Spring 2010

A Nonparametric Method for Ascertaining Change Points in Regression Regimes

Alfreda N. Rogers

A NONPARAMETRIC METHOD FOR ASCERTAINING CHANGE POINTS IN REGRESSION REGIMES

by

ALFREDA N. ROGERS

(Under the direction of Patricia Humphrey)

Abstract

Of interest is the specific model called the joinpoint two regime regression or broken line model composed of one regression line and a horizontal ray. This is a very restricted but highly useful subset of the well-researched change point problem. The usual approach to a more general model was first presented by Quandt (1958) who found the maximum likelihood estimates of the slope, intercept and joinpoint by assuming that the error terms are generated under the usual assumptions, that is, from a normal distribution with constant variance and are uncorrelated. We develop a method that does not rely on this assumption, demonstrate its use on an example of proximity indexes of whale cow and calf pairs, and compare the new method to the Quandt estimates in a simulation study showing this new method performs adequately.

INDEX WORDS: Maximum likelihood, Change point, Moment match

A NONPARAMETRIC METHOD FOR ASCERTAINING CHANGE POINTS IN REGRESSION REGIMES

by

ALFREDA N. ROGERS

B.S., Armstrong Atlantic State University, 2007

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

2010

A NONPARAMETRIC METHOD FOR ASCERTAINING

CHANGE POINTS IN REGRESSION REGIMES

By

ALFREDA N. ROGERS

Major Professor: Patricia Humphrey

Committee:

Martha Abell

Greg Knofczynski

Electronic Version Approved:

May 2010

ACKNOWLEDGEMENTS

What started out as a simple question and blossomed into this thesis, the initial length and quality was greatly underestimated. Without the help of many people along the way, this thesis would not have been possible.

First and foremost, I would like to thank my thesis committee, Dr. Martha Abell, Georgia Southern University; Dr. Patricia Humphrey, Georgia Southern University, and Dr. Greg Knofczynski, Armstrong Atlantic State University. Special thanks are given to Dr. Lorrie Hoffman, Armstrong Atlantic State University. Their efforts and guidance helped me tremendously to endure and complete this thesis. If it were not for their patience, questioning methods, and willingness to work with my mathematical flaws, I would not have made it to this end. I would also like to thank Jaree Hudson, Heather King, and Steve Clark for their previous work using the Hinde methods using actual data from whales in Sea World. Their finished work helped open the door for the question this thesis will answer.

TABLE OF CONTENTS

LIST OF TABLES AND GRAPHS

CHAPTER 1

INTRODUCTION

Somewhere, in the ocean, a mother whale cow has just given birth to a baby whale calf. In the initial stages, one can find the cow closely monitoring her calf. She does so to not only nourish the calf, but also protect it from the many dangers that the calf is too young and naïve to escape. As time passes and the calf gets older, the mother understands that the calf needs to learn how to fend for itself. During a time period one can see that the cow weans the calf slowly, but surely, so it can take its place in the underwater ecosystem. But when is the best time for the mother to start weaning her calf? Is there a point in time that one can expect a cow would completely wean her calf?

There is a natural appeal for a model that relies on a regression equation with a discontinuous derivative function. Thus, the volume of researchers and the totality of their papers number in the hundreds. An annotated bibliography by Khodadadi and Asgharian (2008) includes many of the papers pertinent to our research interest. Ciuperca (2009) has more recent citations in the introduction of his paper. The model we study is a small but useful subset of the "regime(d)" regression, a phrase first used by Quandt (1958) in economic literature, or more currently referred to as a broken-line regression (Gill, 2004). Our interest is in the estimation of all of the straight-line model parameters when exactly two regimes are assumed. The extensive literature discusses solutions pertaining to two or more response functions (either linear or not) defined over their companion independent sets (either univariate or multivariate) with connected or disconnected transition or change point(s), a term first appearing in the literature in the late 1960's, in particular in the 1969 Dagenais paper. More recently for connected

transition points the term "joinpoints" is used (Ghosh, Basu and Tiwari, 2009). Researchers like Koul (2000) study models with either no or few assumptions about error terms. The error terms of the model can be assumed to be normal or non-normal and uncorrelated or correlated. A large number of the papers confine themselves to finding the number of, estimation of, and/or distribution of the change point(s) itself whereas we want to characterize the estimates of all the parameters of our model, a model having two straight-line regression equations with exactly one change point.

Many applications (and here we mention several current ones) lend themselves to analysis under this type of model: quality control (Mahmoud et al., 2006), medical research (Yu et al., 2007), climatology (DeGaetano, 2006) and many other areas. Applications account for the largest percentage of the literature. The balance of the papers propose and/or investigate estimation and testing methods (including maximum likelihood, semi-parametric, non-parametric, bootstrap, Bayesian, wavelet) and their goodness (convergence rates, bias, consistency). We present estimation methods we have derived that are distribution-free and demonstrate our method with an example in the field of biology on a set of paired mother-calf whale data proximity indices collected and aggregated weekly. We first consider the parametric procedure described by Quandt (1958) and updated by him (Quandt, 1960; Quandt, 1972; Goldfeld and Quandt, 1972) since we found assurances by recently published author Gill (2004) of the stable behavior of these estimates for the slope, intercept, and change point for our particular model of interest. The Quandt method has no closed form solutions and is iterative so the process of finding these statistics is computer intensive but due to improvements in processing power it is still in use currently (DeGaetano, 2006). Although the literature offers some

competing approaches to the iterative Quandt method, we found that no authors consider a direct moment-matching scheme coupled with slope estimation via a nonparametric analysis. We develop such an approach and label it MMNPR. Our estimation procedure computes values quickly and performs in a favorable manner when compared to the Quandt method.

Through personal communication with his research colleague L. Hoffman, we are aware that in an extension of his work, Clark (2000) noticed that during the early stages there was a close proximity between the Killer Whale cow/calf dyads. However, as time passed and the mother wanted the child to become more independent, there was a shift in positions between the two. As the cow began to wean its calf, three notions were made: 1) the calf changed how it positioned itself to its mother's dorsal fin, 2) the mother started to move clockwise from the calf, and 3) the distance between the cow and calf started to increase. Numerical and alphabetical representations for each criterion were given to devise a method of measuring and analyzing the data. In this paper, we concentrate on the last of these phenomena, the distances. Using the measurements, the Hinde Index was used and we found linear relationships in order to determine when a cow completely weans her calf. The time when this phenomenon would occur would be called the change point.

The Hinde Index (Hinde, 1970) measures the "approaches" that are the narrowing of the distance between the pair and "leaves" which are a widening between a mother and child. The Hinde Index definition is as follows:

$$\frac{A_m}{A_m + A_i} - \frac{L_m}{L_m + L_i}$$

where A_m = approaches of the mother, L_m = leaves of the mother, A_i = approaches of the infant, and L_i = leaves of the infant. In the case where the total number of leaves by both the mother and infant is the same as the number of approaches by both mother and infant, this equation can be simplified. Therefore, the Hinde Index becomes:

$$\frac{A_m}{A_m + A_i} - \frac{L_m}{L_m + L_i} = \frac{A_m}{A_m + A_i} - \frac{L_m}{A_m + A_i}$$

$$= \frac{A_m - L_m}{A_m + A_i}$$

This expression is easily interpreted since it is positive when the mother's approaches exceed her leaves and negative when she leaves more than she approaches her calf.

In other cases consider the following. Since $A_m < A_m + A_i$, then $0 < \frac{A_m}{A_m + A_i} < 1$. Using the same notion, $0 < \frac{L_m}{L_m + L_i} < 1$. Therefore we know that $-1 < \frac{A_m - L_m}{A_m + A_i} < 1$. Using this index, one may wonder how to tell where the index should be the greatest and least. Table 1.1 below gives a short synopsis of what the trend should look like.

Table 1.1: Values of the Hinde Index for Various Levels of Whale Activity

A_m	A_i	$\frac{A_m}{A_m + A_i}$	L_m	L_i	$\frac{L_m}{L_m + L_i}$	$\frac{A_m}{A_m + A_i} - \frac{L_m}{L_m + L_i}$
High	High	$\approx .5$	High	High	$\approx .5$	≈ 0
High	High	$\approx .5$	High	Low	≈ 1	$\approx -.5$
High	High	$\approx .5$	Low	High	≈ 0	$\approx .5$
High	High	$\approx .5$	Low	Low	$\approx .5$	≈ 0
High	Low	≈ 1	High	High	$\approx .5$	$\approx .5$
High	Low	≈ 1	High	Low	≈ 1	≈ 0
High	Low	≈ 1	Low	High	≈ 0	≈ 1
High	Low	≈ 1	Low	Low	$\approx .5$	$\approx .5$
Low	High	≈ 0	High	High	$\approx .5$	$\approx -.5$
Low	High	≈ 0	High	Low	≈ 1	≈ -1
Low	High	≈ 0	Low	High	≈ 0	≈ 0
Low	High	≈ 0	Low	Low	$\approx .5$	$\approx -.5$
Low	Low	$\approx .5$	High	High	$\approx .5$	≈ 0
Low	Low	$\approx .5$	High	Low	≈ 1	$\approx -.5$
Low	Low	$\approx .5$	Low	High	≈ 0	$\approx .5$
Low	Low	$\approx .5$	Low	Low	$\approx .5$	≈ 0

Using the information in Table 1.1, the Hinde index is largest, or close to 1, when the mother approaches the calf more often than it leaves the calf and the calf leaves more often than it approaches the cow. This also means that although the calf may be trying to become independent, the mother is still keeping up with her calf. On the other hand, the index is smallest when the mother leaves more than she approaches her calf and the calf is approaching the cow more often than it leaves. This lends itself to the notion that the mother is trying to wean the calf to become independent, though the calf is not ready to go on its own.

Using this information, whales, as well as various other animals, are tracked for weeks at a time and a time versus Hinde index plot is formed. Using the plot, one can do an eyeball estimate of the best fit line or lines for the data. However, this method does not lend itself to a mathematical estimate of the true t_0 (the change point, i.e., when one

can expect a whale to have weaned its calf). However, there is another method that would better suit this notion.

We want to find a t_0 for the regime. We need a method that lends itself to finding true change points. The Quandt (1958) method takes into account that one needs to separate the regime into two different groups. The groups must each have at least three observations in order to do a regression analysis. The approach is iterative. Let the number of observations in the first group correspond to the value of t_0. For example, if the first group has 4 observations, then $t_0 = 4$. Hence, one would find the best fit line for the first group ($t \leq t_0$) and for the second group ($t > t_0$) up through the total number of time points T. Using the usual regression estimates for finding standard deviations, one can find the standard deviation for both groups. The standard deviations are then substituted into Quandt's log likelihood function

$$L(t) = -T \log \sqrt{2\pi} - t_0 \log \sigma_1 - (T - t_0) \log \sigma_2 - \frac{T}{2}$$

where T = total number of observations and t_0 = the number of data points in the first of the two groups. The value of t_0 yielding the largest value for $L(t)$ would be considered the true change point. In our case, t_0 would represent the true time point where one can expect the mother to have weaned her calf.

Quandt's (1958) method is a good start; however, it has its flaws. Quandt relies on the assumption that the data come from two separate normal distributions and forms the likelihood function as a product of normal probability density functions. But deriving a closed form function for all five parameters (slopes, intercepts and the change point) presents an intractable problem. Instead, he suggests conditioning on each possible change point and then evaluating $L(t)$. The process of dividing the sample into

two groups depending on the size of t_0 is time consuming. This method may be easy to execute when it comes to small sample sizes, but what about larger ones? Quandt's method uses the MLE (Maximum Likelihood Estimation) method for estimating a true t_0. Unfortunately, MLE may yield biased estimates when it comes to small samples. It is a known fact that as the sample size increases, the need to estimate parameters is minimized and the biasedness of the MLE approaches zero. Since Quandt's method is best used for smaller samples, the estimated t_0 would not necessarily be an accurate representation of the true t_0. Additionally, the method assumes the data come from an underlying normal distribution. What other methods can we use that will minimize the error of estimating the true change point? Can a method be developed that has less distributional assumptions? Would a new method prove to be better than Quandt's method?

CHAPTER 2

PIECE-WISE REGRESSSION ASSUMPTIONS

The model introduced by Quandt in 1958 used the assumption that when (x_t,y_t) are bivariate normal then

$$E(y_t) = B_0 + B_1 x_t \quad t \le t_0$$
$$E(y_t) = B_2 + B_3 x_t \quad t > t_0$$

where B_0 = the original proximity propensity for the first regression, B_1 = the rate of decrease to steady state for the first regression , B_2 = the original proximity of propensity for the second regression, and B_3 = the rate of decrease to steady state for the second regression. We will refer to these as assumption 1.

To better understand this assumption and others to come, real data taken from whales in Sea World (*t*'s and *y*'s) shown in Table 2.1. In this case, *t* = time in weeks and y_t = index of proximity (Hinde Index). We use data acquired from mother-calf pairs of Killer Whales born at one of two SeaWorld parks (California or Florida) between 1999 and 2002. The vagueness preserves the anonymity of the mammals (a directive from SeaWorld administrators). The collection process was immense and covered a total of 66,606 hours with observers stationed both above and below the pools viewing spatial relationship data every 15 minutes via the use of an instantaneous sampling technique (Clark and Odell, 1999). In the initial stages after birth the mother cow closely monitors her calf, to nourish and protect it. As the calf gets older, the mother encourages the calf to begin to fend for itself. For illustrative use of the original Quandt method, the *x*'s were generated in Microsoft Excel to represent an *x* value at time t. We need to identify variables for use with the Quandt method. Since we have three different variables but

wish to rely on simple linear regression, then we use different linear models for the different groupings of the (x_t, y_t) defined by the value t_0.

Table 2.1: Actual Weekly (t) Whale Hinde Index (y) with Simulated x Values

t	x	y
1	-1.00	1.00
2	0.25	0.50
3	0.45	0.20
4	1.00	0.11
5	0.60	0.15
6	-0.05	0.21
7	0.04	0.16
8	0.09	0.12
9	0.30	0.11
10	0.36	0.09
11	-1.00	0.04
12	1.00	0.29
13	-0.60	0.00
14	-0.05	0.11
15	1.00	0.05
16	0.60	0.00
17	0.36	0.02
18	0.12	0.15
19	0.50	0.10
20	0.10	0.09

Quandt's Method used the maximum likelihood estimate approach to inspect all *L(t)* for $t = 1$ to T. There is no assumption that $x_i < x_{i+1}$ for $i = 1, 2, \ldots T$, but a grouping of the x's occurs once t_0 is identified.

In assumption 1, we need to use generated x's versus proximity (Hinde Index). Table 2.2 yields the analysis given using Quandt's method.

Table 2.2: Parameter Estimates, Standard Errors, and Calculated Log Likelihood Value Using Quandt's Method Under Assumption 1

t_0	B_0	B_1	B_2	B_3	s_1	s_2	L(t)
3	**0.517**	**-0.498**	**0.095**	**0.041**	**0.143**	**0.074**	**3.786**
4	0.534	-0.463	0.096	0.046	0.110	0.076	3.758
5	0.516	-0.479	0.095	0.043	0.104	0.079	3.504
6	0.454	-0.443	0.085	0.051	0.173	0.074	2.431
7	0.411	-0.426	0.078	0.054	0.192	0.074	1.769
8	0.378	-0.417	0.075	0.056	0.200	0.076	1.028
9	0.364	-0.424	0.073	0.055	0.192	0.080	0.528
10	0.353	-0.432	0.074	0.056	0.184	0.085	0.072
11	0.265	-0.215	0.067	0.069	0.258	0.090	-2.098
12	0.277	-0.171	0.066	-0.002	0.254	0.060	-1.074
13	0.242	-0.117	0.107	-0.086	0.263	0.048	-1.182
14	0.232	-0.113	0.107	-0.086	0.255	0.053	-2.022
15	0.229	-0.123	0.134	-0.184	0.246	0.053	-2.437
16	0.221	-0.134	0.121	-0.116	0.240	0.060	-3.160
17	0.213	-0.138	0.124	-0.045	0.235	0.043	-3.187

According to the table, the estimated true change point would be when t_0= 3. Using B_0, B_1, B_2 ,and B_3, the graph appears as illustrated in Figure 2.1.

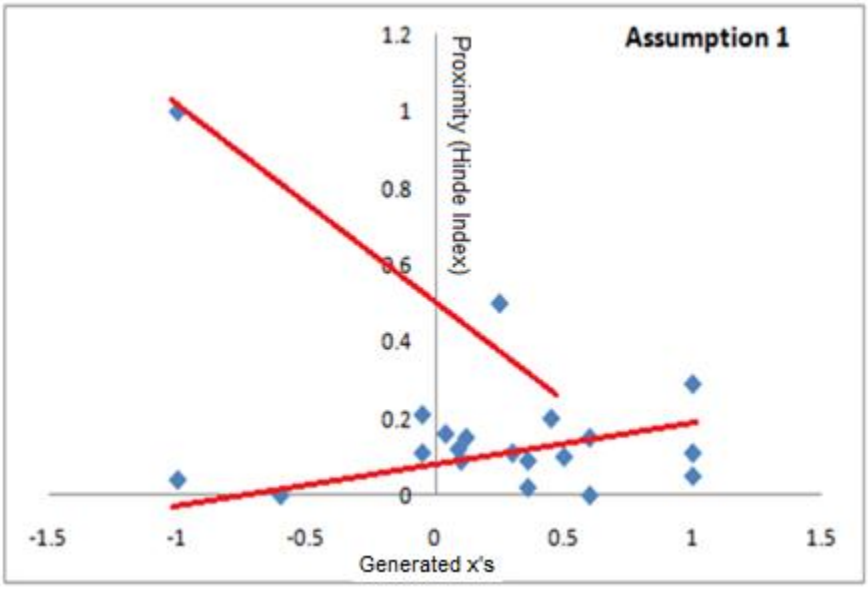


Figure 2.1: Graph of the Fitted Model Using Quandt's Method Under Assumption 1

For assumption 1, it is not assumed that the two regressions must join, nor does the second regression have to have a horizontal slope.

For a more simple model, we assume that t serves as the *x* variable, i.e. we ignore the information provided by *x*, the hypothetical respiratory rate. We now look at time versus proximity. The simplified model has the assumptions

$$E(y_t) = B_0 + B_1 t \qquad t \le t_0$$
$$E(y_t) = B_2 + B_3 t \qquad t > t_0$$

and will be referred to as assumption 2. Using Table 2.1, the parameter estimates, standard errors, and calculated log likelihood values are shown in Table 2.3.

Table 2.3: Parameter Estimates, Standard Errors, and Calculated Log Likelihood Value Using Quandt's Method Under Assumption 2

t_0	B_0	B_1	B_2	B_3	s_1	s_2	L(t)
3	**1.367**	**-0.400**	**0.171**	**-0.005**	**0.082**	**0.072**	**4.724**
4	1.195	-0.297	0.186	-0.006	0.145	0.073	3.522
5	1.019	-0.209	0.188	-0.007	0.200	0.076	2.292
6	0.871	-0.145	0.156	-0.005	0.231	0.077	1.454
7	0.783	-0.113	0.135	-0.003	0.230	0.079	0.814
8	0.724	-0.093	0.127	-0.003	0.223	0.083	0.221
9	0.674	-0.078	0.120	-0.002	0.216	0.087	-0.346
10	0.635	-0.067	0.127	-0.003	0.210	0.092	-0.864
11	0.609	-0.061	0.210	-0.008	0.202	0.095	-1.146
12	0.541	-0.045	-0.112	0.011	0.220	0.053	0.094
13	0.534	-0.044	-0.041	0.007	0.210	0.056	-0.418
14	0.507	-0.038	-0.247	0.018	0.207	0.050	-0.564
15	0.492	-0.035	-0.396	0.026	0.200	0.053	-1.121
16	0.483	-0.034	-0.206	0.016	0.194	0.060	-1.712
17	0.470	-0.032	0.683	-0.030	0.189	0.016	-0.324

We find that the estimate of t_0 is 3. Using B_0, B_1, B_2 ,and B_3, the corresponding graph is illustrated in Figure 2.2.

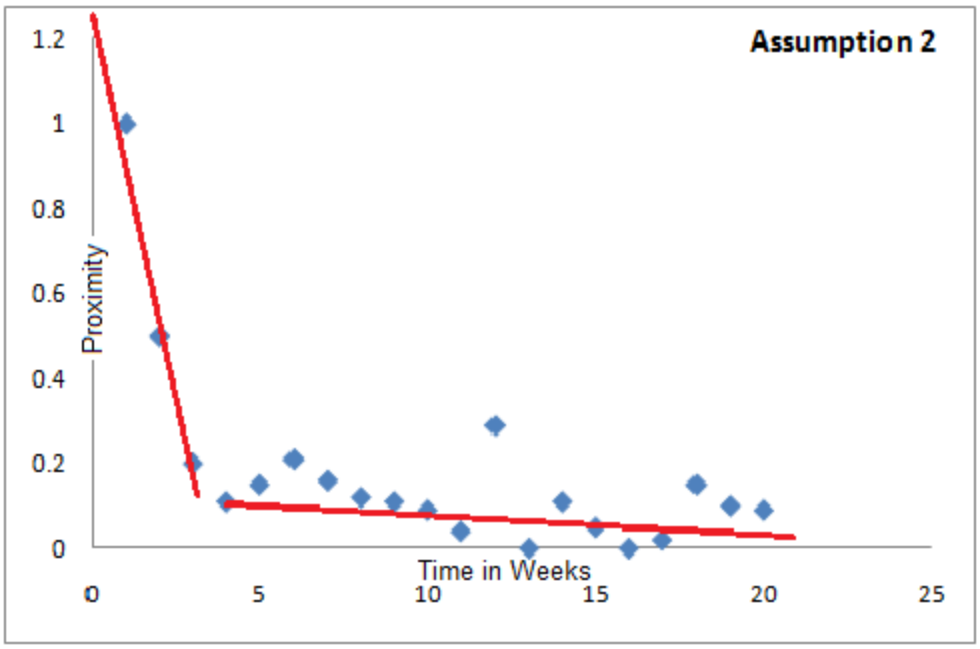


Figure 2.2: Graph of the Fitted Model Using Quandt's Method Under Assumption 2

Once again, we do not assume that the two regressions must join, nor does the second regression have to have a horizontal slope.

As the calf grows older, we assume that the proximity between the mother and calf decreases over time. Because of this, we expect that until a certain time, t_0, there would be a steady decrease in proximity. After we reach t_0, there would be a steady state of proximity, i.e. a horizontal line, between cow and calf. This lends itself to the next model, which will be referred to as assumption 3, (an even more simplified one),

$$E(y_t) = B_0 + B_1 t \quad t \le t_0$$
$$E(y_t) = B_2 \qquad t > t_0$$

In this case, we only look at $t_0 = 3$, B_0, B_1, and B_2. Our graph for assumption 3 is illustrated using Figure 2.3.

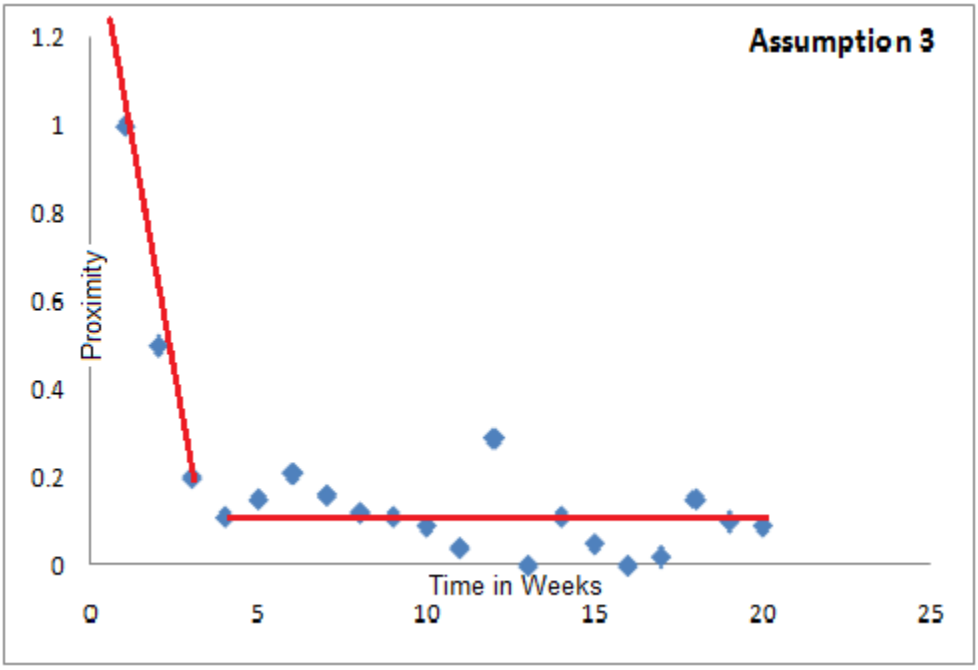


Figure 2.3: Graph of the Fitted Model Using Quandt's Method Under Assumption 3

Notice in the graph, the two lines are disjoint. However, we assume that the second regression must have a horizontal slope, which represents the steady state. For an intersection to occur, we would need to assume that for each $t \geq t_0$, the expected value would have to be the same for each t. Hence, our most simplified model is illustrated in Figure 2.4.

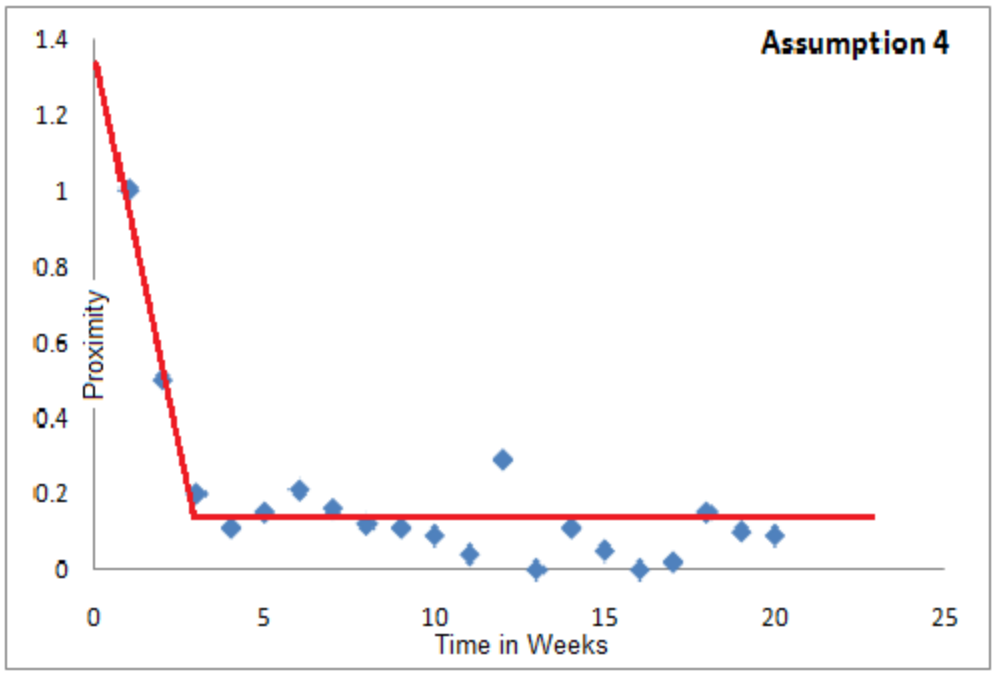


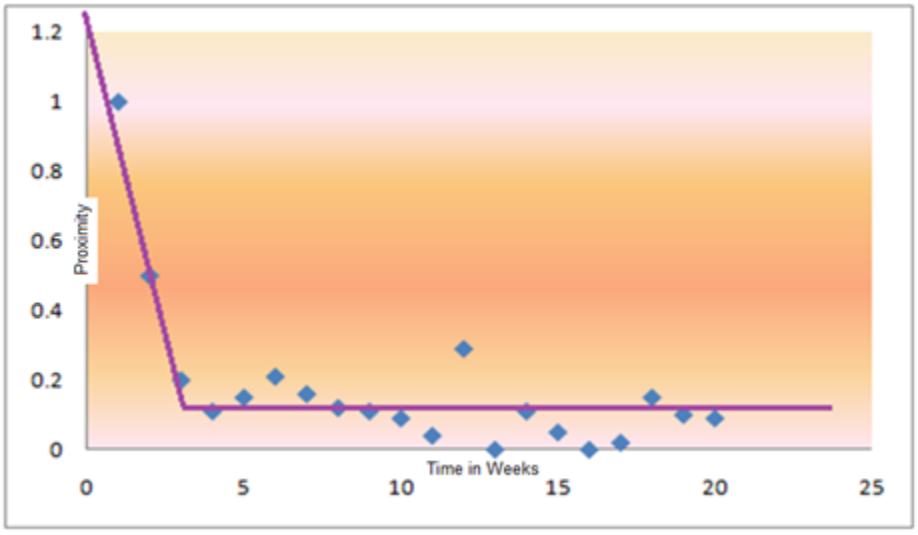
Figure 2.4: Graph of the Fitted Model Using Quandt's Method Under Assumption 4

The model for the graph, referred to as assumption 4, would be

$$E(y_t) = B_0 + B_1 t \quad t \le t_0$$
$$E(y_t) = B_0 + B_1 t_0 \quad t > t_0$$

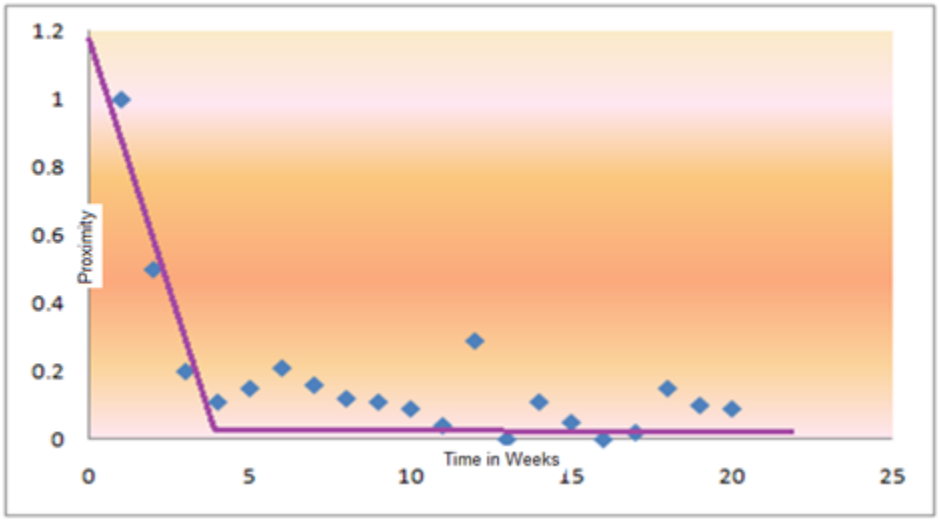
With this data set, we find that all assumptions using Quandt's method yields that the estimated change point would be at $t_0 = 3$.

Though we have estimated that in this case $t_0 = 3$, graphically, how would it compare to other t_0's? Let us use Figure 2.5, Figure 2.6, and Figure 2.7 to visualize what would happen.



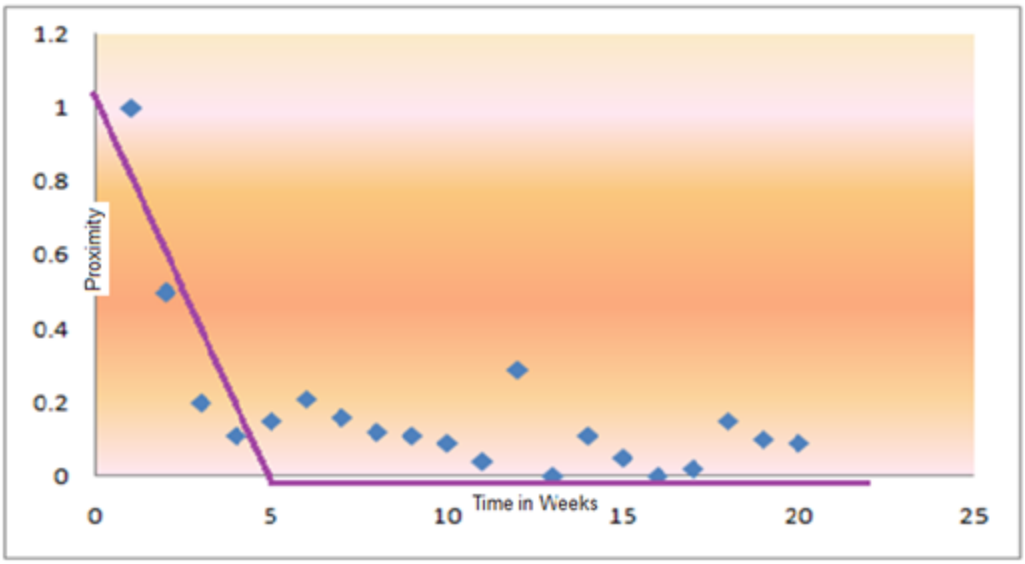
Y_1 = -.400t + 1.367 SSE_1 = 0.007
Y_2 = -.167 SSE_2 = 0.152 $L(t)$ = 20.857

Figure 2.5: Using Whale Hinde Indices to Show Quandt Iterations at t_0=3



Y_1 = -.297t + 1.195 SSE_1 = 0.042
Y_2 = .007 SSE_2 = 0.245 $L(t)$ = 14.158

Figure 2.6: Using Whale Hinde Indices to Show Quandt Iterations at t_0=4



$Y_1 = -.209t + 1.019$ $SSE_1 = 0.119$

$Y_2 = -.026$ $SSE_2 = 0.336$ $L(t) = 9.451$

Figure 2.7: Using Whale Hinde Indices to Show Quandt Iterations at t_0=5

Notice that more of the data points are closest to the two regressions when $t_0 = 3$. When using the Quandt method, one must use the $L(t)$ that is largest. By comparing the $L(t)$ to the sum of squared errors, one finds for this case that the largest $L(t)$ out of the three possible change points has the smallest error.

Quandt's method assumes a normal distribution for a data set with constant variance and uncorrelated error terms for a data set, but real data may not conform to these assumptions. Since most of the points from the data are actually captured by the two lines, this means that the Quandt method does a good job of finding the estimated change point for this data set that is assumed well behaved. But what if there was a set in which most of the points do not follow the regressions because the data set is non-normal? Or the data has non-constant variance? Or the data is correlated over time?

Other data, if not from a normal distribution, could not be modeled as a product of two normal probability distribution functions.

Also, Quandt uses the largest *L(t)* to determine what would be the estimated change point. Unfortunately, there can be more than one local maxima using the *L(t)*. For example, the Table 2.4 and Figure 2.8 both illustrate graphically the points *(t, L(t))* exhibit multiple local maxima.

Table 2.4: Parameter Estimates, Standard Errors, and Calculated Log Likelihood Values Using Quandt's Method Under Assumption 4

t_0	B_0	B_1	sse1	sse2	L(t)
3	**1.367**	**-0.400**	**0.007**	**0.152**	**20.857**
4	1.195	-0.297	0.042	0.245	14.158
5	1.019	-0.209	0.119	0.336	9.451
6	**0.871**	**-0.145**	**0.214**	**0.207**	**11.137**
7	0.783	-0.113	0.264	0.187	10.657
8	0.724	-0.093	0.297	0.204	9.239
9	0.674	-0.078	0.328	0.210	8.314
10	0.635	-0.067	0.353	0.220	7.422
11	0.609	-0.061	0.367	0.267	6.155
12	**0.541**	**-0.045**	**0.485**	**0.055**	**10.820**
13	0.534	-0.044	0.487	0.096	7.990
14	**0.507**	**-0.038**	**0.512**	**0.071**	**8.093**
15	0.492	-0.035	0.523	0.075	7.282
16	0.483	-0.034	0.527	0.095	6.422
17	0.470	-0.032	0.536	0.099	6.117

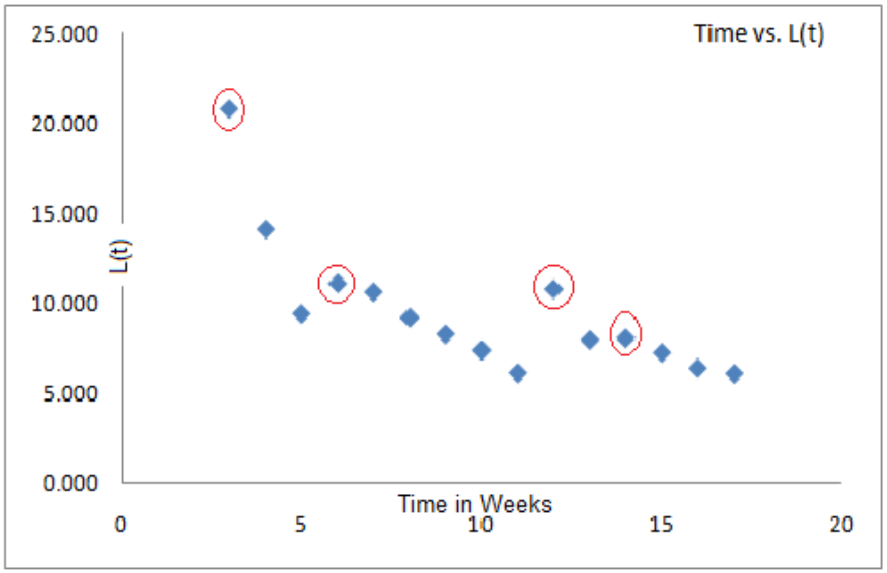


Figure 2.8: Graph of Time versus Log Likelihood Function from Use of Quandt's Method Under Assumption 4

As one can see, there are exactly four local maxima at $t_0 = 3$, $t_0 = 6$, $t_0 = 12$, and $t_0 = 14$. Therefore, anyone of these maximums could be mistaken for the true change point if all iterations are not completed. Therefore, we seek to determine a better method for estimating the true t_0 value independent of the data distribution.

CHAPTER 3

METHODOLOGY

One estimation method that does not rely on distributional assumptions utilizes moment matching in order to find the joint estimate for the slope and t_0. This is accomplished by using the sample data to find the expected value for an unknown parameter. Prior to doing so, an intuitive study of the behavior of the change point in this "broken-line" model led to the insight that points early in time and furthest from the change point were more likely to lie on the regression line rather than on the horizontal ray; thus we chose to make a distributional assumption about the range of the initial regression line (first "regime") in that the probability that the expectation is $B_0 + B_1 t$, $t < t_0$, is proportional to the distance from the unknown actual value t_0.

Since we are dealing with three unknowns (B_0, B_1, and t_0), it would be nice to eliminate one of the unknowns from our model. First consider the $E(y_t)$ as modeled in assumption 4.

$$E(y_1) = B_0 + B_1$$
$$E(y_2) = B_0 + 2B_1$$
$$E(y_3) = B_0 + 3B_1$$
$$\vdots$$
$$E(y_{t_0-1}) = B_0 + (t_0 - 1)B_1$$
$$E(y_{t_0}) = B_0 + t_0 B_1$$
$$\vdots$$
$$E(y_T) = B_0 + t_0 B_1$$

To eliminate B_0, let $w_t = y_{t+1} - y_1$ for $t=1,\ldots,T-1$. This yields

$$E(y_2 - y_1) = (B_0 + 2B_1) - (B_0 + B_1) = B_1$$
$$E(y_3 - y_1) = (B_0 + 3B_1) - (B_0 + B_1) = 2B_1$$
.
.
.
$$E(y_{t_0} - y_1) = (B_0 + t_0 B_1) - (B_0 + B_1) = (t_0 - 1)B_1$$
$$E(y_{t_0+1} - y_1) = (B_0 + t_0 B_1) - (B_0 + B_1) = (t_0 - 1)B_1$$
$$E(y_{t_0+2} - y_1) = (B_0 + t_0 B_1) - (B_0 + B_1) = (t_0 - 1)B_1$$
.
.
.
$$E(y_T - y_1) = (B_0 + t_0 B_1) - (B_0 + B_1) = (t_0 - 1)B_1$$

We then can conclude that

$$E(w_t) = B_1 t \qquad t = 1,\ldots,t_0 - 1$$
$$E(w_t) = B_1(t_0 - 1) \quad t = t_0,\ldots,T - 1$$

Most data is assumed to come from a normal distribution; however, we are not making the assumption that the data we use is normal. Hence, we will conclude that w_t has a mean of $E(w_t)$. We can see that this model has two parameters to estimate: B_1 and t_0. It seems reasonable to assume that w_t follows a distribution that is related to the distance from the actual change point.

We include weights that are proportional to $t_0 - t$. We will use the assumption that when $t > t_0$, the probability is zero. To determine what probability is associated with each time point before t_0 we can discern k by knowing the sum must be 1, i.e.

$$\sum_{t=1}^{t_0-1} k \frac{t_0 - t}{t_0} = 1$$

Therefore,

$$\sum_{t=1}^{t_0-1} k\frac{t_0-t}{t_0} = k\sum_{t=1}^{t_0-1}\left(1-\frac{t}{t_0}\right)$$

$$= k\left((t_0-1)-\frac{1}{t_0}\left(\frac{(t_0-1)(t_0)}{2}\right)\right)$$

$$= k\left((t_0-1)-\left(\frac{t_0-1}{2}\right)\right)$$

$$= k\left(\frac{t_0-1}{2}\right)$$

Solving the equation $k\left(\frac{t_0-1}{2}\right)=1$, we find that $k = \left(\frac{2}{t_0-1}\right)$ and, for example the probability associated with w_1 is $\left(\frac{2}{t_0-1}\right)\frac{t_0-1}{t_0}=\frac{2}{t_0}$. Hence the expected w_t, weighted by these probabilities $\left(\text{i.e.}\left(\left(\frac{2}{t_0-1}\right)\left(\frac{t_0-t}{t_0}\right)\right)\right)$ at each time t is

$$B_1\frac{2(t_0-1)}{t_0(t_0-1)}+2B_1\frac{2(t_0-2)}{t_0(t_0-1)}+3B_1\frac{2(t_0-3)}{t_0(t_0-1)}+\ldots+B_1(t_0-1)\frac{2(1)}{t_0(t_0-1)}$$

$$= \frac{2B_1}{t_0(t_0-1)}\ (t_0-1)+2(t_0-2)+3(t_0-3)+\ldots+(t_0-1)(1)$$

$$= \frac{2B_1}{t_0(t_0-1)}\sum_{i=1}^{t_0-1} i(t_0-i)$$

$$= \frac{2B_1}{t_0(t_0-1)}\left[t_0\left(\frac{(t_0-1)(t_0)}{2}\right)-\frac{(t_0-1)^3}{3}-\frac{(t_0-1)^2}{2}-\frac{t_0-1}{6}\right]$$

$$= \frac{B_1(t_0+1)}{3}$$

Combining this with the remaining $(T - t_0)$'s $B_1(t_0 - 1)$ yields

$$\sum_{i=1}^{T-1} E(w_t) = \frac{(t_0 - 1)\left(\frac{B_1(t_0 + 1)}{3}\right) + (T - t_0)B_1(t_0 - 1)}{T - 1}$$

$$= \frac{\frac{B_1 t_0^{\ 2} - B_1}{3} + TB_1 t_0 + B_1 t_0 - B_1 t_0^{\ 2} - B_1 T}{T - 1}$$

$$= \frac{B_1\ \ t_0^{\ 2} - 1 + 3Tt_0 + 3t_0 - 3t_0^{\ 2} - 3T}{3(T - 1)}$$

$$= \frac{B_1\ \ -2t_0^{\ 2} + (3T + 3)t_0 - (1 + 3T)}{3(T - 1)}$$

Setting $\frac{B_1\ \ -2t_0^{\ 2} + (3T + 3)t_0 - (1 + 3T)}{3(T - 1)} = \bar{w}$, we then have an expression for t_0 and B_1.

Hence

$$\frac{\bar{w}(3(T - 1))}{B_1} = -2t_0^{\ 2} + (3T + 3)t_0 - (1 + 3T)$$

or

$$0 = -2t_0^{\ 2} + (3T + 3)t_0 - (1 + 3T) - \frac{\bar{w}(3(T - 1))}{B_1}$$

Combining the expectations from both the regression line and the horizontal line in a properly weighted manner, results in this formula for the moment match. By re-writing the equation we do produce an estimate of B_1 as a function of t_0. This is helpful as shown in the next section because it provides a restriction on the candidates for the slope estimate. The slope estimate is derived via usual nonparametric regression where we compute all pair wise slopes and rely on the median as the best estimator, but can be

restricted to only a small subset of feasible values. In conducting a nonparametric regression procedure the number of potential estimates number $\frac{T(T-1)}{2}$ since all possible pairs of points are examined. Using $\bar{w}$, we can also solve for B_1

$$\frac{\sum_{t=1}^{T-1} W_t}{T-1} = \frac{(t_0-1)\left(\frac{\beta_1(t_0+1)}{3}\right)+(T-t_0)\beta_1(t_0-1)}{T-1}$$

$$\beta_1 = \frac{3\sum_{t=1}^{T-1} W_t}{-2t_0^{\,2}+3(T+1)t_0-(3T+1)}$$

and we are able to create a lower and upper bound for the estimate of B_1 so that the sort needed to reveal the median of the sample slopes is conducted only on a small subset of the slopes generated from all possible point pairings.

To conduct a non-parametric regression, the slopes of all possible pairings of points slopes are found and the following bounds are used. For $T > 14$, the interval needed to compare all slopes is

$$\frac{3(T-1)}{6T-1} \le B_1 \le \frac{24(T-1)}{(3T-1)^2}$$

The median of the slopes that are within the interval is used as the estimate for the parameter B_1. Using the equation above, we are now able to calculate t_0 after finding a good estimate for B_1 and calculating $\bar{w}$.

CHAPTER 4

MAIN RESULTS

We will illustrate the MMNPR using the Sea World data set. Referring back to Table 2.1, we needed to find all possible pairs of slopes and find the interval for the most influential slopes. Since there are 20 points in the data set, then ${}_{20}C_2 = \frac{20!}{(20-2)!2!} = 190$.

Below in Table 4.1, the chart of the 190 slopes that were found using Excel. SAS was used to find the feasible interval based on bounds developed from the equations in Chapter 3. In this case, there are 10 slopes (which are highlighted) within the interval $-.45 < B_1 < -0.11376$.

Table 4.1: All Possible Slopes Between All Points of Sea World Data Set

-0.50000																		
-0.40000	-0.30000																	
-0.29667	-0.19500	-0.09000																
-0.21250	-0.11667	-0.02500	0.04000															
-0.15800	-0.07250	0.00333	0.05000	0.06000														
-0.14000	-0.06800	-0.01000	0.01667	0.00500	-0.05000													
-0.12571	-0.06333	-0.01600	0.00250	-0.01000	-0.04500	-0.04000												
-0.11125	-0.05571	-0.01500	0.00000	-0.01000	-0.03333	-0.02500	-0.01000											
-0.10111	-0.05125	-0.01571	-0.00333	-0.01200	-0.03000	-0.02333	-0.01500	-0.02000										
-0.09600	-0.05111	-0.02000	-0.01000	-0.01833	-0.03400	-0.03000	-0.02667	-0.03500	-0.05000									
-0.06455	-0.02100	0.01000	0.02250	0.02000	0.01333	0.02600	0.04250	0.06000	0.10000	0.25000								
-0.08333	-0.04545	-0.02000	-0.01222	-0.01875	-0.03000	-0.02667	-0.02400	-0.02750	-0.03000	-0.02000	-0.29000							
-0.06846	-0.03250	-0.00818	0.00000	-0.00444	-0.01250	-0.00714	-0.00167	0.00000	0.00500	0.02333	-0.09000	0.11000						
-0.06786	-0.03462	-0.01250	-0.00545	-0.01000	-0.01778	-0.01375	-0.01000	-0.01000	-0.00800	0.00250	-0.08000	0.02500	-0.06000					
-0.06667	-0.03571	-0.01538	-0.00917	-0.01364	-0.02100	-0.01778	-0.01500	-0.01571	-0.01500	-0.00800	-0.07250	0.00000	-0.05500	-0.05000				
-0.06125	-0.03200	-0.01286	-0.00692	-0.01083	-0.01727	-0.01400	-0.01111	-0.01125	-0.01000	-0.00333	-0.05400	0.00500	-0.03000	-0.01500	0.02000			
-0.05000	-0.02188	-0.00333	0.00286	0.00000	-0.00500	-0.00091	0.00300	0.00444	0.00750	0.01571	-0.02333	0.03000	0.01000	0.03333	0.07500	0.13000		
-0.05000	-0.02353	-0.00625	-0.00067	-0.00357	-0.00846	-0.00500	-0.00182	-0.00100	0.00111	0.00750	-0.02714	0.01667	-0.00200	0.01250	0.03333	0.04000	-0.05000	
-0.04789	-0.02278	-0.00647	-0.00125	-0.00400	-0.00857	-0.00538	-0.00250	-0.00182	0.00000	0.00556	-0.02500	0.01286	-0.00333	0.00800	0.02250	0.02333	-0.03000	-0.01000

The median of the 10 slopes is -0.20375. We use this value as the slope B_1 and use $\sum_{t=1}^{T-1} W_t$ = -16.5 to calculate the corresponding t_0 from the equation

$$\beta_1 = \frac{3\sum_{t=1}^{T-1} W_t}{-2t_0^{\,2} + 3(T+1)t_0 - (3T+1)}$$

$$-.20375 = \frac{3(-16.5)}{-2t_0^{\,2} + 3(21)t_0 - (61)}$$

$$242.945 = -2t_0^{\,2} + 63t_0 - 61$$

$$0 = -2t_0^{\,2} + 63t_0 - 303.945$$

Using the quadratic formula and using the smaller root inside [3, 17], we get t_0 = 5.947. For reporting, we have elected to truncate this to the integer value of 5. Therefore we have estimated t_0 and B_1 and with moment matching, we would use the average of the y's to find B_0.

$$\bar{y} = \frac{\sum_{t=1}^{t_0} (\beta_0 + \beta_1 t) + (T - t_0)(\beta_0 + \beta_1 t_0)}{T}$$

Along with calculating it by hand, a SAS program was used to find all of the values. Table 4.2 shows what values were calculated for both Quandt's method and the MMNPR method.

Table 4.2: SAS Values for Quandt and MMNPR Values

B_0	B_1	t_{0q}	$MMNPB_0$	$MMNPB_1$	$MMNPt_0$
1.36667	-0.400	3.000	1.237	-0.204	5.000

Figure 4.1 below is a visual comparison of the results using the Quandt and MMNPR methods.

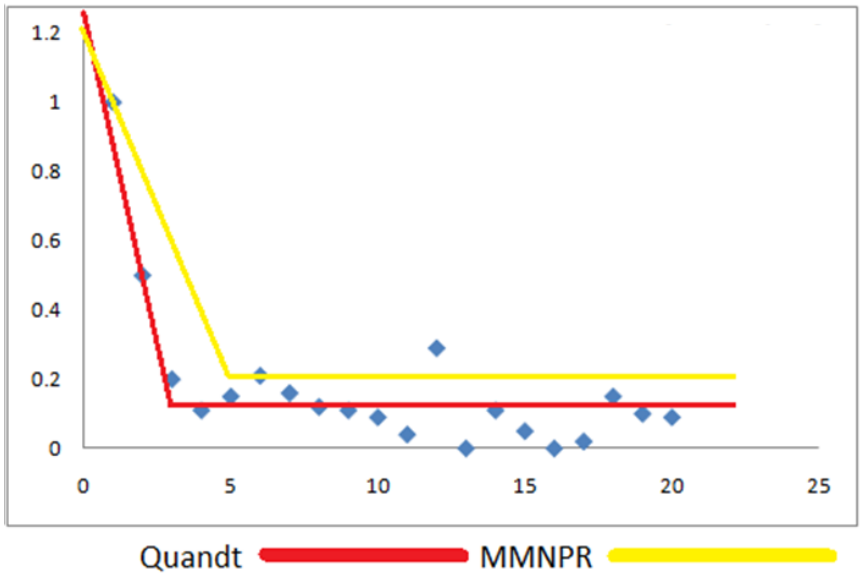


Figure 4.1: Graph of Quandt Method versus MMNPR Method

In this case, Quandt looks to have done a better job of estimating parameters, resulting in regressions having more points closer to the lines. However, we must remember that this sample was a small sample size and this only one case. If we investigate different sample sizes and many more replications, we may observe different results. Using a SAS program, the following Table 4.3 was the output comparing Quandt's method to the MMNPR method.

Table 4.3: SAS Generated Values for Parameters for Quandt and MMNPR Methods

Simulation runs (k=250) comparing Quandt MLE to the MNPR estimates								
CASE	1	2	3	4	5	6	7	8
N	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
t_0	5.00	5.00	5.00	5.00	10.00	10.00	10.00	10.00
β_0	0.90	0.90	0.50	0.50	0.90	0.90	0.50	0.50
β_1	-0.20	-0.20	-0.20	-0.20	-0.10	-0.10	-0.10	-0.10
σ	0.05	0.03	0.05	0.03	0.03	0.02	0.03	0.02
MLE-t_0	4.992 (.237)	4.996 (.063)	5.020 (.228)	5.000 (.000)	9.984 (.390)	9.960 (.196)	9.996 (.375)	9.980 (.166)
MLE-β_0	0.900 (.051)	0.901 (.032)	0.495 (.054)	0.500 (.031)	0.901 (.022)	0.901 (.013)	0.501 (.021)	0.501 (.014)
MLE-β_1	-0.201 (.050)	-0.200 (.010)	-0.199 (.017)	-0.200 (.009)	-0.100 (.004)	-0.100 (.002)	-0.100 (.004)	-0.100 (.002)
MMNPR-t_0	6.224 (.645)	5.956 (.450)	6.280 (.596)	5.940 (.421)	10.556 (.664)	10.896 (.557)	10.620 (.661)	10.932 (.552)
MMNPR-β_0	0.891 (.059)	0.896 (.038)	0.489 (.065)	0.494 (.040)	0.983 (.041)	0.986 (.026)	0.589 (.040)	0.588 (.029)
MMNPR-β_1	-0.156 (.017)	-0.162 (.013)	-0.154 (.017)	-0.162 (.013)	-0.104 (.004)	-0.102 (.002)	-0.104 (.004)	-0.102 (.003)
CASE	9	10	11	12	13	14	15	16
N	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00
t_0	10.00	10.00	10.00	10.00	15.00	15.00	15.00	15.00
β_0	0.90	0.90	0.50	0.50	0.90	0.90	0.50	0.50
β_1	-0.10	-0.10	-0.10	-0.10	-0.07	-0.07	-0.07	-0.07
σ	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.01
MLE-t_0	10.004 (.304)	9.992 (.089)	9.996 (.304)	9.996 (.110)	14.980 (.290)	15.000 (.000)	14.992 (.297)	15.000 (.000)
MLE-β_0	0.899 (.021)	0.901 (.014)	0.502 (.022)	0.500 (.013)	0.900 (.011)	0.900 (.006)	0.501 (.011)	0.500 (.006)
MLE-β_1	-0.100 (.004)	-0.100 (.002)	-0.100 (.003)	-0.100 (.002)	-0.070 (.001)	-0.070 (.001)	-0.070 (.001)	-0.070 (.001)
MMNPR-t_0	10.932 (.646)	10.764 (.503)	10.916 (.571)	10.784 (.492)	16.900 (.729)	17.660 (.546)	16.796 (.762)	17.644 (.578)
MMNPR-β_0	0.948 (.036)	0.946 (.025)	0.550 (.036)	0.550 (.022)	1.003 (.027)	1.009 (.016)	0.602 (.029)	0.609 (.017)
MMNPR-β_1	-0.095 (.004)	-0.096 (.002)	-0.095 (.004)	-0.096 (.002)	-0.072 (.002)	-0.071 (.001)	-0.072 (.002)	-0.071 (.001)

These simulations are based on 250 runs each. The values in the table are the means with their standard deviations in parenthesis. Cases 1-8 are based on sample sizes of 20 and cases 9-16 are based on sample sizes of 30. With the usual normal error

regression assumptions, when comparing the values for Quandt's method (MLE) to the MMNPR method, one finds that both are accurate at estimating the true change point slope and intercept. The MMNPR's estimate of t_0 tends to be larger than the true value of t_0. However, the MMNPR is within 2.5 standard deviations from the true value of t_0. Another advantage of using the MMNPR method is its speed. When running a computer program to find the needed values, Quandt's method takes 2 CPU seconds per replication; the MMNPR uses 1/10 CPU seconds.

CHAPTER 5

CONCLUSION

We have developed a good quick method to find the parameter estimates for the intercept, slope, and change point in the broken line regression model. We illustrated its use by modeling data from mother and calf whale proximity data, comparing our estimates to estimators generated by Quandt (1958). We were able to detect a change in the pairs behavior around week 5 (with our MMNPR method) and week 3 with Quandt's method.

This discovery would indicate that need for early bonding due to nursing declines quite rapidly in Killer Whales. Additionally, we wanted to gain insight into whether our quick MMNPR method was computationally adequate and so we compared our method to the Quandt MLE method. We expected the Quandt method to be superior since our simulations were set up to mimic the assumptions he used to derive his estimators. Yet, the MMNPR performed adequately. Our simulations reveal a repeated small over-estimation of the location of the change point that needs investigation. Further research will involve simulations and include cases where the error terms for the broken line regression are not well behaved. Additionally, we will devise testing procedures and look at their performance under null and alternative hypotheses as suggested in Gregoire and Hamrouni (2002).

APPENDIX A

SAS code

```
/* When using this program please reference via the following citation:
  Hoffman, L.L., Knofczynski, G., Clark, S., Rogers, A., Hudson, J., King, H., and Reiss,
E. 2009. A Nonparametric Method for the
  Estimation of All Parameters in the Joinpoint Two Regime Regression Model.  In JSM
Proceedings, Alexandria, VA: American Statistical
  Association.
*/

data onehinde;

input time index;

cards;

1 1

2 .5

3 .2

4 .11

5 .15

6 .21

7 .16

8 .12

9 .11

10 .09

11 .04
```

```
12 .29
13 0
14 .11
15 .05
16 0
17 .02
18 .15
19 .1
20 .09
;
%let n=20;
data allruns;
/* creating the macro to fit first and second line */
%macro regreg;
/* grabs the first &j set of points */
data first;
set onehinde;
if _n_ gt &j then delete;
/* grabs the second &j set of points */
data second;
set onehinde;
if _n_ le &j then delete;
/* quandt approach */
```

```
/* regression on first &j points to find SSE and b0 and b1 */
proc glm data=first;
model index=time; output out=quandt1 p=yhat r=yresid;
data qvals1;
set quandt1;
retain b10 b0 b1 sse1 sse2;
if _n_=1 then do; b10=yhat; sse1=0;end;
if _n_=2 then do; b1=yhat-b10; b0=b10-b1; end;
sse2=sse1+yresid*yresid;
sse1=sse2;
sseI=sse2;
if _n_ ne &j then delete;
data clnqv1;
set qvals1;
keep b0 b1 sseI;
proc print;
/* finding SSE of the last points on horizontal line */
data qvals2;
merge second clnqv1;
retain bhorz sse1 sse2;
if _n_=1 then do; sse1=0; bhorz=b0+b1*(&j); end;
sse2=sse1+(index-bhorz)* (index-bhorz);
sse1=sse2;
```

```
sseII=sse2;
if _n_ ne &jend then delete;
data clnqv2;
set qvals2;
keep bhorz sseII;
proc print;
/* creating a dataset with all possible &j divisions */
data meshqval;
merge clnqv1 clnqv2;
data allruns;
set allruns meshqval;
data results;
set allruns;
if _n_ = 1 then delete;
ssetot=sseI+sseII;
/*find likelihood value */
Lt=-(&n)*log(sqrt(2*3.1415962))-(_n_+1)*log(sqrt(sseI/(_n_+1)))
  -(&n-(_n_+1))*log(sqrt(sseII/(&n-(_n_+1))))-(&n/2);
t0q=_n_+1;
proc print;
/* keeping only the answers from largest Lt */
proc sort data=results; by descending Lt ;
data quntparm; set results; if _n_ > 1 then delete;
```

```
/* keeping the estimates */
data qntcln;
set quntparm;
keep b0 b1 t0q;
%mend regreg;
/* this is the loop over &j from 3 to n-3 */
%macro loop;
%let nend = &n - 3;
%do j=3 %to &nend;
%let jend=&n - &j;
%regreg;
%end;
%mend loop;
%loop;
/* MMNPR approach */
/* finding sum of w's  */
data prestats;
set onehinde;
array yyval(50);
retain sum y1 yyval;
if _n_=1 then y1=index;
%let m=_n_;
if _n_=1 then sum=0;
```

```
sum=sum+(index-y1);
yyval(&m)=index;
output;
/* nonparametric estimation of b1 slope by creating all
possible slopes and only saving those within feasible bounds */
data stats;
set prestats;
array yyval(50);
array sloper(2500);
retain sloper loopcount;
if _n_ eq &n then loopcount=0;
if _n_ eq &n then do;
  do iclr=1 to 2500;
  sloper(iclr)=99999;
  end;
 end;
if _n_<&n then delete;
/* calculating bounds */
maxb=8*(3*sum)/((3*&n-1)*(3*&n-1));
minb=(3*sum)/(6*&n-10);
midb=(3*sum)/((&n-4)*(&n+7));
put 'key values sum= ' sum ' minb = ' minb
  ' midb = ' midb ' maxb = ' maxb;
```

```
/* calculating slopes */
kkk=1;endn=&n-1;
do k=1 to endn;
kkbeg=k+1;
do kk=kkbeg to &n;
  sloper(kkk)=(yyval(kk)-yyval(k))/(kk-k);
  loopcount=loopcount+1;
  if sloper(kkk)>minb & sloper(kkk)<maxb then
  if sloper(kkk)>minb & sloper(kkk)<maxb then kkk=kkk+1;
end;
end;
sloper(kkk)=99999;
output;
/* finding nonmissing slopes */
data sortslop;
set stats;
array sloper(2500);
keep slopcand;
do k=1 to 2500;
 if sloper(k) ne 99999 then slopcand=sloper(k);
 if sloper(k) ne 99999 then output;
end;
/* finding median etc. of slopes */
```

```
proc print;
proc means mean median min max q1 q3 n var data=sortslop;
  output out=slopeout median=med;
/* singling out sum of w's from previous data */
data feedft0;
set prestats;
if _n_ ne &n then delete;
/* making a dataset with sum of w's and median slope in it
to calculate t0 */
data findt0;
merge slopeout feedft0;
retain med sum;
ourt0 = (.25*(3.*med*&n+3.*med+sqrt(9.*med*med*&n*&n
   -6.*med*med*&n+med*med-24.*med*sum)))/med;
ourb1=med;
output;
data findb0;
set findt0;
ourb0=((sum+&n*y1)-(.5*med*ourt0*(ourt0+1))-(&n*med*ourt0)
  +(med*ourt0*ourt0))/&n;
output;
/* keeping the estimates */
data ourscln;
```

```
set findb0;

keep ourb0 ourb1 ourt0;

data twoests;

merge qntcln ourscln;

data twoest2;

merge qvals2 twoests ;

keep b0 b1 t0q ourb1 ourb0 ourt0 ourt0r;

ourt0r=floor(ourt0); output;

proc print;

run;
```

REFERENCES

Clark, S.T., Odell, D.K., and Lacinak, C.T. (2000), Aspects of Gorwth in Captive Killer Whales (Orcinus Orca), Marine Mammal Science, 16(1), 110-123.

Clark, S. and Odell, D. (1999), "Nursing parameters in captive killer whales (Orcinus orca)," *Zoo Biology*, 18, 373-384.

Chen, X. R. (1987), "Testing and Interval Estimation in a Change-point Model Allowing at most One Change," *Technical Report 87-25*, Center for Multivariate Analysis, University of Pittsburgh.

Ciuperca, G. (2009), "S-Estimator in Change-Point Random Model with Long Memory," pre-print from http://arxiv.org/abs/0906.1710v1, last accessed on September 23, 2009.

Dagenais, M.G. (1969), "A threshold regression model," *Econometrica*, 37, 193-203.

DeGaetano, A. T. (2006), "Attributes of Several Methods for Detecting Discontinuities in Mean Temperature Series," *Journal of Climate*, 19, 838-853.

Ghosh, P., Basu, S., and Tiwari, R.C. (2009), "Bayesian Analysis of Cancer Rates From SEER Program Using Parametric and Semiparametric Joinpoint Regression Models," *Journal of the American Statistical Association*, 104, 439-452.

Gijbels, I., and Goderniaux, A. (2004), "Bandwidth Selection for Changepoint Estimation in Nonparametric Regression," Technometrics, 46, 76-86.

Gill, R. (2004), "Maximum Likelihood Estimation in Generalized Broken-Line Regression," *The Canadian Journal of Statistics*, 32, 227-238.

Goldfeld, S.M. and Quandt, R.E. (1972), *Nonlinear Methods in Econometrics*, Amsterdam: North-Holland Pub. Co.

Gregoire, G., and Hamrouni, Z. (2002), "Two Non-Parametric Tests for Change-Point Problems," *Nonparametric Statistics*, 14, 87-112.

Hinde, R. A. and Atkinson, S., (1970), "Assessing the Roles of Social Partners in Maintaining Mutual Proximity, as Exemplified by Mother-Infant Relations in Rhesus Monkeys," *Animal Behaviour*, 18, 169-176.

Jandhyala, V.K., and Fotopoulos, S.B. (1999), "Capturing the Distributional Behavior of the Maximum Likelihood Estimator of a Changepoint," *Biometrika*, 86, 129-140.

Khodadadi, A., and Asgharian, M. (November, 2008), "Change-point problem and Regression: An Annotated Bibliography," *COBRA Preprint Series, Article 44*, http://biostats.bepress.com/cobra/ps/art44, last accessed on September 13, 2009.

Koul, Hira L. (2000), "Fitting a two phase linear regression model," *Journal of the Indian Statistical Association*, 38, 331-353.

Krishnaiah, P.R., and Maio, B. Q. (1988), "Review About Estimation of Change Points," in *Handbook of Statistics, Volume 7 (Quality Control and Reliability)*, Eds. Krishnaiah, P.R., and Rao, C.R., Amsterdam: Elsevier Science Pub.

Mahmoud, M.A., Parker, P.A., Woodall, W.H., and Hawkins, D.M. (2006), "A Change Point Method for Linear Profile Data," *Journal of Quality and Reliability Engineering International*, 23, 247-268.

Quandt, R. E. (1958), "The Estimation of the Parameters of a Linear-Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53, 873-880.

Quandt, R. E. (1960), "Tests of the Hypothesis That a Linear-Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324-330.

Quandt, R. E. (1972), "New Approach to Estimating Switching Regressions," *Journal of the American Statistical Association*, 67, 306-317.

Yu, B., Barrett, M.J., Kim, H-J, Feuer, E.J. (2007), "Estimating joinpoints in continuous time scale for multiple change-point models," *Computational Statistics & Data Analysis*, 51, 2420-2427.