A Bayesian Hierarchical Poisson Approach to Estimate County Level Suicide Risks in Idaho

A Thesis Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
with a
Major in Statistical Science
in the
College of Graduate Studies
University of Idaho
by
Rosebella Capio

Major Professor: Michelle Wiest, Ph.D.
Committee Members: Timothy Johnson, Ph.D.; Clarissa Richardson, Ph.D.
Department Administrator: Christopher Williams, Ph.D.

May 2019

Authorization to Submit Thesis

This thesis of Rosebella Capio, submitted for the degree of Master of Science with a major in Statistical Science and titled "A Bayesian Hierarchical Poisson Approach to Estimate County Level Suicide Risks in Idaho," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: ________________________________ Date______________
Michelle Wiest, Ph.D.

Committee
Members: ________________________________ Date______________
Timothy Johnson, Ph.D.

________________________________ Date______________
Clarissa Richardson, Ph.D.

Department
Administrator: ________________________________ Date______________
Christopher Williams, Ph.D.

Abstract

Suicide is defined as the act of causing harm to oneself with the goal of dying is a growing public health concern. It tends to claim the lives of both the old and young. According to the World Health Organization, an estimated number of $800,000$ people have been reported to die each year by suicide.

Most published literature are concerned with the rate of suicide but not much work has been done in determining how certain populations or interest groups are at risk of suicide beyond the broad classification of rural and urban. As a result, this work aimed at estimating suicide relative risk levels in all counties in Idaho using a statistical model. Each county's risk obtained is relative to the entire population. Specifically, a Bayesian hierarchical Poisson was used to get these estimates.

Custer and Madison counties were found to have the highest and lowest relative risks respectively when age was unadjusted. After age adjustment, Clark county had the highest relative risk and Ada had the lowest relative risk.

Average unemployment rates were compared with the relative risks the answer the question of possible causation. From the results, unemployment rates cannot be concluded as the sole cause of high/low suicide relative risks as some high risk areas had average or low unemployment rates.

Acknowledgements

First, I would like to express my sincere appreciation to my advisor Dr. Michelle Wiest for her continuous support, for her patience, expertise, time, and for believing in me in the development and completion of this research. Her immense concern and interest in the high rise of suicides in rural Idaho was the initial jot for the research and work behind this thesis. She consistently steered me in the right direction whenever necessary and allowed me to own my work. Her guidance helped me think critically and also prepare me for future research works.

Also, I would like to thank my committer members: Dr. Timothy Johnson and Dr. Clarissa Richardson for their intense contribution and input in this work and also for dedicating their time. Their individual expertise and knowledge on this thesis is highly appreciated and I am extremely grateful.

Dedication

I would express my profound gratitude to my mother, Mrs. Joyce Okrah, for her love, prayers, support and encouragement throughout my school years and during this thesis writing period. I love and appreciate you mum! I will always make you proud!

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Suicide is one of the leading causes of death in the US. In fact, death by suicide was ranked as the 10th leading cause of death with a total of 44,965 deaths across all US states in 2016 [1]. In the US alone, about 750,000 people have died by suicide within the last 25 years and these numbers are twice the lives claimed by homicides [2]. Surprising, every 40 seconds, someone dies by suicide somewhere in the world [3]. In 2017, suicide rates for all states in the US were reported by the CDC and is seen in figure 1 below:

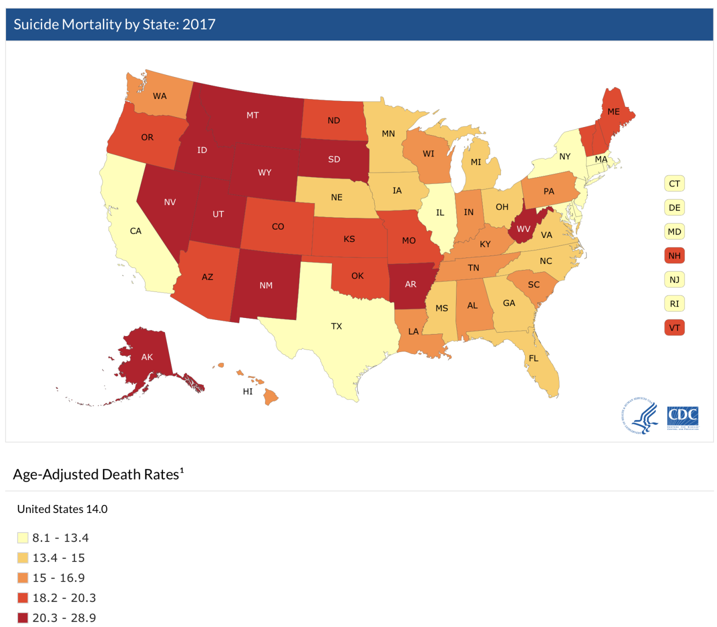

Figure 1.1: Suicide Mortality by State: 2017
Source: Centers for Disease Control & Prevention

In Idaho, suicide is the second leading cause of death for Idahoans aged 15 - 34 and between 10 - 14 for Idahoan males. A total of 1,178 and 323 suicide deaths were recorded from 2010 - 2014 for Idahoan males and females females respectively. 110 school children in Idaho between the ages of 6 and 18 years were reported to have died by suicide with 25 of those deaths being children below 15 years[4].

Suicide rates in the state has consistently been on the rise and an increase from 38% to 58% was realized between 1999 and 2016. A rate of 20.8 (which was a slight decrease from previous years) was recorded in 2016 but this still surpassed that of the national rate by 50% , making it the 8th highest rate in the US for 2016 [**5**].

Comparing Idahoan male and female deaths, the rates have significantly been higher for males than females throughout the years. A rate of 30.5 deaths per $100,000$ males was reported for the period between 2009 and 2013 [**6**]. Again, from 2010 to 2014, male suicide rate was 30.3, which was also significantly higher than that of females [**7**]. These gender rates appear to always exceed that of the national gender suicide rate. Figure 2 shows the male to female suicide rates from 2007 to 2016.

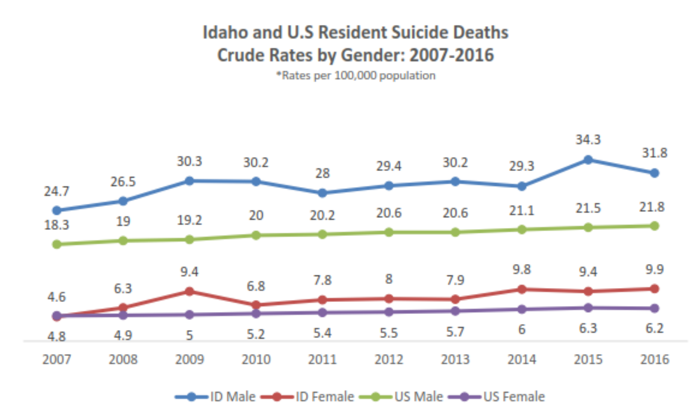

Figure 1.2: Rate per $100,000$ population by gender
Source: Centers for Disease Control & Prevention

1.1 Persons at Risk of Suicide

Most deaths by suicide are by people who have or are experiencing trauma or some form day to day crisis. Most of these deaths are not pre-planned but occur when one experiences a crisis period such as terminal illness or the sudden loss of a close relative [**8**]. High rates of suicide have also been recorded among people in minority groups such the LGBT. Within different races and ethnicities, the non-Hispanic American Indian/Alaska Native and non-Hispanic White populations tend to have the highest suicide rates. For persons between the ages of $10-34$, death by suicide ranks 2^{nd} as the major cause of death, ranks 4^{th} for ages $35-54$, and 8^{th} for persons aged $55-64$ [**9**].

1.2 Some Risk Factors of Suicide

Several factors have been established as leading to suicide. The Suicide Prevention Resource Center have identified the factors below as those that increase the likelihood of a person committing suicide [**10**];

- Prior suicide attempt(s)
- Misuse and abuse of alcohol or other drugs
- Mental disorders, especially depression
- Access to lethal means
- Social isolation
- Chronic disease and disability
- Lack of access to behavioral health care

1.3 Current Method Used In Determining Suicide Rates in Idaho

The Idaho Bureau of Vital Records and Health Statistics, Department of Health and Welfare, calculates suicide rates using 5 year aggregates. These rates are age-adjusted and reported for each health district in Idaho. The method used in obtaining the age-adjusted rates is known as the direct method of estimation. Age-adjusted crude rates are often used to compare rates of different populations by controlling the differences in age distribution. Deaths rates of the population in question are applied to the age distribution of a reference population [**11**]. The 2000 US population by age group is used by the Bureau as its reference population in its calculation of expected deaths rates. Expected death are those deaths which would have occurred in the reference population given the exact same death rates of the reference population in each age group. The age-adjusted death rate is calculated as;

$$\text{Age specific death rate (ASDR)} = \frac{\text{total deaths}}{\text{age specific population}} \times 100000 \tag{1.1}$$

Age-adjusted death rate (AADR) is given:

$$AADR = \sum(\text{ASDR} \times \text{standard proportion}) \tag{1.2}$$

Idaho has seven health districts: Panhandle, North Central, Southwest, South Central, Southeastern, Central and Eastern. Figure 3 shows the counties in each health district.Between $2013-2017$, district 6 had the highest suicide rate with a 24.7, followed by district 2 with a 24.2 rate, and district 5 with 22.3. The health district with the lowest rate was district 7 with 18.3 [**12**].

A drawback of the approach used by the Bureau is that, it is difficult to evaluate trends in suicide deaths per county through out the years. The 5 year aggregate method used here might also result in the loss of information due to the use of direct estimates. Also, more accurate estimates for rates and suicide relative risk estimates could be determined if a model based approach was used in its determination.

1.4 Method Used In This Analysis

A hierarchical Bayes poisson random effects model was used in this analysis. Rate estimates were age-adjusted to the U.S. 2000 standard population obtained from the National Cancer Institute, and expressed per 100,000 persons per year. This model will aim to account for the different suicide case counts for each year as some years may tend to have a higher or lower rates in comparison with other years. These differences in suicide counts could be due to several factors such as the unemployment rate of a particular year [**13**].

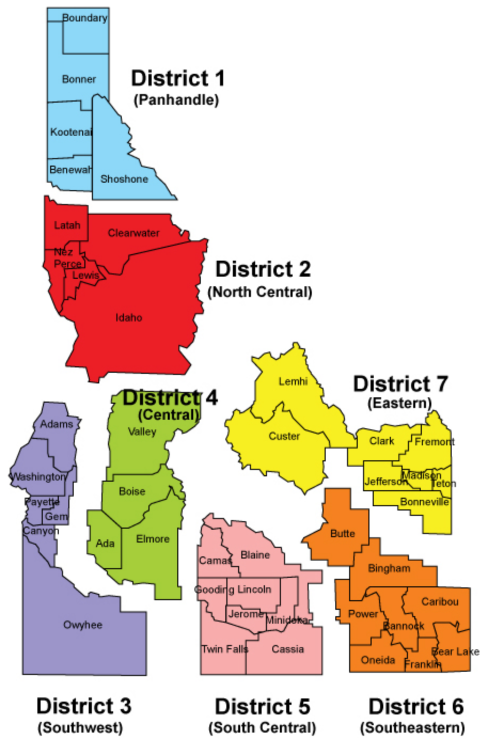

Figure 1.3: Idaho Health Districts
Source: Idaho Bureau of Vital Records and Health Statistics

This work aims to identify counties with low relative suicide risks and those with high relative suicide risks due to the lack of research in terms of suicide relative risks. Through the use of a hierarchical Bayes model, estimates for these risks obtained would aid governments, policy makers, and researchers in the allocation of resources and further studies in suicide risks. With available data on unemployment rates, time series analysis was relied upon to examine the trend in risks and unemployment rates to lay the groundwork for small area estimation per year per age group.

1.5 Standard Estimation

Currently, the suicide rate estimation method being used by the CDC is that of direct estimation. The Center of Diseases Prevention and Control uses age-adjusted death rates which are calculated using the direct method and the 2000 U.S. standard population data [**14**]. The center does not implement any model in its estimation. Similarly, the state of

Idaho employs the use of direct estimation using equations 1.1 and 1.2 and the 2000 U.S. standard population data only [**15**].

1.6 Models Estimating Suicide Rates

There have been several works done in attempts to model suicide rates in the USA. Some of these models are: Suicide rates models during recessions. With the increase in unemployment rates from $5.8\% - 9.6\%$ during the recession period of $2007 - 2010$, time-trend regression models were relied on to evaluate to increase in suicide. The model indicated that the rise in suicide rates by 3.8% corresponding to 1330 suicide deaths resulted from the unemployment caused by the recession for that period [**16**]. Zero-inflated negative binomial regression has also been used to study suicidal behavior. This model was used in a study conducted to determine whether thoughts of hopelessness and burdensomeness were significantly connected to variability to death ideation on 239 adults who were ≥ 60 years. A positive correlation was found to exist. Furthermore, a significant reduction in the probability of suicide ideation was seen when when there is interaction between their needs being met and their perceived burdensomeness and suicide ideators believing their needs to be met [**17**]. In terms of suicide trends, spring was found to have the highest risk of suicide in Finland between $1979-1999$. Poisson Regression was used to model the data and the time series analysis of suicide deaths were analyzed using a seasonal-trend decomposition procedure [**18**]. With the lack of efficient studies suicide rates in small domains, small area estimation methods have been relied upon to provide estimates of suicide rates at the county-level. In an analysis conducted from $2016-2017$ using hierarchical Bayesian odels, which is a small area estimation method, the posterior predicted mean county-level suicide rates rose by $\geq 10\%$ between $2005 - 2015$ for 99% of all US counties. Counties in the western and northwestern US, except Southern California and parts of Washington, were observed to have the highest of these rates [**19**].

CHAPTER 2

Methods

2.1 Introduction

This chapter explains and describes the data source, mathematical methods, how the data was analyzed, and the distribution and models used for the work.

In this work, a procedure for analyzing suicide deaths with the aim recommending to the Idaho Department of Health and Welfare, epidemiologists and public health researchers a reasonable model approach for estimating suicide mortality risks in counties in Idaho.

2.2 Motivation and Objectives

This study was motivated by the lack of research in the high rise of suicide deaths in Idaho especially in its rural parts. The state of Idaho has consistently been among the states with the highest suicide rates, with an estimated $7,100$ of Idahoans aged 18 and above attempting suicide between 2012 to 2016. The Suicide Prevention Action Network of Idaho reported a total of 78 Idaho school children (aged 18 and below) died by suicide in 2018 [**20**].The state's per capita suicide death rate exceeds that of the national rate by more than a third as reported by the Idaho Vital Statistics with rural counties having the highest rates. Despite low population density in rural Idaho which implies low suicide death counts, the rate of suicide far exceed that of the national rate. Elderly men, teenage males, working age males and young native American males have been found to be at the highest risks to die by suicide in Idaho. These deaths are alarming and have detrimental effects on the families of the deceased, the society, economy, friends, etc. This study was not only inspired by the alarming rates of suicide but also by how family and the society are affected by such deaths.

2.3 Data Source and Cleaning

Data on all deaths were first obtained from the Idaho Department of Health and Welfare. The variables contained in this dataset were; year of death, age at death, manner of death, Idaho county where death occurred and gender of deceased person. In order to work with the data properly, death by suicides were extracted since those deaths were the ones of interest. Also, the population of each of the 44 counties were obtained for the different years represented in the data. Data from the United States Census Bureau was used for population estimates. The data from the Bureau contained county population with respect to specified age groups for 2010. Such age specific county population was not available for the 2000 census year. As such, the 2010 data was used as the constant reference for the rest of the years. From here, these two datasets were merged to suit the purpose of the study.

2.4 Small Area Estimation

Usually, counties, a subpopulation of interest or a small geographical area is referred to as a “small area" [**21**]. This method of estimation is primarily concerned with obtaining reliable estimates of a parameter interest in areas where the sample size is too small and estimates obtained from direct surveys can't be relied upon due to sampling errors and large standard errors [**22**]. To address the issues associated with the estimates derived through direct surveys, indirect methods are used. This approach “borrows" strength from neighboring small areas using linking models, census data and available administrative records [**23**]. Empirical best prediction (EBP), hierarchical Bayesian and empirical Bayes are some of the models often used for small area estimation [**24**]

2.5 Hierarchical Bayesian for Small Area Estimation

Hierarchical Bayesian approach has been one of the recent methods for determining rates for small domain. Though this new outlook has some its advantages, its use is still very limited [**25**]. The hierarchical Bayesian models makes use of several models which are structured hierarchically to estimate posterior distributions of the model parameters with the use of the Bayes theorem. The posterior distribution is the probability of the parameter(s) of interest, for example, θ, given some evidence (prior) and likelihood of the observations and expressed mathematically as given in equation (2.1) and often written as a proportionality as stated in equation (2.2) [**26**].

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{2.1}$$

$$\text{Posterior probability} \propto \text{Likelihood} \times \text{Prior probability} \tag{2.2}$$

2.5.1 The Poisson Random Effects Model

A random effects model is often used in hierarchical Bayesian models. Random effects models are often used to account for overdispersion in a dataset. In a scenario where the model set up is hierarchical, the first stage, counts, Y_i is conditional on its parameter, λ_i and is assumed to follow a Poisson distribution with mean $\lambda_i E_i$, such that E_i is the expected deaths for each county. In the second stage, the distribution of the relative risk λ_i is assumed to have a particular probability density function $g(\lambda_i)$ called a prior distribution [**27**].

2.5.2 Likelihood Specification

A Poisson model for the suicide counts is assumed for each county. Let Y_i denote the number of suicide cases observed from $1999-2016$ such that each Y_i is distributed as;

$$Y_i \sim Poisson(\lambda_i E_i),\ i = 1, \cdots, 44 \tag{2.3}$$

where,

- Y_i is assumed to be independent,
- Expected deaths were calculated using both the crude and age-specific rates,
- λ_i is the unknown relative risk associated in county$_i$. The term relative risk is used to refer to the risk of each county relative to the baseline risk. Given $Y_i \sim Poisson(\lambda_i E_i)$, it implies that the mean $\mu_i = \lambda_i E_i$ which is achieved by reparameterization. This is given by;

$$\begin{aligned} \mu_i &= \lambda_i \\ &= \frac{Y_i/n_i}{Y_T/N} \times n_i \times \frac{Y_T}{N} \end{aligned} \tag{2.4}$$

  where $\lambda_i = \frac{Y_i/n_i}{Y_T/N}$ and the baseline relative risk $E_i = n_i \times \frac{Y_T}{N}$

  - Y_i's are the suicide counts
  - n_i is the population size per county,
  - N is total population in Idaho
  - Y_T is the total suicide counts

2.5.3 Prior Specification

In selecting a prior distribution for λ_i, a normal random effects prior is frequently used. This is done to because it can accommodate both spatial dependence and unstructured heterogeneity in a conditional autoregressive setting [**28**]. This implies that the distribution of the relative risk λ_i is assumed to have a certain probability density function $g(\lambda_i)$ called the prior distribution such that,

$$\log \lambda_i = \alpha + \theta_i \tag{2.5}$$

where;

- α is the mean log relative risk ,
- θ_i is the random effect term accounting for county differences.

α, and θ_i in both equations 2.2 and 2.3 are parameters of the prior distribution which are called hyperparameters such that

$$\theta_i \sim \mathcal{N}(0, \tau) \tag{2.6}$$

$$\alpha \sim \mathcal{N}(0, 10^{-3}) \tag{2.7}$$

τ in equation 2.6 is the parameter of the hyperparameter θ_i. Its distribution is known as hyperprior distribution. A gamma distribution is assumed for the hyperprior and is given as a gamma distribution with shape and rate parameters given in equation 2.8,

$$\tau \sim Gamma(10^{-3}, 10^{-3}) \tag{2.8}$$

such that τ is the precision.

2.5.4 Just Another Gibbs Sampler

To ensure the efficient analysis of the data, Just Another Gibbs Sampler (JAGS) was used. JAGS, which was developed by Martyn Plummer in 2003, is an open-source engine for the BUGS language written in C++ which aims to analyze Bayesian hierarchical models through the use of Markov chain Monte Carlo. Through Markov chain Monte Carlo algorithms, JAGS takes samples from probability distributions using the Gibbs sampler. To implement a JAGS model, the 4 essential specifications required are [**29**];

- Data specification
- Model specification
- Compilation of the model and
- Initialization of the model

Relative risks (λ_i) were the parameters of interest in this work. For these risk values to be reported, they were stated in the initialization stage. Summary statistics, diagnostic plots were also obtained. Trace plots and the distributions of the relative risks obtained were also examined.

2.6 Data Manipulation and JAGS Implementation

The total number of deaths for each of the 44 counties for 18 years was first obtained. This was done by getting each county's population across the number of years in question and summing the different suicide count for each county across the 18 year period. Death by suicide rate for the entire state of Idaho was then obtained by the sum all deaths in Idaho divided by Idaho population for the 18 year period and then multiplied by $100,000$. Expected deaths, E_i, are also obtained from the data. This method of using the available data to get expected deaths is termed as internal standardization.

To run a jags model on the data, the total number of counties, suicide deaths, expected number of deaths were specified. A burn-in of 1000 was used and “coda samples" are used to draw samples from the posterior distribution. These samples were drawn because we did not have a closed form for the posterior distribution. A total of $100,000$ samples were drawn and λ_i also being drawn. This gave a total of 46 parameters. That is, 44 relative risk parameters, for all the counties, with the remaining two being α and τ.

CHAPTER 3

Analysis and Results

This section reports the findings from the data. Suicide risk levels for the 44 counties were estimated using the MCMC algorithm. The posterior means reported by the algorithm for each county represents their respective suicide risks for the 18 year period. Credible intervals were generated along with the risk estimate to show how precise the estimates produced are. A relative risk above one implies that the county has a relative risk of suicide above the state average and vice versa for relative risk value less than one. If relative risk is equal to one, then the respective county's relative risk is the same as the state's. A section of the relative risks estimated from the MCMC algorithm is given below with rounded variances.

3.1 Results

To assess convergence of the MCMC algorithm, traceplots and the Gelman-Rubin diagnostic test were used. Four chains with $100,000$ iterations each were used and the traceplots showed good mixing for each estimated relative risk signifying convergence. Density plots obtained appeared to be fairly normal for each estimate. The Gelman-Rubin test statistic for each parameter estimate was approximately one implying that the variances for each parameter was reduced as the chain was run four times.

Tables 3.1 and 3.2 highlight the high risk counties. Custer County in district 7 had the highest suicide risk over the 18 year period with a value of 1.79, which was closely followed by Lemhi County with 1.78, then Shoshone County with 1.59, Clark with 1.42, Boundary County with 1.51 and Bear Lake with 1.49. After age adjustment, relative risks obtained for Custer, Lemhi, Shoshone and Clark were 11.09, 5.98, 4.14 and 28.29 respectively. There is an obvious difference in relative risks after adjusting for ages. Clark county was seen to have the highest relative risk but the estimate is highly uncertain due to it having the lowest population in Idaho.

Table 3.1: High Risk Counties

County	Age-unadjusted λ_i	Variance	95% CI
Bannock	1.13	0.00*	(1.00, 1.26)
Bear Lake	1.49	0.06	(1.05, 2.01)
Bonner	1.26	0.01	(1.08, 1.46)
Boundary	1.51	0.04	(1.15, 1.91)
Custer	1.79	0.10	(1.24, 2.47)
Kootenai	1.10	0.00*	(1.01, 1.21)
Lemhi	1.78	0.06	(1.33, 2.31)
Nez Perce	1.44	0.01	(1.24, 1.67)
Shoshone	1.59	0.04	(1.25, 1.99)

Table 3.2: High Risk Counties

County	Age Adjusted λ_i	Variance	95% CI
Adams	10.08	0.00*	(5.23, 16.69)
Bear Lake	8.50	2.40	(5.72, 11.79)
Boise	5.74	1.10	(3.86, 7.98)
Butte	10.80	15.21	(4.69, 19.74)
Camas	10.90	68.89	(2.98, 34.29)
Clark	28.29	136.19	(10.81, 55.81)
Custer	11.09	4.32	(7.40, 15.55)
Lincoln	8.43	3.92	(5.02, 12.76)

The lowest unadjusted suicide relative risk value estimated over the period was in Madison County. The variance for Madison county is observed to be very low and its effects is seen via its narrow credible interval. Other counties with low variances and narrow credible intervals are listed in table 3.3 below. These low variances are suggestive of less variability in relative risk estimates for respective counties. After adjusting for age, Ada county had the lowest relative risk which was closely followed by Canyon county with risks of 0.89 and 0.94 respectively. Age-adjusted relative risks for counties in table 3.3 are shown in table 3.4 for comparison.

Some of the counties that had above one relative risks and wide credible intervals were; Custer, Lemhi, Shoshone and Clark counties. Their variances were higher in comparison with the rest which resulted in wider credible intervals suggesting more variability in the relative risk estimates. The estimates for each county together with credible intervals are shown in table 3.3.

Table 3.3: Age-unadjusted relative risks

County	Age-unadjusted λ_i	Variance	95% CI
Ada	0.89	0.00*	(0.84,0.94)
Bonneville	1.00	0.00*	(0.89, 1.11)
Canyon	0.87	0.00*	(0.79, 0.95)
Cassia	0.87	0.01	(0.68, 1.09)
Elmore	1.01	0.01	(0.81, 1.22)
Franklin	0.90	0.02	(0.66, 1.18)
Fremont	0.88	0.02	(0.64, 1.15)
Gem	0.99	0.02	(0.76, 1.25)
Gooding	0.90	0.02	(0.67, 1.16)
Jefferson	0.92	0.01	(0.73, 1.12)
Latah	0.83	0.01	(0.68, 1.00)
Madison	0.39	0.00*	(0.28, 0.51)
Twin Falls	1.10	0.00*	(0.97, 1.23)
Washington	0.88	0.02	(0.60, 1.15)

Table 3.4: Age-adjusted relative risks

County	Age-adjusted λ_i	Variance	95% CI
Ada	0.87	0.00*	(0.82,0.92)
Bonneville	1.19	0.00*	(1.06, 1.32)
Canyon	0.94	0.00*	(0.86, 1.02)
Cassia	2.67	0.12	(2.04, 3.04)
Elmore	1.92	0.04	(1.53, 2.36)
Franklin	3.55	0.37	(2.46, 4.85)
Fremont	3.79	0.42	(2.62, 5.18)
Gem	3.29	0.22	(2.44, 4.27)
Gooding	3.44	0.29	(2.46, 4.58)
Jefferson	2.08	0.06	(1.63, 2.59)
Latah	1.48	0.02	(1.19, 1.79)
Madison	1.19	0.04	(0.83, 1.63)
Twin Falls	1.39	0.01	(1.22, 1.57)
Power	0.86	0.02	(0.58, 1.19)
Washington	4.24	0.02	(2.73 6.09)

Map of County Level Age-unadjusted Relative Risks

Figure 3.1 shows the risk levels for all Idaho counties obtained. These risks were grouped into five quantiles. The deeper the color, the higher the relative suicide risk for the county.

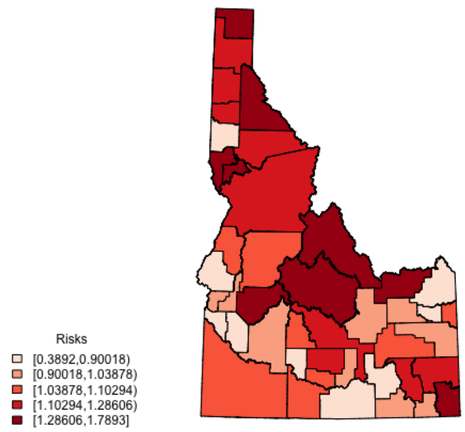

Figure 3.1: Age-unadjusted suicide relative risks

Map of Age-unadjusted Relative Risks Versus Age-unadjusted Relative Risks

A map of age-unadjusted relative risks versus age-unadjusted relative risks is shown is below.

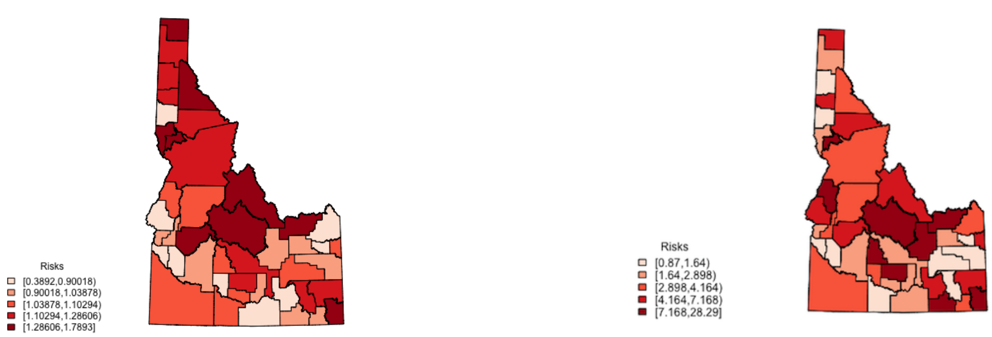

Figure 3.2: Age-unadjusted Relative Risks (left), Age Adjusted Relative Risks

3.2 Age-unadjusted and Adjusted Relative Risks and Average Unemployment Rates Comparison

The average unemployment rates for the period under study was compared with the risks for each county. This was done to check if areas of high risks also experienced high average unemployment rates during the same time frame.

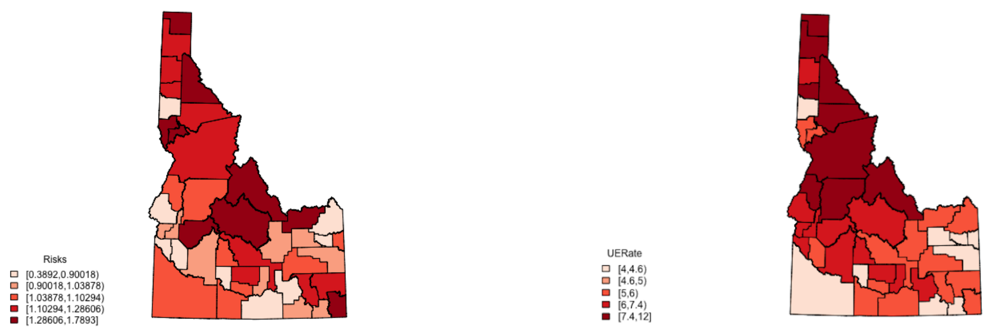

Figure 3.3: Age-unadjusted Relative Risks (left), Average Unemployment Rates (right)

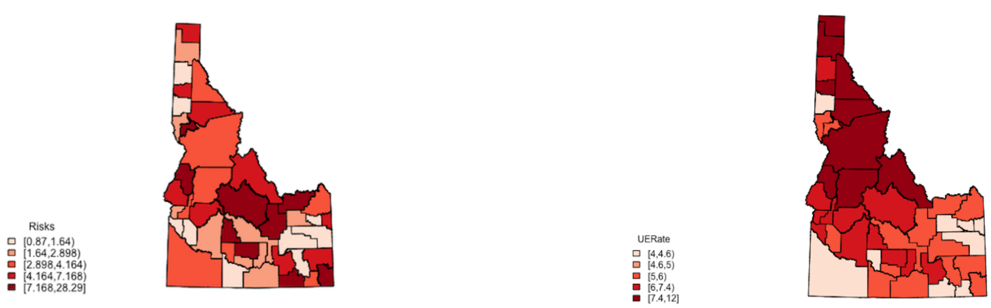

Figure 3.4: Age Adjusted Relative Risks (left), Average Unemployment Rates (right)

From the figure above, some counties that have high risks also appear to have high average unemployment rates. For example, Lemhi, Custer, Clark and Boise counties had very high risk levels and high or moderate unemployment rates. The same can be said for

Boundary, Bonner, and Kootenai counties up north. On the other-hand, Owyhee county can be seen to have a low average unemployment but its risk levels were high. From the two graphs, unemployment seems to have an effect on suicide risk levels but cannot be said to be the sole determinant of relative risk and hence it is suggested that this factor and other possible factors be considered in future works. The scatter plots show that the relationship between relative risks and average unemployment rate is weak. These plots are shown in figures 3.5 and 3.6.

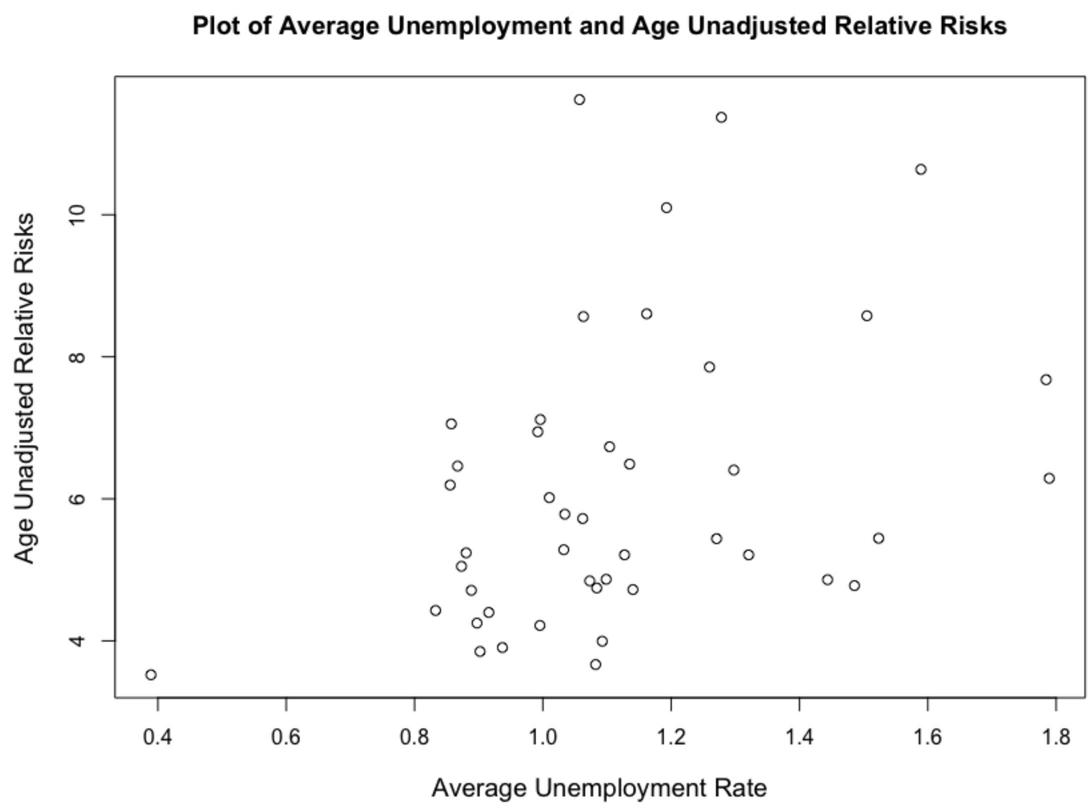

Figure 3.5: Unemployment vs Age-unadjusted Relative Risks

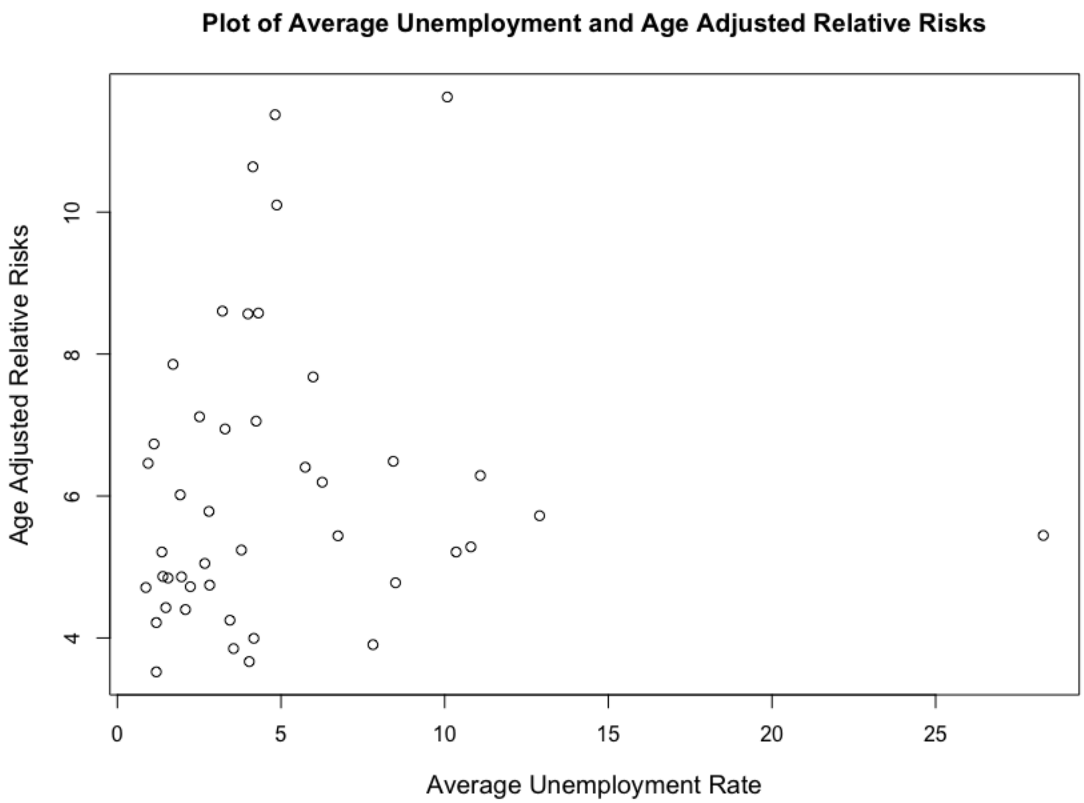

Figure 3.6: Unemployment vs Age-adjusted Relative Risks

3.3 Possible Factors Contributing to High/Low Relative Risks

Relative risk estimates obtained with unadjusted expected deaths appear to have more narrower credible intervals than those intervals obtained after adjustments. Low population counties had particularly high uncertainty in the age adjusted model. The age adjusted model highlighted the differences in relative risks, with Ada county being urban, having the lowest relative risk.

When looking at the age-unadjusted relative risks, one may be lead to conclude that a reason for some counties having narrower credible intervals could be social. For example, Rexburg, the county seat of Madison County, has over 95% of its population Mormon and also home to the Brigham Young University, which is a private university of the Mormon Church. Since this is a close knit community, it may mean that there is more accountability among residents, togetherness and which will lead to less isolation. Residents in this community may have a since of love coming from one another and may cause them know that there is always someone to share their problems with. The reported median household income between 2009 and though the county has been reported to be one of the poorest counties in the country, suicide deaths are at the minimum [**30**].

However, results from the age-adjusted relative risks suggest that these differences are not due social reasons as stated above but rather due to demographics, that is, urban counties appear to have lower relative risks than rural counties.

There could also be the possibility that counties located near other counties with high risks also have high suicide risks. Shoshone and Boundary counties, both in district one have approximately the same risk levels, that is, 1.59 and 1.51 respectively. The same goes for Lemhi, Clark and Custer, which are border counties in district 7. After age-adjustments, relative risks for Shoshone and Boundary counties were 4.14 and 4.31 respectively. Lemhi, Clark and Custer in district seven all had significantly different relative risks; 5.98, 28.29 and 11.09 respectively.

Other factors that have been determined as risk factors of suicide are stress levels, chronic diseases, the local suicide epidemics, access to lethal methods, barriers to assessing to mental health treatment etc. It is therefore imperative these factors be looked into for counties identified as high and low risk counties such as Custer, Lemhi, Shoshone, Bear Lake, Ada, Canyon, Latah and Madison. Comparisons can therefore to made to ascertain the determining factors associated with these relative risks

These risks are indicators that more work need to be put into addressing suicide not only at the state level but also at the county level. The factors that lead to high or low risks such as whether the size of a county in terms of population, should also be studied.

3.4 Recommendation

For further works, it is suggested that other models such as the Zero-inflated Poisson be used and compared with the model used in this analysis. Also, the posterior estimates obtained could be used as prior information in the future. This work should be expanded to include yearly estimates to assess temporal changes and incorporate information from national studies to improve estimability.

References

[1] American Foundation for Suicide Prevention

[2] G. Robert D., Shanghai Archives of Psychiatry, **25** (2013)

[3] World Health Organization (2014) 11

[4] Idaho Department of Health and Welfare: Division of Public Health, Bureau of Vital Records and Health Statistics, (2016)

[5] Idaho Bureau of Vital Records and Health Statistics, Idaho Department of Health and Welfare, Center for Disease Control and Prevention, State Department of Education (2018) 1

[6] Idaho Department of Health and Welfare Bureau of Vital Records and Health Statistics, (2014)

[7] Idaho Department of Health and Welfare: Division of Public Health, Bureau of Vital Records and Health Statistics, (2016)

[8] World Health Organization, Suicide Fact Sheet, (2018)

[9] Centers for Disease Control and Prevention: Preventing Suicide, (2018)

[10] The Suicide Prevention Resource Center: Risk and Protective Factors. Retrieved February 2 2019 from https://www.sprc.org/about-suicide/risk-protective-factors

[11] B. A. Paul, North Caroline Department of Health and Human Services, (2010)

[12] Idaho Department of Health and Welfare: Division of Public Health, Bureau of Vital Records and Health Statistics, Suicide In Idaho: Fact Sheet, September (2018)

[13] Bayesian Analysis pdf

[14] Suicide Rates in the United States Continue to Increase. Retrieved March 7 2019 from https://www.cdc.gov/nchs/products/databriefs/db309.htm

[15] Idaho Department of Health and Welfare Bureau of Vital Records and Health Statistics, (2014)

[16] R. Aaron, S. David ,M. Martin, G. David, C. Shu-Sen , B. Sanjay, The Lancet, **Vol. 380**, (2012), 1813-1814

[17] C. Kelly C. , J. Danielle R. , G. Ryan D. , P. Erin K. , W. Ryan B. , Journal of Abnormal Psychology, **Vol. 122**, (2013), 1021-1030

[18] P. Timo, H. Jari, N. Heikki, L. Jouko, Journal of Affective Disorders **Vol. 81**, (2004), 133 -139

[19] R. Lauren M., H. Holly, K. Diba, W. Margaret, American Journal of Preventive Medicine, **Vol. 55**, (2018), 72-79

[20] The Suicide Prevention Action Network

[21] M. Ghosh and J. N. K. Rao, Statistical Science, **Vol. 9 (1)**, (1994), 55-76

[22] J. N. K. Rao, JIRSS, **Vol. 2 (No. 2)**, (2003), 145-169

[23] J.N.K. Rao, Rivista Internazionale di Scienze Sociali, **Vol. 116 (4)**, (2008), 387-406

[24] J. Jiming, P. Lahiri, Sociedad de Estadı’stica e Investigaciø’n Operativa Test, **Vol. 15 (l)**, (2006), 1-96

[25] T. Matilde, T. Nicola, Hierarchical Bayesian models for small area estimation with count data (2007)

[26] K. Diba, R. Lauren M., H. Holly, W. Margaret, J Data Sci., **Vol. 16(1)**, (2018),147-182.

[27] MD Ugarte, B Ibáñez and AF Militino, Statistical Methods in Medical Research, **Vol. 15**, (2006), 21?35

[28] C. David and K. John, Biometrics, **Vol. 43 (3)**, (1987), 671-681

[29] Sarah Depaoli, James P. Clifton, Patrice R. Cobb, Journal of Educational and Behavioral Statistics, **Vol. 41**, (2016)

[30] The Struggles of Madison County: Idaho's Poorest. Retrieved March 7 2019 from https://www.boiseweekly.com/CityDesk/archives/2015/01/12/the-struggles-of-madison-county-idahos-poorest

Appendix A: County Suicide Risk Table

All relative risks and credible intervals in this table are rounded to 2 decimal places.

Appendix B: Trace Plots and Density Graphs 1

Diagnostic plots showing good mixing for the trace plots and fairly normal posterior distributions for the first 4 lambdas (that is relative risks)

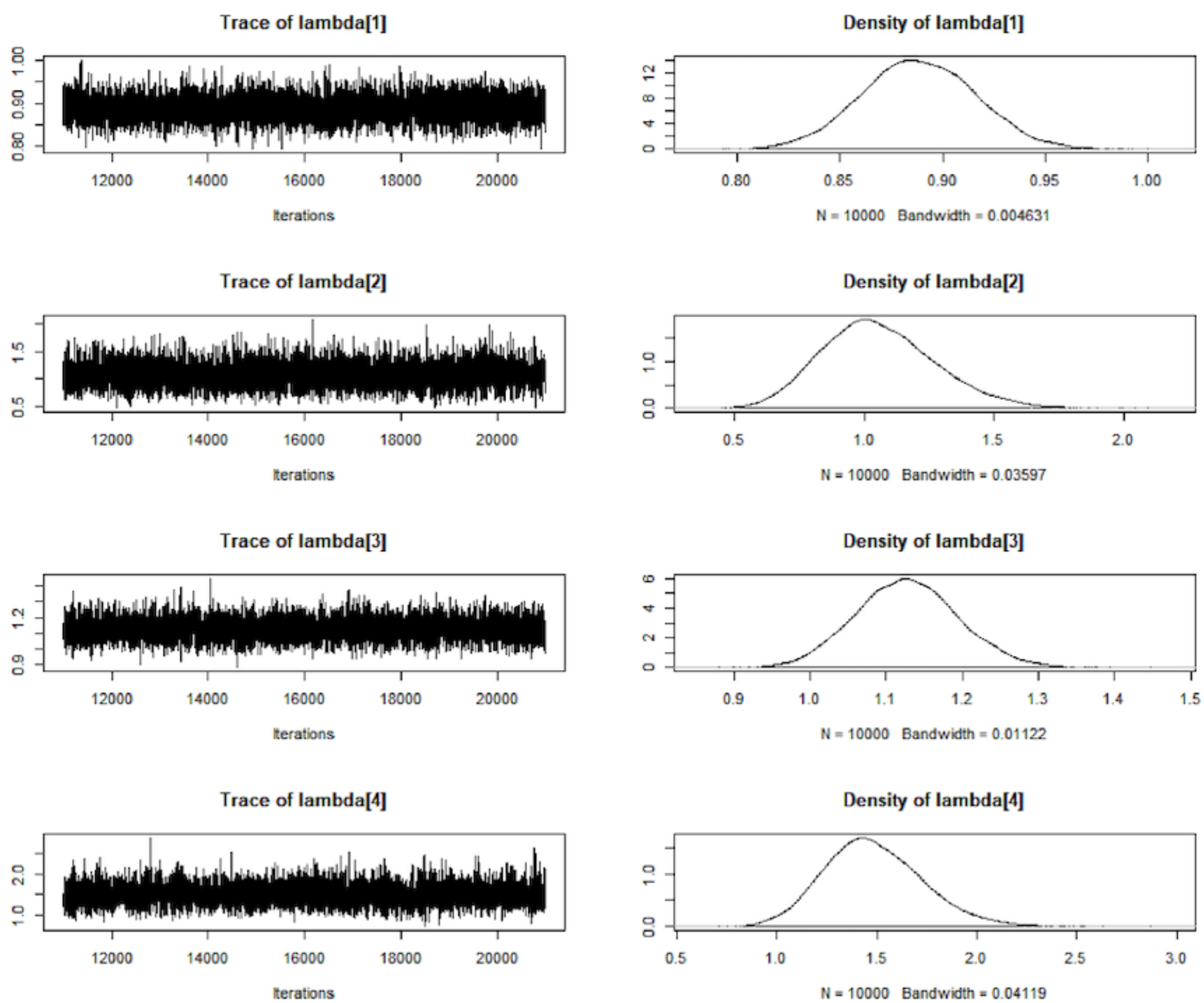

Figure 3.7: Trace and Density Plots1

Appendix B: Trace Plots and Density Graphs 1

Diagnostic plots showing good mixing for the trace plots and fairly normal posterior distributions for $\lambda 5 - \lambda_8$.

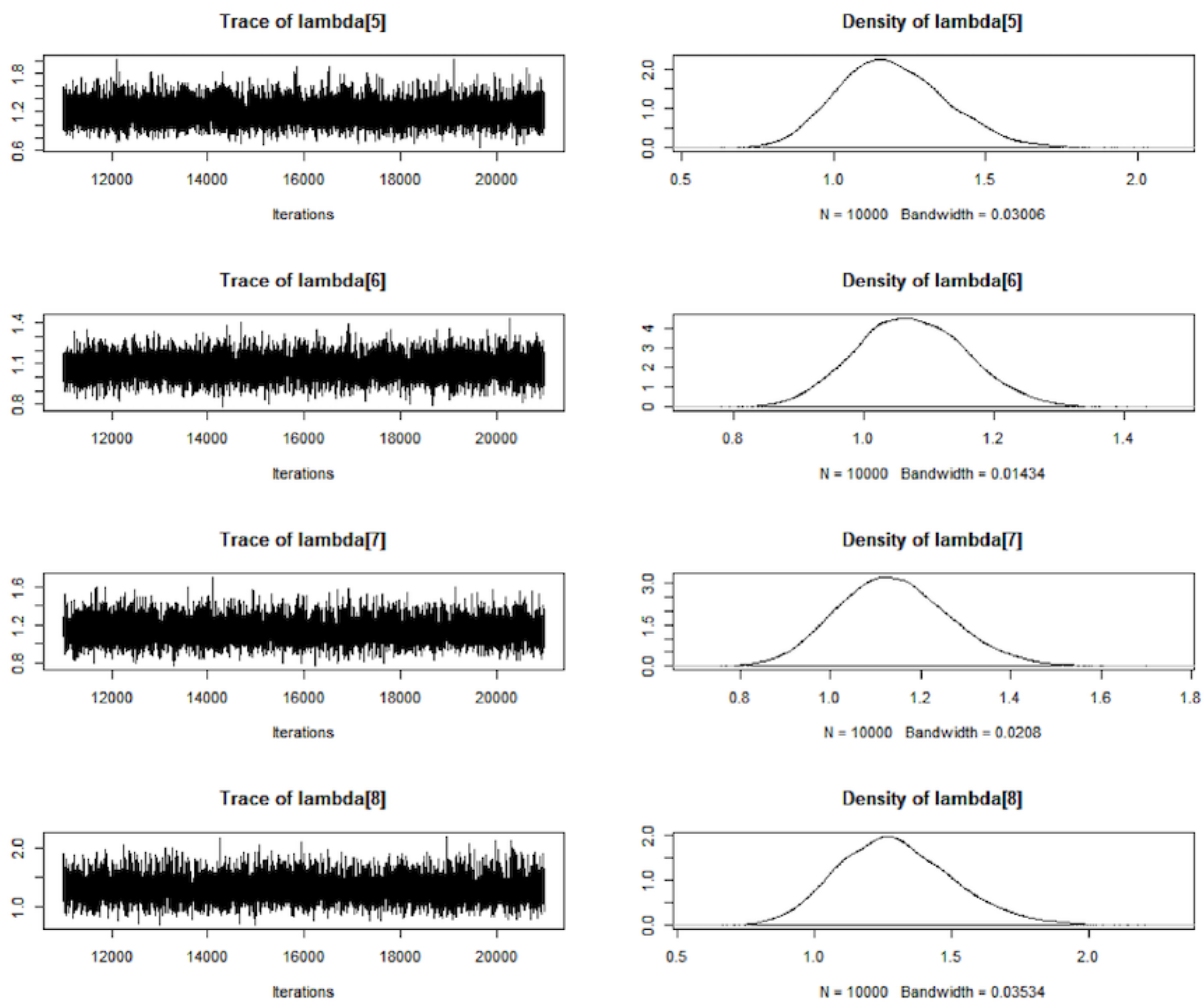

Figure 3.8: Trace and Density Plots2

Appendix B: Trace Plots and Density Graphs 2

Diagnostic plots showing good mixing for the trace plots and fairly normal posterior distributions for $\lambda 9 - \lambda_{12}$.

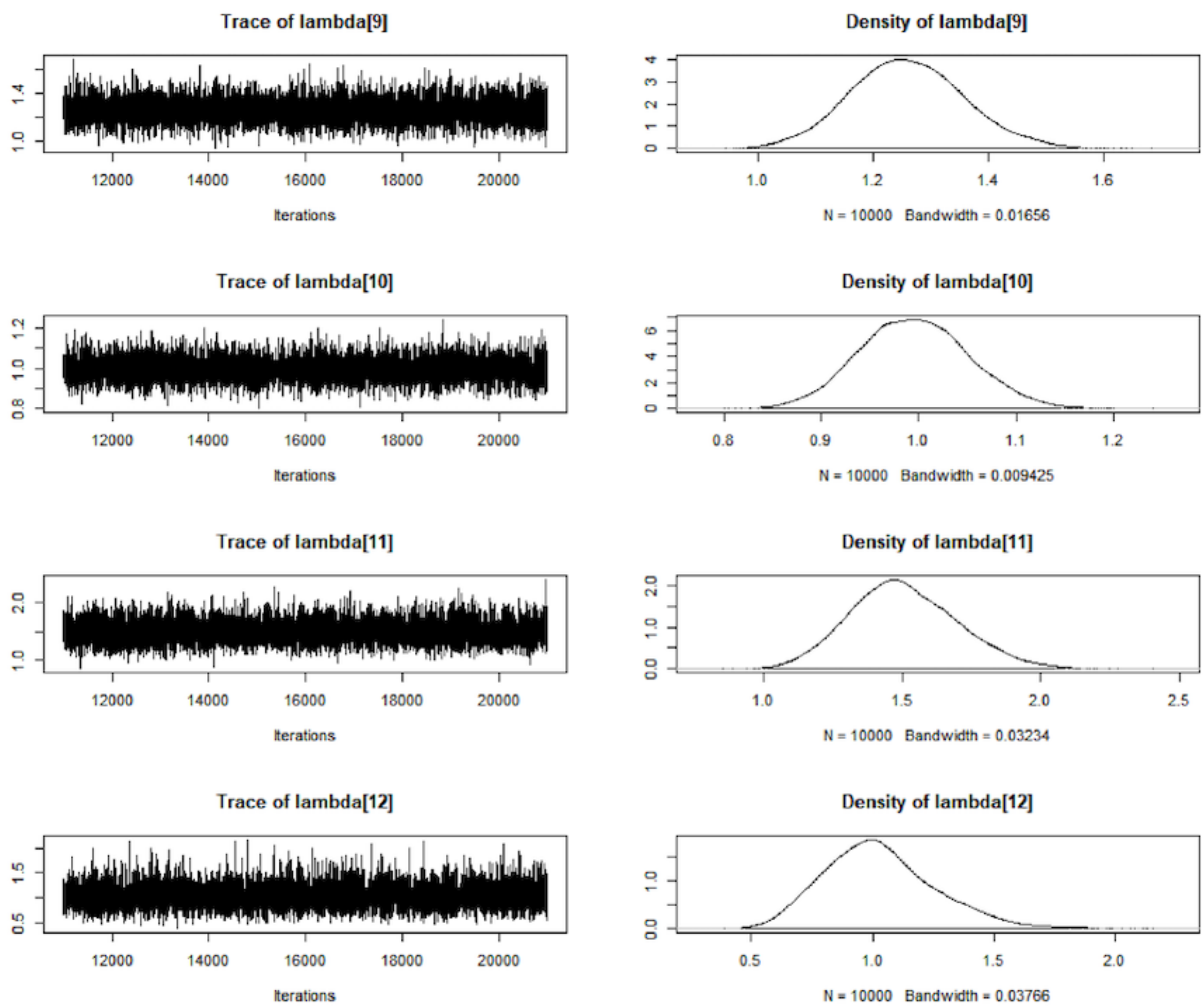

Figure 3.9: Trace and Density Plots3

Appendix B: Trace Plots and Density Graphs 2

Diagnostic plots showing good mixing for the trace plots and fairly normal posterior distributions for $\lambda 13 - \lambda_{16}$.

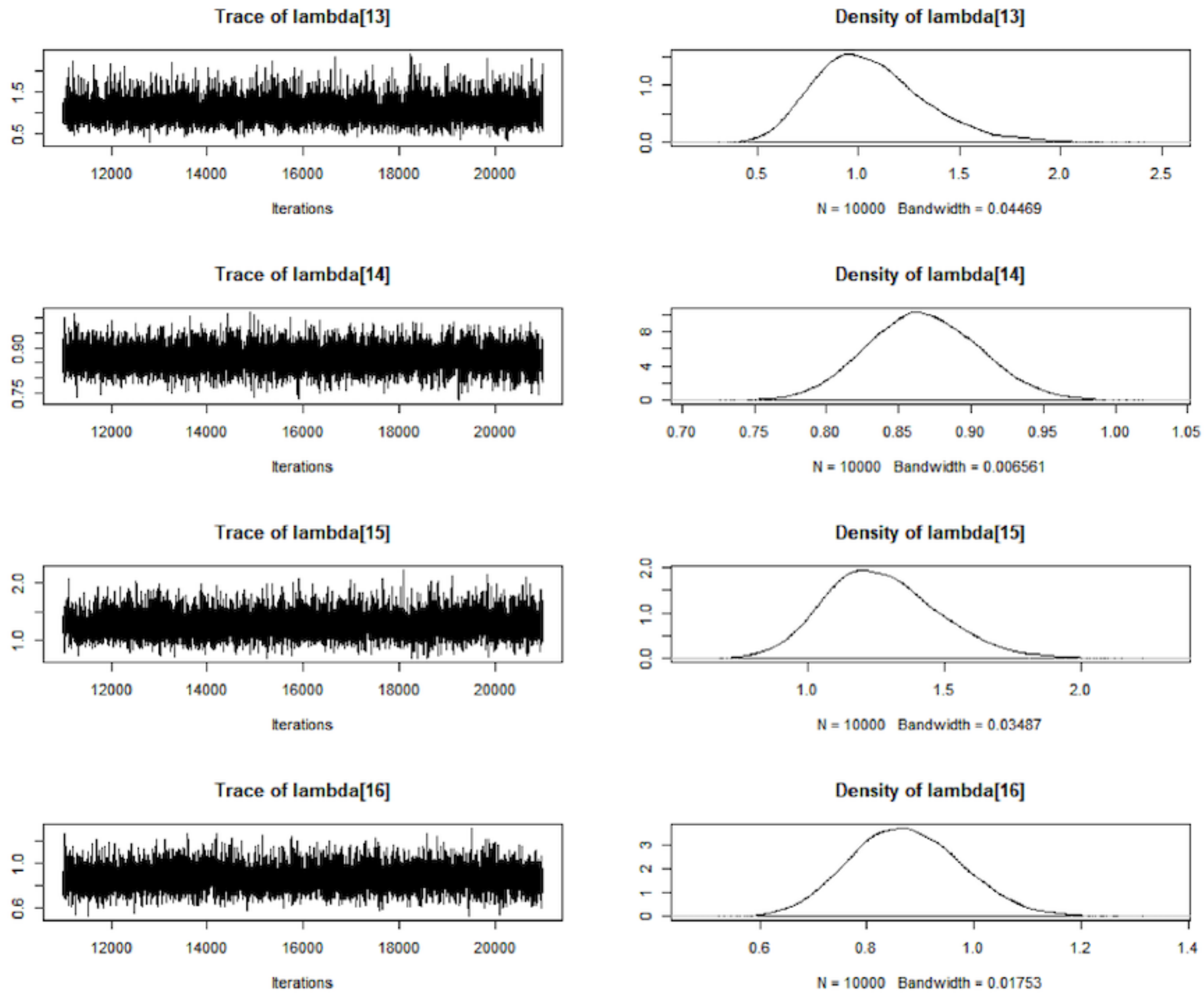

Figure 3.10: Trace and Density Plots4

Appendix B: Gelman-Rubin Test

Diagnostic plots showing that the scale reducation factor is approaching 1 for $\lambda_1 - \lambda_9$ with the exception of λ_2 signifying that there is something happening in the chain which should be investigated.

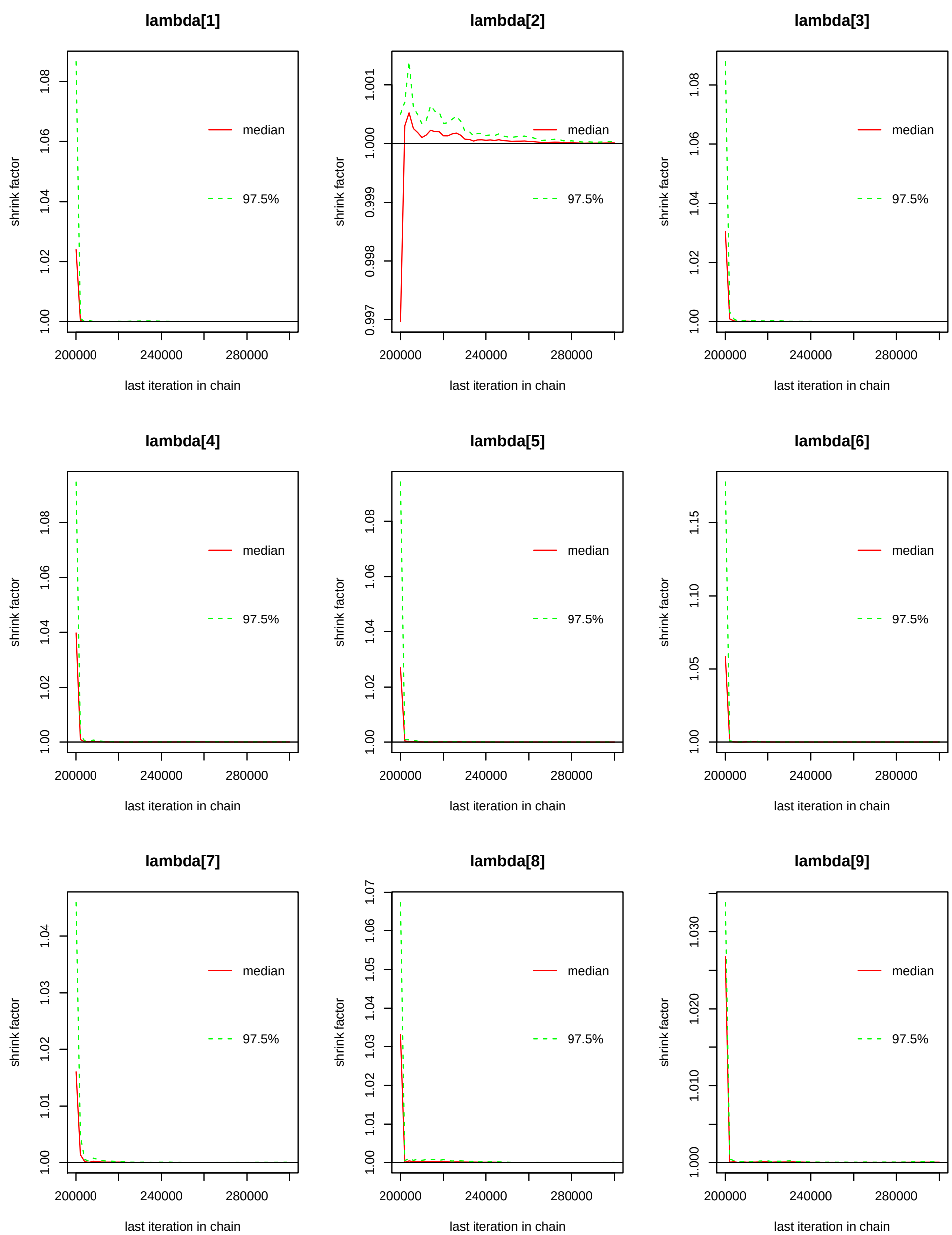

Figure 3.11: Gelman-Rubin Plots on first 9 λ_i's

Appendix B: Gelman-Rubin Test

Diagnostic plots showing that the scale reducation factor is approaching 1 for $\lambda_{10} - \lambda_{18}$ as expected.

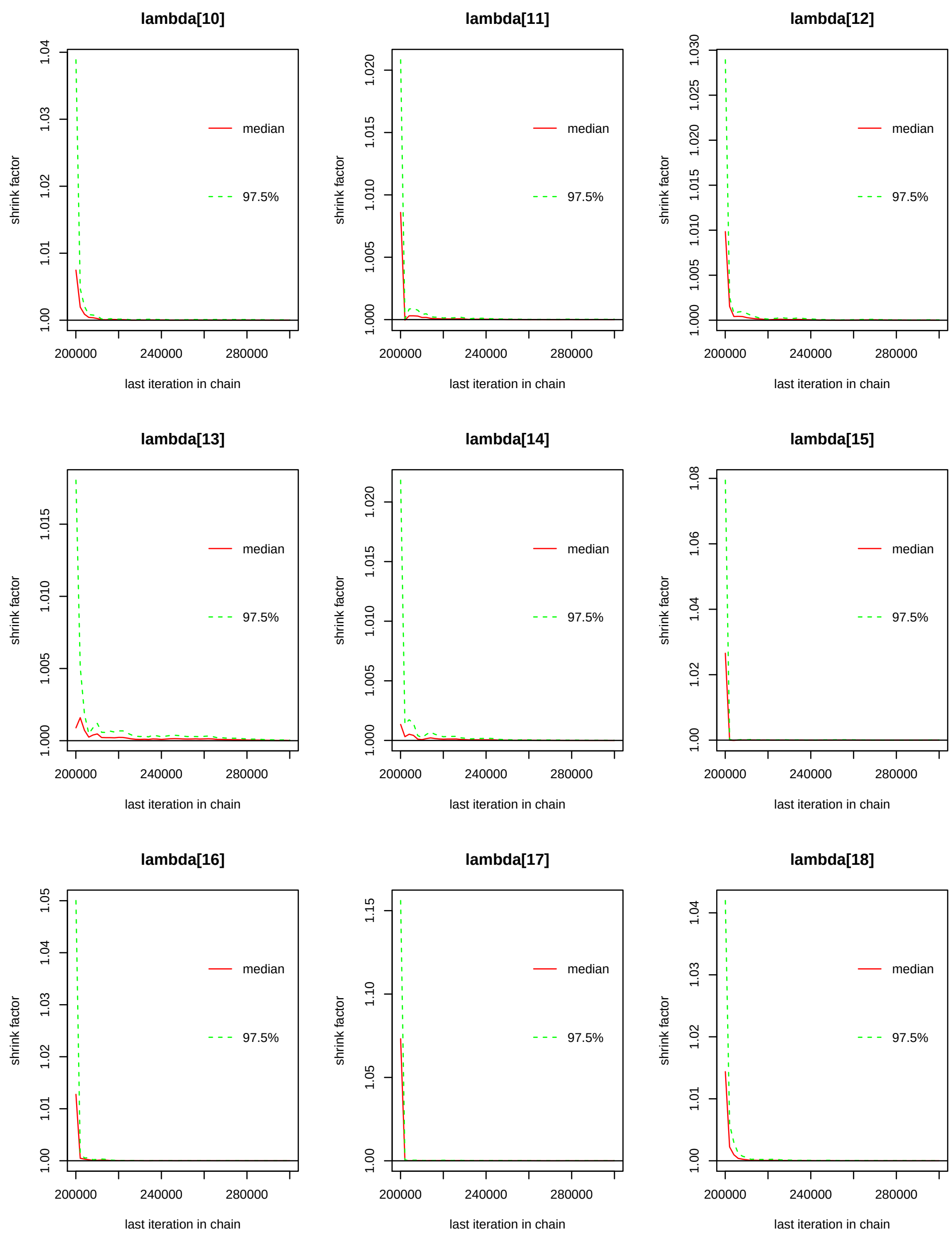

Figure 3.12: Gelman-Rubin Plots for $\lambda_{10} - \lambda_{18}$

Appendix B: Gelman-Rubin Test

Diagnostic plots showing that the scale reducation factor is approaching 1 for $\lambda_{28} - \lambda_{36}$.

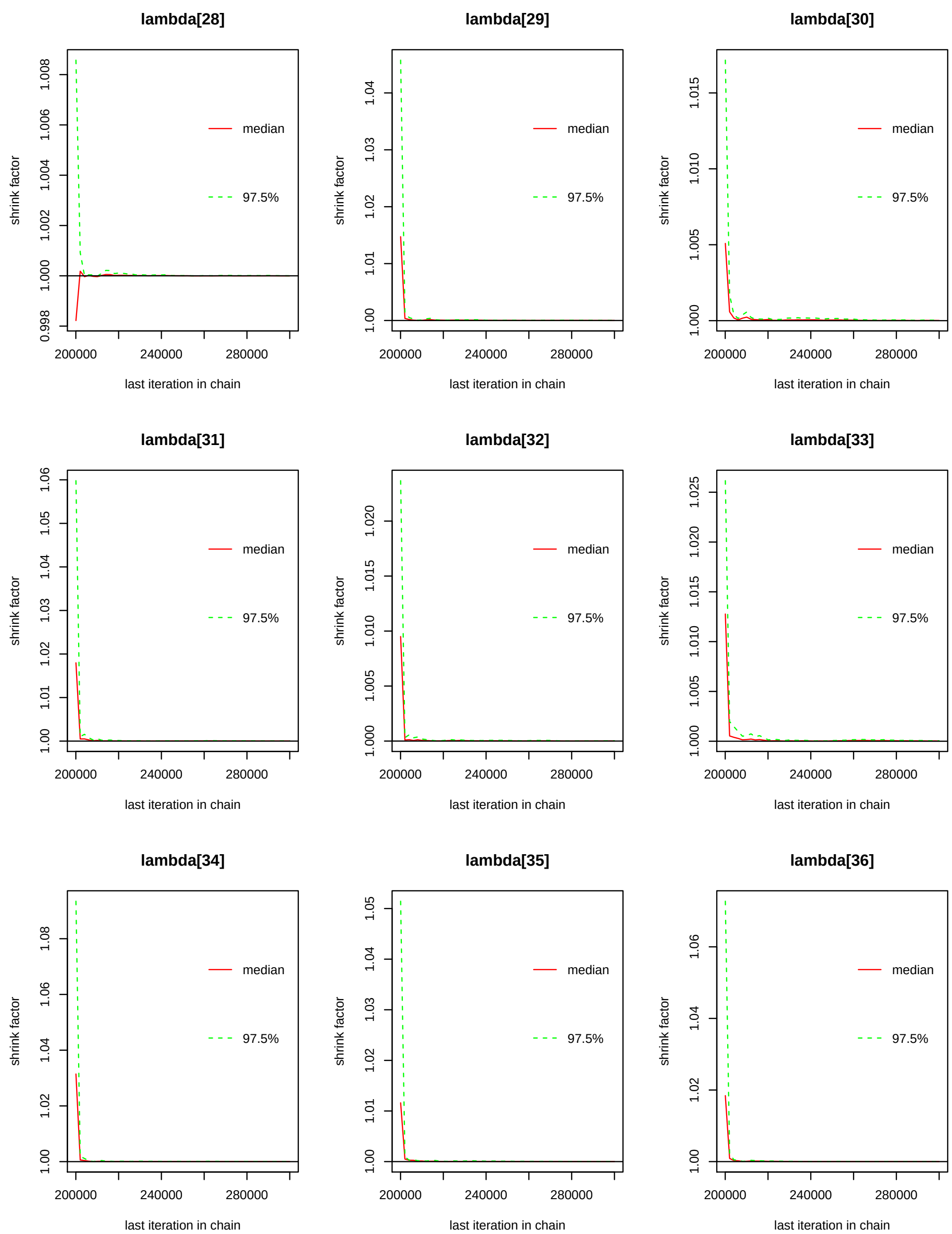

Figure 3.13: Gelman-Rubin Plots for $\lambda_{28} - \lambda_{36}$

Appendix B: Gelman-Rubin Test

Diagnostic plots showing that the scale reducation factor is approaching 1 for the last 8 λ_i's.

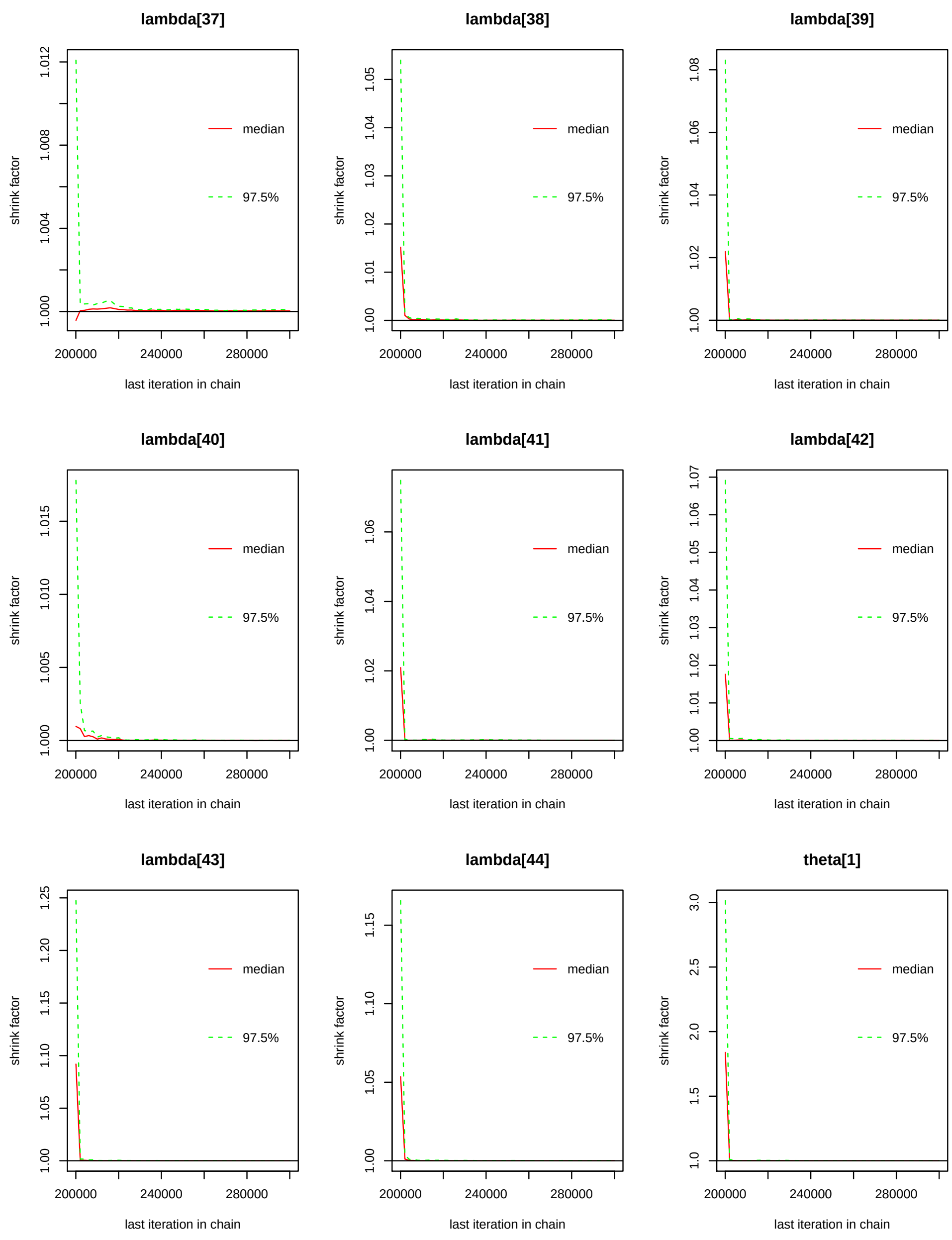

Figure 3.14: Gelman-Rubin Plots of last 8 λ_i's

Appendix C: County Specific Average Unemployment Rate

Average unemployment rate for each county is given in this appendix.

Appendix D: Unemployment Rates

This appendix contains has the plot of unemployment rates

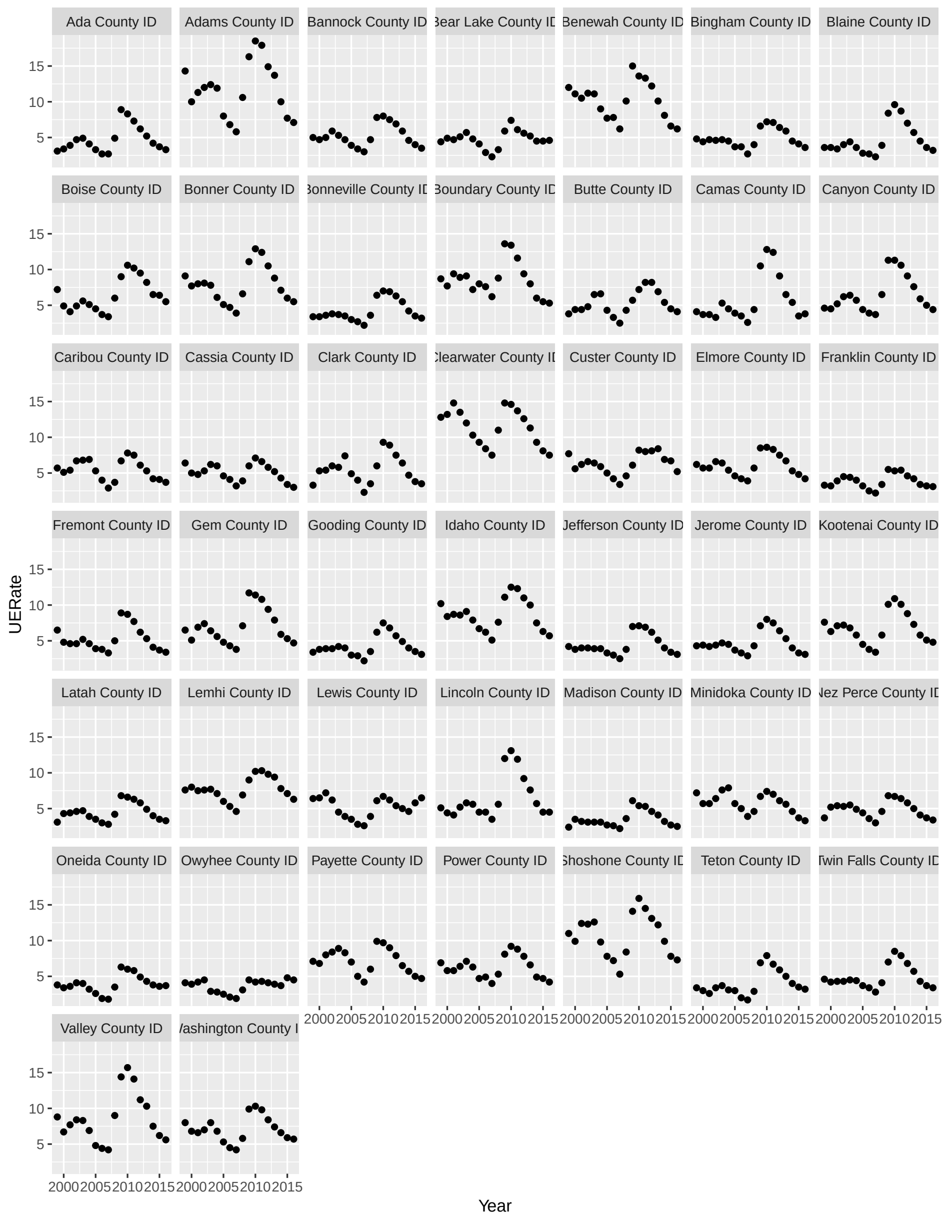

Figure 3.15: Unemployment Rates 1999 − 2016

Appendix E: R Code

This appendix contains the r code used for the analysis.

I = number of counties in Idhao

Code:

```
library(ggplot2)
library(lme4)
library(geepack)
library(reshape2)
library(dplyr)
install.packages("rjags")
library(rjags)
library(plyr)
library(tidyr)
library(tscount)
library(readxl)
my_data = read_excel("WiestData.xlsx")
my_data = read_excel(file.choose())
my_data = read_excel("WiestData.xlsx", sheet = "WiestData")
filter = filter(my_data,my_data$`Manner of death` =="Suicide")
select = dplyr::select(filter,Year, Sex, County, Age,`Manner of death`)
names(select) = c("Year","Sex","County","Age","Manner")
ages.pop = read_excel("Ages.xlsx",sheet = "Sheet1")
ages.pop = read_excel(file.choose())
age.select = select_(ages.pop,"Geographic.area","5","8","11","14","17")
names(age.select) = c("County", "< 18","18-24","25-44","45-64","65+")
merge.pop = merge(select,age.select)
```

```
tab = ftable(select$Year, select$Sex,select$County)
use.dat = melt(tab)
names(use.dat) = c("Year","Sex","County","Freq")
merge = merge(use.dat,age.select)
new.male = male %>%
  dplyr::select(County,Year,Freq)
deaths.male = sum(new.male$Freq)
new.fem = female %>%
 dplyr:: select(County,Year,Freq)
deaths.fem = sum(new.fem$Freq)
deaths.ID = deaths.male + deaths.fem
county.pop = read.csv("AgeGroups.csv")
county.pop = select_(county.pop,"Geographic.area","Total.population")
county.pop = county.pop[-c(45),]
pop.18 = c(county.pop$Total.population)*18
county.pop = cbind(county.pop,pop.18)
pop.ID.18 = sum(county.pop$pop.18)
rate.ID = (deaths.ID/pop.ID.18)*100000
counties = merge%>%
  dplyr::select(County,Year,Freq)
death.county = aggregate(counties$Freq, by=list(County=counties$County),
                         FUN = sum)
names(death.county) = c("County","Deaths")
county.deaths = cbind(death.county,pop.18)
deaths.county.rate = (county.deaths$Deaths/county.deaths$pop.18)*100000
E = (rate.ID/100000)*county.deaths$pop.18
county.exp.deaths = cbind(county.deaths,E, deaths.county.rate)
```

```
# JAGS Model Specification
model {
  for (i in 1 : I) {
    y[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + theta[i]
    lambda[i] <- exp(alpha + theta[i])
    theta[i] ~ dnorm(0,tau)
  }
  #Priors
  alpha ~ dnorm(0, 0.001)
  tau ~ dgamma(0.001,0.001)
  sigma <- 1/sqrt(tau)
}

set.seed(4444)
I = 44
y = c(county.exp.deaths$Deaths)
d = list(y = y, E = E, I = length(y))
m = jags.model(file = "code.txt", d)
update(m, 100000)
samp = coda.samples(m, c( "lambda","theta"), 100000)
samp
summary(samp)

# Unemployment data
e99 <- read.table("laucnty99.csv", skip=554, nrow=44, header=FALSE,
```

```
sep=",")
e00 <- read.table("laucnty00.csv", skip=554, nrow=44, header=FALSE,
 sep=",")
e01 <- read.table("laucnty01.csv", skip=554, nrow=44, header=FALSE,
sep=",")
e02 <- read.table("laucnty02.csv", skip=554, nrow=44, header=FALSE,
 sep=",")
e03 <- read.table("laucnty03.csv", skip=554, nrow=44, header=FALSE,
sep=",")
e04 <- read.table("laucnty04.csv", skip=554, nrow=44, header=FALSE,
sep=",")
e05 <- read.table("laucnty05.csv", skip=554, nrow=44, header=FALSE,
sep=",")
e06 <- read.table("laucnty06.csv", skip=554, nrow=44, header=FALSE,
sep=",")
e07 <- read.table("laucnty07.csv", skip=554, nrow=44, header=FALSE,
 sep=",")
e08 <- read.table("laucnty08.csv", skip=554, nrow=44, header=FALSE,
 sep=",")
e09 <- read.table("laucnty09.csv", skip=554, nrow=44, header=FALSE,
 sep=",")
e10 <- read.table("laucnty10.csv", skip=556, nrow=44, header=FALSE,
sep=",")
e11 <- read.table("laucnty11.csv", skip=556, nrow=44, header=FALSE,
sep=",")
e12 <- read.table("laucnty12.csv", skip=556, nrow=44, header=FALSE,
sep=",")
```

```
e13 <- read.table("laucnty13.csv", skip=556, nrow=44, header=FALSE,
sep=",")
e14 <- read.table("laucnty14.csv", skip=556, nrow=44, header=FALSE,
sep=",")
e15 <- read.table("laucnty15.csv", skip=556, nrow=44, header=FALSE,
sep=",")
e16 <- read.table("laucnty16.csv", skip=556, nrow=44, header=FALSE,
sep=",")

unemply=rbind(e99,e00,e01,e02,e03,e04,e05,e06,e07,
e08,e09,e10,e11,e12,e13,e14,e15,e16)
unemply=unemply[,-c(1,2,3,6)]

names(unemply)=c("County","Year","LaborForce","Employed",
"Unemployed","UERate")

dev.off()
p=ggplot(unemply, aes(Year, UERate)) +
geom_point() + facet_wrap(~County)

x=as.vector(unemply$County)
countylabels=as.vector(sapply(x, FUN=gsub,
pattern = " County ID", replacement= ""))
unemply$County=countylabels
unemply$LaborForce=as.numeric(gsub(",", "", unemply$LaborForce))
unemply$Log.LF=log(unemply$LaborForce)
```

```
U.county=aggregate(. ~County, data=unemply, sum, na.rm=TRUE)
U.county[,-2]      #county and UERate for all counties
county.emp=data.frame(U.county$County,U.county$UERate)
ave.U.county =data.frame(county.emp$U.county.County,
            (county.emp$U.county.UERate/18))
names(ave.U.county)=c("County","Ave.UERate")

data=data.frame(U.county$County, risks)
plot(data$risks,ave.U.county$Ave.UERate)
```

Table 3.5: Age Unadjusted Relative Risks

County	No. of deaths	E_i	Relative Risks	95% CI
Ada	1033	1165	0.89	(0.84 , 0.94)
Adams	12	12	1.06	(0.68, 1.52)
Bannock	278	246	1.13	(1.00, 1.26)
Bear Lake	30	18	1.49	(1.05, 2.01)
Benewah	34	28	1.19	(0.87, 1.58)
Bingham	145	135	1.07	(0.91, 1.24)
Blaine	73	63	1.14	(0.91, 1.40)
Boise	29	21	1.30	(0.92, 1.75)
Bonner	155	121	1.26	(1.08, 1.46)
Bonneville	307	309	0.10	(0.89, 1.11)
Boundary	53	33	1.51	(1.15, 1.91)
Butte	8	9	1.03	(0.64, 1.53)
Camas	3	3	1.06	(0.60, 1.70)
Canyon	484	561	0.87	(0.79, 0.95)
Caribou	28	21	1.27	(0.90, 1.71)
Cassia	57	68	0.87	(0.68, 1.09)
Clark	8	3	1.52	(0.88, 2.46)
Clearwater	35	26	1.28	(0.93, 1.68)
Custer	29	13	1.79	(1.24, 2.47)
Elmore	80	80	1.01	(0.81, 1.22)
Franklin	32	38	0.90	(0.66, 1.18)
Fremont	32	39	0.88	(0.64, 1.15)
Gem	48	50	0.99	(0.76, 1.25)
Gooding	39	46	0.90	(0.67, 1.16)
Idaho	57	48	1.16	(0.91, 1.45)
Jefferson	69	78	0.92	(0.73, 1.12)
Jerome	72	66	1.08	(0.86, 1.33)
Kootenai	454	411	1.10	(1.01, 1.21)
Latah	89	111	0.83	(0.68, 1.00)
Lemhi	48	24	1.78	(1.33, 2.31)
Lewis	17	11	1.32	(0.87, 1.89)
Lincoln	18	15	1.14	(0.77, 1.59)
Madison	31	111	0.39	(0.28, 0.51)
Minidoka	61	60	1.03	(0.81, 1.28)
Nez Perce	172	117	1.44	(1.24, 1.67)
Oneida	10	13	0.94	(0.60, 1.36)
Owyhee	37	34	1.08	(0.80, 1.41)
Payette	66	67	1.00	(0.79, 1.23)
Power	17	23	0.86	(0.58, 1.19)
Shoshone	65	38	1.59	(1.25, 1.99)
Teton	33	30	1.09	(0.80, 1.44)
Twin Falls	252	229	1.10	(0.97, 1.23)
Valley	31	29	1.06	(0.77 , 1.41)
Washington	23	30	0.86	(0.60, 1.15)

Table 3.6: Age Adjusted Relative Risks

County	No. of deaths	E_i	Relative Risks	St.Deviation	95% CI
Ada	1033	1191	0.87	0.03	(0.82, 0.92)
Adams	12	1	10.08	2.94	(5.23, 16.69)
Bannock	278	206	1.36	0.08	(1.20, 1.52)
Bear Lake	30	3	8.50	1.55	(5.72, 11.79)
Benewah	34	7	4.87	0.82	(3.38, 6.60)
Bingham	145	94	1.55	0.13	(1.31, 1.81)
Blaine	73	33	2.23	0.26	(1.76, 2.76)
Boise	29	5	5.74	1.05	(3.86, 7.98)
Bonner	155	92	1.70	0.14	(1.44, 1.97)
Bonneville	307	260	1.19	0.07	(1.06, 1.32)
Boundary	53	12	4.31	0.59	(3.24, 5.53)
Butte	8	1	10.80	3.90	(4.69, 19.74)
Camas	3	0	12.90	8.30	(2.98, 34.29)
Canyon	484	519	0.94	0.04	(0.86, 1.10)
Caribou	28	4	6.74	1.26	(4.49, 9.44)
Cassia	57	21	2.67	0.33	(2.04, 3.40)
Clark	8	0	28.29	11.67	(10.81, 55.81)
Clearwater	35	7	4.82	0.80	(3.38, 6.52)
Custer	29	2	11.09	2.08	(7.40, 15.55)
Elmore	80	42	1.92	0.21	(1.53, 2.36)
Franklin	32	9	3.55	0.61	(2.46, 4.85)
Fremont	32	8	3.79	0.65	(2.62, 5.18)
Gem	48	15	3.29	0.47	(2.44, 4.27)
Gooding	39	11	3.44	0.54	(2.46, 4.58)
Idaho	57	18	3.21	0.42	(2.44, 4.09)
Jefferson	69	34	2.08	0.25	(1.63, 2.59)
Jerome	72	26	2.82	0.33	(2.22, 3.50)
Kootenai	454	408	1.12	0.05	(1.02, 1.22)
Latah	89	61	1.48	0.15	(1.19, 1.79)
Lemhi	48	8	5.98	0.85	(4.43, 7.77)
Lewis	17	1	10.35	2.54	(6.03, 15.91)
Lincoln	18	2	8.43	1.98	(5.02, 12.76)
Madison	31	27	1.19	0.21	(0.83, 1.63)
Minidoka	61	22	2.8	0.35	(2.17, 3.56)
Nez Perce	172	88	1.96	0.15	(1.68, 2.27)
Oneida	10	1	7.81	2.43	(3.84, 13.31)
Owyhee	37	9	4.03	0.65	(2.86, 5.40)
Payette	66	27	2.51	0.30	(1.95, 3.14)
Power	17	3	6.26	1.48	(3.70, 9.49)
Shoshone	65	16	4.14	0.51	(3.20, 5.19)
Teton	33	8	4.17	0.71	(2.90, 5.69)
Twin Falls	252	182	1.39	0.09	(1.22, 1.57)
Valley	31	8	3.99	0.70	(2.73, 5.547)
Washington	23	5	4.24	0.86	(2.73, 6.09)

Table 3.7: County Average Unemployment Rates

County	Unemploy.Rate
Ada	4.71
Adams	11.62
Bannock	5.21
Bear Lake	4.78
Benewah	10.10
Bingham	4.84
Blaine	4.72
Boise	6.41
Bonner	7.86
Bonneville	4.22
Boundary	8.58
Butte	5.28
Camas	5.72
Canyon	6.46
Caribou	5.44
Cassia	5.05
Clark	5.44
Clearwater	11.37
Custer	6.29
Elmore	6.02
Franklin	3.85
Fremont	5.24
Gem	6.94
Gooding	4.25
Idaho	8.61
Jefferson	4.40
Jerome	4.74
Kootenai	6.73
Latah	4.43
Lemhi	7.68
Lewis	5.21
Lincoln	6.49
Madison	3.52
Minidoka	5.78
Nez Perce	4.86
Oneida	3.91
Owyhee	3.67
Payette	7.12
Power	6.19
Shoshone	10.64
Teton	3.99
Twin Falls	4.87
Valley	8.57
Washington	7.06