1-1-2016

On Non-Parametric Confidence Intervals for Density and Hazard Rate Functions & Trends in Daily Snow Depths in the United States and Canada

Yang Xu

On non-parametric confidence intervals for density and hazard rate functions & trends in daily snow depths in the United States and Canada

By

Yang Xu

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Statistics
in the Department of Mathematics and Statistics

Mississippi State, Mississippi

December 2016

On non-parametric confidence intervals for density and

hazard rate functions & trends in daily snow

depths in the United States and Canada

By

Yang Xu

Approved:

Prakash N. Patil
(Co-Major Professor)

Jonathan R. Woody
(Co-Major Professor)

Janice DuBien
(Committee Member)

Mohammad Sepehrifar
(Committee Member)

Tung-Lung Wu
(Committee Member)

Corlis P. Johnson
(Graduate Coordinator)

Rick Travis
Interim Dean
College of Arts & Sciences

Name: Yang Xu

Date of Degree: December 9, 2016

Institution: Mississippi State University

Major Field: Statistics

Major Professor: Dr. Prakash Patil

Title of Study: On non-parametric confidence intervals for density and hazard rate functions & trends in daily snow depths in the United States and Canada

Pages of Study: 68

Candidate for Degree of Doctor of Philosophy

The nonparametric confidence interval for an unknown function is quite a useful tool in statistical inferential procedures; and thus, there exists a wide body of literature on the topic. The primary issues are the smoothing parameter selection using an appropriate criterion and then the coverage probability and length of the associated confidence interval. Here our focus is on the interval length in general and, in particular, on the variability in the lengths of nonparametric intervals for probability density and hazard rate functions. We start with the analysis of a nonparametric confidence interval for a probability density function noting that the confidence interval length is directly proportional to the square root of a density function. That is variability of the length of the confidence interval is driven by the variance of the estimator used to estimate the square-root of the density function. Therefore we propose and use a kernel-based constant variance estimator of the square-root of a density function. The performance of confidence intervals so obtained is studied through

simulations. The methodology is then extended to nonparametric confidence intervals for the hazard rate function.

Changing direction somewhat, the second part of this thesis presents a statistical study of daily snow trends in the United States and Canada from 1960-2009. A storage model balance equation with periodic features is used to describe the daily snow depth process. Changepoint (inhomogeneities features) are permitted in the model in the form of mean level shifts. The results show that snow depths are mostly declining in the United States. In contrast, snow depths seem to be increasing in Canada, especially in north-western areas of the country. On the whole, more grids are estimated to have an increasing snow trend than a decreasing trend. The changepoint component in the model serves to lessen the overall magnitude of the trends in most locations.

Key words: Confidence Interval, Kernel Density, Hazard Rate, Variance Transformation ,Nonparametric Estimation, Snow Trend, Changepoint, Climate Change, Genetic Algorithm, Storage Model.

DEDICATION

To my parents, grandparents, aunt and Lu.

ACKNOWLEDGEMENTS

First of all I would like to express my deep gratitude to my advisors Dr. Prakash Patil and Dr. Jonathan Woody for all their guidance, teaching, mentoring, caring and friendship. Their timely advices and encouragement in the preparation of this thesis were invaluable. I am very fortunate to have advisors who paved a strong foundation through their expertise, supervision, excellent ideas and continuous enthusiasm. Their guidances helped me in all the time of research and writing of this thesis. I could not have imagined having better advisors and mentors for my Ph.D study.

Besides my advisors Dr. Prakash Patil and Dr. Jonathan Woody, I would also like to thank my committee members Dr. Janice DuBien, Dr. Mohammad Sepehrifar and Dr. Tung-Lung Wu for all their support and encouragement. I would also thank all my friends at the department for their company and support.

It also gives me great pleasure to thank our department head Dr. Mohsen Razhaghi for all his support. I am very grateful to our graduate coordinator Dr. Corlis Johnson, for all her care, support and encouragement through out the doctoral program. My sincere gratitude to the Department of Mathematic and Statistics for supporting me with teaching assistantship for the last 5 years. I also would like to thank my teaching mentors Dr. Heyona Lim, Mr. Jacob Tshume and Mr. Jonathan Eaton.

I am very thankful to my parents, grandma and aunt for their love, support and encouragement over the years. I would like to thank my girlfriend for being very understanding supportive and encouraging.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND PRELIMINARIES

1.1 Hazard Function

Let T denote the lifetime of a component that is being manufactured and thus it is a non-negative random variable. One of the obvious questions of interest in reliability analysis is to understand the longevity of the component. For this, choosing an appropriate probability model to describe the random variation in T is important. Thus, one observes the lifetime of n components, and let those lifetimes be $T_1, T_2 \ldots, T_n$. It seems reasonable to assume that T_i's are independent and identically distributed.

The most important question that one may ask is this: what is the probability that a given component that is working until time t will stop functioning at the next moment. That is,

$$P(t < T \leq t + h | T > t)$$

where h is very small. For this let $f(t)$, $F(t)$ denote the probability density and cumulative distribution functions of T, respectively. Then the instantaneous probability of failure can be written as

$$\begin{aligned}
&\lim_{h\to 0} \frac{P(t < T \leq t+h \mid T > t)}{h} \\
=&\lim_{h\to 0} \frac{P(t < T \leq t+h)}{P(T > t)h} \\
=&\lim_{h\to 0} \frac{F(t+h) - F(t)}{h} \frac{1}{1-F(t)} \, .
\end{aligned}$$

Since

$$\lim_{h\to 0} \frac{F(t+h) - F(t)}{h} = f(t)$$

then

$$\lim_{h\to 0} \frac{P(t < T \leq t+h \mid T > t)}{h} = \frac{f(t)}{1-F(t)} \quad \text{and} \quad F(t) < 1 \, .$$

Set

$$\lambda(t) = \frac{f(t)}{1-F(t)} \, ,$$

$\lambda(t)$ is called the hazard rate function, and it plays a vital role in survival and reliability analysis. There is another function which also plays an equally important role and it is the cumulative hazard rate function, defined as

$$\begin{aligned}
\Lambda(x) &= \int_0^x \lambda(t) dt \\
&= \int_0^x \frac{f(t)}{1-F(t)} dt \\
&= -\int_0^x \frac{1}{1-F(t)} d(1-F(t)) \\
&= -\log(1-F(x)) \, .
\end{aligned}$$

The above concepts and their formulation can be given in the multivariate setup. However, in the present thesis we confine ourselves to the univariate failure rate. As mentioned above, the hazard rate is used to model the failure times of manufactured items as well as failure times of repaired items. In medical statistics, the hazard rate is used to model the survival times of individuals following certain treatments. Thus, estimating the hazard rate function is very important in reliability and survival analysis. In broad terms there are two methods to estimate the hazard rate function and those are parametric and nonparametric. We will first introduce the parametric estimation of the hazard rate in the next section. Since it is easier to understand nonparametric kernel-based density estimator, we first discuss kernel density estimator in section 1.3 and then explain the nonparametric hazard rate estimator as an extension to kernel-based density estimator. There is another kernel-based approach to estimate density and hazard rate via nonparametric regression. To define this estimator, we first describe nonparametric regression function estimator in the first two subsections of section 4. Then explain how this approach leads to another nonparametric kernel-based estimation of density and hazard rate in the last two subsections of section 4. In the last section, we briefly describe the problem that we address in this dissertation.

1.2 Parametric Estimation of Hazard Rate

Let $f(t, \underline{\theta})$ be the model which describes the random variation in T and is assumed to be known up to parameter vector $\underline{\theta}$. Then, obviously the hazard rate function $\lambda(t)$ will also depend on the same unknown parameter vector $\underline{\theta}$, and hence we will write it as $\lambda(t, \underline{\theta})$. Now the unknown parameter vector $\underline{\theta}$ can be estimated by the standard estimation methods

like maximum likelihood or method of moments, etc. Let $\hat{\underline{\theta}}$ be such an estimator, then the parametric estimator of $\lambda(t, \underline{\theta})$ is simply $\lambda(t, \hat{\underline{\theta}})$.

In reliability as well as in survival analysis, certain parametric models have been found more useful than others like exponential and Weibull distributions. These distributions admit closed-form expressions for survival probabilities and therefore for the hazard rate functions. For example in the one-parameter exponential distribution, then

$$F(t) = 1 - e^{-\alpha t}\ ,\ f(t) = \alpha e^{-\alpha t},\ \ t \geq 0 \ \text{ and } \ \lambda(t) = \alpha, \ \ t > 0.$$

It means the instantaneous probability of failure, $\lambda(t)$, is independent of t. That is the conditional chance of failure in a time interval of specified length is the same regardless of how long the individual has been under study; this is referred to as the memoryless property of the exponential distribution. If we consider the two-parameter Weibull distribution, then

$$F(t) = 1 - e^{-(\alpha t)^k}\ ,\ f(t) = \alpha k(\alpha t)^{k-1} e^{-(\alpha t)^k},\ \ t \geq 0 \ \text{ and } \ \lambda(t) = \alpha k(\alpha t)^{k-1},\ t > 0.$$

Here, hazard rate function is a polynomial.

In Figure 1.1 we illustrate the shapes of hazard rate for exponential, Weibull and log-normal models. For Weibull hazard rate, the shape parameter k is set to be 1.5 and 0.5, respectively. For the lognormal hazard rate, $\mu = -\log \alpha$ and $\sigma = 1$.

The parametric methods for hazard rate estimation are very powerful when the model assumption correctly describes the variation in T. However, the drawn inference is meaningless if the assumed model is far away from the reality. Therefore it is necessary to search for alternative ways to estimate hazard rate when one is not certain about the true model. This led to the development of nonparametric procedures to estimate the hazard

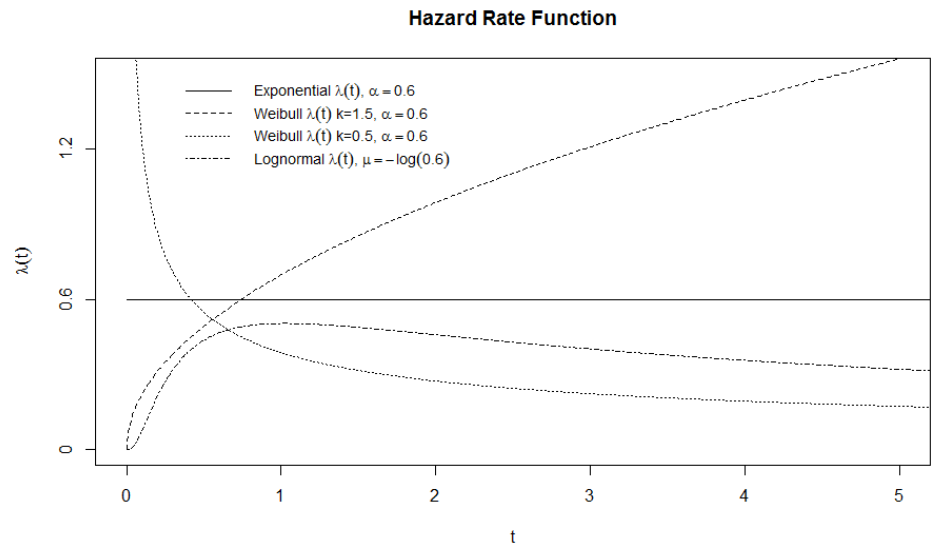


Figure 1.1

Examples of parametric form of hazard rate function

rate function. There are a wide variety of nonparametric methods. For example, spline and wavelet techniques have been used in developing nonparametric models. However, here we will only consider kernel based methods. Since the basic kernel method is much easier to understand for density function estimation, in the next section we first describe the kernel-based nonparametric procedure for density estimation. And then describe the kernel-based estimation of hazard rate.

1.3 Nonparametric Estimation

1.3.1 Kernel Density Estimation

Kernel-based estimation is one of the most commonly used methods in nonparametric curve estimation which we now illustrate. Let $x_1, x_2, \cdots, x_n$ be the realization of

a random sample $X_1, X_2, \cdots, X_n$ from a density function $f(x)$ and let $F(x)$ denote the corresponding cumulative distribution. Then, consider the probability

$$P(x-h < X \leq x+h) = F(x+h) - F(x-h) \ .$$

Note that

$$\lim_{h\to 0} \frac{P(x-h<X<x+h)}{2h} = \lim_{h\to 0} \frac{F(x+h)-F(x-h)}{2h} = f(x) \ .$$

That is

$$f(x) \simeq \frac{F(x+h)-F(x-h)}{2h}$$

when h is very small. Since $F(x+h) - F(x-h)$ can be estimated by the number of observations in the interval $(x-h, x+h)$ divided by the sample size n, then, an estimate of $f(x)$ will be

$$\begin{aligned}
\hat{f}(x) &= \frac{1}{2nh}[\text{Number of } x_1, \ldots, x_n \text{ in } (x-h, x+h)] \\
&= \frac{1}{2nh} \sum_{i=1}^{n} I[x-h < x_i < x+h] \\
&= \frac{1}{nh} \sum_{i=1}^{n} K\big(\frac{x-x_i}{h}\big)
\end{aligned}$$

where $K(\cdot)$ is the probability density function of the uniform random variable over $(-1, 1)$. That is, an estimator of $f(x)$ is

$$\frac{1}{nh} \sum_{i=1}^{n} K\big(\frac{x-X_i}{h}\big) \ .$$

The above estimator is the average of all x_i's in the neighborhood $(x-h, x+h]$ with equal weights. But it makes sense to have higher weights for x_i's closer to x and lower weights

if x_i's are away from x. It means instead of $K(\cdot)$ being a uniform density, one can take K to be any symmetric density such that its mode is at the center and, in both directions away from center, $K(\cdot)$ is a decreasing function. For example

$$K_1(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

or

$$K_2(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2 & \text{if } |x| < 1 \\ 0 & \text{otherwise .} \end{cases}$$

Thus the kernel estimator with kernel K is defined by following

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h}) \ ,$$

where K is a symmetric probability density function and h is referred to as window width. In terms of computation, it means that drawing a kernel density centered at each sample point x_i, scaled by $1/n$ and then adding these densities to get $\hat{f}(x)$.

Figure 1.2 is an example of kernel density estimation with a Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2} \ .$$

for two different values of h. There, the solid line is the estimated function $\hat{f}(x)$, the dashed line functions are the kernels at each observation and the vertical lines at the bottom represent the sampled observations. One is able to observe that for a given sample, different h provides slightly different $\hat{f}(x)$ function.

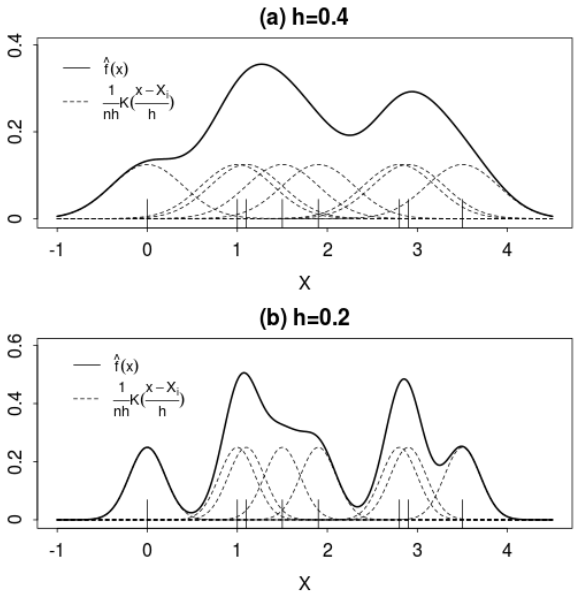


Figure 1.2

kernel estimate with Gaussian kernel. (a) h=0.4, (b) h=0.2

1.3.2 Kernel Hazard Rate Estimation

Once we have a kernel density estimator, it is easy to define kernel based estimate of a hazard rate function $\lambda(x)$. Note that

$$\lambda(x) = \frac{f(x)}{1 - F(x)} \, .$$

Since we have already estimated the numerator $f(x)$, we can plug the kernel estimator of f and an estimator of F in the ratio $f/(1-F)$ to obtain a hazard rate estimator. The cumulative distribution function can be estimated either by integrating the $\hat{f}(x)$ or by using the empirical cumulative distribution function. Let

$$\hat{F}_1(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} \, ,$$

$$\hat{F}_2(x) = \int_{-\infty}^{x} \hat{f}(u) du$$

where I is the indicator function. Since the life time variables are always positive, the integral in the definition of $\hat{F}_2$ is actually taken from 0. Thus a kernel estimate of the hazard rate as a ratio is

$$\hat{\lambda}(x) = \frac{\hat{f}(x)}{1 - \hat{F}(x)}$$

where $\hat{f}(x)$ is the kernel density estimate and $\hat{F}(x)$ could be either $\hat{F}_1$ or $\hat{F}_2$. Note that if we let $\hat{F} = \hat{F}_1$ then hazard rate estimate $\hat{\lambda}(x)$ is undefined for $X_{(n)} = x$, where $X_{(n)}$ is the nth order statistic. So generally a little adjustment is made to $\hat{F}$ by defining

$$F_n(x) = \frac{n}{n+1}\hat{F}_1(x) \ .$$

The hazard rate estimator mentioned above can then be rewritten as

$$\hat{\lambda}_1(x) = \frac{\hat{f}(x)}{1 - F_n(x)} \ .$$

In the above construction of estimates of λ, we treated $\lambda(x)$ as a ratio of f to $1 - F$. However, there is another estimator of λ discussed in the literature, which is obtained by treating $\lambda(x)$ as a function by itself. For that, define the approximattion to $\lambda(x)$ by

$$\lambda(x) \simeq \int \frac{1}{h} K(\frac{x-u}{h})\lambda(u)du = \int \frac{1}{h} K(\frac{x-u}{h})d\Lambda(u) \ ,$$

where K is a kernel function as before. Let $\Lambda_n(u)$ be the empirical version of $\Lambda(u)$, for example

$$\Lambda_n(x) = -\log(1 - F_n(x)) \ .$$

Then we write

$$\begin{aligned}\lambda(x) &\simeq \int \frac{1}{h}K(\frac{x-u}{h})\lambda(u)du \\ &= \int \frac{1}{h}K(\frac{x-u}{h})d\Lambda(u) \\ &= \int \frac{1}{h}K(\frac{x-u}{h})d\Lambda_n(u) + \int \frac{1}{h}K(\frac{x-u}{h})d[\Lambda(u) - \Lambda_n(u)] \,. \end{aligned} \tag{1.1}$$

The integral of the first term in (1.1) provides an estimator of $\lambda(x)$ say, $\hat{\lambda}_2(x)$ and is given as

$$\hat{\lambda}_2(x) = \int \frac{1}{h}K(\frac{x-u}{h})d\Lambda_n(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{h}K(\frac{x-X_i}{h})}{1-F_n(X_i)} = \sum_{i=1}^{n}\frac{\frac{1}{h}K(\frac{x-X_i}{h})}{n-i+1} \,.$$

The second term in (1.1) represents the noise.

1.4 Nonparametric Regression

As mentioned in section 1.1, one can estimate the hazard rate using a nonparametric regression approach. To define such an estimator, we first describe nonparametric kernel-based regression estimators.

1.4.1 Nadaraya-Watson Estimation

There exist several approaches to nonparametric regression estimation, for example, kernel smoothing, spline smoothing and wavelets. Each of these approaches has its own particular strengths and weaknesses, although for being easy to understand and for its mathematical simplicity compared to other methods, kernel methods are most commonly used.

Suppose we have a bivariate sample (x_i, Y_i), $i = 1, 2, \cdots, n$ on bivariate random variable (X, Y) and is modeled as

$$Y_i = m(x_i) + \epsilon_i, \qquad i = 1, 2, \cdots, n$$

where $m(\cdot)$ is a smooth function, conditioning on $X_1, \cdots, X_n$, ϵ_i's are independent random variables such that

$$E(\epsilon_i) = 0, \qquad Var(\epsilon_i) = \sigma_\epsilon^2 \, .$$

The interest here is to estimate $m(x_i) = E[Y|X = x_i]$.

The Nadaraya-Watson approach is to estimate $m(x)$ at point x_0 by taking means of all those $Y_i's$ corresponding $x_i's$ which are in a close neighborhood of x_0 assuming that $m(x)$ is constant in that small neighborhood . That is Nadaraya-Watson estimator at x_0 defined by

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n I[x_0 - h < X_i < x_0 + h] Y_i}{\sum_{i=1}^n I[x_0 - h < X_i < x_0 + h]} \, . \tag{1.2}$$

Here, in the local average around point x_0, each observation is given equal weight regardless of its distance from the point of estimation x_0. Again as explained in kernel density estimation, it seems more natural to have weighted average with more weight to the observation closer to x_0. That leads to the modified estimator which is defined as

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n Y_i K_h(x_i - x_0)}{\sum_{i=1}^n K_h(x_i - x_0)} \tag{1.3}$$

where $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ is taken to be a symmetric probability density function.

1.4.2 Local Linear Fitting

With the same model and assumptions described in 1.4.1, assuming $m(x)$ to be a smooth function and by using Taylor expansion, we can write that

$$m(x) \simeq m(x_0) + (x - x_0) m'(x_0) + \frac{(x - x_0)^2}{2} m''(x_0) + \ldots + \frac{(x - x_0)^p}{p!} m^{(p)}(x_0) \, .$$

Now define

$$Q = \sum_{i=1}^{n} \big(Y_i - m(x_0) - m'(x_0)(x_i - x_0) - \cdots - \frac{m^{(p)}(x_0)}{p!}(x_i - x_0)^p\big)^2 K_h(x_i - x_0) \ .$$

Let $\beta_r = \frac{m^{(r)}(x_0)}{r!}$, and rewrite the above equation as

$$Q = \sum_{i=1}^{n} \big(Y_i - \beta_0 - \beta_1(x_i - x_0) - \beta_2(x_i - x_0)^2 - \cdots - \beta_p(x_i - x_0)^p\big)^2 K_h(x_i - x_0)$$

where p is the degree of the polynomial being fit. So now the task is to minimize Q and find least squares estimators of $\beta_0, \beta_1, \cdots, \beta_p$. In that, one can easily see that the Nadaraya-Watson estimator described in the last subsection is obtained by letting $p = 0$. That is

$$\hat{\beta}_0 = \hat{m}(x) = \frac{\sum_{i=1}^{n} Y_i K_h(x_i - x_0)}{\sum_{i=1}^{n} K_h(x_i - x_0)} \ , \tag{1.4}$$

which is exactly the same as (1.2). If we set $p = 1$ then

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1(x_i - x_0))^2 K_h(x_i - x_0) \tag{1.5}$$

then the estimator of $m(x)$ will be $\hat{m}(x_0) = \hat{\beta}_0$ referred as the local linear estimator. The estimator of $m'(x)$ will be $\hat{m}'(x_0) = \hat{\beta}_1$. Now by minimizing Q with respect to β_0 and β_1 we have

$$\hat{\beta}_0 = \hat{m}(x) = n^{-1} \sum_{i=1}^{n} \frac{\{\hat{s}_2(x_0; h) - \hat{s}_1(x_0; h)(x_i - x_0)\} K_h(x_i - x_0) Y_i}{\hat{s}_2(x_0; h)\hat{s}_0(x_0; h) - \hat{s_1}(x_0; h)^2}$$

where

$$\hat{s}_r(x_0; h) = n^{-1} \sum_{i=1}^{n} (x_i - x_0)^r K_h(x_i - x_0) \ .$$

For a detailed discussion including the properties of the local polynomial smoothing estimators, see Fan and Gijbels (1998) [14].

1.4.3 Density Estimation Through Regression

One can use regression methodology, described in the last subsection to estimate a probability density function. This approach is of interest because of the nice properties such as no boundary bias that local polynomial estimators possess. To illustrate the local linear regression approach to estimate a probability density function, suppose that $X_1, \ldots, X_N$ is a random sample from a distribution with the probability density function f and that the density function f is compactly supported on an interval, say the unit interval $[a, b]$. Divide this interval into t subintervals $\{I_k, k = 1, \cdots, t\}$ with equal length $\Delta = (b - a)/t$ and let x_k be the center of I_k and q_k be the proportion of the data $\{X_i, i = 1, \cdots, N\}$ in the interval I_k, divided by the bin length Δ. Then it is obvious that the bin counts $N\Delta q_k$ follow a binomial distribution

$$N\Delta q_k \sim \text{Binomial}(N, p_k) \text{ with } p_k = \int_{I_k} f(x)dx \, . \tag{1.6}$$

When $N \to \infty$, $\Delta \to 0$ from (1.6) we can show that

$$E(q_k) \approx f(x_k), \quad Var(q_k) \approx \frac{f(x_k)}{N\Delta} \, . \tag{1.7}$$

So, with data $(q_k, x_k), k = 1, \ldots, t$ we can treat the density estimation problem as a nonparametric regression problem such that

$$q_k = m(x_k) + \sigma(x_k)\epsilon_k, \quad k = 1, 2, \cdots, t$$

with $m(x) = f(x)$ and $\sigma(x) = \frac{f(x)}{N\Delta}$. Thus, the nonparametric regression techniques introduced in the last section can be used to obtain a density estimator.

1.4.4 Hazard Rate Estimation Using Local Linear Fit

In the last subsection we have explained how to use the local linear regression methodology to estimate a probability density function. One can apply the same technique to estimate a hazard rate function as well. Suppose that we are interested in estimating the hazard rate function $\lambda(x)$ on an interval $[0, b]$ where $b > 0$ given the sample points $X_1, \cdots, X_N$ from density function f. First, we divide the interval into n bins, say $I_1, I_2, \cdots, I_n$ each of width $\Delta = b/n$ and let x_i, $i = 1, \cdots, n$, be the center of the interval I_i. Then construct appropriate estimators for the hazard rate, say Y_i, at each bin center x_i such that

$$E(Y_i) \simeq \lambda(x_i), \;\; Var(Y_i) \simeq \sigma(x_i).$$

Then we have the data (x_i, Y_i) which can be modeled as

$$Y_i = m(x_i) + \sigma(x_i)\epsilon_i, \;\; i = 1, 2, \cdots, n$$

with $m(x_i) = \lambda(x_i)$ and ϵ_i being independent $r.v.$ with mean 0 and variance 1.

There are a number of ways to construct empirical estimators Y_i of the hazard rate. The most natural one is the hazard histogram estimator which is the ratio of relative frequency to the empirical survival function. By choosing the bin center x_i which is the point of estimation, and the hazard histogram estimator Y_i, we could treat the hazard estimation problem as an heteroscedastic nonparametric regression problem based on the data $\{x_i, Y_i, i = 1, \cdots, n\}$.

Associated with each x_i, the center of the interval I_i, $i = 1, 2, \cdots, n$, define

$$Y_i = \frac{\frac{\{\#\text{ of } X_i \in I_i\}}{N\Delta}}{1 - \sum_{j=1}^{i} \frac{\{\#\text{ of } X_i \in I_j\}}{N}} = \frac{f_i/\Delta}{N - \sum_{j=1}^{i} f_j}$$

where $f_i = \{\#\text{ of } X_i \in I_i\}$. Note that if $i = n$ then

$$N - \sum_{j=1}^{n} f_j = 0$$

In order to avoid denominator to be zero, we redefine Y_i as

$$Y_i = \frac{1}{\Delta} \frac{f_i}{N - \sum_{j=1}^{i} f_j + 1}, \quad i = 1, 2, \cdots, n.$$

and also define

$$c_i = \frac{f_i}{N - \sum_{j=1}^{i} f_j + 1}, \quad i = 1, 2, \cdots, n.$$

Note that,

$$f_i \sim \text{binomial}(N, p_i), \quad p_i = \int_{I_i} f(x)dx.$$

It is easy to see that

$$E(Y_i) = E(\frac{1}{\Delta} c_i) \simeq \lambda(x_i)$$
$$Var(Y_i) = Var(\frac{1}{\Delta} c_i) \simeq \frac{1}{N\Delta} \frac{\lambda(x_i)}{1 - F(x_i)}$$

Thus the data$\{x_i, Y_i, i = 1, \cdots, n\}$ can be used to provide a hazard rate estimator based on the local linear method.

1.5 Outline of the Dissertation

The nonparametric confidence interval for an unknown function is quite an useful tool in statistical inferential procedures and thus there exists a wide body of literature on the

topic. The primary issues are the smoothing parameter selection using appropriate criterion and then the coverage probability and length of the associated confidence interval. In this dissertation our focus is on the interval length in general and, in particular, on the variability in the length of nonparametric intervals for probability density and hazard rate functions.

In Chapter 2 we start with the analysis of a nonparametric confidence interval for a probability density function. The bias and standard error of a nonparametric kernel density estimator is of the same order when mean square error optimal bandwidth is used. It is, therefore, known that a standard confidence interval of the form 'density estimate plus or minus constant multiple of its standard error provides a confidence for the density function plus bias rather than the density function only. To remedy the situation Hall (1992) [17] has proposed two different solutions C bias corrected and under-smoothing methods. Of course, interest there is coverage accuracy. Our starting point is these solutions but instead of coverage probability we address the issue of the variability in the length of the confidence interval. This variability depends on the variability of the estimator of the square-root of the density function. To control the variability we propose to use an estimator of the square-root of a density function which has constant variance. For that we use a regression based approach to estimate the unknown density function. Once the data is created to estimate the density function using a regression approach, we transform the response variable so that we get an estimated square-root of the density function with constant variance, irrespective of the point of estimation. We then use this estimator in the confidence intervals proposed by Hall (1992) [17] and then through simulation investigate its effect on the variability of the confidence intervals.

In Chapter 3 , we extend the methodology to the settings of hazard rate estimation. Very similar to the case of kernel density estimation, standard deviation of the standard kernel based nonparametric hazard rate estimator depends on the square root of hazard rate divided by cumulative distribution function. To estimate this standard deviation we propose two estimators. In the first, using the approach used in Chapter 2, we find an estimate of the square-root of the hazard rate which has a constant variance. Then to obtain the estimate of the standard deviation of the standard kernel-based hazard rate estimator we divide the estimate of the square-root of hazard rate by the square-root of the empirical distribution function. In the other, we obtain the constant variance estimator of the square root of the ratio of hazard rate to cumulative distribution function. These estimators are then used in the existing nonparametric confidence intervals for hazard rate function to investigate their effect on the variability of the confidence intervals.

CHAPTER 2

NONPARAMETRIC CONFIDENCE INTERVAL FOR AN UNKNOWN DENSITY FUNCTION

2.1 Introduction

The main focus of this chapter is the nonparametric confidence intervals for a density function $f(x)$. The problem of constructing a nonparametric confidence interval has been widely discussed in the literature and the main two issues associated with it are the coverage accuracy and the length of the confidence interval. Here our interest is in the latter issue and for that we will confine our attention to the bias correction and under-smoothing methods discussed in Hall (1992) [17]. First, in section 2.2, we describe these two methods in detail. We know that the length of the confidence interval depends on the standard deviation of the density estimator. But, its estimate, that is the standard error of density estimator, varies with the point of estimation. Thus the variability of the interval length depends on the variance of the standard error of density $f(x)$ at point x. Therefore, it is essential to have an estimate of the standard error of $\hat{f}(x)$ with "nice" properties, e.g. constant variance.

Towards this end, in section 2.3, we consider an estimator of the standard error of $\hat{f}(x)$ which has a constant variance. The advantage is that the variability of confidence interval length is a constant regardless of the estimation point x. If we were to use such an

estimator, we show that it will lead to the confidence intervals with the property that the variances of their lengths is smaller than the initial estimator mentioned in section 2.2.

2.2 Constructing Density Confidence Intervals

2.2.1 Estimators of Bias and Variance

Let $\hat{f}$ denote the kernel density estimator mentioned in (1.3.1)

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) \ .$$

The standard calculations show that

$$E(\hat{f}(x)) = f(x) + \frac{1}{2}h^2 f''(x) \int y^2 K(y)dy + o(h^2) \tag{2.1}$$

and

$$Var(\hat{f}(x)) = \frac{1}{nh} f(x) \int K^2(y)dy + o(\frac{1}{nh}), \tag{2.2}$$

see for example Silverman (1988) [36]. If we assume that

$$\frac{\hat{f} - E(\hat{f})}{\sqrt{Var(\hat{f})}} \sim N(0,1)$$

and if $SE(\hat{f})$ denotes the standard error of $\hat{f}$, then the nominal $100(1-\alpha)\%$ confidence interval of true density f can be given as

$$\left(\hat{f} - SE(\hat{f}) * z_{\alpha/2}, \hat{f} + SE(\hat{f}) * z_{\alpha/2}\right) , \tag{2.3}$$

where $z_{\alpha/2}$ is such that $\Phi(z_{\alpha/2}) = 1 - (1/2)\alpha$ with Φ being the distribution function of the standard normal random variable.

From (2.1) it is clear that since $\hat{f}$ is not an unbiased estimator of f, the confidence interval in (2.3) is not exactly for $f(x)$; instead it provides confidence interval for "$f(x)$ + bias". Hall (1992) [17] addressed this problem and proposed two methods to deal with it. One of them is referred to as bias correction and the other under-smoothing. Of the two methods, we first describe the bias correction method.

Let K be a rth order kernel $r \geq 1$, then we have the following

$$\int y^i K(y)dy \begin{cases} = 1, & \text{if } i = 0, \\ = 0, & \text{if } 1 \leq i \leq r-1, \\ \neq 0, & \text{if } i = r. \end{cases}$$

For $r = 2$, the kernel K will usually be a symmetric probability density function, for example the normal density. The bias of $\hat{f}$ is given by

$$\begin{aligned} b(x) &= E\hat{f}(x) - f(x) \\ &= \int h^{-1} K\{(x-u)/h\} f(u)du - f(x). \end{aligned}$$

If we let

$$y = \frac{x-u}{h}$$

then the bias can be written as

$$\begin{aligned} b(x) &= \int K(y) f(x-hy)dy - f(x) \\ &= \int K(y)\{f(x-hy) - f(x)\}dy. \end{aligned}$$

A Taylor series expansion gives

$$f(x-hy) = f(x) - hyf'(x) + \frac{1}{2}h^2y^2f''(x) + \cdots$$

so that,

$$b(x) = \kappa_r h^r f^{(r)}(x) + o(h^r) \tag{2.4}$$

as $h \to 0$, assuming $f^{(r)}$ is bounded and continuous where

$$\kappa_r = (-1)^r (r!)^{-1} \int y^r K(y) dy.$$

Because $\hat{f}$ is a biased estimator of f as mentioned before, the interval in (2.3) is actually a confidence interval for "$f(x)$ + bias". Then the first method to get confidence interval for $f(x)$ requires one to estimate the bias in $f(x)$ and then adjust the interval and thus it is called bias correction method. Now if the bias correction method is to be used, then one needs to estimate $b(x)$. Typically then one estimates the dominant term in the (2.4). For example, let $\tilde{f}^{(r)}(x)$ be an estimate of $f^{(r)}(x)$ where

$$\tilde{f}^{(r)}(x) = \frac{1}{n h_1^{r+1}} \sum_{i=1}^{n} L^{(r)}\{\frac{x - X_i}{h_1}\} ,$$

where L is an sth order kernel having at least r derivatives. So the estimator of $b(x)$ in (2.4) is

$$\hat{b}(x) = \kappa_r h^r \tilde{f}^{(r)}(x) . \tag{2.5}$$

The explicit bias correction method uses the estimator of the bias in (2.5) by way of subtracting $\hat{b}(x)$ from both end points of the interval in (2.3).

For estimator of the standard error of $\hat{f}(x)$, first note that the variance of $\hat{f}(x)$ is

$$\begin{aligned}\sigma^2(x) =& (nh^2)^{-1} \int K^2\{(x-y)/h\} f(y) dy \\ & - n^{-1}[h^{-1} \int K\{(x-y)/h\} f(y) dy]^2.\end{aligned}$$

We consider two different estimators of $\sigma^2(x)$. The first of the two, denoted by σ_1^2 is obtained by replacing the first term in the $\sigma^2(x)$ by its moment estimator and noting the fact that the second term is $n^{-1}\hat{f}^2(x)$. That is

$$\sigma_1^2(x) = \frac{1}{nh^2}E[K^2(\frac{x-y}{h})] - n^{-1}\hat{f}^2(x)$$

and the estimator is written as

$$\hat{\sigma}_1^2(x) = \frac{1}{nh}[\frac{1}{nh}\sum_{i=1}^{n} K^2\{\frac{x-X_i}{h}\} - h\hat{f}^2(x)] \tag{2.6}$$

and then the standard error of $\hat{f}(x)$ can be written as

$$SE(\hat{f}(x)) = \sqrt{\hat{\sigma}_1^2(x)}\ .$$

Alternatively, note that for large n, we may approximate $\sigma^2(x)$ by using Taylor Series to obtain that

$$\sigma_2^2(x) = \frac{1}{nh}f(x)\int K^2(t)dt + o(\frac{1}{nh}),$$

and then we define our second estimator as

$$\hat{\sigma}_2^2(x) = \frac{1}{nh}\hat{f}(x)\int K^2(t)dt. \tag{2.7}$$

Then by using $\hat{\sigma}_2^2$, the standard error of $\hat{f}(x)$ is

$$SE(\hat{f}(x)) = \sqrt{\hat{\sigma}_2^2(x)}.$$

Thus the bias corrected confidence interval will be

$$\big(\hat{f}(x) - \hat{b}(x) - \sqrt{\hat{\sigma}^2(x)} * z_{\alpha/2}, \hat{f}(x) - \hat{b}(x) + \sqrt{\hat{\sigma}^2(x)} * z_{\alpha/2}\big) \tag{2.8}$$

where $\sqrt{\hat{\sigma}^2(x)}$ is either $\sqrt{\hat{\sigma}_1^2(x)}$ or $\sqrt{\hat{\sigma}_2^2(x)}$.

Now to describe the under-smoothing method, from equation (2.4), note that the smaller the h the smaller will be the bias. Therefore, if we choose a relatively small value of h such that the bias term in (2.4) is negligible, then the kernel density estimator $\hat{f}(x)$ will be very close to unbiased estimator of $f(x)$. Then the confidence interval defined in (2.3) is written as

$$\big(\breve{f}(x) - \sqrt{\hat{\sigma}^2(x)} * z_{\alpha/2}, \breve{f}(x) + \sqrt{\hat{\sigma}^2(x)} * z_{\alpha/2}\big), \tag{2.9}$$

where $\breve{f}(x)$ is the under-smoothing kernel density estimator with a smaller bandwidth, say h', which makes (2.4) negligible. Again, in (2.9), $\sqrt{\hat{\sigma}^2(x)}$ is either $\sqrt{\hat{\sigma}_1^2(x)}$ or $\sqrt{\hat{\sigma}_2^2(x)}$.

For the critical value in (2.3), we initially used standard normal distribution. Alternatively, one can use bootstrap to find the critical value. Hardle and Bowman (1988) [18] have analyzed properties of bootstrap confidence intervals in nonparametric regression in terms of the consistency problem which ensures that the coverage error converged to 0. However, Hall (1992) [17] is the first to use bootstrap to address the effect of both the bias estimation and bandwidth choice on coverage accuracy of a two sided confidence interval. We are using the same strategy here to find the critical value for the confidence interval instead of the standard normal distribution.

2.2.2 Bootstrap Confidence Intervals

First, we describe the construction of bootstrap confidence intervals for $E(\hat{f})$. Let $\{X_1^*, \ldots, X_n^*\}$ denote a resample drawn randomly, with replacement, from $\mathbb{X} = \{X_1, \ldots, X_n\}$. We have

$$\hat{f}^*(x) = \frac{1}{nh} \sum_{i=1}^{n} K\{\frac{x - X_i^*}{h}\},$$

$$\hat{\sigma}^*(x)^2 = \frac{1}{nh}[\frac{1}{nh}\sum_{i=1}^{n} K\{\frac{x - X_i^*}{h}\}^2 - h\hat{f}^*(x)^2],$$

$$S^*(x) = \frac{\hat{f}^*(x) - \hat{f}(x)}{\hat{\sigma}^*(x)}.$$

Here, for convenience, we shall use $f, \hat{f}, \hat{\sigma}$ for $f(x), \hat{f}(x), \hat{\sigma}(x)$ respectively.

To construct a confidence interval for $E(\hat{f})$, we would ideally wish to know the distribution of $S = (\hat{f} - E\hat{f})/\hat{\sigma}$, from which we would compute the quantile u_α defined by $P(S \leq u_\alpha) = \alpha$ and the bootstrap estimate of u_α could be defined as $\hat{u}_\alpha$ such that

$$P(S^* \leq \hat{u}_\alpha | \mathbb{X}) = \alpha.$$

Thus, one can use u_α instead of $z_{\alpha/2}$ in (2.3) to give the confidence interval for both bias correction and under-smoothing.

From the discussion in 2.2.1, we construct the two sided confidence interval of $\hat{f}$ by using both bias correction and under-smoothing methods. Let $\hat{f}_1$ be the kernel density estimator used in the under-smoothing method with bandwidth h_{us} and $\hat{f}_2$ be the kernel density estimator used in the bias correction method with bandwidth h_{bc}. For the standard error term, we can use either $\hat{\sigma}_1$ or $\hat{\sigma}_2$ on both methods mentioned above. That is $\hat{\sigma}_u = \hat{\sigma}_1$ or $\hat{\sigma}_2$ and similarly $\hat{\sigma}_b = \hat{\sigma}_1$ or $\hat{\sigma}_2$. That leads to the confidence intervals

$$\hat{J}_{\text{us}} = (\hat{f}_1 - \hat{\sigma}_u \hat{u}_{(1+\alpha)/2}, \hat{f}_1 - \hat{\sigma}_u \hat{u}_{(1-\alpha)/2}) \tag{2.10}$$

$$\hat{J}_{\text{bc}} = (\hat{f}_2 - \hat{\sigma}_b \hat{u}_{(1+\alpha)/2} - \hat{b}, \hat{f}_2 - \hat{\sigma}_b \hat{u}_{(1-\alpha)/2} - \hat{b}), \tag{2.11}$$

where $\hat{J}_{\text{us}}$ refers to the confidence interval obtained using under smoothing and $\hat{J}_{\text{bc}}$ refers to the confidence interval obtained by the bias correction method; and both confidence intervals given above have nominal coverage probability α.

2.2.3 Variability of the Confidence Interval

From (2.10) and (2.11), one can notice that the variability of the interval length stems from the standard error terms for both the bias correction and under-smoothing methods. For example, the length of the confidence interval for the bias correction approach is

$$\hat{\sigma}_b \hat{u}_{(1+\alpha)/2} - \hat{\sigma}_b \hat{u}_{(1-\alpha)/2} \, .$$

Since $\hat{u}_{(1+\alpha)/2}$ and $\hat{u}_{(1-\alpha)/2}$ are the critical values and $\hat{\sigma}_b$ is the only term which affects the interval length and the same is true for the under-smoothing approach. That is, $\hat{\sigma}_u$ is the only term which affects the length of the interval in (2.10).

We noted that $\hat{\sigma}_b$ could be either $\hat{\sigma}_1$ or $\hat{\sigma}_2$ which are defined in (2.6) and (2.7), respectively, and the same is true for $\hat{\sigma}_u$. In either the case of $\hat{\sigma}_1$ or $\hat{\sigma}_2$, the variance of the standard error of $\hat{f}$ varies depending on the estimated point. In the next section, we use a variance stabilization transformation to propose an estimate of the standard error of $\hat{f}$ such that the variance of the standard error of $\hat{f}$ is constant for all $x_1, \cdots, x_n$.

2.3 Variance Stabilization Methodology

2.3.1 Variance Stabilizing Transformation

In applied statistics, a variance stabilizing transformation is a data transformation that is specifically chosen to simplify either on consideration in data analysis or the ap-

plication of regression based analysis. Variance stabilizing transformations, and closely related transformations to approximate normality, have been used in many statistical contexts. Suppose a r.v. X has mean 0 and variance equal to $\sigma^2 q(\theta)$. The goal behind the choice of a variance stabilizing transformation is to find a simple function g such that variability of the transformed r.v. $Y = g(X)$ does not vary with θ.

2.3.2 Root-Unroot Transformation

The literature gives us a number of ways to stabilize variance by using the transformation technique. Bartlett (1936) [5] was the first to propose the root transformation $\sqrt{X}$ in a homoscedastic linear model. Anscombe (1948) [2] proposed improving the variance stabilizing properties by using $\sqrt{X + \frac{3}{8}}$ instead of $\sqrt{X}$. The constant $\frac{3}{8}$ is chosen to optimally stabilize the variance using the Taylor expansion. Brown and Cai (2009) [7] have shown that if X $\sim$ Poisson(λ) with $\lambda > 0$, then

$$E\big(\sqrt{X + \frac{1}{4}}\big) = \lambda^{\frac{1}{2}} - \frac{1}{64}\lambda^{-\frac{3}{2}} + O(\lambda^{-\frac{5}{2}}) \tag{2.12}$$

$$Var\big(\sqrt{X + \frac{1}{4}}\big) = \frac{1}{4} + \frac{1}{32}\lambda^{-1} + \frac{3}{64}\lambda^{-2} + O(\lambda^{-3}). \tag{2.13}$$

That is, variance of X varies with λ but variance of $\sqrt{X + 1/4}$ is almost constant for large λ. We will apply this technique to the density estimation via nonparametric regression mentioned in section 1.4.3. But before that notice that if we use $\hat{\sigma}_2$ for say $\hat{\sigma}_u$ then

$$\begin{aligned}\hat{\sigma}_u(x) &= \sqrt{\frac{1}{nh_{\text{us}}}\hat{f}(x)\int K^2(t)dt} \\ &= \sqrt{\frac{1}{nh_{\text{us}}}\int K^2(t)dt}\sqrt{\hat{f}(x)} \\ &= C_u \times \sqrt{\hat{f}(x)} \end{aligned} \tag{2.14}$$

where

$$C_u = \sqrt{\frac{1}{nh_{\text{us}}}\int K^2(t)dt}.$$

Table 2.1

Variance Comparison

X	-1.5	-0.75	0	0.75	1.5
Variance of Square Root of Kernel Density	0.011	0.0048	0.0061	0.0095	0.017
Variance via Transformation & Nonparametric Regression	0.00021	0.00018	0.00016	0.00018	0.00022

To illustrate the effect of the the point estimation on $\hat{\sigma}_u(x)$, suppose r.v. X follows the standard normal distribution and take 300 samples each of size 100 from $f(x)$. We will have 300 estimates of f at the designated points, say $x_1, x_2, \ldots, x_d$. One is able to calculate $Var(\sqrt{\hat{f}(x_i)})$, $i = 1, 2, \ldots, d$ according to the 300 samples. Table 2.1 shows the simulation numbers of $Var(\sqrt{\hat{f}(x_i)})$, $i = 1, 2, \ldots, 5$. First row lists all the estimation points and the second row shows the variance of $\sqrt{\hat{f}(x)}$ by using kernel density estimation

mentioned above. The third row shows the variance of $\widehat{\sqrt{f(x)}}$ through nonparametric regression estimator of $\sqrt{f(x)}$ as discussed below.

Apparently, the variances introduced in the third row are close to each other and also smaller compared to the second row. We then illustrate this method by applying the root-unroot transformation to estimate $\sqrt{f}$ via regression and have the new estimator $\widehat{\sqrt{f}}$ so that the variance of this term will be a constant as we see in table 2.1, and it makes the variance of confidence interval length constant as well.

We now refer to the density estimation through nonparametric regression approach discussed in section 1.4.3. Let y_i and x_i be the observations and the center of the bins respectively. N is the number of total observations and Δ is the bin length and suppose q_i is the proportion of the observations in the bin divided by the bin length Δ. We have shown that

$$N\Delta q_i \sim \text{Binomial}(N, p_i) \text{ with } p_i = \int_{I_i} f(x)dx$$

where I_i is the bin. When $N \to \infty$, $\Delta \to 0$ we have

$$E(q_i) \approx f(x_i), \quad Var(q_i) \approx \frac{f(x_i)}{N\Delta}\,.$$

So, here we define

$$M_i = \sqrt{\frac{y_i + \frac{1}{4}}{N\Delta}}\,.$$

Then,

$$E(M_i) \approx \sqrt{f(x_i)}, \quad Var(M_i) = \sigma^2(x_i),$$

Thus, the problem of estimating $\sqrt{f(x)}$ can be seen as a nonparametric regression problem based on data (x_i, M_i) with the model,

$$M_i = m(x_i) + \sigma(x_i)\epsilon_i, \;\; i = 1, \cdots, n.$$

where $m(x_i) = \sqrt{f(x_i)}$ and ϵ_i being independent $r.v.$ with mean 0 and unit variance.

$$\hat{m}(x_0) = \widehat{\sqrt{f(x_0)}} = n^{-1} \sum_{i=1}^{n} \frac{\{\hat{s}_2(x_0; h) - \hat{s}_1(x_0; h)(x_i - x_0)\} K_h(x_i - x_0) Y_i}{\hat{s}_2(x_0; h)\hat{s}_0(x_0; h) - \hat{s_1}(x_0; h)^2}$$

where

$$\hat{s}_r(x_0; h) = n^{-1} \sum_{i=1}^{n} (x_i - x_0)^r K_h(x_i - x_0) \; .$$

One may notice that from (2.13), asymptotically,

$$\sigma^2(x_i) = \frac{1}{4} * C_\sigma \tag{2.15}$$

where C_σ is a constant depending on the sample size. Thus, based on this nonparametric regression, we have $\hat{m} = \widehat{\sqrt{f}}$ as an estimate of $\sqrt{f}$, and by taking advantage of this variance stabilization technique it essentially gives us a confidence interval of f with a nice variance property of interval length. We write

$$\widehat{\sqrt{f(x)}} = \hat{m}(x) \; . \tag{2.16}$$

In the nonparametric local linear regression, if one considers the fixed equally spaced design, it has been shown by Wand and Jones (1995) [37] that

$$Var(\hat{m}(x)) \simeq \frac{1}{nh} \int K^2(z) dz \; \sigma^2(x)$$

which is a constant based on (2.15). Thus instead of $\hat{\sigma}_u(x)$ we can use the estimator with constant variance, that is,

$$\hat{\sigma}_{vs} = C_u * \hat{m}(x) \ . \tag{2.17}$$

Note that the standard error acquired in (2.17) is not related to the kernel density estimator at point x and also the variance of standard error is a constant, thus we can use $\widehat{\sqrt{f(x)}}$ for bias correction method as well and one is able to rewrite both (2.10) and (2.11) as

$$\tilde{J}_{\text{us}} = (\hat{f}_1 - C_u\widehat{\sqrt{f(x)}}\hat{u}_{(1+\alpha)/2}, \hat{f}_1 - C_u\widehat{\sqrt{f(x)}}\hat{u}_{(1-\alpha)/2}) \tag{2.18}$$

$$\tilde{J}_{\text{bc}} = (\hat{f}_2 - C_b\widehat{\sqrt{f(x)}}\hat{u}_{(1+\alpha)/2} - \hat{b}, \hat{f}_2 - C_b\widehat{\sqrt{f(x)}}\hat{u}_{(1-\alpha)/2} - \hat{b}). \tag{2.19}$$

2.4 Numerical Results

2.4.1 Summary

In this section we report details of a simulation study which exhibits performance of the variance stabilization technique on both under-smoothing and bias correction methods. We first simulate the confidence intervals which are mentioned in (2.10) and (2.11) and compute the variability in the lengths of the confidence intervals. Then we compute the confidence intervals in (2.10) and (2.11) again but replacing $\hat{\sigma}_\mu$ or $\hat{\sigma}_b$ by the new estimator $\hat{\sigma}_{vs}$, and also compute the variability in the lengths of those intervals. Throughout, the intervals are two-sided. Bootstrap approximations are based on $B = 299$ resamples, and the approximate coverage probabilities are based on 300 independent samples.

2.4.2 Performance of Variance Stabilization Method on both Undersmoothing and Explicit Bias Correction

For the bias correction confidence interval, let kernel K denote Epanechnikov's kernel

$$K(u) = \begin{cases} (3/4)5^{-1/2}\{1-(u^2/5)\} & \text{if } |u| \leq \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$$

for which $\kappa_r = 1/2$ and $r = 2$ in the bias formula (2.4). To estimate $f^{(2)}$ we use the $(2,4)th$ order optimal kernel suggested by Gasser, Muller, and Mammitzsch given by

$$L^{(2)}(u) = \begin{cases} (105/16)(6u^2 - 5u^4 - 1) & \text{if } |u| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

We take $h = 1.05\hat{\tau}n^{-1/5}$ where $\hat{\tau}^2$ denotes sample variance and $h_1 = 2.70\hat{\tau}n^{-1/9}$, which are asymptotically optimal if the data are normal.

For the under-smoothing confidence interval, we take K to be the optimal fourth order kernel defined by

$$K_1(u) = \begin{cases} (15/32)(7u^4 - 10u^2 + 3) & \text{if } |u| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

When the true density f is that of an $N(0,\tau^2)$ distribution, the bandwidth which asymptotically minimizes mean integrated squared error is $5.05\tau n^{-1/5}$. This suggests the empirical version, $5.05\hat{\tau}n^{-1/5}$, where $\hat{\tau}^2$ denotes sample variance. To assess the effect of under-smoothing we take $h = 5.05c\hat{\tau}n^{-1/5}$, where $0 < c \leq 1$. The numerical performance is summarized in Table 2.2 and Table 2.3.

In Table 2.2, the simulation is based on the distribution $N(0,1)$ and sample size is 50 by using R [28]. Coverage probabilities is p; w is the interval widths and s is the width

standard deviations. The first three row in the first block shows the performance of bias correction without specifying the value of c. From block 2 to block 7, it shows the outcomes of under-smoothing by assigning different values of c of nominal 95% confidence intervals constructed by the methods described above. In the table, at each point x, left column is the number generated by (2.10) and the column on the right is the number generated by the variance stabilization methodology. From Table 2.2, for the bias correction method, it seems like the variance stabilization method does not improve the standard deviation of the interval width. However, one can see that as c becomes smaller and smaller, the standard deviation of interval length is much stable than the original method especially when $c = 0.1$ even though we sacrifice some accuracy on the coverage probabilities.

In table 2.3, we use the normal mixture $1/2 \times N(0,1) + 1/2 \times N(3,1)$. From both bias correction and variance stabilization method, we can see that the variance of the interval length is smaller than the original method which explained the proposal we show in the previous sections.

Table 2.2

Simulation on N(0,1)

N(0,1) n=50							
		x=0		x=0.75		x=1.5	
	p	87.6	83.6	87	84	89.6	82.3
	w	0.17577	0.17318	0.1758	0.17479	0.15137	0.14287
	s	0.02067	0.02111	0.02042	0.02233	0.01962	0.01913
c=1	p	26.7	33.3	90	92	71	63
	w	0.07755	0.0739	0.09762	0.09559	0.10841	0.10772
	s	0.01122	0.0105	0.01449	0.01444	0.01141	0.00986
c=0.75	p	79	83.3	97.6	95.3	89	84.3
	w	0.13539	0.13032	0.14704	0.1444	0.14035	0.13759
	s	0.01745	0.01564	0.01911	0.01819	0.01382	0.01455
c=0.5	p	94.6	92.3	94	93	97.3	91.7
	w	0.22945	0.22365	0.22666	0.22158	0.18453	0.1749
	s	0.02881	0.0264	0.02797	0.02627	0.02332	0.02534
c=0.3	p	95.3	93.6	95	92.3	96.3	89.6
	w	0.36136	0.34865	0.33522	0.32134	0.28054	0.22691
	s	0.04773	0.04901	0.04582	0.04353	0.07309	0.05256
c=0.2	p	96.7	95	97.3	94.6	93.3	88.6
	w	0.4899	0.45588	0.44859	0.40267	0.44381	0.27811
	s	0.0684	0.06774	0.07159	0.06809	0.27583	0.07943
c=0.1	p	96.3	91	95.6	88.3	82.6	79.7
	w	0.86331	0.67029	0.89485	0.58228	1.51034	0.36371
	s	0.22732	0.15289	0.39229	0.16388	5.85617	0.17278

Table 2.3

Simulation on $1/2 \times N(0,1) + 1/2 \times N(3,1)$

1/2*N(0,1)+1/2*N(3,1) n=50							
		x=0		x=0.75		x=1.5	
	p	88.6	87.3	41	33.3	46.3	37.3
	w	0.13922	0.13859	0.13707	0.13827	0.11276	0.11343
	s	0.1575	0.01532	0.017	0.0162	0.01168	0.01256
c=1	p	50.6	65	19.6	10	0	0
	w	0.0791	0.07907	0.05759	0.05731	0.03916	0.03979
	s	0.00707	0.00757	0.00785	0.00805	0.00655	0.00608
c=0.75	p	86.3	85.7	58	40.3	0	0
	w	0.11296	0.1123	0.10018	0.09961	0.08914	0.08854
	s	0.00917	0.00923	0.0102	0.01111	0.00951	0.00941
c=0.5	p	92.7	88	70.6	60.7	71.3	61
	w	0.15602	0.15624	0.14912	0.15001	0.15016	0.15273
	s	0.01314	0.01329	0.01562	0.0155	0.01561	0.01607
c=0.3	p	97.3	94.3	90.3	89.3	87	86
	w	0.26024	0.2465	0.26518	0.25915	0.21942	0.21479
	s	0.03483	0.03483	0.0355	0.0356	0.02797	0.02789
c=0.2	p	97.3	92.6	92	91.3	92.3	92.7
	w	0.35368	0.31195	0.35143	0.3176	0.29307	0.26317
	s	0.06619	0.06022	0.06133	0.05794	0.0465	0.0467
c=0.1	p	94	89.3	94	90.6	93.7	91.3
	w	0.81412	0.44217	0.96729	0.13139	0.63585	0.35949
	s	1.53151	0.13322	2.83142	0.4445	0.49083	0.10488

CHAPTER 3

CONFIDENCE INTERVAL ON HAZARD RATE

3.1 Introduction

A variety of methods of constructing nonparametric confidence intervals for the hazard rate have been discussed in the literature. See for example, Cheng and Hall (2006) [10],who provide the nonparametric confidence intervals for hazard rate function under censorship model and references in there. The main issues discussed in those articles include choice of smoothing parameter, coverage accuracy and the length of a confidence intervals. Here, however, we confine our attention to the length of confidence interval of hazard rate.

In the last chapter, we discussed the effect of an estimator of the standard error of $\hat{f}$ which has the constant variance, on the length of the nonparametric confidence intervals for f obtained by bias correction and under smoothing methods. In this chapter, our primary aim is to carry out similar investigation for the nonparametric confidence intervals for hazard rate function. For that, first we propose an estimator of square root of the hazard

rate which will have a constant variance. For that, in section 3.2, we consider the following two estimators of hazard rate functions

$$\hat{\lambda}_1(x) = \frac{\hat{f}(x)}{1 - F_n(x)} \tag{3.1}$$

$$\hat{\lambda}_2(x) = \sum_{i=1}^{n} \frac{\frac{1}{h}K(\frac{x-X_i}{h})}{n-i+1} \, . \tag{3.2}$$

For their motivation see [29]. We then use regression approach to obtain an estimator of $\sqrt{\lambda}$ and then the confidence interval of hazard rate. In section 3.3, we introduce the hazard rate confidence interval proposed by Cheng and Hall (2006) [10]. In the section 3.4, we show the performance of the confidence intervals which use constant variance estimator of the standard deviation of the hazard rate estimator.

3.2 An Estimator of Square Root of Hazard Rate

In section 2.3, we have discussed root-unroot transformation in nonparametric regression context. There by using the local linear regression, we applied root-unroot transformation to estimate $\sqrt{f}$ which is in the standard error term the nonparametric kernel density estimator. We now apply this methodology to estimate the standard error of hazard rate estimator. For that note that asymptotic variance of the either estimator in (3.1) or (3.2) is given by

$$\frac{1}{nh}\frac{\lambda(x)}{1 - F(x)} \, .$$

Notice that, n is the number of observations, h is the bandwidth and F is the cumulative distribution function which can be estimated by the empirical distribution function. Therefore standard error will depend on estimate of $\sqrt{\lambda(x)}$, and hence here we will obtain an

estimate of $\sqrt{\lambda(x)}$ with constant variance. In section 1.4.4, we have discussed the hazard rate estimation by using local linear fit. Define

$$Y_i = \frac{1}{\Delta}\frac{f_i}{N - \sum_{j=1}^{i} f_j + 1}, \quad i = 1, 2, \cdots, n \tag{3.3}$$

where Δ is the length of the interval and f_i is the number of observations in ith interval. N is the total counts of observations. We have shown that

$$E(Y_i) \simeq \lambda(x_i)$$
$$Var(Y_i) \simeq \frac{1}{N\Delta}\frac{\lambda(x_i)}{1 - F(x_i)} .$$

However, in order to take advantage of the root-unroot transformation to stabilize the variance of the confidence interval length, one need to normalize Y_i in (3.3). After the normalization the new estimator will be similar to the density estimator discussed in the chapter 2. It is mentioned in section (1.1) that

$$\Lambda(x) = \int_0^x \lambda(t)dt = -\log(1 - F(x)) . \tag{3.4}$$

If we let $\Lambda(x)$ be the integral of hazard rate on the entire data range, say $[0, x_{max}]$ then we can rewrite $\Lambda(x_{max}) = C_h$ where C_h is a constant. Define

$$\begin{aligned} H_i &= \sqrt{Y_i/C_h + 1/4} \\ &= \sqrt{\frac{1}{\Delta}\frac{f_i}{N - \sum_{j=1}^{i} f_j + 1}\frac{1}{C_h} + \frac{1}{4}} \\ &= \sqrt{G_i + \frac{1}{4}} \end{aligned} \tag{3.5}$$

where

$$G_i = \frac{1}{\Delta} \frac{f_i}{N - \sum_{j=1}^{i} f_j + 1} \frac{1}{C_h}$$

is the normalized hazard rate estimator. Apply local linear nonparametric regression to the binned center x_i and transformed data H_i and note that

$$E(H_i) \simeq \sqrt{\frac{\lambda(x_i)}{C_h}}, \quad Var(H_i) = \sigma^2(x_i).$$

One can show that asymptotically,

$$Var(H_i) = \sigma^2(x_i)$$

and it is constant. In this fixed design nonparametric regression problem, we have data (x_i, H_i) such that

$$H_i = m(x_i) + \sigma(x_i)\epsilon_i, \;\; i = 1, \cdots, n$$

where

$$m(x_i) = \sqrt{\frac{\lambda(x_i)}{C_h}}$$

and ϵ_i is independent r.v. with mean 0 and unit variance. Now let $\hat{m}(x_0)$ be usual local linear estimator of $m(x_0)$ then

$$\hat{m}(x_0) = \widehat{\sqrt{\frac{\lambda(x_0)}{C_h}}}\,.$$

Notice that C_h is a constant and the estimator of $\hat{\lambda}$ can be written as

$$\widehat{\sqrt{\lambda(x_0)}} = \hat{m}(x_0) * \sqrt{C_h}$$

where $\hat{m}(x_0)$ is local linear estimator at x_0 mentioned in section 1.4.2. Similarly to what we discussed in (2.17). One can show that

$$\widehat{Var}\big(\sqrt{\lambda(x_0)}\big) \simeq \frac{1}{nh}\int K^2(z)dz\ \sigma^2(x_0) * C_h\ .$$

3.2.1 Estimation of the Standard Deviation of Nonparametric Hazard Rate Estimator

Recall that asymptotically, standard deviation of either $\hat{\lambda}_1$ or $\hat{\lambda}_2$ is determined by

$$\sqrt{\frac{\lambda(x)}{1-F(x)}}\ . \tag{3.6}$$

Thus instead of obtaining a constant variance estimator of $\sqrt{\lambda(x)}$, we propose a constant variance estimator of

$$\sqrt{\frac{\lambda(x)}{1-F(x)}}\ .$$

Recall the empirical estimators of f and F by using the local linear fit in section 1.4.4. At point x_i, we have

$$\hat{f}(x_i) = \frac{f_i}{N\Delta},\ \ \hat{F}(x_i) = \sum_{j=1}^{i}\frac{f_j}{N}\ ,$$

so one can write the estimator of (3.6) at x_i as following

$$\begin{aligned}
\sqrt{\frac{\hat{\lambda}(x_i)}{1-\hat{F}(x_i)}} &= \sqrt{\frac{\hat{f}(x_i)}{(1-\hat{F}(x_i))^2}} \\
&= \sqrt{\frac{f_i/N\Delta}{(1-\sum_{j=1}^{i}\frac{f_j}{N})^2}} \\
&\simeq \sqrt{\frac{f_i N/\Delta}{(N-\sum_{j=1}^{i}f_j+1)^2}} \\
&= \sqrt{S_i}\ ,
\end{aligned} \tag{3.7}$$

where

$$S_i = \frac{f_i N/\Delta}{(N - \sum_{j=1}^{i} f_j + 1)^2}$$

and it is an empirical estimator of

$$\frac{\lambda(x)}{1 - F(x)} .$$

Intuitively, we want to normalize the empirical estimator term in (3.7) to utilize variance stabilization transformation. Thus define

$$\begin{aligned} \beta(x) &= \int_0^x \frac{\lambda(t)}{1 - F(t)} dt \\ &= \int_0^x \frac{f(t)}{(1 - F(t))^2} dt \\ &= \frac{1}{1 - F(x)} . \end{aligned} \tag{3.8}$$

Suppose the integral above is on the data range from 0 to x'_{max}, hence $\beta(x'_{max})$ is a constant and let $C'_h = \beta(x'_{max})$. In order to apply it to the local linear regression, set

$$\begin{aligned} D_i &= \sqrt{S_i/C'_h + 1/4} \\ &= \sqrt{\frac{f_i N/\Delta}{(N - \sum_{j=1}^{i} f_j + 1)^2} \frac{1}{C'_h} + \frac{1}{4}} \\ &= \sqrt{E_i + \frac{1}{4}} , \end{aligned} \tag{3.9}$$

where

$$E_i = \frac{f_i N/\Delta}{(N - \sum_{j=1}^{i} f_j + 1)^2} \frac{1}{C'_h}$$

is the normalized estimator of

$$\frac{\lambda(x)}{1 - F(x)} .$$

Therefore, one can use the data (x_i, D_i) where x_i is the binned center and D_i is the transformed count to fit the local linear nonparametric regression. Note

$$E(D_i) \simeq \sqrt{\frac{\lambda(x_i)}{1-F(x_i)}\frac{1}{C_h'}}, \quad Var(D_i) = \sigma^2(x_i),$$

and asymptotically one can show that $Var(D_i)$ is constant. For this local linear regression, one can write

$$D_i = m(x_i) + \sigma(x_i)\epsilon_i, \;\; i = 1, \cdots, n$$

where

$$m(x_i) = \sqrt{\frac{\lambda(x_i)}{1-F(x_i)}\frac{1}{C_h'}}$$

and ϵ_i is independent r.v. with mean 0 and unit variance. Then the local linear estimator of $m(x)$ is $\hat{m}(x_0)$ and

$$\hat{m}(x_0) = \widehat{\sqrt{\frac{\lambda(x_i)}{1-F(x_i)}\frac{1}{C_h'}}}\,.$$

Thus, the estimator of (3.6) is

$$\widehat{\sqrt{\frac{\lambda(x_i)}{1-F(x_i)}}} = \hat{m}(x_0) * \sqrt{C_h'}, \tag{3.10}$$

and the variance of (3.10) is a constant.

3.3 Hazard Rate Confidence Interval

3.3.1 Hazard Rate Estimator

For the hazard rate confidence interval estimation, a number of methods have been discussed in the literature. Even if we confine our attention to smoothing, there still exists a wide range of different methods. We focus our attention to the recently proposed method which forms the basis for our investigation.

Let X denote the true survival time and write f and F for the density and distribution functions, respectively, of X. The hazard rate is given by

$$\lambda(x) = \frac{f(x)}{1 - F(x)} \, .$$

Also, let $\hat{F}$ be the standard empirical estimator of F, computed from the data $x_1, \cdots, x_n$. Then writing h for a bandwidth and taking the kernel, K, to be a bounded, compactly supported, symmetric probability density, we define

$$\hat{\lambda}(x) = \frac{1}{h} \sum_{i=1}^{n} K\Big(\frac{x - X_{(i)}}{h}\Big) \frac{1}{n - i + 1} = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big) \frac{1}{1 - \hat{F}(X_i) + n^{-1}} \quad (3.11)$$

to be our hazard estimator.

3.3.2 Hazard Confidence Interval

In nonparametric estimation of density confidence interval described in the last chapter, we focus on two approaches, namely explicit bias correction and under-smoothing to reduce the effects of bias. To construct the confidence interval for the hazard rate, exactly similar approach can be followed.

In addition, Hall (1992) [17] showed that under-smoothing generally gives better performance than bias correction. To tackle the problem of choosing the extent of under-smoothing, Cheng and Hall (2006) [10] suggest a simple, practical method for locally-adaptive bandwidth choice.

Assume that $t \in (0, T]$ and if K is standardized so that $\int K^2 = 1$. It has been shown in the literature that for the estimator define in (3.11)

$$E\big(\hat{\lambda}(x)\big) = \lambda + O(h^2) \tag{3.12}$$

$$\mathit{Var}\big(\hat{\lambda}(x)\big) = \frac{1}{nh}\frac{\lambda(x)}{1-F(x)} + o(\frac{1}{nh})\,. \tag{3.13}$$

Muller and Wang (1990) [27] show that the distribution of $(\hat{\lambda} - E(\hat{\lambda}))/\sqrt{\mathit{var}(\hat{\lambda})}$ converges to the standard normal as n increases. Let Φ denote the standard normal distribution function and assume $0 < \alpha < 1/2$. We can define $\Phi(z_{\alpha/2}) = 1 - (1/2)\alpha$. The two sided confidence interval for $\lambda(x)$ discussed above can be written as

$$J_{hazard} = (\hat{\lambda} - \frac{1}{\sqrt{nh}}\sqrt{\frac{\hat{\lambda}(x)}{1-\hat{F}(x)}}z_{\alpha/2},\ \ \hat{\lambda} + \frac{1}{\sqrt{nh}}\sqrt{\frac{\hat{\lambda}(x)}{1-\hat{F}(x)}}z_{\alpha/2}), \tag{3.14}$$

with nominal coverage $1 - \alpha$. It has been discussed that the coverage error of J_{hazard} is minimized by choosing the bandwidth, h, to be of size of $n^{-1/3}$. Hence, one can treat bandwidth $h_{opt} = Bn^{-1/3}$ where B is a constant. If we consider both bias-ignored and bias-corrected version of confidence intervals discussed in the literature, given a bandwidth, h, these two confidence intervals with coverage of $(1-\alpha)100\%$ can be written as

$$(\hat{\lambda} - \frac{1}{\sqrt{nh}}\sqrt{\frac{\hat{\lambda}(x)}{1-\hat{F}(x)}}z_{\alpha/2},\ \ \hat{\lambda} + \frac{1}{\sqrt{nh}}\sqrt{\frac{\hat{\lambda}(x)}{1-\hat{F}(x)}}z_{\alpha/2}), \tag{3.15}$$

$$(\hat{\lambda} - \widehat{Bias} - \frac{1}{\sqrt{nh}}\sqrt{\frac{\hat{\lambda}(x)}{1-\hat{F}(x)}}z_{\alpha/2},\ \ \hat{\lambda} - \widehat{Bias} + \frac{1}{\sqrt{nh}}\sqrt{\frac{\hat{\lambda}(x)}{1-\hat{F}(x)}}z_{\alpha/2}), \tag{3.16}$$

where $\widehat{Bias}$ is the bias estimator introduced by Muller and Wang (1990) [27]. Now to stabilize the variance of the confidence interval length. From either (3.15) or (3.16), the length is only related to the standard error term

$$\frac{1}{\sqrt{nh}}\sqrt{\frac{\hat{\lambda}(x)}{1-\hat{F}(x)}}\,. \tag{3.17}$$

Thus now can get confidence intervals if we replace

$$\sqrt{\frac{\lambda}{1-F}} \quad \text{by} \quad \frac{\widehat{\sqrt{\lambda}}}{\sqrt{1-F_n}}$$

obtaining the confidence interval

$$(\hat{\lambda}-\frac{1}{\sqrt{nh}}\frac{\widehat{\sqrt{\lambda(x)}}}{\sqrt{1-\hat{F}(x)}}z_{\alpha/2},\ \hat{\lambda}+\frac{1}{\sqrt{nh}}\frac{\widehat{\sqrt{\lambda(x)}}}{\sqrt{1-\hat{F}(x)}}z_{\alpha/2})\,, \tag{3.18}$$

or if we replace

$$\sqrt{\frac{\lambda}{1-F}} \quad \text{by} \quad \widehat{\sqrt{\frac{\lambda}{1-F}}}$$

obtaining the confidence interval

$$(\hat{\lambda}-\frac{1}{\sqrt{nh}}\sqrt{\frac{\widehat{\lambda(x_i)}}{1-F(x_i)}}z_{\alpha/2},\ \hat{\lambda}+\frac{1}{\sqrt{nh}}\sqrt{\frac{\widehat{\lambda(x_i)}}{1-F(x_i)}}z_{\alpha/2})\,. \tag{3.19}$$

In the next section, we perform the simulation by using the three confidence intervals mentioned above in (3.15), (3.18) and (3.19) to show the effect of using constant estimator on the variability of the length of the confidence interval.

3.4 Numerical Results

In this section, we report the numeric performance of the variance stabilization in the nonparametric hazard confidence interval discussed above. In order to show the property

of variance stabilization in the confidence interval length, we first discussed the confidence interval introduced by Cheng and Hall [10] in (3.15) denoted as M1 and then compare to both confidence intervals of variance stabilization on the standard error introduced in (3.18) and (3.19) denoted as M2 and M3 respectively. The coverage of confidence intervals is based on nominal 95% with $z_{0.025}$ and 1000 simulated sets of data with sample size 200.

For the $\hat{\lambda}$ described in (3.15), we use Epanechnikov kernel mentioned in section 2.4.2 and choose the bandwidth to be of size $n^{-1/3}$. Consider the binned data when applying nonparametric local linear regression and let f_n be the number of observations in the last bin. Then one can use f_n/N to estimate $1 - F(x)$ getting the constant C_h and C_h' in (3.5) and (3.9) respectively. The simulation performance is summarized in Table 3.1 and Table 3.2. In both tables, p is the coverage probabilities; w is the average interval width based on 1000 simulated data sets and s is the standard deviation of the interval length. M1, M2 and M3 are the three confidence interval methodologies mentioned in (3.15), (3.18) and (3.19) respectively.

In table 3.1, the result is based on exponential distribution with $\lambda = 0.8$. At these four points $(x = 0.6, x = 0.8, x = 1, x = 1.2)$, p, s and w have the same trend from M1 to M3. The initial method (3.15) has a very nice coverage probabilities compared to the two other confidence intervals. Since we apply the variance stabilization transformation to the standard error in the confidence interval, M2 and M3 have a smaller standard deviation on the length of the confidence intervals and also the average lengths of the confidence interval are shorter than M1. In table 3.2, the performance is from weibull distribution with scale parameter equals 0.5 and shape parameter equals 5. The conclusion we have

from Table 3.2 is very similar to Table 3.1 that is M2 and M3 have a better performance on the variance of the interval length and both intervals are shorter than the one from M1. However, from both Table 3.1 and Table 3.2, one can see that we sacrifice some coverage probabilities to stabilize the confidence interval length by using the variance stabilization transformations.

Table 3.1

Simulation on Exponential Distribution

Exponential Distribution						
	x=0.6			x=0.8		
	M1	M2	M3	M1	M2	M3
p	94.3	93.4	90.5	95	93.4	93
s	0.03531	0.02594	0.02411	0.04321	0.03036	0.02962
w	0.54573	0.45906	0.36006	0.62861	0.49335	0.37431
	x=1.0			x=1.2		
	M1	M2	M3	M1	M2	M3
p	95	94.7	87.3	95	92	88.3
s	0.05389	0.03767	0.03734	0.06269	0.04481	0.04291
w	0.68153	0.52852	0.40942	0.73728	0.56837	0.45025

Table 3.2

Simulation on Weibull Distribution

Weibull Distribution						
	x=1.2			x=1.4		
	M1	M2	M3	M1	M2	M3
p	94.7	93.8	92.1	95	92.7	91.2
s	0.16947	0.04917	0.04656	0.18373	0.06007	0.06214
w	0.98002	0.81919	0.65624	1.43731	1.01628	0.87375
	x=1.6			x=1.8		
	M1	M2	M3	M1	M2	M3
p	95	94	94	95	80.6	84.7
s	0.22646	0.09087	0.08442	0.38313	0.17468	0.16385
w	2.11295	1.36125	1.27534	3.33795	2.07941	2.07328

CHAPTER 4

NORTH AMERICAN SNOW TRENDS

4.1 Introduction to Snow Trend Estimation

The second part of this thesis provides a statistical study of temporal trends in daily snow depths in a gridded data set covering the United States and Canada. Snow is an vital environmental and geophysical quantity and is sensitive to climate change since its magnitude depends on both temperature and precipitation [4] [20]. Climate change is believed to be most prominent during the cold season in the mid and high latitudes [25]. Global climate models indicate that snow cover changes will considerably impact the cryopsheric portion of the water budget [3].

A daily snow depth is the measured (or estimated) depth of the snow pack at a particular location, whereas snow cover refers to the regions in the study that are not snow-bare. Satellite data over the northern hemisphere suggest that snow cover has lessened since the mid-1980s [33] [35] [34] [16] [19]. It is not known how this influences snow depths. Snow depth analyses complement snow cover change studies, providing further information on hydrological resources, surface energy, soil processes, and ecological systems [13]. Snow depth trend estimates over Canada have been previously made [8] and related to climate variability [1] [9]; however, no statistically detailed study of the daily snow depths in the US and Canada as a whole has been conducted.

Changepoint issues arise and are important in snow trend studies. A changepoint is a time where the structural pattern of a time series shifts, often attributed to station location moves or instrumentation (measuring) changes. Changepoints can profoundly affect temperature and snow trend estimates [39]. While our data has been quality checked for outliers and scaling issues [34], changepoints and their effects will still play a key role in trend inferences. Mitchell (1953) [26] estimates that US stations experience six location or instrumentation changes per century on average. Any detailed snow trend study account should consider changepoint (inhomogeneity) effects. This paper represents a first attempt to include changepoints into a region-wide assessment of snow depth trends.

Station moves and instrumentation change times are not included in our data; these times need to be estimated before making trend inferences. Estimation of multiple changepoint times via genetic algorithms (GA) has become increasingly popular in climate homogenization pursuits [22]. Here, a GA is applied to the snow depth time series at each data grid to estimate times where changepoints occurred. These changepoints are subsequently used in a stochastic storage model to estimate snow depth trends.

Estimation of daily snow depth trends is difficult for several reasons. First, snow is seasonal, mostly absent during the summer months at all but Arctic or high alpine locations. Second, daily snow depths are highly correlated in time: tomorrow's snow depth depends on today's snow depth. Statistical inferences that ignore correlation will yield artificially high levels of confidence [24]. Third, as pointed out by Marsh (1999), snow depths cannot be negative — this "zero modified support set issue" must be addressed. Fourth, as will become apparent here, one should allow for changepoints in trend analyses.

The storage model approach used here was introduced in Woody et al. (2009) [39] and stems from queuing theory. Snow is a natural storage phenomena: the snow depth today is the snow depth yesterday, plus any new snow that has fallen, minus any melt off or compaction. Perona et al. (2007) [30] also use storage models to describe inter-annual snow depths, but do not allow for trends, melting in the fall and winter, or snow increases during spring ablation.

4.2 The Data

Snow depths for each grid were generated from US and Canadian daily snow depth observations taken from the US National Climatic Data Center (NCDC; `http://www.ncdc.noaa.gov`) and the Meteorological Service of Canada [6]. Further information on these data sources, including station distribution, station density, and measurement procedures, can be found in Dyer and Mote (2006) [13]. The data were quality controlled via the methods in Robinson (1989) [32], which involved comparing daily snow depths with associated extreme daily snow depths, snow fall, maximum and minimum temperature, and precipitation data for the associated state or province. In this pursuit, extremes were based on existing quality controlled data (Environment Canada, Canadian climate normal or averages 1971-2000, available at `http://climate.weatheroffice.ec.gc.ca/climate_normals/index_e.html`; National Climatic Data Center, Local climatological data annual summary, `http://www.ncdc.noaa.gov`). Any snow depth that did not meet all quality control criteria were flagged as inconsistent and excluded in our analysis.

To maintain geographic consistency in the data, a spherically-based inverse-distance algorithm was used to interpolate depths before projecting onto a Cartesian plane (Willmott et al., 1984) [38]. Our algorithm uses a maximum of 25 different station observations to calculate snow depths at each grid and time, with a maximum search radius of 100 km from each grid center. If less than 25 station observations were available within the 100 km radius, only those depths within the radius were used for interpolation. Should no observations reside within the 100-km search radius, the nearest available observation was used as the snow depth at that grid point and time. This approach is advantageous in high station density areas, such as the Eastern US, where the search radius needed to find 25 stations is typically small. In areas with lower station density, such as Central and Northern Canada, the 100-km maximum search radius minimizes biases from using non-local observations. Our interpolation algorithm was used to construct daily 0.250×0.250 degree longitude by latitude snow depth grids over an extent of $530 - 1680$ W longitude and $200 - 710$ N latitude, which were then averaged to develop daily snow depth grids. This scheme minimizes the impact of inconsistent observations on interpolated grid depths. The primary disadvantages of this approach include terrain influences within mountainous regions, where snow depth is dependent on slope, aspect, and altitude, and low station density on interpolated values.

4.3 Changepoint Detection via Genetic Algorithm

This section discusses our multiple changepoint homogenization methods. As an example, the methods are applied to a series from a grid centered at latitude 46.5N, Longitude -99.5W (near Wishek, North Dakota).

Genetic algorithms (GA) have been successfully used in recent climate homogenization studies [22] [21]. GAs use principles of genetic selection and biological mutation to intelligently search for the best possible (as defined below) changepoint configuration. While trends in daily depths are our objective, changepoints in yearly average snow depth series are first sought. Homogenization for daily series is significantly more involved due to long series lengths [31]. Any changepoint detected in the yearly series is assigned to the first day of the year.

Let $\{X_t\}_{t=1}^{N}$ denote the snow depth series at a fixed grid for days $t = 1, \ldots, N$. Winter centered years (WCY) are analyzed here. The starting observation for a WCY is 1 July; 30 June of the subsequent calendar year is the last day. WCYs prevent a single winter season from straddling two distinct years. To have 1960 commence our study, $t = 1$ will correspond to 1 July 1959; $t = N = 14,600$ corresponds to 30 June 2000, the study end date. Leap year data are handled by deleting the 30 June observation within the same calendar year that 29 February occurs — this has virtually no impact on results. There are $n = 41$ years in our study.

For a periodic notation, write time t as $t = dT + \nu$ where $d \in \{0, \ldots, n-1\}$ is the WCY and $\nu \in \{1, \ldots, T = 365\}$ represents the day of the WCY. The actual data set spans 1 January 1960 – 31 December 2000.

At a given grid, define $Y_d = \#(\mathcal{W})^{-1} \sum_{\nu \in \mathcal{W}} X_{dT+\nu}$ as the average depth for WCY $d \in \{0, \ldots, n-1\}$, where $\#(A)$ is the number of elements in the set A and $\mathcal{W}$ denotes the snow season for this grid. The snow season is taken to start on the first day in the Fall/Winter on which snow is present on at least 20 percent of the WCY years in the study; the last day of the WCY is the latest day in the Winter/Spring on which at least 20 percent of the WCY years report some snow. Since the entire winter is not observed when $d = 0$ and $d = 40$, Y_0 and Y_{40} are considered missing. For the grid centered near Wishek, North Dakota, the snow season spans 1 Nov to 27 April, a length of 117 days.

We now describe our changepoint homogenization methods. From the yearly averaging (the central limit theorem), Y_d is approximately normally distributed. Hence, the changepoint techniques for normal data from Li and Lund (2012) [22] apply. Trend components must be included in the changepoint detection methodology to avoid identifying spurious changepoints caused by the presence of non-zero trends [15]. Our basic model for the yearly average snow depths $\{Y_d\}_{d=0}^{n-1}$ is the time series regression

$$Y_d = \alpha + \gamma d + \mu_d + \epsilon_d. \tag{4.1}$$

Here, α models the overall mean, γ is a linear trend parameter, μ_d is a changepoint effect described further below, and $\{\epsilon_d\}_{d=0}^{n-1}$ is a zero-mean first order autoregressive (AR(1)) error process that allows for temporal correlation in the yearly averages. The AR(1) errors obey $\epsilon_d = \phi \epsilon_{d-1} + Z_d$, where $\{Z_d\}$ is a zero-mean white noise process with variance σ^2, and $\phi \in (-1, 1)$ is the correlation between consecutive years of average snow depths. The changepoint effect μ_d entails k mean (level) shifts at times $\tau_1 < \tau_2 < \ldots < \tau_k$:

$$\mu_d = \begin{cases} \Delta_0 = 0, & \tau_0 < d < \tau_1 \\ \Delta_1, & \tau_1 \leq d < \tau_2 \\ \vdots & \vdots \\ \Delta_k, & \tau_k \leq d < \tau_{k+1} \end{cases},$$

where $\tau_0 = 0$ and $\tau_{k+1} = n + 1$ by convention. Here, Δ_ℓ is interpreted as the changepoint effect of the ℓth regime.

This model contains the regression parameters $\alpha, \gamma, k, \Delta_1, \ldots, \Delta_k$, the changepoint times $\tau_1, \tau_2, \ldots, \tau_k$, and the time series parameters ϕ and σ^2. For a given configuration of k changepoints occurring at the times $\tau_1, \ldots, \tau_k$, the regression parameters are estimated using standard linear regression methods; Yule-Walker estimators are fitted to linear regression residuals to estimate the time series parameters ϕ and σ^2.

Estimation of the best changepoint configuration of a time series is a statistical model selection problem. Popular model selection criteria include the Akaike Information Criterion (AIC), the Bayesian Information (BIC), and Minimum Description Lengths (MDLs). To date, MDL methods have produced superior empirical results [11] [23]. MDL methods minimize an objective function of form MDL $= -\ln(L_{\text{opt}}) + P$ over all possible changepoint configurations. Here, $-\ln(L_{\text{opt}})$ is an optimal model likelihood given the number of changepoints k and their locations $\tau_1, \ldots, \tau_k$, and P, which depends on the number of changepoints are their location times, is a penalty term to prevent overfitting. Unlike AIC and BIC penalties, MDL penalties are more complex than a multiple of the number of

changepoints. The MDL penalty, in fact, penalizes changepoint times that are relatively closer together more than sparsely spaced changepoints.

The Innovations form of the multivariate Gaussian likelihood is used:

$$L(\phi, \sigma, \alpha, \gamma, \Delta_1, \ldots, \Delta_k) = (2\pi)^{-\frac{N}{2}} \left(\prod_{t=0}^{N-1} v_t \right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \sum_{t=1}^{N} \frac{(Y_t - \hat{Y}_t)^2}{v_{t-1}} \right). \quad (4.2)$$

Here, $\hat{Y}_t$ is the one-step-ahead predictor of Y_t and $v_t = E[(Y_{t+1} - \hat{Y}_{t+1})^2]$ is its mean squared prediction error. For an AR(1) series, $\hat{Y}_t = \hat{m}_t + \hat{\phi}(Y_{t-1} - \hat{m}_{t-1})$, where $\hat{m}_t = \hat{\alpha} + \hat{\gamma} d + \hat{\mu}_d$, and $v_t \equiv \hat{\sigma}^2$. Parameter estimates for ϕ and σ are calculated from the Yule-Walker equations for $\{\hat{\epsilon}_d\}_{d=0}^{n-1}$. Substituting the estimators of $\phi, \sigma, \alpha, \gamma; k, \Delta_2, \ldots, \Delta_{k+1}$ into (4.2) numerically gives $\ln(L_{\text{opt}})$ for the changepoint configuration with k changepoints at locations $\tau_1, \ldots, \tau_k$.

The penalty term for the changepoint configuration is the same as that in [21]:

$$P = P(k, \tau_1, \ldots, \tau_k) = \frac{1}{2} \sum_{\ell=2}^{k+1} \log(\tau_\ell - \tau_{\ell-1}) + \sum_{\ell=2}^{k} \log(\tau_\ell) + \log_2(k+1).$$

The best changepoint model is estimated as the one that minimizes the MDL score. As an exhaustive search for the best changepoint configuration(s) requires evaluation of 2^{N-1} MDL scores, an arduous task on the world's fastest computers for even moderately large N. Here, a GA was devised to perform the minimization. GAs have recently been employed for multiple changepoint detection in climate data [22] [21]. See Davis (1991) [12] for more on genetic algorithms.

As an example, when these methods are applied to the Wishek, North Dakota grid, one changepoint is declared in 1969. Panel 1 of Figure 1 graphically depicts the fit to $\{Y_d\}_{d=0}^{n-1}$.

4.4 The Storage Model

Woody et al. (2009) [39] introduce a storage model to estimate trends in daily snow depths in the presence of changepoints. This model was fitted to a time series from Napoleon, North Dakota. For each fixed grid, estimates of the number of mean shifts k and their times $\tau_1, \ldots, \tau_k$ are now in place. Our model for the daily snow depths is based on the storage equation

$$X_t = \max\{X_{t-1} + C_t, 0\},$$

where C_t quantifies the random change in the snow pack from day $t-1$ to day t. The $\max$ term prevents the snow depths from becoming negative. Here, the depth change C_t is assumed statistically independent of $X_1, \ldots, X_{t-1}$.

For features, we posit that $E[C_t] = m_t$ has a periodic component, possibly multiple changepoints, and a linear trend. The variance $\text{Var}(C_t) = w_t^2$ is assumed periodic with period $T = 365$. For computational convenience, C_t is assumed normally distributed: $C_t \sim N(m_t, w_t^2)$; other distributions can be easily used if desired. Specifically, m_t is parameterized as

$$m_t = P_\nu \left[A + B\cos\left(\frac{2\pi(\nu - \zeta)}{T}\right) + \delta_t + \gamma t \right],$$

where the mean shifts are estimated as in the last section:

$$\hat{\delta}_t = \begin{cases} 0, & t \in \{1, \ldots, \hat{\tau}_1 T - 1\} \\ \hat{\Delta}_1, & t \in \{\hat{\tau}_1 T, \ldots, \hat{\tau}_2 T - 1\} \\ \vdots & \vdots \\ \hat{\Delta}_{\hat{k}}, & t \in \{\hat{\tau}_{\hat{k}} T, \ldots, dn\} \end{cases}.$$

The term P_ν is a zero/one indicator that is unity when $\nu \in \mathcal{W}$.

To estimate the storage model parameters (this is not the same task as estimating changepoint parameters), we minimize the weighted sum of squared one-step-ahead prediction errors

$$S(\boldsymbol{\theta}) = \sum_{d=0}^{n-1} \sum_{\nu=1}^{T} \frac{(X_{dT+\nu} - \hat{X}_{dT+\nu})^2}{w_\nu^2}.$$

Here, $\boldsymbol{\theta} = (A, B, \zeta, \gamma, \Delta_2, \ldots, \Delta_{k-1})'$ is a vector containing all model parameters and $\hat{X}_t := E_{\boldsymbol{\theta}}[X_{t+1}|X_t]$ is the one-step-ahead prediction, calculated in Woody et al. (2009) [39] as

$$E_{\boldsymbol{\theta}}[X_{t+1}|X_t] = [X_t + m_{t+1}] \left[1 - \Phi\left(\frac{-X_t - m_{t+1}}{w_{t+1}}\right)\right] + w_{t+1}\phi\left(\frac{X_t + m_{t+1}}{w_{t+1}}\right), \quad (4.3)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution and density functions, respectively, of a standard normal random variable. Because snow is highly seasonal, weighted least squares is used, where the weight w_ν is set to Var(Z_ν) (while it is important to weight, optimal weightings are not needed). To estimate w_ν, first calculate a point estimate of the mean change from day $\nu-1$ to day ν via $\hat{e}_\nu = n^{-1} \sum_{d=0}^{n-1} (X_{dT+\nu} - X_{dT+\nu-1})$. The estimate of w_ν is then simply

$$\hat{w}_\nu^2 = \sum_{d=0}^{n-1} \left[(X_{dT+\nu} - X_{dT+\nu-1}) - \hat{e}_\nu \right]^2 .$$

We smooth these daily variances via the Matlab function "cfit" before using them to minimize variability. The bottom panel in Figure 1 shows w_ν for the Wishek grid. Here, the variance peaks in the late snow season at the end of March. This method appears visually reasonable, capturing the increased volatility in the snow pack in the early and late portions of the snow season.

While the parameter γ controls the trend and is the object of our study, it is not easily interpretable within the storage equation. For example, because of the maximum in the storage equation, $\hat{\gamma} = 50$ does not imply a depth change of 50 units per time. A quantity that does have interpretable units of depth change per time is the linear trend statistic in the presence of changepoints:

$$\hat{\beta} = \frac{\sum_{r=1}^{k} \sum_{n \in \mathcal{H}_r} \sum_{\nu=1}^{T} \left(X_{nT+\nu} - \bar{X}_r(\nu) \right) \left(nT + \nu - \bar{t}_r(\nu) \right)}{\sum_{r=1}^{k} \sum_{n \in \mathcal{H}_r} \sum_{\nu=1}^{T} \left(nT + \nu - \bar{t}_r(\nu) \right)^2}, \tag{4.4}$$

where $\mathcal{H}_r = \{t : \tau_{r-1} T \leq t < \tau_r T\}$ denotes the set of times where the series was in the r^{th} segment (regime), $\bar{X}_r(\nu)$ denotes the average snow depth on day ν of regime r, and $\bar{t}_r(\nu)$ denotes the average time during the r^{th} regime and day ν. See Lund et al. (2001) [24] for more on this statistic. In this context, $\hat{\beta}$ estimates the mean change in the snow depth per unit of time (scaled to cm per century below). To obtain standard errors for $\hat{\beta}$, we simply simulate 1000 independent realizations of the storage model with parameters as estimated from the data for each grid, and then compute the sample mean and standard deviations of

the 1000 trend estimates in (4.4). Simulation is used since no explicit forms exist for the mean of a storage equation model.

At Wishek, parameter estimates are $\hat{A} = -0.5510$, $\hat{B} = 0.7501$, $\hat{\zeta} = 182.96$, $\hat{\gamma} = -0.0638$, and $\hat{\Delta}_1 = 0.4386$. The negative $\hat{\gamma}$ suggests declining snow; the positive $\hat{\Delta}_1$ moved the station to a snowier location in 1969. The thousand simulations of the storage model with these parameters report an average trend statistic of $\hat{\beta} = -1.4904$ cm/Century, with a standard error of 0.7947 cm/Century. This yields a z-score of -1.8756 when testing the null hypothesis that $\beta = 0$ (a one-sided p-value of 0.0304).

4.5 Results

Results are described for two cases: 1) changepoints ignored and 2) changepoints included. The average number of changepoints in the 2,831 grids with a non-empty WCY is 1.593. More specifically, 197 (7.0%) of the grids are changepoint free and 1176 grids (41.5%) have only one changepoint. The maximum number of changepoints in a grid is five (this occurred in six grids). The average grid trend is 3.774 cm/Century with changepoints and 9.742 cm/Century without changepoints. More grids have an increasing trend than decreasing trend: 1690 grids (59.7%) have a positive trend when changepoints are taken into account and 2254 grids (79.6%) have a positive trend when changepoints are ignored. The average trend standard deviation is 0.947 cm/Century when changepoints are taken into account and 1.91 cm/Century when changepoints are ignored. The standard error is smaller when changepoints are accounted for because this component significantly explains much of the data (i.e., a better model is obtained).

Figure 4.2 shows spatially smoothed trends, in cm/Century, with and without changepoints. Elaborating, the $\hat{\beta}$s from each grid are spatially smoothed with the head-banging algorithm with 30 triples. Head-banging is a robust median-polished smoother, capable of identifying rough structure (ridges and boundaries) while eschewing outliers. Figure 4.3 depicts head-banging smoothed z-scores. The z-scores are computed for each grid assuming a null hypothesis that the trend is zero: this is the trend $\hat{\beta}$ at each station divided by the standard error of the trend estimate.

Overall, the estimated snow depth trends over North America without changepoints indicate a general increase from northern Alaska and Western Canada (above roughly 55N) through eastern Canada and into Maine in the US (Figure 2). There are no areas with a strong negative trend, although much of the US shows a slight decrease in snow depths. When changepoints are included in the trend estimates, the primary area of increasing trends is northwestern Canada, centered over The Northwest Territories, with lower positive trends (and some slight negative trends) through central Canada and into the US. More importantly, however, with the inclusion of changepoints, a substantial area of negative trends in western Quebec through Ontario (Figure 2) arises. This is likely a result of the poor station density in this region [13]. The low station density increases sensitivity to individual stations during the generation of the gridded data, which leads to a numerically unstable interpolation if stations have intermittent missing data. Inclusions of changepoints addresses this issue, leading to a more reasonable estimation of trends.

When changepoints are included, the trend estimates show an overall shift of snow depth from east to west across Canada, which could stem from a change in extra-tropical

cyclone tracks and/or intensity over the study period. Such a westward shift of cyclone tracks coming off the North Pacific would lead to an increase in snow depth over the northwestern edges of North America as the systems carry oceanic moisture inland. At the same time, a decrease in snowfall through Eastern Canada would result as the cyclones traverse further into the Great Lakes region of the US. The increasing trends over New England and the far eastern Canadian provinces could be a result of cyclone track variations, specifically a change in frequency and/or intensity of NorEasters. Such an interpretation was offered by Dyer and Mote (2006) [13]; however, it is interesting to note that the results from that analysis, which utilized the same snow depth dataset (albeit with a simpler linear trend estimate), showed somewhat different trend patterns over North America. Additionally, the spatial distribution of areas with significant snow depth trends (p-value less than 0.05) in the Dyer and Mote (2006) [13] study was somewhat different than those found in this study. This lends credence to the necessity for the more robust statistical methodology utilized in this research. The z-scores in Figure 3 provide additional information regarding the spatial distribution of positive and negative snow depth trends across North America. Namely, they highlight the location of relatively weak (though still significant) trends, which is useful in areas where trends exist but are small due to an inherently shallower snow pack. This is especially true in the US, where small changes in the snow pack can have drastic influences on spring and summer water budgets despite having lower overall snow depths relative to regions in northern Canada.

Interpretation of the z-scores across the US without changepoints indicates regionally varying depth trends across the mountainous west, with generally negative z-scores in the

Pacific Northwest (Figure 3). It is difficult to offer definite justification for these patterns, however, as the mountainous regions in Canada show no such pattern. This discrepancy is possibly a result of the greater station density used in the generation of the snow depth grids in the US, as illustrated in Dyer and Mote (2006) [13]. Focusing on the Eastern US, there are large negative z-scores from Tennessee through Pennsylvania, with the exception of weak positive z-scores through the southern Appalachians.

Taking changepoints into account produces high positive z-scores through Northwest Territories and far eastern Canada and into New England, while the highest negative z-scores occur in western Quebec and northern Ontario (Figure 3). Again, the inclusion of changepoints in the trend estimates minimizes the sensitivity of the calculations resulting from interpolation bias related to factors such as station density; therefore, the extent and magnitude of extreme high z-scores is drastically reduced while the extent and magnitude of negative z-scores is increased. Of interest, however, is the area of higher positive z-scores through the central Great Plains of the US and negative z-scores along the East Coast from Virginia through Pennsylvania. Since this area roughly coincides with the southern extent of the continental winter snow pack, the results indicate a decrease in the extent of the snow pack over the study period through coastal locations and an increase through interior continental locations with respect to the maximum seasonal extent of snow cover over North America.

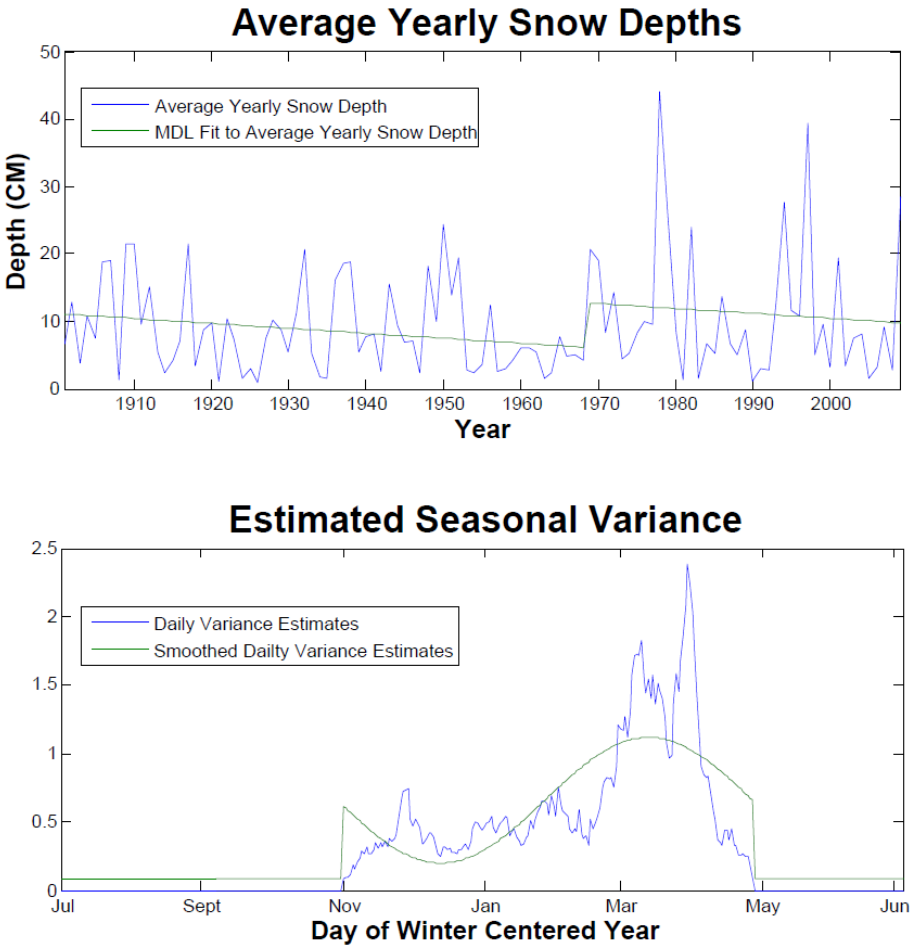
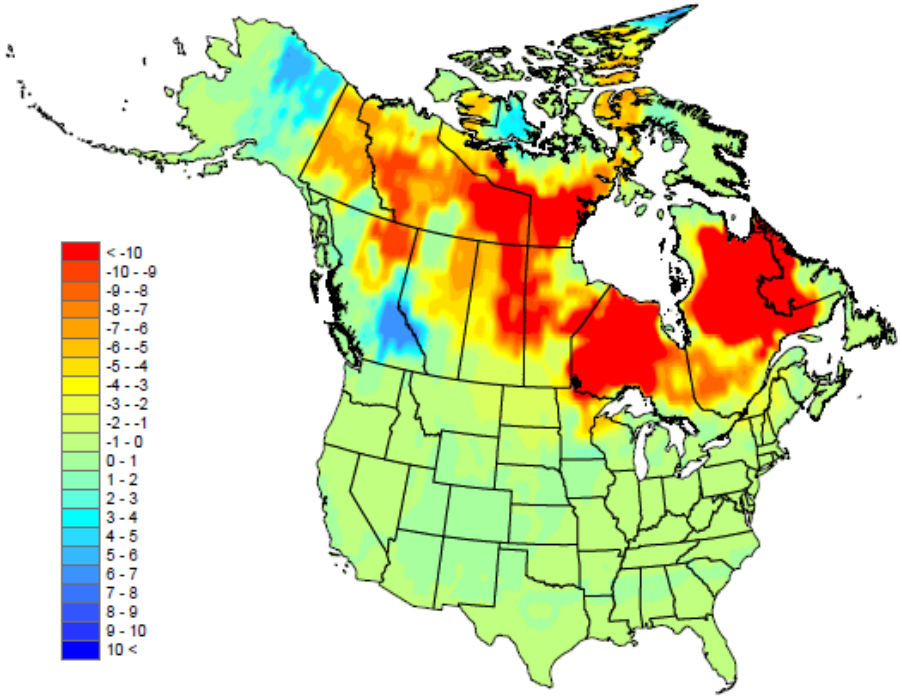
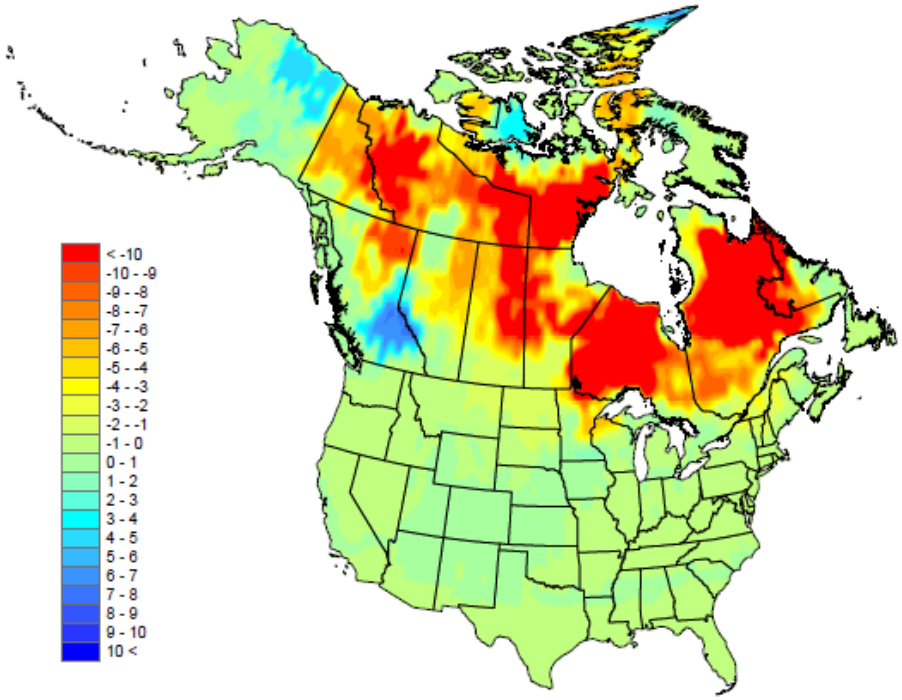


Figure 4.1

The Wishek, North Dakota grid. Top: the MDL fit to the average yearly snow depths. Bottom: seasonal variance estimates.



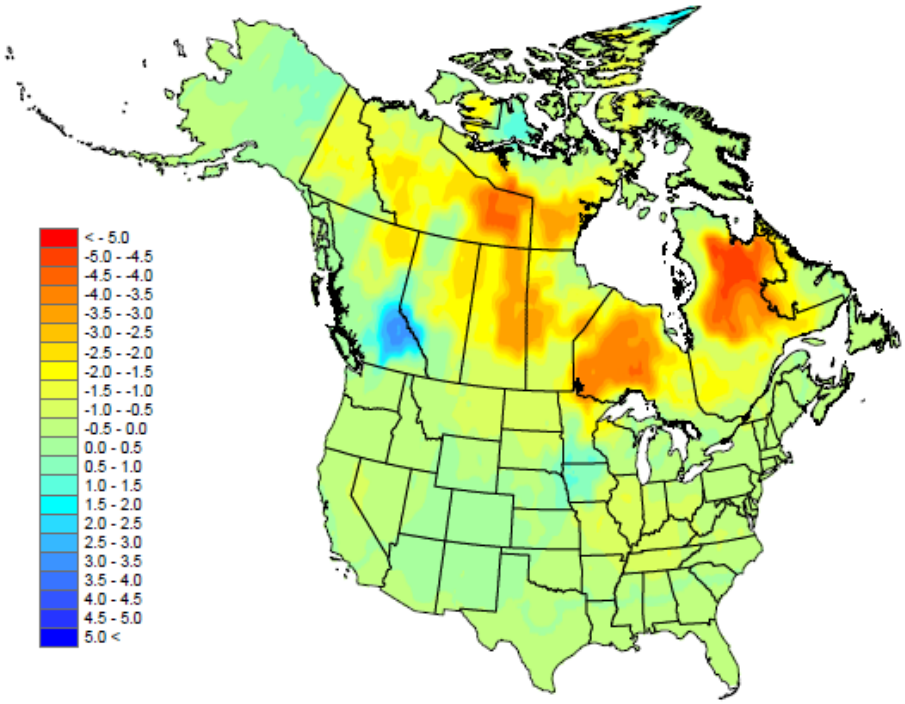
(a) Trend estimates: changepoints ignored.



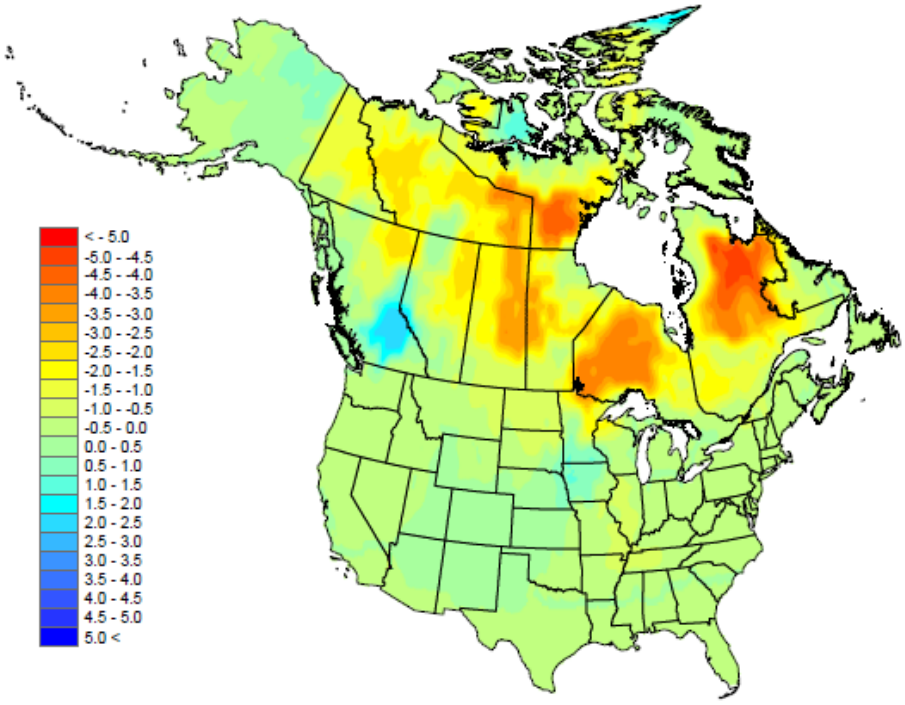
(b) Trend estimates: changepoints taken into account.

Figure 4.2

North America Trend Estimates.



(a) z-score: changepoints ignored.



(b) z-score: changepoints taken into account

Figure 4.3

North America Trend Z-Score Estimates.

REFERENCES

[1] E. Aduado, D. Cayan, L. Riddle, and M. Roose, "Climatic fluctuations and the timing of West Coast streamflow," *J. Clim*, vol. 5, 1992, pp. 1468–1483.

[2] F. Anscombe, "The transformation of Poisson, binomial and negative binomial data," *Biometrika*, vol. 35, 1948, pp. 246–254.

[3] T. Barnett, J. Adam, and P. Lettenmaier, "Potential impacts of a warming climate on water availability in snow dominated regions," *Nature*, vol. 438, no. 303-309, 2005.

[4] R. Barry, "Evidence of recent changes in global snow and ice cover," *GeoJournal*, vol. 20, 1990, pp. 121–127.

[5] M. Bartlett, "The square root transformation in analysis of variance," *J. Roy. Statist. Soc. Suppl.*, vol. 3, 1936, pp. 68–78.

[6] R. Braaten, "An updated Canadian snow depth database, contract report," *Atmos. Environ. Serv.*, 1996.

[7] L. Brown, T. Cai, R. Zhang, L. Zhao, and H. Zhou, "The root-unroot algorithm for density estimation as implemented via waveletblock thresholding," *Probability theory and related fields*, vol. 146, 2010, pp. 401–433.

[8] R. Brown and R.O.Bratten, "Spatial and temporal variability of Canadian monthly snow depths," *Atmos. Ocean*, vol. 36, 1998, pp. 37–45.

[9] D. Cayan, T. Mckee, and N. J. Doesken, "Annual snowpack patterns across the Rockies: Long-term trends and associated 500-mb synoptic patterns," *Mon. Weather Rev.*, vol. 121, 1996, pp. 633–647.

[10] M. Cheng, P. Hall, and D. Tu, "Confidence bands for hazard rates under random censorship," *Biometrika*, vol. 93, 2006, pp. 357–366.

[11] a. T. L. Davis, R.A. and G. Rodriguez-Yam, "Structural break estimation for nonstationary time series models," *J. Amer. Stat. Assoc.*, vol. 101, 2006, pp. 223–239.

[12] L. Davis, *Handbook of Genetic Algorithm*, Van Nostrand Reinhold, 1991.

[13] J. Dyer and T. Mote, "Spatial variability and trends in observed snow depth over North America," *Geophys. Res. Letters*, vol. 33, 2006.

[14] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*, Chapman and Hall, London, 1996.

[15] C. Gallagher, R. Lund, and M. Robbins, "Changepoint detection in climate series with long-term trends," *Journal of Climate*, vol. 26, 2013, pp. 4994–5006.

[16] P. Groisman, T. Karl, and R. Knight, "Changes of snow cover, temperature,and radiative heat balance over the Northern Hemisphere," *J. Clim.*, vol. 7, 1994, pp. 1633–1656.

[17] P. Hall., "Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density," *Ann. Statist.*, vol. 20, 1992, pp. 675–694.

[18] W. Härdle and A. W. Bowman, "Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands," *J. Amer. Statist. Assoc.*, vol. 83, 1988, pp. 102–110.

[19] T. Karl, R. P.Y.Groisman, and R. Jr., "Recent variations of snow cover and snowfall in North America and their relation to precipitation and temperature variations," *J. Glimatol.*, vol. 6, 1993, pp. 1327–1344.

[20] G. J. Kulka, "Climatic role of snow covers, Sea Level, Ice and Climatic Change," *IAHS Publ.*, vol. 131, 1979, pp. 79–107.

[21] J. Lee, S. Li, and R. Lund, "Trends in extreme United States temperatures," *Journal of Climate*, vol. 27, 2014, pp. 4209–4225.

[22] S. Li and R. Lund, "Multiple changepoint detection via genetic algorithms," *J. Clim.*, vol. 25, 2011, pp. 674–686.

[23] Q. Lu, R. Lund, and T. Lee, "An MDL approach to the climate segmentation problem," *Ann. Appl. Stat.*, vol. 4, 2007, pp. 299–319.

[24] R. Lund, P. Seymour, and K. Kafadar, "Temperature trends in the United States," *Environmetrics*, vol. 12, 2001, pp. 673–690.

[25] S. M. and J. Framcos, "The arctic ampli

cation debate," *Climatic Change*, vol. 76, 2006, pp. 241–264.

[26] J. Mitchel, "On the causes of instrumentally observed secular temperature trends," *J. Meteor.*, vol. 10, 1953, pp. 244–261.

[27] H.-G. Müller and J.-L. Wang, "Locally adaprive hazard smoothing," *Prob. Theory. Rel. Fields*, vol. 85, 1990, pp. 523–538.

[28] S. Nadarajah and S. Kotz, "Programs in R for computing truncated t distributions," *Qual. Reliab. Engng. Int.*, vol. 4, no. 3, 2007, pp. 407–412.

[29] P. Patil, "Bandwidth choice for nonparametric hazard rate estimation," *Journal of Statistical Planning and Inference*, vol. 35, no. 1, 1993, pp. 15–30.

[30] P. Perona, P. Porporato, and L. Ridolfi, "A stochastic process for the interannual snow storage and melting dynamics," *Cent.J. Geophys. Res.*, vol. 112, 2007.

[31] H. Priyadarshani, Y. Y.Li, R. Lund, and J. Rennie, "Homogenization of daily temperature data," *Journal of Climate*, 2016.

[32] D. Robinson, "Construction of a United States historical snow data base," *Proc. Eastern Snow Conf.*, vol. 45, 1989, pp. 50–59.

[33] D. A. Robinson, "Northern Hemisphere snow extent drning the satellite era," *Am. Meteorol. Soc.*, 1999.

[34] D. A. Robinson, K. F. Dewey, and R. Heim, "Global snow cover and monitoring: An update," *Bull. Am. Meteorol. Soc*, vol. 74, 1993, pp. 1689–1696.

[35] D. A. Robinson and A. Frei, "A Northern Hemisphere snow cover climatology using satellite information," *Am. Meteorol. Soc.*, 1997.

[36] B. W. Silverman, *Density Estimation for Statistics and Data Anylysis*, Chapman and Hall, London, 1988.

[37] M. Wand and M. Jones, *Kernel Smoothing*, Chapman and Hall, London, 1995.

[38] C. J. Willmott, C. M. Rowe, and W. D. Philpot, "Spheremap," *Cent.For Clim. Res.*, 1984.

[39] J. Woody, R. Lund, and M. T. Grundstein, A., "A storage model approach to the assessment of snow depth trends," *Water Resources Research*, vol. 45, 2009.