-covariance matrix can be obtained by the inverse of negative of the Hessian matrix. Then, by using the diagonals, which correspond to variances of the parameters β and α_js, confidence intervals can be computed.

4.5 Estimation of the Model Parameters for the Aggregate Claim Amount Model

The aggregate claim amounts are assumed to be approximately Normal, which is the member of exponential family. So, the aggregate claim amounts can be modeled with the GLM like the number of claims. As stated earlier, if $logS_i \sim Normal$ $(S_i = \sum_{i=1}^{N_i} X_i)$ then $S_i \sim Lognormal$ is used in the calculations by taking the natural logarithm of the aggregate claim amounts. The probability density of Normal Distribution is

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty$$

where the parameters mean μ and standard deviation σ.

The probability density of Normal distribution can be written in an exponential family form,

$$f(x;\mu) = \exp\left(\frac{x\mu}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right), \quad -\infty < x < \infty.$$

The likelihood function of this density is in the following form,

$$L(x;\mu) = \prod_{i=1}^{n} \exp\left(\frac{x_i\mu}{\sigma^2} - \frac{{x_i}^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right),$$

where x_i is the claim amount in this study. The log-likelihood function can be written as,

$$logL(x;\mu) = \sum_{i=1}^{n}\left(\frac{x_i\mu}{\sigma^2} - \frac{{x_i}^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}log(2\pi\sigma^2)\right). \tag{4.20}$$

In parameter estimations, the exponential kernel function is used instead of the mean μ parameter for the claim amount, which is assumed to be approximately Normal.

$$\mu = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \tag{4.21}$$

If we located (4.21) in (4.20), the log-likelihood is

$$logL(x;\mu) = \sum_{i=1}^{n}\left(-\frac{{x_i}^2}{2\sigma^2} + \frac{x_i(\alpha_0 + \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+})}{\sigma^2} - \frac{(\alpha_0 + \sum_{j=1}^{k}\alpha_j e^{-\beta(t_i-s_j)|+})^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right). \tag{4.22}$$

Estimation of β Parameter

The first derivative of log-likelihood function (4.22) by β is,

$$\frac{\partial logL}{\partial\beta} = \sum_{i=1}^{n}(-\frac{x_i t_i \sum_{j=1}^{k}\alpha_j e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} + \frac{x_i \sum_{j=1}^{k}\alpha_j s_j e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} - \frac{\alpha_0 t_i \sum_{j=1}^{k}\alpha_j e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}} + \frac{\alpha_0 \sum_{j=1}^{k}\alpha_j s_j e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}} - \frac{t_i(\sum_{j=1}^{k}\alpha_j e^{\beta s_j})^2}{2\sigma^2(e^{\beta t_i})^2} + \frac{\left(\sum_{j=1}^{k}\alpha_j e^{\beta s_j}\right)(\sum_{j=1}^{k}\alpha_j s_j e^{\beta s_j})}{2\sigma^2(e^{\beta t_i})^2}) \tag{4.23}$$

and second partial derivative of (4.23) by β is

$$\frac{\partial^2 logL}{\partial\beta^2} = \sum_{i=1}^{n}(\frac{x_i {t_i}^2 \sum_{j=1}^{k}\alpha_j e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} - \frac{x_i t_i \sum_{j=1}^{k}\alpha_j s_j e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} + \frac{x_i \sum_{j=1}^{k}\alpha_j {s_j}^2 e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}}$$

$$- \frac{\alpha_0 {t_i}^2 \sum_{j=1}^{k}\alpha_j e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}} + \frac{2\alpha_0 t_i(\sum_{j=1}^{k}\alpha_j s_j e^{\beta s_j})^2}{2\sigma^2 e^{\beta t_i}} - \frac{\alpha_0 \sum_{j=1}^{k}\alpha_j {s_j}^2 e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}} - \frac{2{t_i}^2(\sum_{j=1}^{k}\alpha_j e^{\beta s_j})^2}{2\sigma^2(e^{\beta t_i})^2}$$

$$+ \frac{4t_i\left(\sum_{j=1}^{k}\alpha_j e^{\beta s_j}\right)(\sum_{j=1}^{k}\alpha_j s_j e^{\beta s_j})}{2\sigma^2(e^{\beta t_i})^2} - \frac{(\sum_{j=1}^{k}\alpha_j s_j e^{\beta s_j})^2}{2\sigma^2(e^{\beta t_i})^2} - \frac{\left(\sum_{j=1}^{k}\alpha_j e^{\beta s_j}\right)(\sum_{j=1}^{k}\alpha_j {s_j}^2 e^{\beta s_j})}{2\sigma^2(e^{\beta t_i})^2}.$$

Estimation of α Parameter

The linear parameter $\alpha_j, j = 1, 2, \ldots, k$, which represents the big earthquakes is obtained from the log-likelihood in (4.22) by taking the first derivative (score function) and second derivative (Hessian) for the exponential kernel function.

$$\frac{\partial logL}{\partial\alpha_j} = \sum_{i=1}^{n}(\frac{x_i \sum_{j=1}^{k} e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} - \frac{\alpha_0 \sum_{j=1}^{k} e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} - \frac{(\sum_{j=1}^{k}\alpha_j e^{\beta s_j})(\sum_{j=1}^{k} e^{\beta s_j})}{\sigma^2(e^{\beta t_i})^2}), \tag{4.24}$$

and the Hessian, second derivative, is,

$$\frac{\partial^2 logL}{\partial {\alpha_j}^2} = \sum_{i=1}^{n}\left(-\frac{(\sum_{j=1}^{k} e^{\beta s_j})^2}{\sigma^2 (e^{\beta t_i})^2}\right). \tag{4.25}$$

When we take derivative of (4.24) in respect to α_0

$$\frac{\partial^2 logL}{\partial \alpha_j \alpha_0} = \sum_{i=1}^{n}\left(-\frac{\sum_{j=1}^{k} e^{\beta s_j}}{\sigma^2 e^{\beta t_i}}\right).$$

The first derivative, score function, of α_0 can be derived from (4.22)

$$\frac{\partial logL}{\partial \alpha_0} = -\frac{n\alpha_0}{\sigma^2} + \left(\sum_{i=1}^{n}\left(\frac{x_i}{\sigma^2} - \frac{\sum_{j=1}^{k} e^{\beta s_j}}{\sigma^2 e^{\beta t_i}}\right)\right), \tag{4.26}$$

and the Hessian function is,

$$\frac{\partial^2 logL}{\partial {\alpha_0}^2} = -\frac{n}{\sigma^2} \ .$$

Second partial derivative for $j = j' = 1, \dots, k$ $\left(\frac{\partial^2 logL}{\partial \alpha_j \partial \alpha_{j'}}; \frac{\partial^2 logL}{\partial \alpha_1 \partial \alpha_2}, \frac{\partial^2 logL}{\partial \alpha_2 \partial \alpha_3}, \dots\right)$ is,

$$\frac{\partial^2 logL}{\partial \alpha_j \partial \alpha_{j'}} = \left(-\sum_{i=1}^{n}\frac{e^{-\beta(t_i - s_j)|+} e^{-\beta(t_i - s_{j'})|+}}{\sigma^2}\right).$$

The Hessian can be completed by taking derivates according to α_j, α_0 and β parameters. If we differentiate (4.23) with α_j

$$\frac{\partial^2 logL}{\partial \beta \partial \alpha_j} = \sum_{i=1}^{n}(-\frac{x_i t_i \sum_{j=1}^{k} e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} + \frac{x_i \sum_{j=1}^{k} s_j e^{\beta s_j}}{\sigma^2 e^{\beta t_i}} - \frac{\alpha_0 t_i \sum_{j=1}^{k} e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}}$$

$$+ \frac{\alpha_0 \sum_{j=1}^{k} s_j e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}} - \frac{2t_i\left(\sum_{j=1}^{k} \alpha_j e^{\beta s_j}\right)(\sum_{j=1}^{k} e^{\beta s_j})}{2\sigma^2 (e^{\beta t_i})^2} + \frac{(\sum_{j=1}^{k} e^{\beta s_j})(\sum_{j=1}^{k} \alpha_j s_j e^{\beta s_j})}{2\sigma^2 (e^{\beta t_i})^2}$$

$$+ \frac{(\sum_{j=1}^{k} \alpha_j e^{\beta s_j})(\sum_{j=1}^{k} s_j e^{\beta s_j})}{2\sigma^2 (e^{\beta t_i})^2}).$$

$\frac{\partial^2 logL}{\partial \alpha_j \partial \beta}$ is same with $\frac{\partial^2 logL}{\partial \beta \partial \alpha_j}$ because of the symmetry.

If (4.23) is differentiated by α_0 , the Hessian is obtained as

$$\frac{\partial^2 logL}{\partial \beta \partial \alpha_0} = \sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{k} \alpha_j s_j e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}} - \frac{t_i \sum_{j=1}^{k} \alpha_j e^{\beta s_j}}{2\sigma^2 e^{\beta t_i}}\right).$$

$\frac{\partial^2 logL}{\partial \alpha_0 \partial \beta}$ is also same with $\frac{\partial^2 logL}{\partial \beta \alpha_0}$ because of the symmetry.

Confidence intervals of the parameters β and α_js can be computed by using the diagonals of inverse of negative of the Hessian matrix.

CHAPTER 5

ANALYSES AND MODELING

5.1 Explanatory Data Analysis

The data used in this thesis, which is collected in the TCIP between 2000 and 2008. The starting date is December 15, 2000 and continues until July 20, 2008. The study contains earthquake insurance claims, raw data earthquake realization dates, hours, paid claim amounts for each claim, the provinces and towns, earthquake size (magnitude) and also the risk zones. As the variables of the data; the claim amount, the number of claims and magnitude are organized in terms of weeks by R 2.15.1 software and used for analysis. In total, 12,075 number of claims data are used and 396 weeks from the beginning until the date of the last earthquake including 4 risk zones (risk zone 1, 2, 3 and 4), which are the degree of risk for the earthquake regions. The number of claims and the aggregate claim amount are the two main variables of the analysis. The number of claims variable (N_i) is the total number of arrived earthquake claims in terms of week. Aggregate claim amount (S_i) corresponding to the number of claims that is paid to the insured is used as the sum of the individual claim amounts for each week. Many sudden changes (rise and fall) are observed in claims data. For example, while no earthquake is observed in some weeks, maximum 2,913 earthquake claims are reported in one week. In addition, the magnitude, which is the measure of earthquakes, is used by its modes for each week. Also, the highest observed magnitude is 6.5 in 9 years.

Structure of the data:

W_i	N_i	S_i	M_i	R_i
1	$N_1 = n_{11} + n_{12} + \cdots + n_{1n}$	$S_1 = X_1 + X_2 + \cdots + X_{N_1}$	M_1	R_1
2	$N_2 = n_{21} + n_{22} + \cdots + n_{2n}$	$S_2 = X_1 + X_2 + \cdots + X_{N_2}$	M_2	R_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
396	$N_{396} = n_{396\,1} + n_{396\,2} + \cdots + n_{396\,n}$	$S_{396} = X_1 + X_2 + \cdots + X_{N_{396}}$	M_{396}	R_{396}

The week of the first occurred event, December 15, 2000 is chosen as the first week and the last event July 20, 2008 is chosen as the last week of the analysis. The 51st week in 2000 at Afyonkarahisar/Bolvadin with a magnitude of 5.8 in risk zone 1 is the first claim data and the last claim data occurred in the 29th week of 2008 in Izmir/Karaburun with magnitude 4.0 in risk zone 1. In the data, the highest magnitude is observed on January 27, 2003 in Tunceli/ Pülümür (takes place in risk zone 1) with magnitude 6.5 (see table 5.1). This region is on the area of the active North Anatolian and East Anatolian faults. Since many residential buildings are close to the fault lines, magnitude earthquakes whether large or small can be observed and arrived total losses to insurance companies can be still be high. In the following Table 5.1, the claim numbers that are greater than 100 and magnitudes, which are greater than 5.0 are selected and listed.

Table 5.1 The Significant Earthquake Claims Data from the TCIP

Date	Corresponding weeks in R	Place	Magnitude	Risk Zone	Number of Claims
25/06/2001	28	Osmaniye/Merkez	5.5	1	128
31/10/2001	46	Osmaniye/Merkez	5.2	1	139
03/02/2002	59	Afyonkarahisar/ Sultandağı	6.0	1	1,471
27/01/2003	111	Tunceli/Pülümür	6.5	1	168
10/04/2003	121	İzmir/Urla	5.6	1	1,731
01/05/2003	124	Bingöl/Merkez	6.4	1	470
28/03/2004	171	Erzurum/Aşkale	5.3	2	269
06/06/2005	235	Bingöl/Karlıova	5.7	1	105
21/10/2005	254	İzmir/Seferihisar	5.9	1	2,913
21/02/2007	323	Elazığ/Sivrice	5.9	1	169

The Izmir/Seferihisar Earthquake on October 21, 2005 with the highest claim number reaches to the TCIP according to the data. Since the earthquake happened at night and also people felt it strongly in and around Izmir, it caused panic and fear. Therefore, many citizens were injured but fortunately, nobody died. The reported claim number is an indicator and shows the awareness of the citizens who have the CEI. It is vital to take precautions in Izmir since it is placed in risk zone 1.

The main variables of the analysis are the number of claims N_i ($N_i \sim Pois(\Lambda_i)$) (section 3.2) and the aggregate claim amount S_i ($S_i = X_1 + X_2 + \cdots + X_{N_i}$) (section 3.1). In many insurance applications, the aggregate claim amount is assumed to be distributed as Lognormal, Gamma, Log-gamma or Negative Binomial. In this thesis, the aggregate claim amount S_i is assumed to be Lognormal therefore, logarithm transformed, $logS_i$ is assumed to be Normal. After transformation of S_i, Figure 5.1 shows that $logS_i$ is roughly symmetric and represents approximately Normal distribution and Figure 5.2 Q-Q plot also supports the normality assumption. In Figure 5.2, small changes are seen at the tails but most of the points lie on the line. Since the earthquake insurance claims data are extreme values, these leaving points can be observed. Extreme values will be another study subject. A close form of this figure is also derived in the study of Achieng (2000) for fitting distribution who accepts the assumption as well. Moreover, normality assumption tests are done in EasyFit distribution fitting program, which is usually used in engineering, actuarial science, medicine etc. and R 2.15.1 software. The test results are illustrated in Table 5.2, where the null hypothesis equals the data follow the specified distribution (Normal distribution). While Shapiro-Wilk normality test rejects the null hypothesis, the Anderson Darling and Kolmogrov-Smirnov and Pearson Chi-square tests do not reject the normality. Since, Shapiro-Wilk is a powerful test and the number of the data is high, it rejects normality even as a small change is observed between fitting and actual values. Therefore, the normality assumption does not rejected and then the logarithm transformed of aggregate claim amount is used in the analysis. Also, the number of claims has a nonhomogeneous Poisson process since the time variation of rate Λ_i(see section 3.2). In modeling of the aggregate claim amount and the claim number the exponential kernel function is preferred to represent the sudden jumps after big earthquakes.

H_o: Data follow Normal distribution
H_1: Data do not follow Normal distribution

Table 5.2 The Test Results of the Normality Assumption

Test	Reject H_o	Not Reject H_o
Shapiro-Wilk	√	
Anderson Darling		√
Kolmogrov-Smirnov		√
Pearson Chi-square		√

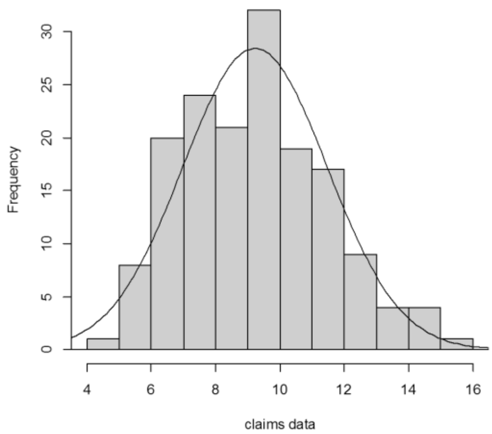


Figure 5.1 The Histogram of
the log of the Aggregate Claim Amount

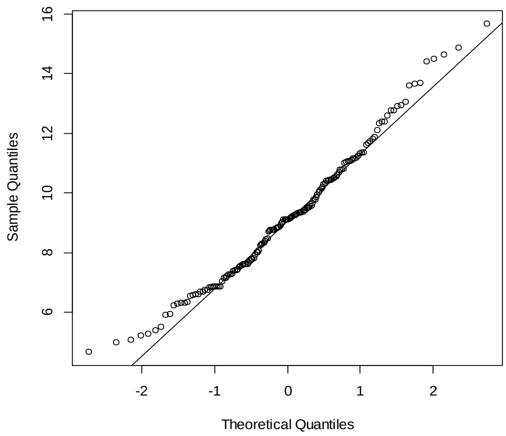


Figure 5.2 The Q-Q Plot of
the log of the Aggregate Claim Amount

The relations between the magnitudes, the number of claims and the aggregate claim amount (after transformation) in all risk zones in terms of weeks can be seen by graphical analysis in Figures 5.3, 5.4, 5.5 and 5.6. These figures show different properties of the data. The main interest of this thesis is to represent the sudden jumps when big earthquakes occur in the country that affect both the number of claims and the aggregate claim amount. As it was mentioned in Section 4.3, the special mathematical function, the exponential kernel function, is used in modeling section and for parameter estimation (α_j) in the empirical selection of the knots with sudden jumps occurring at big earthquakes are used. These sudden changes of claims can be easily seen in Figure 5.3. Three top points are observed at weeks 59 (Afyonkarahisar/ Sultandağı with 1,471 claims), 121(İzmir/Urla with 1,731 claims) and 254 (İzmir/Seferihisar with 2,913 claims), respectively.

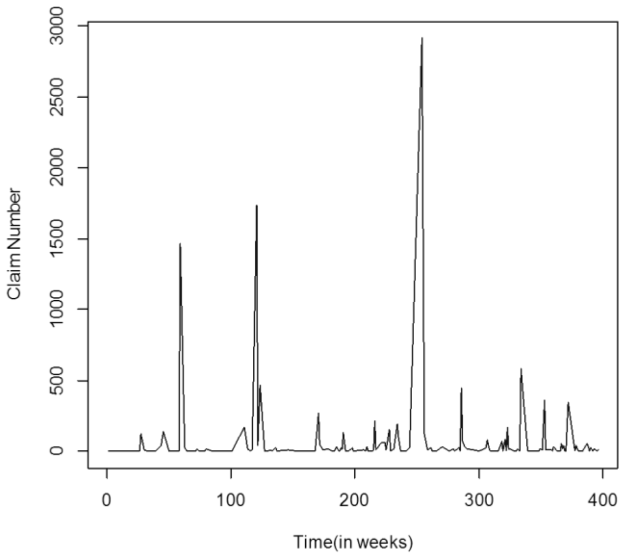


Figure 5.3 The Plot of Claim Number and Time (in weeks)

In both Figures 5.4 and 5.5, the magnitude and the aggregate claim amount demonstrate a homogeneous scatter in time. The magnitude varies mostly between 3.5 and 5.5. The highest claim payment is observed in week 254, the Izmir /Seferihisar earthquake with a magnitude of 5.9 and 2,913 number of claims.

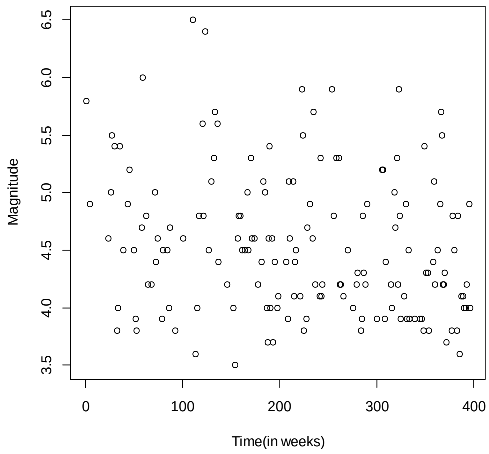


Figure 5.4 The Plot of Magnitude and Time (in weeks)

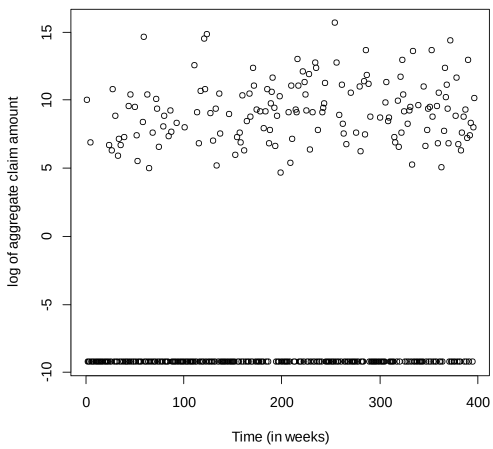


Figure 5.5 The Plot of the log of the Aggregate Claim Amount and Time (in weeks)

There is a 93% correlation between the claim amount and the number of claims. This means that, as the number of claims increases, the total payment of the TCIP increases. Also, the correlation of the magnitude with the aggregate claim amount and the number of claims are 23% and 24%, respectively, as expected.

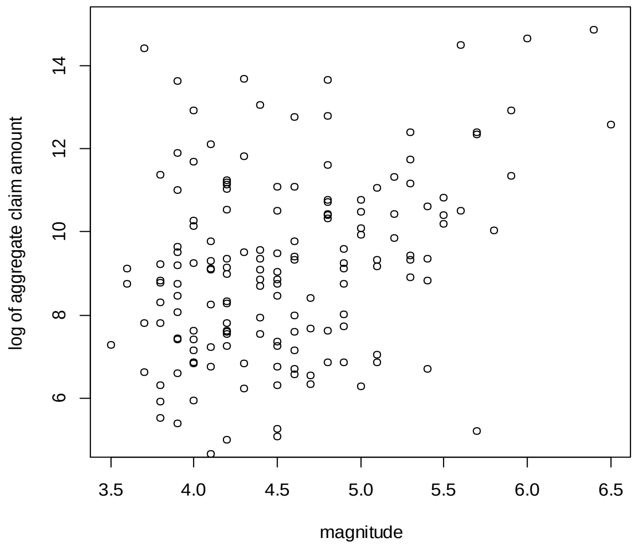


Figure 5.6 The Plot of Magnitude and the log of Aggregate Claim Amount

The plot of the magnitude and the log of the aggregate claim amount give a more scattered pattern than a shaper linear relation (Figure 5.6). However, even if the magnitude is large, such as 5.0, high aggregate claim amounts are not expected because of the resistance of construction or awareness of the people to the insurance in the region. It can be observed that the magnitude increases exponentially and faster up to 5.0 then becomes almost linear after 5.0. Also, the frequency of claims is higher until magnitude 5.0 and payments are moderate. High payments are observed at high magnitudes.

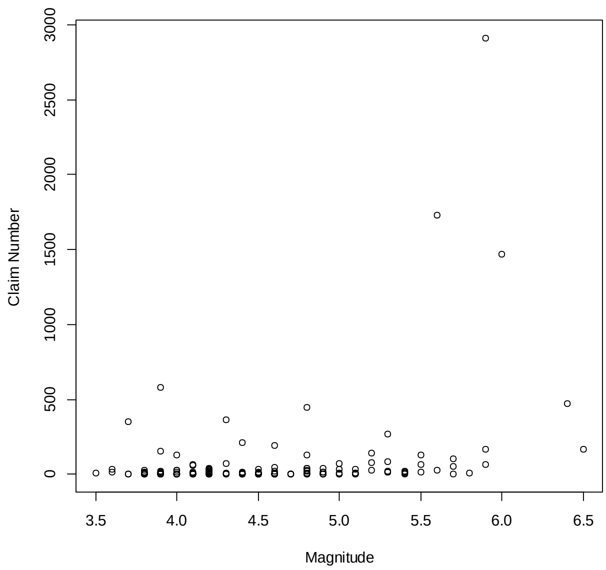


Figure 5.7 The Plot of Magnitude and Claim Number

Mostly, when the magnitude is low, the number of claims is low, too. However, there are some exceptions. For instance, the Izmir /Seferihisar earthquake had 2,913 claims with a magnitude 5.9 but the Tunceli/Pülümür earthquake has 168 claims with a 6.5 magnitude.

In the following sections, the relationships between the number of claims, the aggregate claim amount and the magnitude are presented separately in two parts: risk zone 1 and risk zone 2, 3 and 4. Since most of the thesis data is observed in risk zone 1, other risk zones (2, 3 and 4) are considered together. The graphical analyses illustrate the differences among the risk zones. For example, the higher the number of claims the higher the corresponding claim amounts is generally observed in risk zone 1, which has the highest earthquake risk.

Graphical Analysis of Risk Zone 1 Claims Data

The relationships among the aggregate claim amount, the number of claims, the magnitude and time in terms of weeks are given in Figures 5.8, 5.9 and 5.10. Since weeks 28, 46, 59,111, 121, 124, 235, 254 and 323 with high claims are gathered in risk zone 1; significant changes are observed in this area. The magnitudes of these claims generally exist between 5.0 and 6.5. These observations are evidence for decreasing risks in constructions, taking precautions by municipalities and having CEI. In addition, sudden jumps after big earthquakes are shown in Figure 5.9.

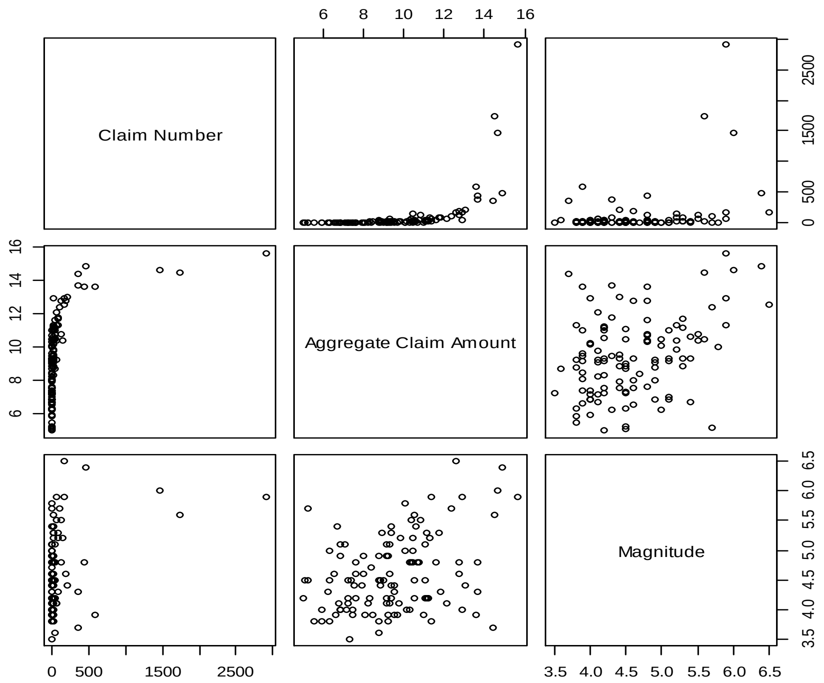
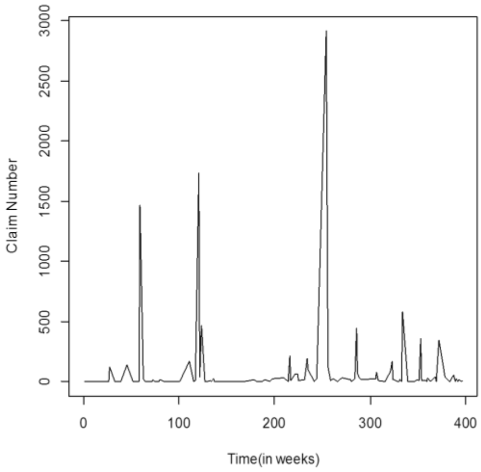


Figure 5.8 The Scatter Plot Matrices of Risk Zone 1 Claims Data



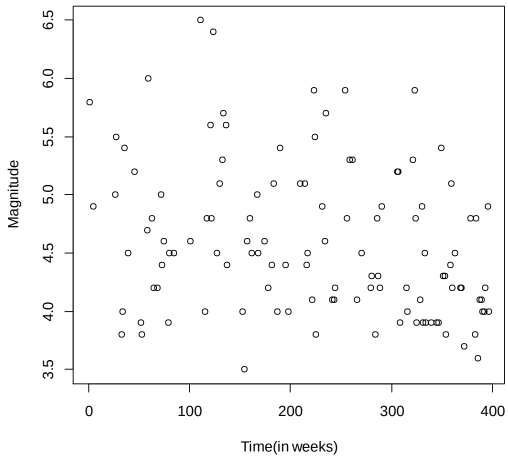


Figure 5.9 The Plot of Claim Number & Time (in weeks) Risk Zone 1

Figure 5.10 The Plot of Magnitude & Time (in weeks) in Risk Zone 1

Graphical Analysis of Risk Zones 2, 3 and 4 Claims Data

In risk zones 3 and 4, there are minor claims, where magnitudes are less than 5.0 and the claim numbers are less than 5 that arrive to the TCIP. Therefore, in graphical analysis, they are considered with risk zone 2. The relations among the aggregate claim amount, the number of claims, the magnitude and time in terms of weeks in the earthquake risk zones 2, 3, and 4 are represented in Figures 5.11, 5.12 and 5.13.

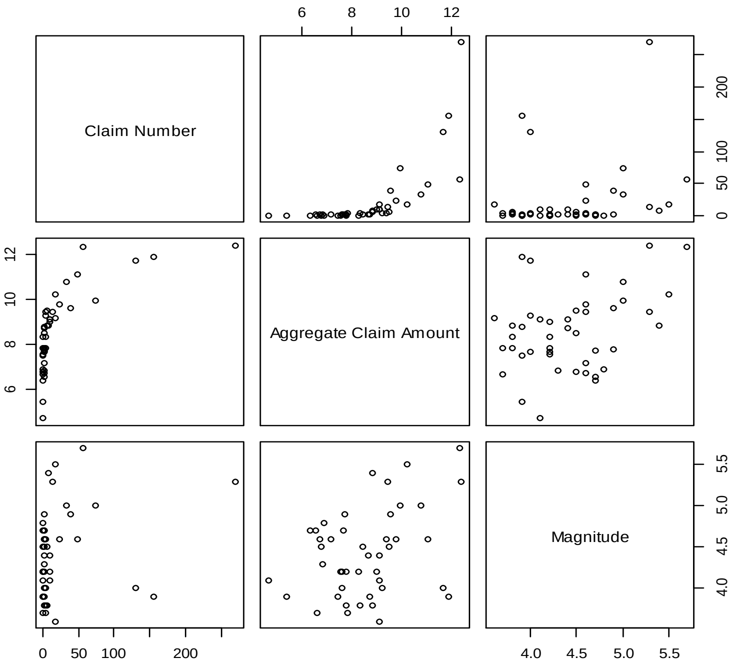


Figure 5.11 The Scatter Plot Matrices of Risk Zones 2, 3, 4 Claims Data

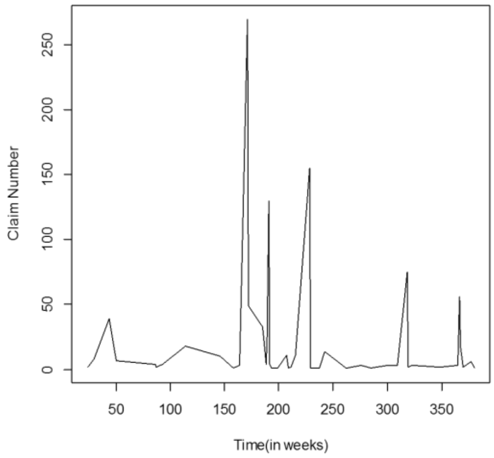


Figure 5.12 The Plot of Claim Number & Time (in weeks) in Risk Zones 2, 3, 4

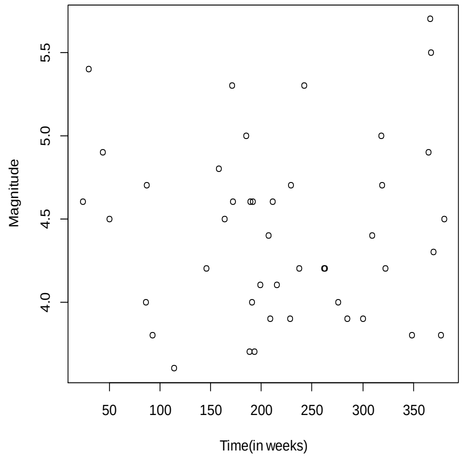


Figure 5.13 The Plot of Magnitude & Time (in weeks) in Risk Zones 2, 3, 4

Figure 5.12 denotes the jumps in claim numbers again, which are observed by risk zones 2, 3 and 4 claims data. At weeks 171, 191 and 228 more than 100 claims are observed. The highest magnitude 5.7 is observed at week 366 in risk zone 2. Magnitude also gathers between 4.0 and 5.0 and shows a homogeneous scatter in weeks in Figure 5.13.

5.2 Modeling

In this section, the main variables of this study, the claim number N_i and the aggregate claim amount $S_i$are modeled by the GLM. The TCIP claims data between 15/12/2000 and 20/07/2008 are studied. As mentioned before, sudden changes in the number of claims when a big earthquake strikes in Turkey are represented by the exponential kernel function. For this study, significant claim numbers $N \geq 30$ are empirically chosen in weeks including all risk zones. There are 41 weeks that are empirically chosen for the analysis. These claims can be seen in Appendices A. Same chosen weeks and their corresponding kernel knots are used for both modeling the claim number and the aggregate claim. In some weeks significant increases are observed after or before ordinary weeks. The Poisson regression is used to model the claim number N_i $(N_i \sim Poisson(\Lambda_i))$ by using log-link function and the aggregate claim amount S_i ,which is assumed to be Lognormal ($logS_i \sim Normal$), is modeled as Gaussian by the GLM with identity link function. It is observed that these big earthquakes are picked by their α_j coefficients of the exponential kernel function during the modeling. Models include all risk zones (risk zone 1, 2, 3 and 4) and magnitude is used as a covariate in models.

Before modeling, the data, which have 12,075 claims, are organized in terms of weeks. In total, there are 396 weeks. However, 236 weeks (more than 50% of the weeks) of the data do not have any claims. Therefore, data include zero values for the claim number and the corresponding aggregate claim amount as well as magnitude. None of the earthquake data are excluded in order not to break the time chain of the data. In cases, where zero values are more than expected for Poisson distribution, there are alternative models to model claim numbers. (Ridout et al., 1998) discuss some alternative models and compare the Mixed Poisson distribution, Zero-Modified distributions (Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) distributions in their study. However, in our case we prefer to use the Poisson distribution for the claim number N_i . However, mentioned models are alternatives for counting data with zero values and it can be analyzed in further studies.

In modeling, the aim is to obtain estimates of β and α_j parameters, where the β parameter represents the features of the each region and α_j parameters pick the sudden jumps of the big earthquakes. Additionally, the nuisance parameter α_0 refers to ordinary claim arrivals, which occur due to small tremors, is the constant during the modeling. The use of these parameter estimates helps us to calculate net premium ($\Pi(S_t) = E(S_t)$) and aggregate claim amount (S_t) that have an important use in actuarial work. The following steps are taken during the modeling of the number of claims and the aggregate claim in R 2.15.1:

Algorithm

1. Choose empirical weeks with large claim numbers ($N \geq 30$) and corresponding kernel knots,

2. Use the exponential kernel function to model the claim number N_i with rate $\Lambda_i = e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}}$, the aggregate claim amount with $\mu_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}$ for $i = 1, \dots, n$ and $j = 1, \dots, k$

3. Use GLM,

4. Randomly select a β value, which represents the features of regions,

5. Model the claim number with Poisson regression with log-link function ($N_i \sim Poisson(\Lambda_i)$) and model the aggregate claim amount as Gaussian with identity link function by empirically chosen kernel knots,

6. Add magnitude as a linear explanatory variable if it gives better results,

7. Choose the β value, which gives the smaller Residual deviance and Akaike's Information Criterion (AIC) in the model and use maximum likelihood estimation of β as check,

8. Run the model with the chosen β,

9. Obtain α_j parameters' estimates from model, which pick the significant earthquakes,

10. Obtain Hessian matrix, which consists of the second partial derivatives of log-likelihood (derived in Chapter 4) with estimations of β and α_j,

11. By using inverse of negative of the Hessian matrix (var-cov matrix), Observed Information Matrix, construct the confidence interval of β parameter, (if it is interested α_j's confidence interval can also be derived),

By following this algorithm, the claim number and the aggregate claim amount models are studied.

Modeling the Claim Number (N_i)

In this section, as a counting data the number of claims N_i , is modeled with Poisson regression with log-link function by using the exponential kernel function (see Section 4.3). The result of the suggested model is given in terms of weeks. We do not separate the data into risk zones (1, 2, 3 or 4), all risk zones are considered in models, while most of the claims are arrived from risk zone 1. While modeling N_i ($N_i \sim Poisson(\Lambda_i)$) former algorithm is followed. Good results were obtained as expected by using the exponential kernel function with empirically chosen time (in weeks) and corresponding kernel knots. α_j ($j = 1, \dots, k$) parameters successfully represent the sudden jumps when big earthquakes occur. Also, the

α_0 the parameter represents the ordinary claims in weeks due to small tremors and the non-linear parameter β represents the different characteristics of the earthquake regions.

In modeling, many other covariates can be added to the model. However, in this thesis only magnitude (M) is used as a linear explanatory variable and added to Model 1 in the following form:

$$N_i \sim Poisson(\Lambda_i),$$

where log-linear Poisson count rate $log\Lambda_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} + M_i$ for the exponential kernel function with $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

The suggested model results are illustrated in Table 5.3. The suggested model is chosen according to smaller residual deviance and AIC among the other models. In the following table, the non-linear parameter β estimation, AIC and residual deviance, which are the main criterions for the GLM analysis for reasonability of the model, and a 95% confidence interval of $\hat{\beta}$ are given. The main idea of the modeling is approved by α_j parameters, which pick the big earthquakes effects while several α_j coefficients are given in Table 5.4. Even small significant jumps are picked by the exponential kernel function in modeling.

Table 5.3 Results of Model 1

$\hat{\beta}$	**Residual Deviance**	**Akaike's Information Criterion (AIC)**	**Confidence Interval for $\hat{\beta}$ (95%)**
10.77	890.89	1,638.1	(10.51597; 11.02403)

Table 5.4 Coefficients of Model 1

Week	**Number of Claims**	α_j **Coefficients of Model 1**
59	1,471	4.08469
121	1,731	4.60574
185	33	1.18331
254	2,913	4.85751
286	448	3.97070

Figures 5.14 and 5.15 verify the validity of the fitted model. As can be seen, actual claim numbers and fitted values of the model show close variability in weeks.

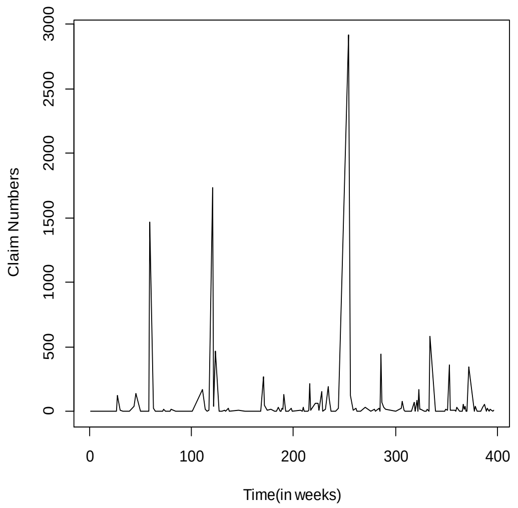


Figure 5.14 The Plot of the Claim Number in Terms of Weeks

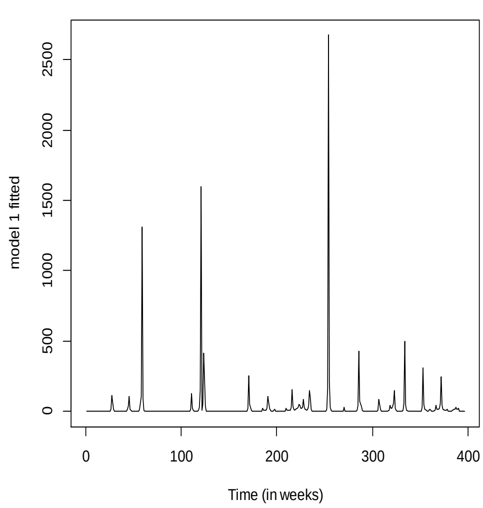


Figure 5.15 The Plot of Fitted Values of Model 1 in Terms of Weeks

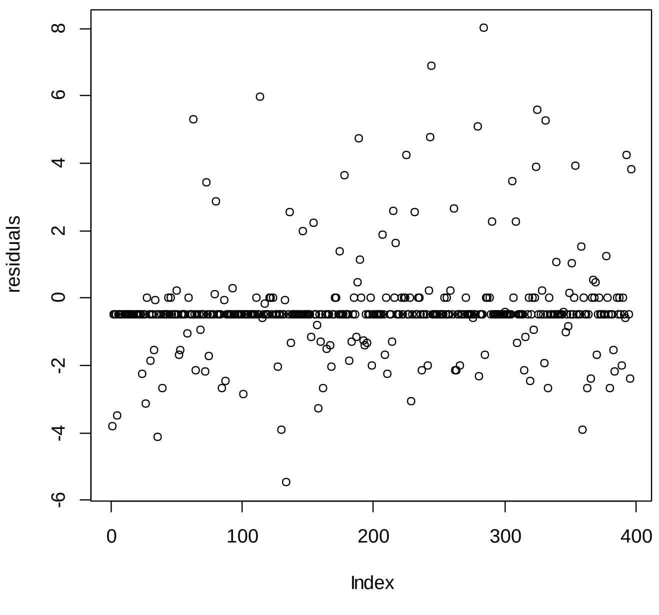


Figure 5.16 The Plot of Residuals of Model 1

Figures 5.16 supports the suggested Model 1 by using the exponential kernel function in terms of weeks. Residuals give the differences between the actual values and the fitted values. It is an alternative way to interpret a good model selection. It also supports the suggested Model 1, which uses the exponential kernel function to fit the model.

Modeling the Aggregate Claim Amount (S_i)

The aggregate claim amount S_i ($S_i = \sum_{i=1}^{N_i} X_i = X_1 + X_2 + \cdots + X_{N_i}$), which is assumed to be Lognormal ($logS_i \sim Normal(\mu_i, \sigma^2)$), is modeled by the GLM with identity link function. The assumption of X_i 's are independently and identically distributed is used while modeling. The individual claim amounts are totaled for each week and log transformation is applied for the aggregate claim amounts for the assumption of normality. The exponential kernel function is used in modeling the aggregate claim amounts with the same idea for the claim numbers. It picks the sudden increases when a big earthquake strikes the country as in the number of claim modeling.

Model 2 with linear explanatory variable Magnitude (M) can be written as

$$logS_i \sim Normal(\mu_i, \sigma^2),$$

where $\mu_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} + M_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k$ is for the use of the exponential kernel function.

The same algorithm is followed for the aggregate claim amount with the modeling of the claim numbers N_i. AIC and residual deviance are used again in model selection. The model with minimum AIC and residual deviance is chosen. The results of Model 2 are given in Table 5.5. The aggregate claim amount S_i infers the aim of the modeling as well. $\alpha_j s$ coefficients of the exponential kernel function represent the big earthquakes like in Model 1. In Table 5.6, estimations of α_j parameters that correspond to the significant claim arrivals are given that are computed by S_i modeling.

Table 5.5 Results of Model 2

$\hat{\beta}$	Residual Deviance	Akaike's Information Criterion (AIC)	Confidence Interval for $\hat{\beta}$ (95%)
0.99	698.12	1,436.3	(0.9899025; 0.9900975)

Table 5.6 Coefficients of Model 2

Week	Number of Claims	α_j Coefficients of Model 2
185	33	0.50876
256	129	2.36839
287	74	2.19065
334	581	5.22886
387	58	0.98665

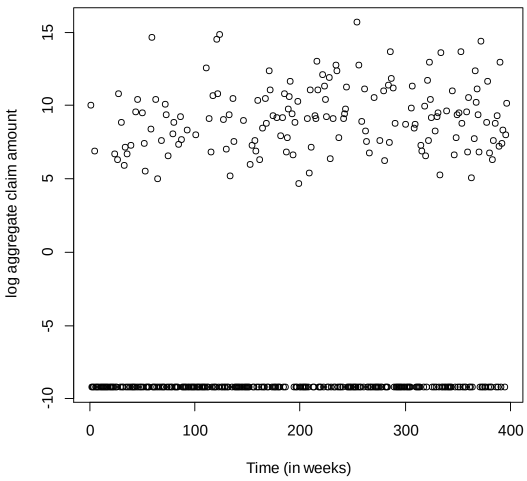


Figure 5.17 The Plot of the log Aggregate Claim Amount in Terms of Weeks

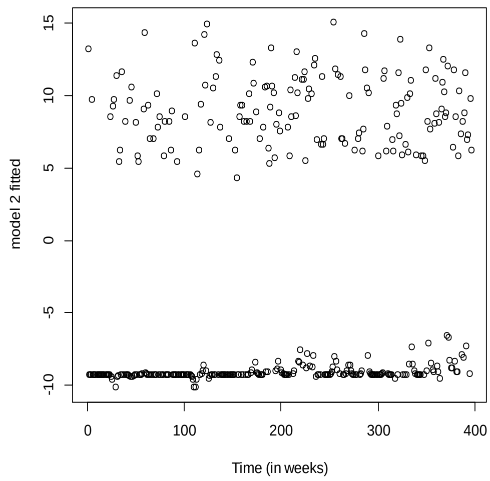


Figure 5.18 The Plot of the Fitted Values of Model 2 in Terms of Weeks

The plot of the fitted values in terms of weeks (Figures 5.18) indicates close scatter with the actual observation values (Figure 5.17). In Figure 5.19, the residuals plot also supports the validity of Model 2.

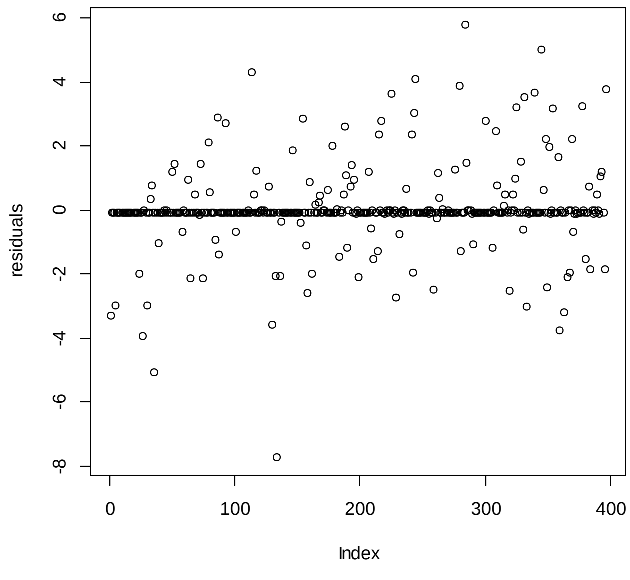


Figure 5.19 The Plot of the Residuals of Model 2

In modeling, the better models that represent the main interest of the study are given. Model 1 and Model 2 with magnitude covariate verify the sudden jumps affects. The α_j coefficients of the exponential kernel function represent sudden jumps when a big earthquake strikes the country.

A Sample Study

In this section, the first 200 weeks of the TCIP data are studied to verify the main idea of the study. That is, the exponential kernel function picks the sudden jumps after big earthquakes. By using the same algorithm, the number of claims (N_i) and the aggregate claim amount (S_i) are modeled. The Poisson regression is used to model the number of claims, N_i, which has a Poisson distribution with Λ_i via log-link function and the aggregate claim amount S_i is assumed to be Lognormal ($logS_i \sim Normal$) and modeled as Gaussian by GLM with identity link function. The weeks corresponding to the number of claims, which are greater or equal to 30 are empirically chosen. α_j ($j = 1, \ldots, k$) parameters of the exponential kernel function represent the sudden jump effects. α_0 and the non-linear parameter β respectively represent the ordinary claim arrivals due to small tremors and the different features of the each earthquake region.

In sample study modeling, the magnitude (M) is used as a linear explanatory variable and added to the Sample Model and the suggested model is

$$N_i \sim Poisson(\Lambda_i),$$

where log-linear Poisson count rate $log\Lambda_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} + M_i$ for the exponential kernel function with $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

The GLM results of the number of claims N_i are denoted in Table 5.7. Smaller non-linear β parameter estimation, residual deviance and AIC are obtained against Model 1. The model with minimum residual deviance and AIC are chosen among other models.

Table 5.7 Results of Sample Model N

$\hat{\boldsymbol{\beta}}$	**Residual Deviance**	**Akaike's Information Criterion (AIC)**	**Confidence Interval for $\hat{\boldsymbol{\beta}}$ (95%)**
8	353.43	652.59	(6.562558; 9.437442)

Table 5.8 Coefficients of Sample Model N

Week	**Number of Claims**	$\boldsymbol{\alpha_j}$ **Coefficients of Sample Model N**
44	39	1.69165
59	1,471	4.30955
121	1,731	4.83980
171	269	3.25398

Table 5.8 gives various α_j ($j = 1, \ldots, k$) coefficients of the exponential kernel function, which successfully represents the sudden jumps after a big earthquake. Moreover, the following Figure 5.20 and 5.21 illustrate close changing with the first 200 week claims and the fitted values of Sample Model N. Figure 5.22 also supports the validity of Sample Model N.

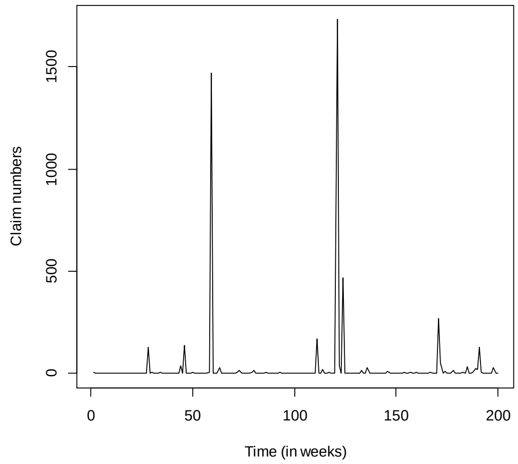


Figure 5.20 The Plot of the First 200 Week Claims

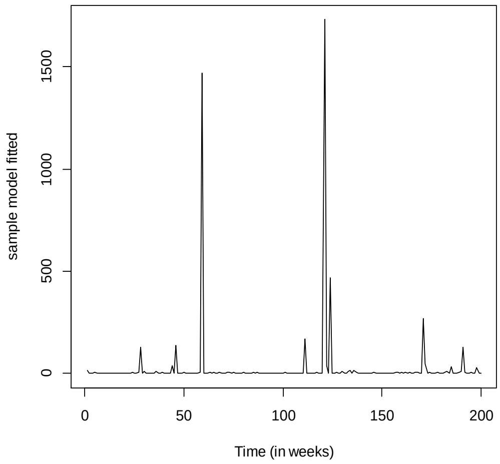


Figure 5.21 The Plot of the Fitted Values of Sample Model N

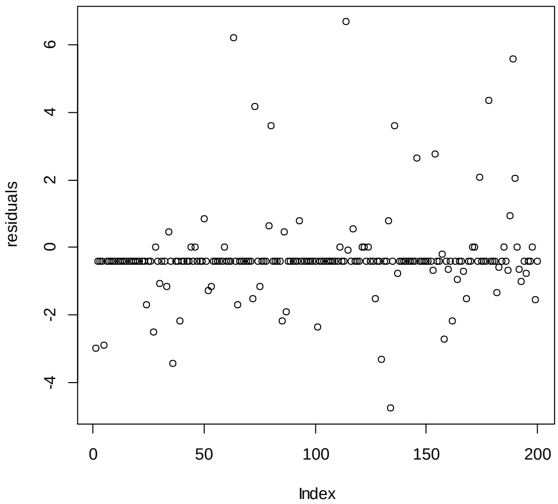


Figure 5.22 The Plot of Residuals of Sample Model N

By using the same argument to use the exponential kernel function for the aggregate claim amount $S_i (S_i = \sum_{i=1}^{N_i} X_i = X_1 + X_2 + \cdots + X_{N_i})$ of the first 200 week claims, the suggested model is

$$logS_i \sim Normal(\mu_i, \sigma^2),$$

where $\mu_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} + M_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k$. The GLM with identity link function is used and the magnitude (M) is used again as a linear explanatory variable. Table 5.9 expresses the results of the chosen model for S_i by using the exponential kernel function. The confidence interval of the non-linear parameter β estimates is very narrow because of very low variance. Lower residual deviance and AIC are obtained for

Sample Model S than Model 2. α_js coefficients represents the jumps after big earthquakes strike like the other models and estimates given Table 5.10.

Table 5.9 Results of Sample Model S

$\hat{\beta}$	Residual Deviance	Akaike's Information Criterion (AIC)	Confidence Interval for $\hat{\beta}$ (95%)
0.7	266.89	657.28	(0.6999025; 0.7000975)

Table 5.10 Coefficients of Sample Model S

Week	Number of Claims	α_j Coefficients of Sample Model S
44	39	0.16536
59	1,471	0.77896
121	1,731	1.67089
171	269	0.18185

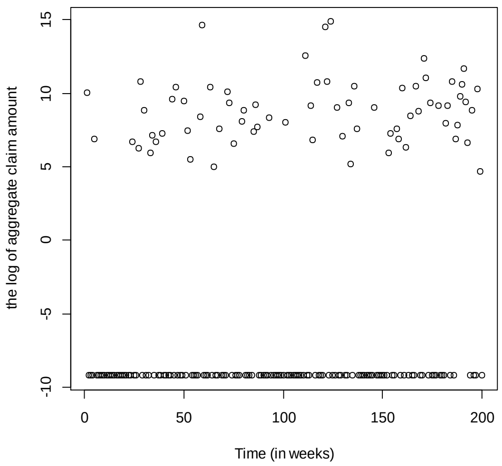


Figure 5.23 The Plot of the log of Aggregate Claim Amount (first 200 week claims)

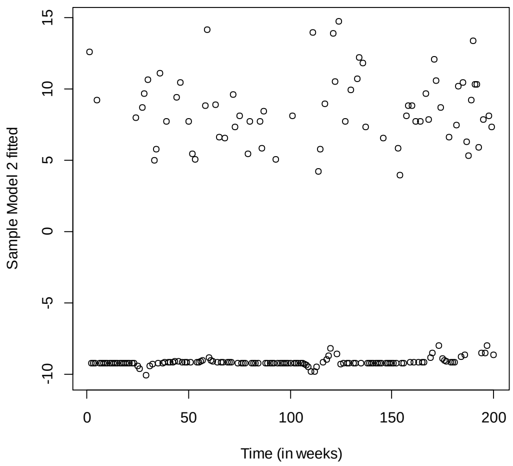


Figure 5.24 The Plot of the Fitted Values of Sample Model S

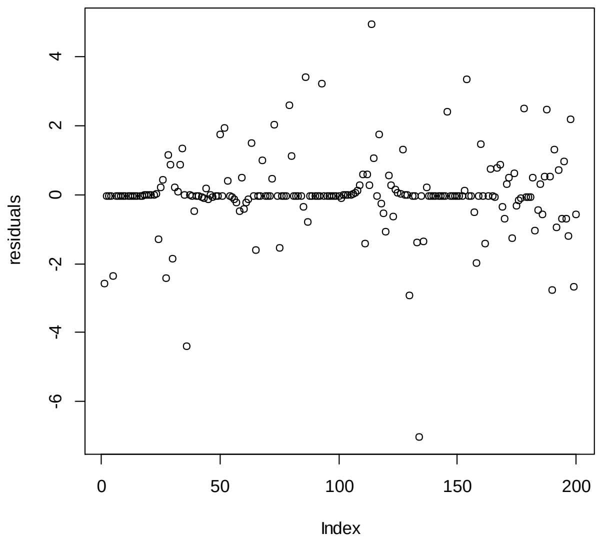


Figure 5.25 The Plot of Residuals of Sample Model S

Figures 5.23, 5.24 and 5.25 are plots that show the usage of the Sample Model S in the behavior of the aggregate claim amount when the exponential kernel function is used.

All models (Model 1, Model 2, Sample Model N and Sample Model S) successfully represent the aim of the study by using the exponential kernel function. α_j s can easily pick the sudden jumps in each model. Results and diagnostic plots support the reasonability of the suggested models.

CHAPTER 6

CONCLUSION AND SUGGESTIONS

It is not possible to predict the exact place, time and magnitude of earthquakes anywhere in the world. A future big earthquake that occurs in Turkey might cause significant losses to the TCIP reserves. However, economic losses due to an earthquake might be reduced by certain precautions and policy procedures to the CEI premium. Therefore, in this study some suggestions on the expected aggregate claim amount are made for the TCIP.

The calculation of aggregate claim amount (S_t) and the expected aggregate claim amount ($E(S_t)$) have a wide use in actuarial context. The number of claims (N_t) and individual claims (X_i) are important elements for these calculations. The estimates of the rates μ_i and Λ_i are needed in the computation of $E(S_t)$ ($E(S_t) = \mu\lambda(t)$). In net premium calculations, interest rate, inflation and expenses are excluded. Net premium equals to the expected value of the aggregate claim amount ($\Pi(S_t) = E(S_t)$). The required μ_i and Λ_i rate are estimated by using the maximum likelihood estimates of the α_j and β parameters of suggested GLM. The exponential kernel function that is used in models for the number of claims and the aggregate claim amount are respectively,

$$\Lambda_i = e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+}} \text{ and } log\Lambda_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \qquad j = 1,\ldots,k \text{ and } \mathrm{i} = 1,\ldots,\mathrm{n},$$

$$\mu_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} \quad j = 1,\ldots,k \text{ and } \mathrm{i} = 1,\ldots,\mathrm{n},$$

where s_j, $j = 1,..,k$ is the empirical kernel knots that the earthquake takes place with α_j effect and t_i is the corresponding kernel knots time, which are chosen to observe the jump effect. The different features of each earthquake region of different years as an explanatory parameter for the GLM is represented by the non-linear parameter β. Also, α_j parameter represents the sudden jumps after big earthquakes and α_0 gives ordinary claim arrivals due to small tremors. Daily small tremors and earthquakes, which are less than five can cause damage to constructions. The parameter estimations of α_j and β will be used for the net premium and the aggregate claim amount of the TCIP. The parameter estimations (α_j and β) are important to express the idea of the modeling. According to the results, the TCIP can check its reserves for unexpected high claims by considering the aggregate claim amount (expected claim payments).

In Chapter 2, some technical information on earthquakes is explained. Moreover, some basic definitions on natural disasters especially on earthquakes are given. Some statistics are given about the natural disaster profile of Turkey and the history of earthquakes is also discussed. The TCIP is explained in detail and some important existing facts of the claims are given. Then, the methodology used in the analysis of the aggregate claim amount and the number of claims are given in Chapter 3. As the main interest of the thesis, the exponential kernel function, the GLM and log-likelihood estimations of the parameters of the functions' computations are derived in Chapter 4. Explanatory data analysis, graphical analysis, suggested models for the number of claims and the aggregate claim amount and their results are discussed in Chapter 5. In addition, a sample study is done for the first 200 week claims and we verify the idea of picking sudden jumps with exponential kernel function by using the GLM.

The exponential kernel function that is used in modeling, the aggregate claim amount, and the claim number, successfully represent the sudden jumps after big earthquakes as expected with α_j parameter. For further studies, different kernel functions can be chosen and studied, and better results can be obtained. The power kernel function was used as an alternative to the exponential kernel function in Başbuğ (2007) for the TCIP claims data that are obtained between years 2000 and 2003. It successfully represented the jump effects of the big earthquakes as well. A further study will be done by using the same power kernel function again with recent claims data. The power kernel function was in the following form,

$$\Lambda_i = e^{\alpha_0 + \sum_{j=1}^{k} \alpha_j \ ((t_i - s_j)|+)^{-\beta}} \qquad j = 1, \ldots, k \text{ and } i = 1, \ldots, n.$$

In addition, the power exponential kernel function (Bozdoğan et al., 2003)

$$f(x; \mu, \sigma, \beta) = \frac{1}{\sigma\Gamma\left(1+\frac{1}{2\beta}\right)2^{1+\frac{1}{2\beta}}} \exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right),$$

where μ, $-\infty < \mu < \infty$ and $\sigma > 0$ are, respectively location and scale parameters and $\beta \epsilon (0, \infty)$ is located on the Kurtosis parameter can be chosen for modeling the aggregate claim amount and the number of claims.

In this thesis, magnitude of an earthquake is used as a covariate in models but for further studies, alternative covariates can be added to the models. Additional features of the regions, population numbers, age of the buildings and structure types of buildings can be added to the models as well. In addition, if the total number of the data is large enough, models can be studied for each risk zone (risk zone 1, 2, 3, 4 and 5). Better models can be obtained with large claims data of the TCIP.

In summary, this thesis suggests the following models for variables of interest, which are number of claims and aggregate claim amount, respectively:

$$N_i \sim Poisson(\Lambda_i),$$

where log-linear Poisson count rate $log\Lambda_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} + M_i$ for the exponential kernel function with $i = 1, \ldots, n$, and $j = 1, \ldots, k$.

$$S_i \sim Lognormal(\mu_i, \sigma^2),$$

where $\mu_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j e^{-\beta(t_i - s_j)|+} + M_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k$ is for the use of the exponential kernel function.

As a further study, the studies of Yücemen (2005) and Yücemen et al. (2009) on insurance premium rates for reinforced concrete and masonry buildings can be combined with thesis premium calculations. Hence, new model suggestions can be given.

In the sample study, the first 200 weeks of the TCIP claim data are studied to verify the aggregate claim amount and the number of claims models. The October 23, 2011 Van/ Erciş earthquake with a magnitude 7.2 caused a peak in the TCIP claims with its magnitude, claim numbers and total payment.

For a further study, the Zero-Inflated Poisson or the Zero-Inflated Negative Binomial distributions can be used instead of the Poisson distribution to model claim numbers.

The TCIP should not limit its coverage to only residential buildings. It should also include business premises and state buildings. A destructive earthquake might cause significant losses in Istanbul, which is the center of the industry and financial sector of Turkey. Therefore, displacement of the financial sector such as the general management of banks is considerable.

Earthquakes cause losses in social, economic and cultural areas. In an emerging economy like Turkey, these losses become much more significant. Mitigation plans should be prepared to reduce impacts of disasters. Insurance is one of these mitigation mechanisms to recover the economy and return it back to the pre-disaster conditions.

The TCIP has become a significant example for other countries by being a private-public partnership in a developing economy for thirteen years. The CEI was revised with Disaster Insurance Law No. 6305 of May 5, 2012 and Decree Law No. 587 was annulled. During the services of water and electricity for homeowners, the CEI becomes a must. With this application, the penetration rate of the TCIP will increase.

The TCIP experienced the October 23, 2011 Van/ Erciş earthquake with a magnitude 7.2. The TCIP paid 15 million USD between 2000 and 2008 for claims, where it paid 56 million USD for the Van/ Erciş earthquake alone. With this example, the TCIP proved that its financial system works. Therefore, if new arrangements are made according to the age of the residential buildings, location, the magnitude of the earthquakes, claim numbers and aggregate claims, losses of the TCIP can be reduced. Different statistical models will support these arrangements by analyzing the available data. However, it should not be forgotten that there can always be better models for the system.

REFERENCES

[1] Achieng, O. M. (2010). *Actuarial Modeling For Insurance Claim Severity In Motor Comprehensive Policy Using Industrial Statistical Distributions.* Bosom Insurance Brookers Ltd., Kenya. Retrieved from http://www.ica2010.com/docs/22_final_paper_Oyugi.pdf. (accessed on January 18, 2013).

[2] Akın, M. (2008). *A Case Study of Public/ Private Partnership: The Turkish Catastrophe Insurance Pool.* MIR Cover Story: Nat CATs in the Middle East, www.meinsurancereview.com.

[3] Alexander, D. (1999). *Natural Disasters.* Dordrecht; Boston: Kluwer Academic Publishers.

[4] Başbuğ, B.B. (2006). *The Mandatory Earthquake Insurance Shame in Turkey.* Retrieved from http://webarchive.iiasa.ac.at/Research/RAV/conf/IDRiM06/pres/basbugpaper.pdf. (accessed on January 22, 2010).

[5] Başbuğ, B. B. (2007). *Modeling of the Turkish Catastrophe Insurance Pool Data, 2000-2003*. (PHD Thesis).

[6] Bickerstaff, D. R. (1972). *Automobile Collision Deductibles and Repair Cost Groups: The Lognormal Model.* Retrieved from http://www.casact.org/pubs/proceed/proceed72/72068.pdf. (accessed on October 13, 2010)

[7] Boland, P.J. (2007). *Interdisciplinary Statistics Statistical and Probabilistic Methods in Actuarial Science,* Boca Raton, FL: Chapman & Hall/CRC.

[8] Boucher,J., Denuit, M. and Guillèn, M. (2008). *Models of Insurance Claim Counts with Time Dependence Based on Generalization of Poisson and Negative Binomial Distributions*. Casuality Actuarial Society. Volume 2/ Issue 1. Retrieved from http://www.variancejournal.org/issues/02-01/135.pdf. (accessed on January 21, 2013).

[9] Bowers, Jr, N. L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997). *Actuarial Mathematics.* 2nd ed. Schaumburg, Ill.: Society of Actuaries.

[10] Bozdoğan, H. and Liu, Z. (2003). *RBF Neural Networks for Classification Using New Kernel Functions.* Neural, Parallel Scientific Computations II (2003), pp. 41-52.

[11] Bülhmann, H. (2005). *Mathematical Methods in Risk Theory.* second printing, Berlin; New York: Springer, 1996, c1970.

[12] Casella, G. and Berger, R. L. (1990). *Statistical Inference*. United States: Brooks/Cole.

[13] Cornell University Earth and Atmospheric Sciences, URL: http://www.geo.cornell.edu/geology/classes/Geo101/eq/EqkWebsite/earthquakev2.html (accessed on April 11, 2012).

[14] Daykin, C. D., Pentikäinen, T. and Pesonen, M. (1994). *Practical Risk Theory for Actuaries,* Monographs on Statistics and Applied Probability. London; New York: Chapman & Hall.

[15] Denuit, M., Maréchal, X., Pitrebois,S. and Wolhin,J.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Clasification, Credibility and Bonus-Malus Systems.* Hoboken: John Wiley & Sons, Incorporated.

[16] Dickson, D. C. M. (2005). *Insurance Risk and Ruin*. Cambridge, UK; New York: Cambridge University Press.

[17] Diggle, P. J., Heagerty, P.J., Liang K. and Zeger, S.L. (2002). *Analysis of Longitudinal Data.* 2nd ed. Oxford; New York: Oxford University Press.

[18] Doğal Afet Sigortaları Kurumu, URL:http://www.dask.gov.tr, page visited at various times in 2011and 2012.

[19] Doğal Afet Sigortaları Kurumu, URL:http://www.dask.gov.tr/240.html (accessed on August 21, 2012).

[20] "EM-DAT: The OFDA/CRED International Disaster Database, www.em-dat.net - Université Catholique de Louvain - Brussels - Belgium", accessed on July 7, 2012.

[21] Gurenko, E., Lester, R., Mahul, O., Gonulal, S.O. (2006). *Earthquake Insurance in Turkey*, *History of Turkish Catastrophe Insurance Pool*. The World Bank Washington, D.C.

[22] Gutenberg, R., and Richter C. F. (1944). *Frequency of earthquakes in California*. Bull. Seismological Society of America , 34, pp. 185-188.

[23] Heckman, P.E. and Meyers, G.G. (1983). *The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distribution.,* (PCAS LXX, 1983) appear in the subsequent volume of the Proceedings (PCAS LXXI, 1984). Retrieved from http://casact.net/pubs/proceed/proceed83/83022.pdf . (accessed on April 10 ,2012).

[24] Hogg, R. V. and Klugman, S. A. (1984). *Loss Distributions*. New York: Wiley.

[25] İlk, Ö. (2011). *R Yazılımına Giriş.* Ankara: Odtü Yayıncılık.

[26] Jong, P.de and. Heller, G.Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge ; New York : Cambridge University Press.

[27] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2001). *Modern Actuarial Risk Theory.* Boston : Kluwer Academic.

[28] Keller, E. A. and Blodgett, R. H. (2006). *Natural Hazards:earth's processes as hazards, disasters, and catastrophes.* Upper Saddle River, NJ: Pearson Prentice Hall.

[29] Klugman, S.A., Panjer, H. H. and Willmot, G.E. (2008). *Loss Models From Data to Decisions.* 3rd ed. Hoboken, N.J. : John Wiley & Sons.

[30] Mosley, R. C., (2008). *Upgrading the Claims Process through Predictive Modeling*. Pinnacle Actuarial Resources, Inc. The Firm of Choice. Retrieved from http://www.pinnacleactuaries.com/Files/Publications/mon-PinnacleMonographJan08ClaimsPredictiveModeling.pdf. (accessed on January 21, 2013).

[31] Özmen, B. and Nurlu, M. (1999). *Deprem Bölgeleri Haritası ile İlgili Bazı Bilgiler*. Retrieved from http://www.deprem.gazi.edu.tr/posts/download?id=43446. (accessed on April 20, 2011).

[32] Pampal,S. and Özmen, B. (2007). *Türkiye Deprem Bölgeleri Haritalarının Gelişimi.* Retrieved from http://www.e-kutuphane.imo.org.tr/pdf/2629.pdf (accessed on November 18, 2010).

[33] Pampal,S. and Özmen,B. (2009). *Depremler Doğal Afet Midir? Depremlerle Baş Edebilmek.* Ankara: Eflatun Yayınevi.

[34] Papush, D.E., Ph.D., FCAS, Patrik, G., FCAS and Podgaits, F.. *Approximations of the Aggregate Loss Distribution*. Retrieved from http://www.casact.org/pubs/forum/01wforum/01wf175.pdf. (accessed on March 15, 2010).

[35] Rejda, G. E. (2005). *Principles of risk management and insurance.* Boston, MA : Addison-Wesley.

[36] Ridout M., Demétrio C.G.B. and Hinde J. (1998). *Models for count data with many zeros*. International Biometric Conference, Cape Town, December 1998. Retrieved from https://www.kent.ac.uk/smsas/personal/msr/webfiles/zip/ibc_fin.pdf. (accessed on September 27, 2012).

[37] Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J. (1999). *Stochastic processes for Insurance and Finance*. Chicester ; New York : J. Wiley.

[38] Rotar, V. I. (2007). *Actuarial Models: the mathematics of insurance*, Boca Raton, FL: Chapman & Hall/CRC.

[39] T.C. Başbakanlık Afet ve Acil Durum Yönetimi Başkanlığı, URL: http://www.deprem.gov.tr/sarbis/shared/DepremHaritalari.aspx (accessed on June 18, 2012).

[40] U.S. Geological Survey, URL: http://earthquake.usgs.gov/earthquakes/eqarchives/year/eqstats.php (accessed on July 7, 2012).

[41] U.S. Geological Survey, URL: http://earthquake.usgs.gov/learn/topics/mercalli.php (accessed on July 7, 2012).

[42] U.S. Geological Survey, URL: http://soundwaves.usgs.gov/2009/11/fieldwork2.html (accessed on July 7, 2012).

[43] Utkucu, M., Çetin, C. and Alptekin, Ö. (2005). *Deprem Zararlarının Azaltılmasında Artçı Depremlerin Önemi*. Deprem Sempozyumu Kocaeli. Retrived from http://kocaeli2007.kocaeli.edu.tr/kocaeli2005/deprem_sempozyumu_kocaeli_2005/8_deprem_ve_insan/d_49_deprem_tehlikesi_riski_ve_deprem_senaryolari/deprem_zararlarinin_azaltilmasinda_artci_depremlerin_onemi.pdf . (accessed on July 26, 2012).

[44] Yücemen, M. S. (2005). *Probabilistic Assessment of Earthquake Insurance Rates for Turkey*. Springer. Natural Hazards Vol. 35, pp. 291-313.

[45] Yücemen, M. S. and Deniz, A. (2009). *Assessment of the earthquake insurance rates for the Turkish Catastrophe Insurance Pool.* Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, Vol. 3 No.2, pp. 67-74.

APPENDIX A

THE FREQUENCIES OF THE NUMBER OF CLAIMS IN WEEKS

Table A.1 The Frequencies of the Number of Claims in Weeks

Week	Frequency	Week	Frequency	Week	Frequency	Week	Frequency
1	6	134	1	**224**	**67**	**323**	**169**
5	1	136	29	225	14	324	22
24	2	137	3	**228**	**155**	325	19
27	2	146	10	229	1	328	5
28	**128**	153	2	231	18	330	4
30	8	154	7	**234**	**191**	331	18
33	1	157	5	**235**	**105**	333	1
34	4	158	1	237	1	**334**	**581**
36	2	160	5	241	1	339	6
39	1	162	1	242	14	345	3
44	**39**	164	3	243	18	346	2
46	**139**	167	6	244	27	348	2
50	7	168	2	**254**	**2913**	349	15
52	1	**171**	**269**	**256**	**129**	351	8
53	1	**172**	**49**	259	14	**353**	**365**
58	5	174	11	261	24	354	13
59	**1471**	178	15	262	1	358	10
63	28	182	2	263	1	359	1
65	1	183	7	266	1	**360**	**35**
68	3	**185**	**33**	**270**	**34**	363	1
72	4	187	2	276	3	365	3
73	16	188	4	279	20	**366**	**56**
75	3	189	23	280	1	367	18
79	4	190	19	284	27	**368**	**42**
80	15	**191**	**130**	285	1	369	6
85	1	192	4	**286**	**448**	370	2
86	4	193	1	**287**	**74**	**372**	**349**
87	2	195	3	**288**	**36**	377	6
93	4	**198**	**30**	290	17	**378**	**43**
101	1	199	1	300	3	380	1
111	**168**	207	11	306	26	383	1
114	18	209	1	**307**	**80**	384	3
115	3	**210**	**36**	308	9	**385**	**34**
117	8	211	2	309	3	**387**	**58**
121	**1731**	214	7	315	1	389	1
122	**39**	215	11	316	2	**390**	**30**
124	**470**	**216**	**214**	**318**	**75**	392	3
127	2	217	11	319	2	393	17
130	1	**221**	**64**	**321**	**87**	395	3
133	13	**223**	**66**	322	3	396	14

* The weeks, where the claim numbers are equal or greater than 30 ($N \geq 30$) (signed with bold) are illustrated. The weeks without any claims are excluded in the table.

APPENDIX B

AN EXAMPLE OF COMPULSORY EARTHQUAKE INSURANCE POLICY

Seri No.: 01415004

DOĞAL AFET SİGORTALARI KURUMU
ZORUNLU DEPREM SİGORTASI POLİÇESİ

99900000000135062437

Poliçe No : 22879772

ANA POLİÇE

Sigorta Şirketi:	GSDASK	**Telefon:**	(212)368-08-88
Acente Adı:	DENEME ACENTE	**Telefon:**	(212)368-08-88
Poliçe No:	22879772	**Tanzim Tarihi:**	23/07/2010
Yenileme/Ek Belge No:	0/0	**Başlangıç-Bitiş Tarihi:**	23/07/2010-23/07/2011
Grup No:	0	**İndirim Tipi:**	

Sigortalı Bilgileri

Adı Soyadı:	ÖRNEK ÖRNEK		
Uyruk:	Yabancı	**Pasaport No:**	11111111111
Sabit Telefonu:		**Cep Telefonu:**	(111)111-11-11
İletişim Adresi:	YILDIZ MAH. BARBAROS CAD. CAD. CİHAN APT. APT. NO: 113 DAIRE: 5 KAT : 5 YILDIZ ISTANBUL BEŞİKTAŞ BEŞİKTAŞ		

Sigorta Ettiren Bilgileri

Adı Soyadı:	ÖRNEK ÖRNEK	**Sıfatı:**	MAL SAHIBI
Uyruk:	Yabancı	**Pasaport No:**	11111111111
Sabit Telefonu:		**Cep Telefonu:**	(111)111-11-11
E-Posta:			

Sigortalı Yere İlişkin Bilgiler

İl/İlçe/Belde:	ISTANBUL/BEŞİKTAŞ/BEŞİKTAŞ		
Adres:	YILDIZ MAH. BARBAROS CAD. CAD. CİHAN APT. APT. NO :113 DAIRE: 5 KAT : 5 YILDIZ 0		
Ada/Pafta:	434/23	**Parsel/Sayfa:**	2/15
Bina İnşa Tarzı:	ÇELIK,BETONARME KARKAS	**Bina İnşa Yılı:**	2000 VE SONRASI
Toplam Kat Sayısı:	05-07 ARASI KAT	**Daire Kullanım Şekli:**	MESKEN
Hasar Durumu:	HASARSIZ	**Evrak Tarih Sayı:**	
Daire Yüzölçümü:	90	**Tarife Fiyatı:**	0,001550
Sigorta Bedeli:	49.500,00 TL	**Poliçe Primi:**	91,73 TL
Oto.Yeni.Süresi:	0		

Genel Şartlar ve Klozlar

Zorunlu Deprem Sigortası tarife ve talimatıyla belirlenen asgari prim tutarından (25 TL) düşük olmamak üzere hesaplanan iş bu poliçede yer alan prim tutarına, anılan tarife ve talimatta yer alan maktu prim (İstanbul ili için 15 TL, İstanbul ili dışı rizikolar için 10 TL) dahildir.

Özel Şartlar

(*) Genel şartlar uyarınca her bir hasarda, sigorta bedelinin %2'si oranında tenzili muafiyet uygulanır. Doğal Afet Sigortaları Kurumu hasarın bu şekilde bulunan muafiyet miktarını aşan kısmından sorumludur. Muafiyet uygulaması açısından, her bir 72 saatlik dönemde meydana gelen bütün hasarlar bir hasar sayılır.

(**) 27.12.1999 tarihinden sonra inşa edilmiş binaların sigortalanabilmesi için, ilgili mevzuat çerçevesinde inşaat ruhsatının bulunması gerekmektedir.

(***) "Orta" hasarlı olarak belirlenen konutların sigortalanabilmesi için, gerekli onarımın yapıldığı belgelendirilmelidir.

Doğal Afet Sigortaları Kurumu, sigortalı/sigorta ettirenin beyanı doğrultusunda bu poliçede yazılı olan bağımsız bölümü/meskeni, bu poliçeye ekli matbu genel şartlar ve özel şartlar dahilinde, yukarıda yazılı olan prim karşılığında yine yukarıda yazılı olan sigorta bedeli üzerinden sigorta eder. Bu poliçe yukarıda yazılı olan başlangıç ve bitiş tarih ve saatleri arasında geçerli primi peşin olarak tahsil eden yetkili sigorta şirketi/acente yetkilisi tarafından imzalanmış olması şartı ile, bu asıl belge makbuz yerine geçer.

"Poliçe bilgilerinize www.dask.gov.tr adresinden ulaşabilirsiniz."

Sigorta sözleşmesini Doğal Afet Sigortaları Kurumu nam ve hesabına düzenleyen yetkili sigorta şirketi/acente	Sigortalı/Sigorta Ettirenin
Adı :	Adı :
İmzası :	Soyadı :
Kaşesi :	İmzası :

Dikkat: Bu poliçenin zemininde mavi renkte küçük harf "dask" filigranı bulunması gerekmektedir.

DOĞAL AFET SİGORTALARI KURUMU Rumeli Caddesi No:48 Kat:7 Şişli 34365 İSTANBUL
Çağrı Merkezi: (212) 368 0 800 Faks: (212) 219 71 88 Internet Adresi: www.dask.gov.tr e-posta: webadmin@dask.gov.tr

Figure B.1 The Compulsory Earthquake Insurance Policy [19]

APPENDIX C

R CODES OF ORGANIZATION OF THE DATA IN TERMS OF WEEKS

```
# 'tarih.txt' includes date, day, month, year, week and event hour
# 'yer.txt' includes city and township
# 'mag.txt' includes claim, magnitude and risk zone

date<-read.table("D:/Profil/gozde.saribekir/tarih.txt",as.is = TRUE, header = TRUE)
place<-read.table("D:/Profil/gozde.saribekir/yer.txt",as.is = TRUE, header = TRUE)
mag<-read.table("D:/Profil/gozde.saribekir/mag.txt",as.is = TRUE, header = TRUE)
data<-cbind(date,place,mag)
data<-data.frame(date,place,mag)
data[,12]<-c(rep(1,12075)) # number of claim equals to 1 for each event
names(data)<-c ( 'Date', 'Day', 'Month', 'Year', 'Week', 'EventHour', 'City', 'Township',
'Claim','Magnitude','Risk Zone','Number of Claim')

# Year 2000

co2000<-c();    # the vector that includes week, claim amount, magnitude, risk zone and
number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2000") # the year that is studied
co2000<-c(co2000,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11]))}
t.2000<-table(co2000);
dt2000<-matrix(c(co2000),ncol=6);
t.dt2000<-t(dt2000);
data_2000<-c(51,6,sum(t.dt2000[,3]),5.8,1);
names(data_2000)<- c('week','N','S','M','R.Z')

d2000=diag(0,52,5)
m=1:52
d2000[,1]=t(m)
colnames(d2000)<- c('week','N','S','M','R.Z')
for(i in 1:length(data_2000))
for(j in 1:dim(d2000)[1])
if (data_2000==d2000[j,1])
d2000[j,]=data_2000


# Year 2001

comb2001<-c();
co2001<-c();  # the vector that includes week, claim amount, magnitude, risk zone and
number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2001")  # the year that is studied
co2001<-(c(co2001,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2001),nrow=5);
a<-t(k);
data_2001<-c();
```

```
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);  # the vector that includes all week numbers
n.nr<-c();
n.nr<- as.vector(table(a[,1]));            #number of claims for each week
data_2001<-c(week.nr,n.nr);
data_2001<-matrix(c(as.matrix(data_2001)),ncol=2);

m<-c(a[1,4]);        # the vector of magnitude
for (i in 2:dim(a)[1]-1)
{
        if (a[i,1]!= a[i+1,1]){
        m<-c(m,a[i,4]);
        }

}

if(m[1]==m[2])
m<-m[-1];
if( ((a[dim(a)[1],4]) !=(m[length(as.matrix(m))])))
m<-c(m,a[dim(a)[1],4]);
m[4]=5.5
m[12]=4.5
dim(as.matrix(m))

r<-c(a[1,5]);         # the vector of risk zones
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);
}
}
if(r[1]==r[2])
r<-r[-1];
if( ((a[dim(a)[1],5]) !=(r[length(as.matrix(r))])))
r<-c(r,a[dim(a)[1],5]);
dim(as.matrix(r))
r[14]=1

x.nr<-c();           # aggregate claim amount of each week
for(j in 1:dim(data_2001)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2001[j,1]==a[i,1])
sum=sum+a[i,3];
}
x.nr<-c(x.nr,sum);
```

```
}

data_2001<-c(data_2001,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2001<-matrix(c(data_2001),ncol=5);
colnames(data_2001)<- c('week','N','S','M','R.Z')

d2001=diag(0,52,5)
m=1:52
d2001[,1]=t(m)
colnames(d2001)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2001)[1])
        for(j in 1:dim(d2001)[1])
if (data_2001[i,1]==d2001[j,1])
d2001[j,]=data_2001[i,]
```

Year 2002

```
comb2002<-c();
co2002<-c();     # the vector that includes week, claim amount, magnitude, risk zone and number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2002")  #  the year that is studied
co2002<-(c(co2002,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2002),nrow=5);
a<-t(k);
data_2002<-c();
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);      # the vector that includes all week numbers
n.nr<-c();
n.nr<- as.vector(table(a[,1]));              # number of claims for each week
data_2002<-c(week.nr,n.nr);
data_2002<-matrix(c(as.matrix(data_2002)),ncol=2);

m<-c();                      # the vector of magnitude
for (i in 2:dim(a)[1]-1)
{
        if (a[i,1]!= a[i+1,1]){
        m<-c(m,a[i,4]);
}
}

if( ((a[dim(a)[1],4]) !=(m[length(as.matrix(m))])))
m<-c(m,a[dim(a)[1],4]);
m[1]=4.7
m[3]=4.8
m[8]=4.6
dim(as.matrix(m))
```

```
r<-c(a[1,5]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);    # the vector of risk zones
}
}
if(r[1]==r[2])
r<-r[-1];
dim(as.matrix(r))
r[1]=1
r[12]=2
r[13]=2
r[14]=3
r[15]=1

x.nr<-c();    # aggregate claim amount of each week
for(j in 1:dim(data_2002)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2002[j,1]==a[i,1])
sum=sum+a[i,3];
}
x.nr<-c(x.nr,sum);
}
dim(as.matrix(x.nr))

data_2002<-c(data_2002,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2002<-matrix(c(data_2002),ncol=5);
colnames(data_2002)<- c('week','N','S','M','R.Z')

d2002=diag(0,52,5)
m=1:52
d2002[,1]=t(m)
colnames(d2002)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2002)[1])
        for(j in 1:dim(d2002)[1])
if (data_2002[i,1]==d2002[j,1])
d2002[j,]=data_2002[i,]
```

 # Year 2003

```
comb2003<-c();
co2003<-c();    # the vector that includes week, claim amount, magnitude, risk zone and number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2003")  # the year that is studied
co2003<-(c(co2003,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2003),nrow=5);
a<-t(k);
data_2003<-c();
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
```

```
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);               # the vector that includes all week numbers
n.nr<-c();
n.nr<- as.vector(table(a[,1]));                     # number of claims for each week
data_2003<-c(week.nr,n.nr);
data_2003<-matrix(c(as.matrix(data_2003)),ncol=2);
m<-c(a[1,4]);     # the vector of magnitudes
for (i in 2:dim(a)[1]-1)
{
          if (a[i,1]!= a[i+1,1]){
          m<-c(m,a[i,4]);
}
}
if(m[1]==m[2])
m<-m[-1];
if( ((a[dim(a)[1],4]) !=(m[length(as.matrix(m))])))
m<-c(m,a[dim(a)[1],4]);
m[7]=6.4
dim(as.matrix(m))

r<-c(a[1,5]);   # the vector of risk zones
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);
}
}
if(r[1]==r[2])
r<-r[-1];
if( ((a[dim(a)[1],5]) !=(r[length(as.matrix(r))])))

r<-c(r,a[dim(a)[1],5]);
dim(as.matrix(r))

x.nr<-c();                              # aggregate claim amount for each week
for(j in 1:dim(data_2003)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2003[j,1]==a[i,1])
sum=sum+a[i,3];
}
x.nr<-c(x.nr,sum);
}
dim(as.matrix(x.nr))

data_2003<-c(data_2003,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2003<-matrix(c(data_2003),ncol=5);
colnames(data_2003)<- c('week','N','S','M','R.Z')
```

```
d2003=diag(0,52,5)
m=1:52
d2003[,1]=t(m)
colnames(d2003)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2003)[1])
        for(j in 1:dim(d2003)[1])
if (data_2003[i,1]==d2003[j,1])
d2003[j,]=data_2003[i,]
```

 # Year 2004

```
comb2004<-c();
co2004<-c();              # the vector that includes week, claim amount, magnitude, risk zone
and  number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2004")                # the year that is studied
co2004<-(c(co2004,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2004),nrow=5);
a<-t(k);
data_2004<-c();
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);#the vector that includes all week numbers
n.nr<-c();
n.nr<- as.vector(table(a[,1]));   # number of claims for each week
data_2004<-c(week.nr,n.nr);
data_2004<-matrix(c(as.matrix(data_2004)),ncol=2);
m<-c(a[1,4]);
for (i in 2:dim(a)[1]-1)
{
        if (a[i,1]!= a[i+1,1]){
        m<-c(m,a[i,4]);
}
}
if(m[1]==m[2])
m<-m[-1];
if( ((a[dim(a)[1],4]) !=(m[length(as.matrix(m))])))

m<-c(m,a[dim(a)[1],4]);
dim(as.matrix(m))
m[12]=5
m[17]=4

r<-c(a[1,5]);   # the vector of risk zones
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);
```

```
}
}
if(r[1]==r[2])
r<-r[-1];
if( ((a[dim(a)[1],5]) !=(r[length(as.matrix(r))])))

r<-c(r,a[dim(a)[1],5]);
dim(as.matrix(r))
r[12]=2
r[17]=2

x.nr<-c();    # aggregate claim amount of each week
for(j in 1:dim(data_2004)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2004[j,1]==a[i,1])
sum=sum+a[i,3];
}
x.nr<-c(x.nr,sum);
}
dim(as.matrix(x.nr))

data_2004<-c(data_2004,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2004<-matrix(c(data_2004),ncol=5);
colnames(data_2004)<- c('week','N','S','M','R.Z')

d2004=diag(0,53,5)
m=1:53
d2004[,1]=t(m)
colnames(d2004)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2004)[1])
        for(j in 1:dim(d2004)[1])
if (data_2004[i,1]==d2004[j,1])
d2004[j,]=data_2004[i,]

# Year 2005

comb2005<-c();
co2005<-c();                # the vector that includes week, claim amount, magnitude, risk zone and  number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2005")          # the year that is studied
co2005<-(c(co2005,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2005),nrow=5);
a<-t(k);
data_2005<-c();
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
```

```
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);                    #  the vector that includes all week numbers
n.nr<-c();
n.nr<- as.vector(table(a[,1]));                         # number of claims for each week
data_2005<-c(week.nr,n.nr);
data_2005<-matrix(c(as.matrix(data_2005)),ncol=2);

m<-c(a[1,4]);                              # the vector of magnitudes
for (i in 2:dim(a)[1]-1)
{
        if (a[i,1]!= a[i+1,1]){
        m<-c(m,a[i,4]);
}
}
if(m[1]==m[2])
m<-m[-1];
for (i in 2:dim(as.matrix(m))-1)
m[i]=m[i+1]
m[1]=5.1
m[2]=4.1
m[3]=4.4
m[11]=4.9
m[17]=4.1
m[18]=4.2
m[20]=4.8
dim(as.matrix(m))

r<-c(a[1,5]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);
}
}
if(r[1]==r[2])
r<-r[-1];
if( ((a[dim(a)[1],5]) !=(r[length(as.matrix(r))])))

r<-c(r,a[dim(a)[1],5]);
r[2]=2
r[17]=1
r[18]=1
r[20]=1
r[24]=2
dim(as.matrix(r))

x.nr<-c();                        # aggregate claim amount of each week
for(j in 1:dim(data_2005)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2005[j,1]==a[i,1])
sum=sum+a[i,3];
}
```

```
x.nr<-c(x.nr,sum);
}
dim(as.matrix(x.nr))

data_2005<-c(data_2005,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2005<-matrix(c(data_2005),ncol=5);
colnames(data_2005)<- c('week','N','S','M','R.Z')

d2005=diag(0,52,5)
m=1:52
d2005[,1]=t(m)
colnames(d2005)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2005)[1])
        for(j in 1:dim(d2005)[1])
if (data_2005[i,1]==d2005[j,1])
d2005[j,]=data_2005[i,]
d2005[52,1]=52
d2005[52,2]=1
d2005[52,3]=1875
d2005[52,4]=4.2
d2005[52,5]=2

# Year 2006

comb2006<-c();
co2006<-c();                    # the vector that includes week, claim amount, magnitude,
risk zone and  number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2006")     # the year that is studied
co2006<-(c(co2006,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2006),nrow=5);
a<-t(k);
data_2006<-c();
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);    # the vector that includes all week numbers

n.nr<-c();
n.nr<- as.vector(table(a[,1]));#number of claims for each week
data_2006<-c(week.nr,n.nr);
data_2006<-matrix(c(as.matrix(data_2006)),ncol=2);

m<-c(a[1,4]);   # the vector of magnitudes
for (i in 2:dim(a)[1]-1)
{
        if (a[i,1]!= a[i+1,1]){
        m<-c(m,a[i,4]);
}
```

```
}
if(m[1]==m[2])
m<-m[-1];
if( ((a[dim(a)[1],4]) !=(m[length(as.matrix(m))])))
m<-c(m,a[dim(a)[1],4]);
m[3]=4
m[6]=3.8
m[9]=4.3
m[15]=3.9
dim(as.matrix(m))

r<-c(a[1,5]);           # the vector of risk zones
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);
}
}
if(r[1]==r[2])
r<-r[-1];
if( ((a[dim(a)[1],5]) !=(r[length(as.matrix(r))])))
r<-c(r,a[dim(a)[1],5]);
dim(as.matrix(r))

x.nr<-c();
for(j in 1:dim(data_2006)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2006[j,1]==a[i,1])
sum=sum+a[i,3];
}
x.nr<-c(x.nr,sum);
}
dim(as.matrix(x.nr))

data_2006<-c(data_2006,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2006<-matrix(c(data_2006),ncol=5);
colnames(data_2006)<- c('week','N','S','M','R.Z')

d2006=diag(0,52,5)
m=1:52
d2006[,1]=t(m)
colnames(d2006)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2006)[1])
        for(j in 1:dim(d2006)[1])
if (data_2006[i,1]==d2006[j,1])
d2006[j,]=data_2006[i,]
```

 # Year 2007

```
comb2007<-c();
co2007<-c();                    # the vector that includes week, claim amount, magnitude, risk zone and  number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2007")          # the year that is studied
co2007<-(c(co2007,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2007),nrow=5);
a<-t(k);
data_2007<-c();
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);      # the vector that includes all week numbers
n.nr<-c();
n.nr<- as.vector(table(a[,1]));               # number of claims for each week
data_2007<-c(week.nr,n.nr);
data_2007<-matrix(c(as.matrix(data_2007)),ncol=2);

m<-c(a[1,4]);            # the vector of magnitudes
for (i in 2:dim(a)[1]-1)
{
        if (a[i,1]!= a[i+1,1]){
        m<-c(m,a[i,4]);
}
}
if(m[1]==m[2])
m<-m[-1];
if( ((a[dim(a)[1],4]) !=(m[length(as.matrix(m))])))
m<-c(m,a[dim(a)[1],4]);
m[8]=3.9
m[9]=4.1
m[11]=3.9
m[20]=4.3
m[24]=4.2
m[26]=4.9
m[28]=5.5
dim(as.matrix(m))

r<-c(a[1,5]);                 # the vector of risk zones
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);
}
}
if(r[1]==r[2])
r<-r[-1];
```

```
if( ((a[dim(a)[1],5]) !=(r[length(as.matrix(r))])))
r<-c(r,a[dim(a)[1],5]);
r[26]=4
r[28]=2
dim(as.matrix(r))

x.nr<-c();                    # aggregate claim amount
for(j in 1:dim(data_2007)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2007[j,1]==a[i,1])
sum=sum+a[i,3];
}
x.nr<-c(x.nr,sum);
}
dim(as.matrix(x.nr))

data_2007<-c(data_2007,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2007<-matrix(c(data_2007),ncol=5);
colnames(data_2007)<- c('week','N','S','M','R.Z')

d2007=diag(0,52,5)
m=1:52
d2007[,1]=t(m)
colnames(d2007)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2007)[1])
        for(j in 1:dim(d2007)[1])
if (data_2007[i,1]==d2007[j,1])
d2007[j,]=data_2007[i,]
```

 # Year 2008

```
comb2008<-c();
co2008<-c();                  # the vector that includes week, claim amount, magnitude, risk
zone and  number of claim
for(i in 1:dim(data)[1])
{
if (as.matrix(data[i,4])=="2008")      # the year that is studied
co2008<-(c(co2008,c(data[i,5],data[i,12],data[i,9],data[i,10],data[i,11])))
}
k<- matrix(c(co2008),nrow=5);
a<-t(k);
data_2008<-c();
week.nr<-c(a[1,1]);
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
week.nr<-c(week.nr,a[i,1]);
}
}
if(week.nr[1]==week.nr[2])
week.nr<-week.nr[-1];
if( ((a[dim(a)[1],1]) !=(week.nr[length(as.matrix(week.nr))])))
week.nr<-c(week.nr,a[dim(a)[1],1]);       # the vector that includes all week numbers
n.nr<-c();
```

```
n.nr<- as.vector(table(a[,1]));                # number of claims for each week
data_2008<-c(week.nr,n.nr);
data_2008<-matrix(c(as.matrix(data_2008)),ncol=2);
m<-c();                                         # the vector of magnitudes
for (i in 2:dim(a)[1]-1)
{
        if (a[i,1]!= a[i+1,1]){
        m<-c(m,a[i,4]);
}
}
if( ((a[dim(a)[1],4]) !=(m[length(as.matrix(m))])))
m<-c(m,a[dim(a)[1],4]);
m[4]=3.7
m[6]=4.8
m[15]=4.2
m[17]=4
dim(as.matrix(m))

r<-c();            # the vector of risk zones
for (i in 2:dim(a)[1]-1)
{
if (a[i,1]!= a[i+1,1]){
r<-c(r,a[i,5]);
}
}
if( ((a[dim(a)[1],5]) !=(r[length(as.matrix(r))])))
r<-c(r,a[dim(a)[1],5]);
r[4]=1
r[6]=1
r[15]=1
r[17]=1
dim(as.matrix(r))

x.nr<-c();     # aggregate claim amount
for(j in 1:dim(data_2008)[1])
{
sum=0;

for(i in 1: dim(a)[1])
{
if (data_2008[j,1]==a[i,1])
sum=sum+a[i,3];
}
x.nr<-c(x.nr,sum);
}
dim(as.matrix(x.nr))

data_2008<-c(data_2008,as.matrix(x.nr),as.matrix(m),as.matrix(r))
data_2008<-matrix(c(data_2008),ncol=5);
colnames(data_2008)<- c('week','N','S','M','R.Z')

d2008=diag(0,52,5)
m=1:52
d2008[,1]=t(m)
colnames(d2008)<- c('week','N','S','M','R.Z')
for(i in 1:dim(data_2008)[1])
        for(j in 1:dim(d2008)[1])
if (data_2008[i,1]==d2008[j,1])
```

```
d2008[j,]=data_2008[i,]
```

Combination of all years as 'all_data'

```
all_data<-data.frame(rbind(d2000,d2001,d2002,d2003,d2004,d2005,d2006,d2007,d2008))
all_data[,1]=1:469
all_data=all_data[51:446]
all_data[,1]=1:396
```

APPENDIX D

R CODES OF THE NUMBER OF CLAIMS AND THE AGGREGATE CLAIM AMOUNT MODELING

 # The Number of Claims Modeling

```
time<-c(1:396)
beta<- 10.77   # derived β estimator by Maximum Likelihood Estimation

# exponential kernel functions:

kernel28<-(exp((-beta)*abs(time-28)))
kernel44<-(exp((-beta)*abs(time-44)))
kernel46<-(exp((-beta)*abs(time-46)))
kernel59<-(exp((-beta)*abs(time-59)))
kernel111<-(exp((-beta)*abs(time-111)))
kernel121<-(exp((-beta)*abs(time-121)))
kernel122<-(exp((-beta)*abs(time-122)))
kernel124<-(exp((-beta)*abs(time-124)))
kernel171<-(exp((-beta)*abs(time-171)))
kernel172<-(exp((-beta)*abs(time-172)))
kernel185<-(exp((-beta)*abs(time-185)))
kernel191<-(exp((-beta)*abs(time-191)))
kernel198<-(exp((-beta)*abs(time-198)))
kernel210<-(exp((-beta)*abs(time-210)))
kernel216<-(exp((-beta)*abs(time-216)))
kernel221<-(exp((-beta)*abs(time-221)))
kernel223<-(exp((-beta)*abs(time-223)))
kernel224<-(exp((-beta)*abs(time-224)))
kernel228<-(exp((-beta)*abs(time-228)))
kernel234<-(exp((-beta)*abs(time-234)))
kernel235<-(exp((-beta)*abs(time-235)))
kernel254<-(exp((-beta)*abs(time-254)))
kernel256<-(exp((-beta)*abs(time-256)))
kernel270<-(exp((-beta)*abs(time-270)))
kernel286<-(exp((-beta)*abs(time-286)))
kernel287<-(exp((-beta)*abs(time-287)))
kernel288<-(exp((-beta)*abs(time-288)))
kernel307<-(exp((-beta)*abs(time-307)))
kernel318<-(exp((-beta)*abs(time-318)))
kernel321<-(exp((-beta)*abs(time-321)))
kernel323<-(exp((-beta)*abs(time-323)))
kernel334<-(exp((-beta)*abs(time-334)))
kernel353<-(exp((-beta)*abs(time-353)))
kernel360<-(exp((-beta)*abs(time-360)))
kernel366<-(exp((-beta)*abs(time-366)))
kernel368<-(exp((-beta)*abs(time-368)))
kernel372<-(exp((-beta)*abs(time-372)))
kernel378<-(exp((-beta)*abs(time-378)))
kernel385<-(exp((-beta)*abs(time-385)))
kernel387<-(exp((-beta)*abs(time-387)))
kernel390<-(exp((-beta)*abs(time-390)))
```

```
modelN396<-glm (all_data[,2]~kernel28+ kernel44+ kernel46+ kernel59+ kernel111+
kernel121+ kernel122+ kernel124+ kernel171+ kernel172+ kernel185+ kernel191+
kernel198+ kernel210+ kernel216+ kernel221+ kernel223+ kernel224+ kernel228+
kernel234+ kernel235+ kernel254+ kernel256+ kernel270+ kernel286+ kernel287+
kernel288+ kernel307+ kernel318+ kernel321+ kernel323+ kernel334+ kernel353+
kernel360+ kernel366+ kernel368+ kernel372+ kernel378+ kernel385+ kernel387+
kernel390+ all_data[,4], family=poisson (link="log"),data=all_data)

summary(modelN396)
```

 # Maximum Likelihood Estimation of the Number of Claims

```
library("maxLik")
loglikFun<-function(param)
{
t<-c(1:396)
s<-c (28, 44, 46, 59, 111, 121, 122, 124, 171, 172, 185, 191, 198, 210, 216, 221, 223, 224,
228, 234, 235, 254, 256, 270, 286, 287, 288, 307, 318, 321, 323, 334, 353, 360, 366, 368,
372,378,385,387,390)  # s is the emprically selected kernel knots vector (the week with claim
number N>=30)

alfa0=param[1]
alfa=param[2:42]
b=param[43]
Lsum=0
x<-all_data[,2]  # claim number column of the all_data

for(i in 1:396){

top=alfa0+sum(alfa+exp(-b*abs(t[i]-s)))
Lsum=Lsum+x[i]*top-exp(top)
}
Lsum
}
param=c(1,alfa,b)

ml<-maxLik(loglikFun,start=param,method="nm")
```

 # The Aggregate Claim Amount Modeling

```
time<-c(1:396)
beta<-0.99      # derived β estimator by Maximum Likelihood Estimation

# exponential kernel functions:

kernel28<-(exp((-beta)*abs(time-28)))
kernel44<-(exp((-beta)*abs(time-44)))
kernel46<-(exp((-beta)*abs(time-46)))
kernel59<-(exp((-beta)*abs(time-59)))
kernel111<-(exp((-beta)*abs(time-111)))
kernel121<-(exp((-beta)*abs(time-121)))
kernel122<-(exp((-beta)*abs(time-122)))
kernel124<-(exp((-beta)*abs(time-124)))
kernel171<-(exp((-beta)*abs(time-171)))
kernel172<-(exp((-beta)*abs(time-172)))
kernel185<-(exp((-beta)*abs(time-185)))
kernel191<-(exp((-beta)*abs(time-191)))
```

```
kernel198<-(exp((-beta)*abs(time-198)))
kernel210<-(exp((-beta)*abs(time-210)))
kernel216<-(exp((-beta)*abs(time-216)))
kernel221<-(exp((-beta)*abs(time-221)))
kernel223<-(exp((-beta)*abs(time-223)))
kernel224<-(exp((-beta)*abs(time-224)))
kernel228<-(exp((-beta)*abs(time-228)))
kernel234<-(exp((-beta)*abs(time-234)))
kernel235<-(exp((-beta)*abs(time-235)))
kernel254<-(exp((-beta)*abs(time-254)))
kernel256<-(exp((-beta)*abs(time-256)))
kernel270<-(exp((-beta)*abs(time-270)))
kernel286<-(exp((-beta)*abs(time-286)))
kernel287<-(exp((-beta)*abs(time-287)))
kernel288<-(exp((-beta)*abs(time-288)))
kernel307<-(exp((-beta)*abs(time-307)))
kernel318<-(exp((-beta)*abs(time-318)))
kernel321<-(exp((-beta)*abs(time-321)))
kernel323<-(exp((-beta)*abs(time-323)))
kernel334<-(exp((-beta)*abs(time-334)))
kernel353<-(exp((-beta)*abs(time-353)))
kernel360<-(exp((-beta)*abs(time-360)))
kernel366<-(exp((-beta)*abs(time-366)))
kernel368<-(exp((-beta)*abs(time-368)))
kernel372<-(exp((-beta)*abs(time-372)))
kernel378<-(exp((-beta)*abs(time-378)))
kernel385<-(exp((-beta)*abs(time-385)))
kernel387<-(exp((-beta)*abs(time-387)))
kernel390<-(exp((-beta)*abs(time-390)))


all_data[all_data[,3]==0,3]=0.0001    # put 0.0001 instead of 0 values of aggregate claim amount to take logarithm
all_data[,3]=log(all_data[,3])
x=all_data[,3]   # aggregate claim amount column of the all_data

modelS<-glm (all_data[,3]~ kernel28+ kernel44+ kernel46+ kernel59+ kernel111+
kernel121+ kernel122+ kernel124+ kernel171+ kernel172+ kernel185+ kernel191+
kernel198+kernel210+kernel216+ kernel221+ kernel223+ kernel224+ kernel228+
kernel234+ kernel235+ kernel254+ kernel256+ kernel270+ kernel286+ kernel287+
kernel288+ kernel307+ kernel318+ kernel321+ kernel323+ kernel334+
kernel353+kernel360+ kernel366+kernel368+ kernel372+kernel378+ kernel385+kernel387+
kernel390 + all_data[,4],family=gaussian(link = "identity"),data=all_data)

summary(modelS)


# Maximum Likelihood Estimation of the Aggregate Claim Amount

library("maxLik")
loglike<-function(param)
{
t<-c(1:396)
s<-c (28, 44, 46, 59, 111, 121, 122, 124, 171, 172, 185, 191, 198, 210, 216, 221, 223, 224,
228, 234, 235, 254, 256, 270, 286, 287, 288, 307, 318, 321, 323, 334, 353, 360, 366, 368,
372,378,385,387,390) # s is the emprically selected kernel knots vector (the week with claim number N>=30)
```

```
alfa0=param[1]
alf=param[2:42]
b=param[43]
top=0
x<-all_data[,3]

for(i in 1:396){

top1=(x[i])^2/(2*sigma^2)
top2=x[i]*(alfa0+sum(alf*exp(-b*abs(t[i]-s))))/sigma^2
top3=(alfa0+sum(alf*exp(-b*abs(t[i]-s))))^2/(2*sigma^2)
top=top-top1+top2-top3
}
top
}
param=c(1,alf,b)

mls<-maxLik(loglike,start=param,method="nm")
```

APPENDIX E

R CODES OF THE HESSIAN MATRICES FOR THE NUMBER OF CLAIMS AND THE LOG OF THE AGGREGATE CLAIM AMOUNT

```
# Derivatives of log-likelihood of the number of claims

modelN=summary(modelN)    # glm model result
alfa<-modelN$coeff[2:42,1]     # alfa (j) coefficients
t<-c(1:396)              # t is number of weeks
s<-
(28,44,46,59,111,121,122,124,171,172,185,191,198,210,216,221,223,224,228,234,235,254,
256,270,286,287,288,307,318,321,323,334,353,360,366,368,372,378,385,387,390)
# s is the emprically selected kernel knots vector (the week with claim number N>=30)
b=beta                    # selected beta parameter
alfa0=1                   # α0 parameter
x<-all_data[,2]          # claim number column of the data

#  2nd derivative of alfa(zero)   #

c=0
for (i in 1:396)
{
  sumk=sum(alfa*(exp(-b*abs(t[i]-s))))
  c = c - exp(alfa0+sumk)
}

# derivative of alfa(zero) and alfa[j]  #

d=0
for(i in 1:396)
{
  sumk1=sum(alfa*(exp(-b*abs(t[i]-s))))
  sumk2=sum(exp(-b*abs(t[i]-s)))
  d = d - exp(alfa0+sumk1)*sumk2
}

# derivative of beta and alfa(zero)  #

e=0
for(i in 1:396)
{
  sumk1=sum(alfa*(exp(-b*abs(t[i]-s))))
  sumk2=sum(alfa*exp(-b*abs(t[i]-s))*abs(-t[i]+s))
  e = e - sumk1*sumk2
}

# derivative of beta and alfa[j]  #

f=0
for(i in 1:396)
{
  sumk1=-x[i]*sum(exp(-b*abs(t[i]-s))*abs(-t[i]+s))
```

```
sumk2=-exp(alfa0+sum(alfa*(exp(-b*abs(t[i]-s)))))*sum(exp(-b*abs(t[i]-s)))*sum(alfa*exp(-b*abs(t[i]-s))*abs(-t[i]+s))
  sumk3=-sum(alfa*exp(-b*abs(t[i]-s))*abs(-t[i]+s)*exp(alfa0+sum(alfa*exp(-b*abs(t[i]-s)))))
  f=f+sumk1+sumk2+sumk3
  }
```

 # 2nd derivative of beta #

```
g=0
for(i in 1:396)
{
   sumk1=-x[i]*sum(alfa*exp(-b*abs(t[i]-s))*(abs(-t[i]+s))^2)
   sumk2=-exp(alfa0+sum(alfa*(exp(-b*abs(t[i]-s)))))*(sum(alfa*exp(-b*abs(t[i]-s))*abs(-t[i]+s)))^2
   sumk3=-exp(alfa0+sum(alfa*(exp(-b*abs(t[i]-s)))))*(sum(alfa*exp(-b*abs(t[i]-s))*(abs(-t[i]+s))^2))
   g=g+sumk1+sumk2+sumk3
}
```

 # 2nd derivative of alfa[j] #

```
h=0
for(i in 1:396){
sumk=-exp(alfa0+sum(alfa*(exp(-b*abs(t[i]-s)))))*(sum(exp(-b*abs(t[i]-s))))^2
h=h+sumk
}
```

 # derivative of alfa[j]and alfa[j'] #

```
m=0
for(i in 1:396){
  for(j in 1:41){
   for(n in 1:41){
    if (j!=n){
    sumk1=alfa0+sum(alfa*(exp(-b*abs(t[i]-s[j]))))
    sumk2=exp(-b*abs(t[i]-s[j]))
    sumk3=exp(-b*abs(t[i]-s[n]))
    m=m-exp(sumk1)*sumk2*sumk3
}
}
}
}
```

 # Hessian matrix (43X 43)

```
hesN=matrix(1,nrow=43,ncol=43)
myhesN=hesN*m
diag(myhesN)=h
myhesN[1,1]=c
myhesN[1,2:42]=d
myhesN[2:42,1]=d
myhesN[1,43]=e
myhesN[43,1]=e
myhesN[43,2:42]=f
myhesN[2:42,43]=f
myhesN[43,43]=g
```

 # Confidence Interval

```
varcovn=solve(-myhesN)
lowerlimn<- b-(1.96)* sqrt(varcovn[43,43]) # varcovs[43,43] variance of β parameter
upperlimn<- b+(1.96)*sqrt(varcovn[43,43])
```

 # Derivatives of the log-likelihood of log of the aggregate claim amount

```
t<-c(1:396)      # t is number of weeks
s<-
c(28,44,46,59,111,121,122,124,171,172,185,191,198,210,216,221,223,224,228,234,235,25
4,256,270,286,287,288,307,318,321,323,334,353,360,366,368,372,378,385,387,390)
# s is the emprically selected kernel knots vector (the week with claim number N>=30)
b=beta           # selected beta parameter
alfa0=1          # α0 parameter
models=summary(modelS)
alf<-models$coefficients[2:42,1]      # alfa (j) coefficients
sigma<-sqrt(var(x))
```

 # 2nd derivative of beta #

```
y=0
for (i in 1:396){                            # i is number of week
for (j in 1:41) {                            # j is the number of kernel knots
denom=sigma^2*exp(b*t[i])
denom2=(sigma^2)*(exp(b*t[i]))^2
sum1=x[i]*t[i]^2*sum(alf*exp(b*s[j]))/denom
sum2=-(x[i]*t[i]*sum(alf*s*exp(b*s[j])))/(denom)
sum3=(x[i]*sum(alf*s[j]^2*exp(b*s[j])))/(2*denom)
sum4=-(alfa0*t[i]^2*sum(alf*exp(b*s[j])))/(2*denom)
sum5=(alfa0*t[i]*(sum(alf*s*exp(b*s[j]))))/denom
sum6=-(alfa0*sum(alf*s[j]^2*exp(b*s[j])))/denom
sum7=-(t[i]^2*(sum(alf*exp(b*s[j])))^2)/denom2
sum8=(2*t[i]*sum(alf*exp(b*s[j]))*sum(alf*s[j]*exp(b*s[j])))/denom2
sum9=-((sum(alf*s[j]*exp(b*s[j])))^2)/(2*denom2)
sum10=-(sum(alf*exp(b*s[j]))*sum(alf*s[j]^2*exp(b*s[j])))/(2*denom2)
y=y+sum1+sum2+sum3+sum4+sum5+sum6+sum7+sum8+sum9+sum10
}
}
```

 # derivative of beta and alfa[j] #

```
p=0
for (i in 1:396){
for (j in 1:41) {
sum1=(x[i]*t[i]*sum (exp(b*s[j])))/denom
sum2=(x[i]*sum(s[j]*exp(b*s[j])))/denom
sum3=(alfa0*t[i]*sum(exp(b*s[j])))/(2*denom)
sum4=(alfa0*sum(s[j]*exp(b*s[j])))/(2*denom)
sum5=(2*t[i]*(sum(alf*exp(b*s[j])))*(sum(exp(b*s[j]))))/(2*denom2)
sum6=((sum(exp(b*s[j])))*(sum(alf*s[j]*exp(b*s[j]))))/(2*denom2)
sum7=((sum(alf*exp(b*s[j])))*(sum(s[j]*exp(b*s[j]))))/(2*denom2)
p=p-sum1+sum2-sum3+sum4-sum5+sum6+sum7
}}
```

 # derivative of beta and alfa0 #

```
r=0
for(i in 1:396){
  for (j in 1:41) {
 sum1=-(sum(s[j]*alf*exp(b*s[j]))/(2*sigma^2*exp(b*t[i])))
sum2=t[i]*sum(exp(b*s[j]))/(2*sigma^2*exp(b*t[i]))
r=r+sum1+sum2
}}
# 2nd derivative of alfa[j] #
z=0
for(i in 1:396){
for (j in 1:41) {
  sum1=(sum(exp(b*s[j])))^2
  sum2=sigma^2*(exp(b*t[i]))^2
  z=z-(sum1/sum2)
}}
```

 # derivative of alfa[j] and alfa0 #

```
v=0
for(i in 1:396){
for (j in 1:41) {
   sum1=sum(exp(b*s[j]))
   sum2=sigma^2*(exp(b*t[i]))
v=v-(sum1/sum2)
}
}
```

 # derivative of alfa0 and alfa0 #

```
n=396
w=-n/sigma^2
```

 # derivative of alfa[j]and alfa[j'] #

```
k=0
for(i in 1:396){
  for (j in 1:41) {
  for (n in 1:41) {
    if (j!=n){
        sum1=exp(-b*abs(t[i]-s[j]))
        sum2=exp(-b*abs(t[i]-s[n]))
        k=k+sum(sum1*sum2)
}}}}
k=-k/(sigma^2)
```

 # Hessian Matrix (43X43)

```
hes=matrix(1,nrow=43,ncol=43)
hes=hes*k
diag(hes)=z
hes[1,1]=w
hes[1,2:42]=v
hes[2:42,1]=v
hes[43,1]=r
hes[1,43]=r
hes[43,2:42]=p
```

```
hes[2:42,43]=p
hes[43,43]=y
```

Confidence Interval

```
varcovs=solve(-hes)

lowerlims<-b-(1.96)*sqrt (varcovs[43,43])   # varcovs[43,43] variance of β parameter
upperlims<-b+(1.96)*sqrt(varcovs[43,43])
```