THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENTS OF STATISTICS AND ECONOMICS

ANALYSIS OF US PER CAPITA INCOME GROWTH USING MARKOV CHAINS

BRYNNE MCGARRY
SPRING 2013

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Statistics and Economics
with interdisciplinary honors Statistics and Economics

Reviewed and approved* by the following:

David Hunter
Statistics Department Head
Thesis Supervisor

John Fricks
Associate Professor in Statistics
Honors Adviser

Russell Paul Chuderewicz
Senior Lecturer in Economics
Faculty Reader

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Despite its constant presence in the news, the discussion of growth theory in economics rarely emphasizes its deep mathematical aspects. While our nation's economic history combined with recent social and political events provide the driving forces behind this growth, the story can be verified quantitatively. This paper considers one method for looking at economic growth in the US through a technical lens; namely, we treat changes in income growth over the last 82 years as a Markov chain and use maximum likelihood estimation to find time points where changes appear most drastic. After exploring the data and discussing some of the finer points of Markov processes, the algorithm for finding one change point is discussed in depth and extended to two change points and the general case.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. David Hunter and Dr. John Fricks for advising me on this thesis and working with me throughout the entire process.

Chapter 1

Introduction

In today's information-driven society, economic forecasting is probably one of the most important, and also most criticized, forms of prediction. Economists' failure to foresee the financial and economic crisis of 2007 has shed an unfortunate light on the profession – in particular people want to know what caused experts to miss such obvious signs of trouble. With sophisticated software programs, various data collecting methods, and well-trained analysts, one would think that the predictive power of economic forecasting would be improving through time, but recent events tell a different story. My interest in such data and the various analyses that can be applied to it forms the basis for this honors thesis.

In his book "The Signal and the Noise," economist Nate Silver examines various fields that utilize data and prediction. These range from weather forecasting to baseball statistics to politics, and he also includes a chapter on economic forecasting. While each of these fields presents unique questions for experts to explore, all experts face the same problem of distinguishing the signal from the noise; in other words, finding trends that most closely match real-world events. The problem, Silver believes, is that our world is so full of information and we are so focused on always gathering more that the signal is constantly swallowed by the noise. While it may seem intuitive that more information should lead to better results, the best predictive models adhere to parsimony.

In his chapter on economic forecasting, Silver interviews Jan Hatzius, chief economist of the investment bank Goldman Sachs (Silver 184-187). Hatzius names two to three dozen economic variables that contain the most "substance," or hold the most predictive power. These variables fall into a few major categories, one of which is wage and income. For this paper, I have gathered annual data from the Bureau of Economic Analysis on per capita personal income for each US state and for each year going back to 1929. I will use these data to identify "change points" in the US economy over the last 82 years. By finding points in time where the data undergoes significant changes and possibly matching those to major economic events, I will be able to show that economic data does in fact have merit and in more sophisticated analyses, can be utilized to predict major events. I certainly cannot claim that I will make any major breakthroughs, but from an economic perspective, a thorough understanding of the data cannot be oversold. Through this analysis, I hope to better comprehend the often-misunderstood ideas behind economic growth.

The rest of the paper is organized as follows: first, I introduce the data, as well as a few time plots to illustrate how the data change over time. Then I delve into properties of stochastic processes with probability transition matrices that show a state's likelihood of growing or shrinking economically when beginning in some given income category. Finally, I attempt to use more sophisticated statistical practices to identify the "change points" previously mentioned and analyze how they fit into our nation's economic history.

I gathered my data from the Bureau of Economic Analysis website (bea.gov). The data consist of per capita personal income for each of the continental states of the United States and the District of Columbia as its percentage of the average US value for each year from 1929 to 2011. Throughout this paper, I use the term "growth" often. I am always referring to a how state's personal per-capita income as a percentage of the US average has changed over a certain time period. Additionally, statistical analyses and calculations are carried out using R (R Core Team 2012).

To begin working with the data, I calculated each state's change in per capita personal income from each year to the next. Manipulating the data in this way allows me to work with transition values, rather than levels at each year. The data range from approximately -28% to 50%. In other words, at some point over the 82 year sample, some state's economy grew 50% from one year to the next, and at another point, a state's economy shrunk by nearly 30%. While the data seem somewhat volatile over this range, this statistic proved to be the best way to universally compare growth, both geographically and in time.

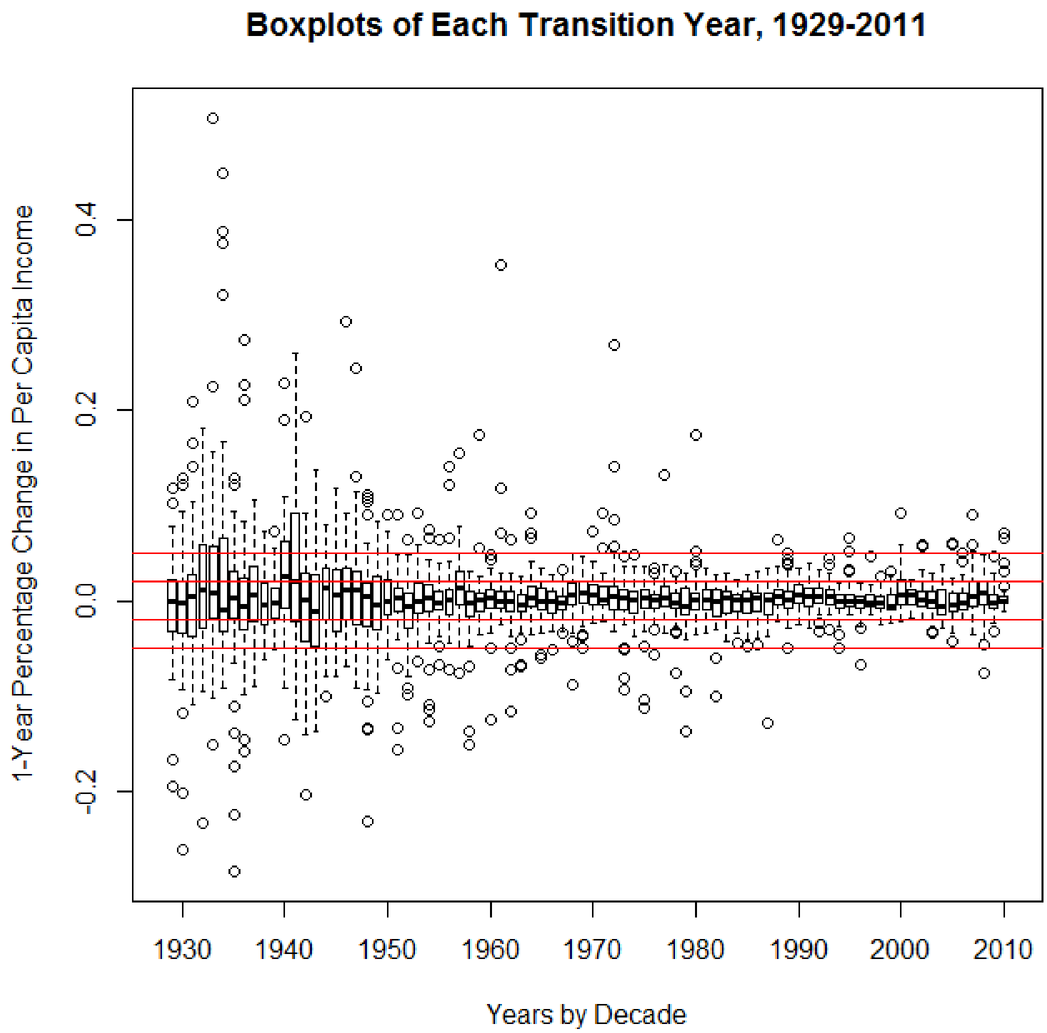

Figure 1-1: These boxplots show the distribution of states' percentage change in per-capita personal income for each year in the dataset. The red lines at -5%, -2%, 2%, and 5% help to show how the data change less and less over time.

Figure 1-1 shows a boxplot for each year in the dataset. The sizes of the boxes show that for the first 30 or so years, states experienced more variation in per-capita income changes. As the boxes shrink from left to right on the timeline, the US appears to enter a more steady state (the boxes shrink because the gap between the first and third quartile shrinks, meaning most observations for that year are contained in a smaller interval). This trend is emphasized by the red cutoffs, in that from about 1960 on, the boxes for each year are entirely contained between the line at -2% and the line at 2%. Furthermore, after about 1990, no state experiences a one-year change in income of more than 10% in either direction.

The fan shape of the boxes suggests that the pattern of the data changes over time, but when and where do these changes occur? The change appears to happen gradually, but is it possible to pinpoint a year, or multiple years, in which the change is most significant? These questions are explored further in chapter 3.

The maps in Figures 1-2 and 1-3 show each state's change in per-capita personal income for the first transition year and the last transition year. The intervals used to categorize the states are described in detail in the next chapter.

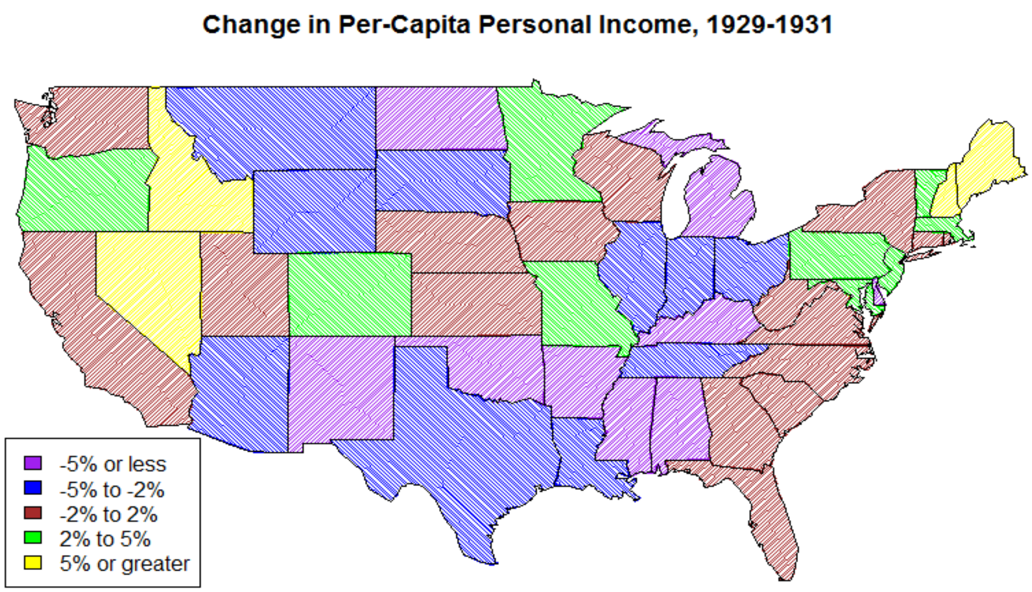

Figure 1-2: This map visually represents how each state's wealth changed over the first two years of data

Clearly over the first two transition years in the sample, the economy was more volatile than in later years. Each transition category is represented about equally, meaning that across all states, the probability for no growth was comparable to the probabilities of major gains or losses. By contrast, Figure 1-3 shows that changes among states are much more homogeneous in 2009-11 than in 1929-31. The results of these maps are consistent with those from the above boxplots. The fracking boom in North Dakota following the recent financial crisis explains the higher levels of growth only four states not in the middle category, which represents little change. Beginning in late 2008, ongoing oil extraction in North Dakota has led to rapid growth and the lowest unemployment rate in the US, recently calculated at 3.2% and not touching 5% in over 25 years (Oldham 2012).

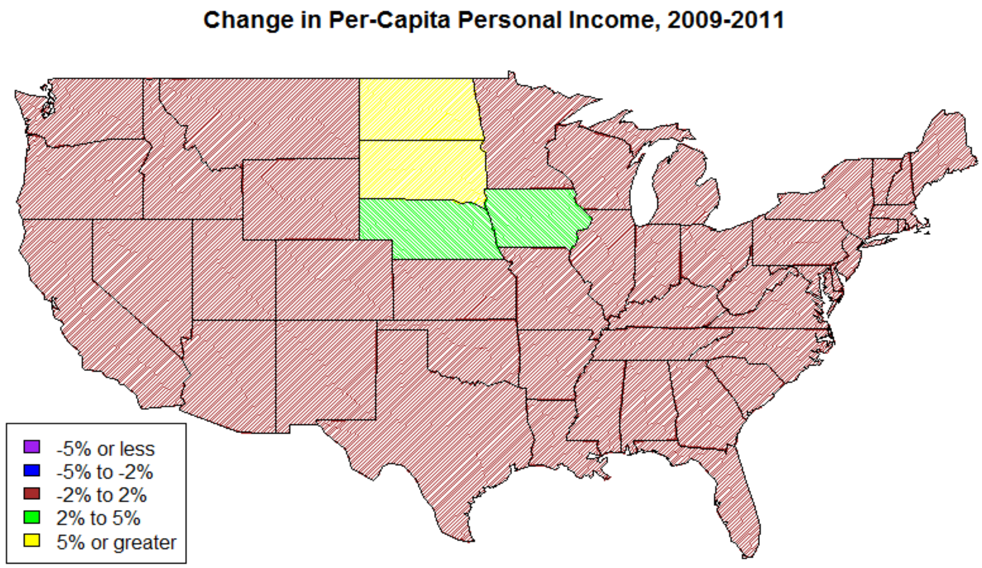

Figure 1-3: This map shows the same as Figure 1-2, but for the two most recent years of data

Chapter 2

Growth Viewed as a Markov Chain

This thesis examines economic trends by treating time series data of this sort as a stochastic process. Variations on this idea will serve as the topic for the rest of this paper. A stochastic process is a probabilistic way to represent data as they evolve over time.

Specifically, I will treat the data here as a discrete-time, finite space Markov chain. For our purposes, suppose $\{X_n, n=1,\ldots,82\}$ is a stochastic process that describes a state's growth change in the years from 1929-2011. If $X_n = i$, the process is experiencing some growth level i at time n and can move to a different level of growth j at time $n+1$. The probability that this transition occurs is denoted P_{ij} and is given by equation (2.1)

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij} \qquad (2.1)$$

Each P_{ij} depends only on a state's current state rather than its history over multiple time points. Discrete time refers to analyzing transitions from year to year, rather than at continuous time points. Finite space means that X_n can only take on finitely many values.

Treating the data as a discrete-time, finite space Markov chain relies on a major assumption. Markov chains are memoryless, that is, the conditional probability distribution of transitions between categories depends only on the process' current category, rather than on its entire history of transitions. This is also called the Markov Property (Ross, 1972, p. 185). In order to use a Markov chain to describe this data, we must assume that a state's per-capita personal income growth for a given year is determined only on the year immediately preceding it, ignoring all other past years. In economics, this situation is somewhat unrealistic, but for the purposes of this model, we will continue on with this assumption. The Markov chain can be described by a matrix P of probabilities that the chain moves from one category to another given its current situation, known as a probability transition matrix or PTM (Ross, 1972, p. 185). Entries in P are calculated as the probability of moving from category i to category j. These probabilities depend only on a state's current category rather than its entire transition history. Finally, we must assume time homogeneity, or that the joint probability distribution of the process remains the same over time. Therefore, we can start the process in any year and use the PTM to describe movements in growth the following year.

In the context of this paper, the PTMs illustrate how likely a state is to experience changes in growth, given their current growth. The categories for the PTMs in this paper will be the same ones that I used in the maps in chapter 1: a state's income can decrease by more than 5% from one year to another, decrease from 5% to 2%, change somewhere between -2% and 2%, grow between 2% and 5%, or grow more than 5%. In the table below, these categories are numbered from 1 to 5. I chose these cutoffs simply by analyzing the data using various categories and seeing which ones yielded the most interesting results. In the context of this paper, interesting results occur when there are more movements across intervals from year to year. For instance, when increasing the middle intervals to -10% to -5%, -5% to 5%, and 5% to 10%, changes are much rarer across the entire dataset. Subtly shrinking the intervals made it easier to generalize where changes may occur as a whole and should help in finding distinct change points, as I experiment with in chapter 3.

From this point forward, I will rely heavily on probability transition matrices to describe how the data change according to a Markov chain. Because the PTM is a matrix of probabilities, each entry is nonnegative. Further, the rows must sum to one, as the entries in the i^{th} rows are the probabilities of the process moving to all possible categories *j* given that it is currently in state *i*. Given the current category of some state, the state must either stay in the same category (an event whose probability is on the diagonal of the matrix) or move to a new category (an event whose probability is off the diagonal).

When using the data to estimate the PTM, each entry in the estimated matrix is a maximum likelihood estimator (MLE) of the actual (population) probability and is calculated using the sample data and techniques in statistical inference. Specifically, the MLE for each p_{ab} in the matrix maximizes some likelihood function. The likelihood function is the joint density function evaluated at the observed data and viewed as a function of the model parameters. It is expressed mathematically as

$$L(\Theta) = P(y|\Theta), \tag{2.2}$$

where we suppress the dependence of Lon the data y in equation (2.2).

The process of computing each MLE follows, but first, I will make some assumptions. First, I assume that the number of states in each of the five transition categories (determined in the EDA section of this paper) in the first year, 1929, is fixed and known. I also assume that memberships in these categories are determined by a homogeneous Markov chain, as discussed earlier in this chapter, which means that each state's transition from year to year is dictated by the same probability transition matrix and that these movements are independent of anything that happened in the past. This assumption is stated explicitly in the definition of a Markov chain, as expressed in equation (2.1). Finally, I assume that the data for each geographic state are independent of each other.

The likelihood function L(p) is given by

$$L(p) = \prod_{i=1}^{82}\prod_{j=1}^{5}\left\{\binom{Y_{ij.}}{Y_{ij1}\ \dots\ Y_{ij5}}\left(\prod_{k=1}^{4} p_{jk}^{Y_{ijk}}\right)(1 - p_{j1} - p_{j2} - p_{j3} - p_{j4})^{Y_{ij5}}\right\} \tag{2.3}$$

where each Y_{ijk} represents the number of states that move from category j to category k in some year i. The variable $Y_{ij.}$ refers to the number of states in category j in some year i when summing over all k transition categories. Each p_{jk} is the jk^{th} entry of the probability transition matrix. Because the rows of the matrix sum to one, we only count p_{jk} up to k =4 and use one minus the other 4 probabilities to equal $p_{j5.}$ We count i from 1 to 81 to represent 81 transitions over 82 years of data, and j from 1 to 5 to represent the five transition categories. Note that the joint probability of y given in equation (2.3) is the result of conditioning: we begin with Y_1 (the income growth of each state for the first year), then multiply it by the conditional probability of Y_2 given Y_1, then multiply that by the conditional probability of Y_3 given Y_1 and Y_2, and so on. We know that by the Markov property, these conditional probabilities actually do not depend on the data for all years preceding Y_i, but only on the data for the most recent year. Additionally, equation (2.3) does not explicitly include the density of Y_1 because we are not actually interested in the initial parameters. Instead, we condition on Y_1 throughout and treat these initial probabilities as nuisance parameters.

In order to find the maximum likelihood estimates for the p_{jk}, we take the log likelihood function as

$$\log L(p) = \sum_{i=1}^{82}\sum_{j=1}^{5}\left\{\log\binom{Y_{ij.}}{Y_{ij1}\ \dots\ Y_{ij5}} + \sum_{k=1}^{4} Y_{ijk}\log(p_{jk}) + Y_{ij5}(1 - p_{j1} - \dots - p_{j4})\right\}. \tag{2.4}$$

This function is maximized where its derivative

$$\frac{\partial}{\partial P_{ab}}\log L(p) = \frac{\partial}{\partial p_{ab}}\left[\sum_k \sum_j (\log p_{jk})(Y_{.jk}) + \sum_j \{\log(1 - p_{j1} \dots - p_{j4})(Y_{.j5})\}\right] \tag{2.5}$$

equals zero. Setting equation (2.5) equal to zero, we end up with

$$\frac{Y_{.ab}}{p_{ab}} = \frac{Y_{.a5}}{p_{a5}}\text{, which implies } \widehat{p_{ab}} = \frac{Y_{.a5}}{Y_{.a.}} \tag{2.6}$$

for all $1 \leq a \leq 5$ and $1 \leq b \leq 5$. So, we can find the maximum likelihood estimate of each p_{ab} using the number of times any state transitions from category *a* to category 5 divided by the number of total transitions to category 5 over all 81 transition years.

Using the categories depicted in Figures 1-2 and 1-3, the empirical probability transition matrix for the entire dataset is shown below. Each entry is the maximum likelihood estimator of each p_{ab}. The first column of the matrix gives the total number (count) of transitions that begin in each interval.

Number		1	2	3	4	5
175	1	0.109	0.171	0.263	0.166	0.291
450	2	0.051	0.189	0.529	0.158	0.073
2599	3	0.017	0.083	0.770	0.104	0.027
504	4	0.046	0.129	0.573	0.187	0.065
241	5	0.232	0.187	0.228	0.137	0.216

Table 2-1: This is the estimated probability transition matrix that communicates a state's likelihood of moving between growth categories, given where they currently stand

For instance, we see the largest probability, 0.77, in the p_{33} cell. This means that over the entire 82 year period, if a state has just experienced a growth change somewhere between 2% and -2%, the probability that they experience the same change in the following year is about 77%. In fact, the p_{i3} probabilities are all fairly high, with the exception of the first row, where p_{15} is the greatest. It makes sense that these probabilities would be fairly high, as the EDA showed that most growth changes happen in this range and become more and more common towards the present. Some of the extreme transition probabilities are rather high; for instance, a state whose economy has just shrunk by more than 5% has a 29.1% probability of growing by 5% or more the next year. In light of the high volatility across the board in the early years, this probability may not be very surprising.

The variability of the estimator for p_{ab} is a question worth exploring. While the above probabilities are calculated using maximum likelihood estimation, and are therefore considered the "best" estimates for the p_{ab},, they are still subject to error. One way to assess this error is to construct confidence intervals for each probability. A simple way to do this would be to treat the probabilities as binomial proportions and use a normal approximation interval. The formula for the interval is given by

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \tag{2.7}$$

where $\hat{p}$ is equivalent to $\hat{p}_{ab}$ above, that is, the number of times a transition from category *a* to category *b* occurs divided by the number of times the process is in category *a; n* is the number of times the process is in category *a*; 1-α is the confidence level, and *z* is the corresponding percentile of the standard normal distribution. In this case, a transition from category *a* to category *b* is considered a "success," and therefore $\hat{p}$ is the proportion of successes in *n* Bernoulli trials, thus creating the binomial situation. Applying this method to the (1,1) entry in table 2-1 yields an interval of (0.063, 0.155). Therefore, we can say with 95% confidence that $p_{1,1}$ could range from 6.3% to 15.5%. It is important to remember that this interpretation is only valid if all assumptions of the model are met.

After actually calculating one of these intervals using a normal approximation of a binomial proportion, it is easy to see where this method could encounter some difficulty. First, these intervals ignore the fact that the rows of the matrix must still sum to one. This is most easily accomplished when pinpointing one value for each p_{ab}. Second, small values of $\hat{p}$ can lead to problems; namely, a slower convergence to the normal approximation and intervals that contain zero, which cannot be applied to nonnegative probabilities. Finally, in situations with small sample sizes, it is not necessary that *n* be greater than zero. If the process is never in a certain category, the interval for $\hat{p}$ is zero for that category. We do not explore other methods of accounting for these issues in this paper.

Figure 2-1 visually communicates the information represented in table 2-1:

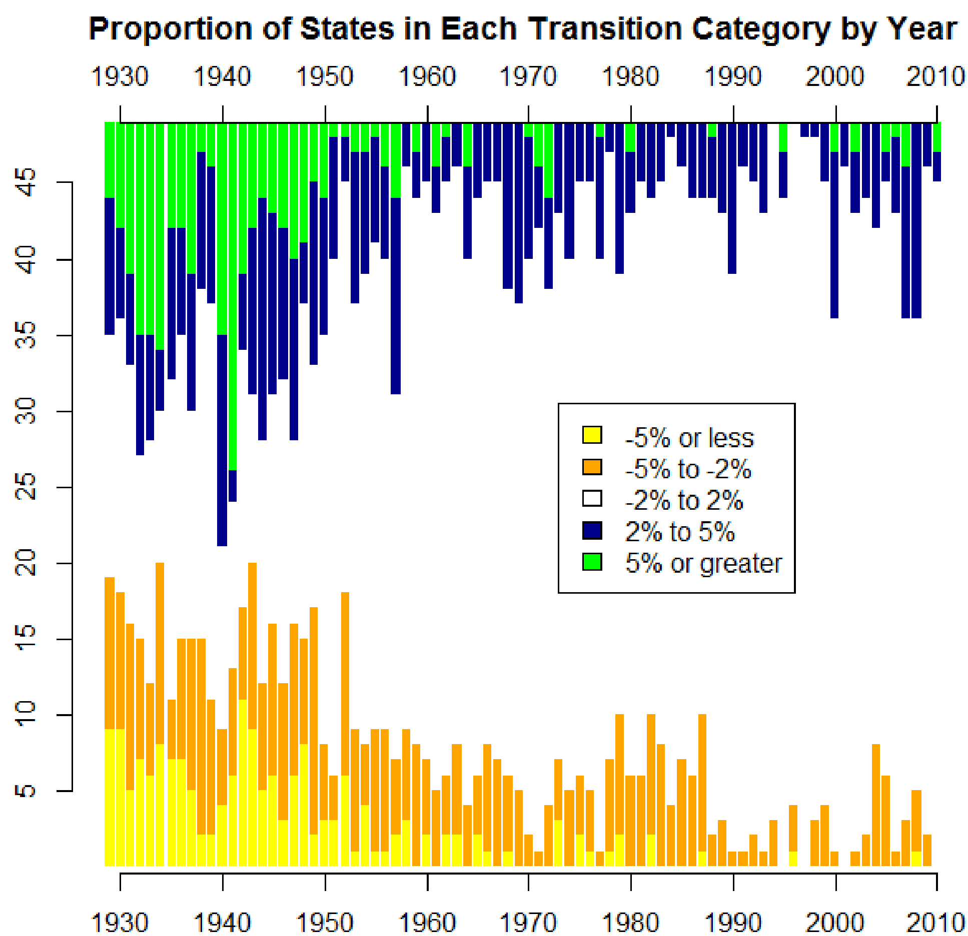

Figure 2-1: The heights of the colored bars are proportional to the number of states in each transition category for the years 1929-2011. The legend defines these categories.

Figure 2-1 categorizes the data by the five intervals defined above, representing a state's change in per-capita personal income from the previous year. The intervals are re-stated in the legend above. The changes in the heights bars show how the data have changed in the period of study. Clearly, the white portion of each bar (representing the -2% to 2% growth range) increases across time, while the extreme yellow and green categories die out and are barely present at all by 1990. These results may suggest that using a model with a fixed PTM operating over time is invalid. The discussion of the stationary distribution that follows helps explain Figure 2-1 more appropriately.

I also examine the stationary distribution to determine the long-run proportions of time spent in each interval. These proportions may be summarized in a single vector whose dimension is the number of categories; we call this vector the stationary distribution of the Markov chain and denote it by π. It satisfies the equation

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}.$$

(2.8),

where p_{ij} is the single time-independent matrix described in Table 3-1, and the π_j are non-negative and sum to one (Ross, 1972, p. 205). For the stationary distribution to be unique, the Markov chain must be irreducible, that is, it must be possible to move from any category to any other category. The process must also be aperiodic, which means that there are only finitely many *n* such that the process cannot get back to *i* from *i* in exactly *n* steps (Ross, 1972, p. 204).

Under the above assumptions, the j^{th} entry in the stationary distribution vector is calculated as the limit of P_{ij}^n as *n* approaches infinity, no matter what value *i* takes (Ross, 1972, p. 204). That is, we can find the probability (independent of where the process begins) that a state will be in category *j* in the long run by multiplying the original matrix by itself over and over. Using this method in R, I calculated the stationary distribution as it appears in table 2-2.

1	2	3	4	5
0.039	0.110	0.667	0.125	0.059

Table 2-2, Stationary Probabilities: This table shows the long-run proportion of time that US states spend in each income category

Assuming we have a chain that is finite, irreducible, and aperiodic, as discussed above, these are the probabilities that US states spend in each interval in the long run. The stationary distribution does not depend on the starting distribution; under these assumptions, the chain will converge to the above distribution independent of where it begins.

According to these results, in the long run, US states spend the most time in the middle interval, less time in the ones on either side of the middle, and the least time in the extreme intervals. This makes intuitive sense based on what we have seen so far. Across the entire sample, we have seen that the data converge to a situation where states are most likely to stay in the middle interval.

One should be able to recreate similar analyses using smaller subsets of the data. For instance, how likely are these transitions over the first half of the data as compared to the last half? As it turns out, this question is not difficult to answer. Using the same method as above, I create an estimated probability transition matrix for the first half of the transition years (1929-1970) and another for the last half (1971-2011) in Tables 2-3 and 2-4, respectively.

Number		1	2	3	4	5
322	1	0.105	0.168	0.280	0.155	0.292
607	2	0.072	0.183	0.464	0.183	0.098
2006	3	0.039	0.126	0.636	0.140	0.059
602	4	0.072	0.146	0.521	0.173	0.088
432	5	0.236	0.194	0.213	0.134	0.223

Table 2-3, Probability Transition Matrix: The above matrix estimates the same information as Table 2-1, but only over the years 1929-1970.

Number		1	2	3	4	5
28	1	0.143	0.214	0.071	0.286	0.286
288	2	0.007	0.201	0.667	0.104	0.021
3205	3	0.003	0.055	0.856	0.080	0.006
396	4	0.005	0.102	0.655	0.208	0.030
52	5	0.200	0.120	0.360	0.160	0.160

Table 2-4, Probability Transition Matrix: The above matrix estimates the same information as Table 2-1, but only over the years 1971-2011.

The category counts in the first column of Table 2-4 show that by 1971, states had begun to converge into the middle category; that is, it became less likely that states would begin a transition year in one of the more extreme categories. We also observe that in Table 2-4, three of the five estimated probabilities in the first column amount to less than 1%. Intuitively, this means that by 1971, it had become rare for states to transition into the lowest income category from the middle three categories (in other words, moderate or small changes in income one year did not often predict large decreases in income the next year). The corresponding estimated probabilities in Table 2-3 are relatively small as well (all less than 10%), but this is still an interesting change over time. We notice that the same can be said for the middle estimated probabilities in the fifth column of the second matrix. While each one was less than or equal to 10% in the beginning years, they are all less than or equal to 3% over the second half. Therefore, the chance of a state's economy significantly shrinking not only decreased, but the chance of a state's economy significantly growing decreased as well.

Table 2-5 shows the corresponding stationary distributions for the first and second halves of data.

	1	2	3	4	5
1929-1970	0.074	0.148	0.523	0.152	0.103
1971-2011	0.007	0.072	0.811	0.097	0.013

Table 2-5, Stationary Probabilities: This table shows the long-run proportion of time that US states spend in each income category for the two first halves of the dataset represented in Tables 2-3 and 2-4

Given Figures 1-1 and 2-1, it seems reasonable that we would see more uniform stationary probabilities for the years 1929-1970 than for the years 1971-2011. Over the earlier time period, changes between categories are less homogeneous than in the later time period and this distribution suggests that while states do spend the most time in category three over these early years, the other categories have not completely died out yet. For the years 1970-2011, we would expect the stationary probabilities for the extreme intervals to decrease, giving more weight to the middle categories. Table 2-5 shows that time spent in category one does decrease (nearly to zero) during this period and states spend by far the most time in category 3 over these years. The probability associated with category five also decreases greatly to only about 1%, and those corresponding to two and four decrease as well.

Another interesting way to analyze subsets of the data is to look at differences in the probabilities associated with eastern and western states. The intuition here relies on the economy of the West being more agriculturally focused and therefore subject to fluctuations in growth due to factors like the weather, while the East is more industrial and ideally sees more stable growth. Of course, there exist exceptions to this idea. Certainly the West has modernized over the 82 years of sample data, but this general idea will form the basis for this particular analysis.

Tables 2-6 and 2-7 describe the estimated probability transition matrices for the East and West, respectively. For simplicity, I have defined the east and west as the states on either side of the Mississippi River (Minnesota falls in the West for the purposes of this analysis). Another interesting comparison would be between, say, agricultural and urban states by population density; however, I do not cover that in this paper.

Number		1	2	3	4	5
69	1	0.145	0.203	0.391	0.159	0.102
206	2	0.049	0.174	0.583	0.136	0.058
1573	3	0.016	0.073	0.803	0.092	0.016
255	4	0.035	0.082	0.620	0.216	0.047
84	5	0.119	0.190	0.262	0.131	0.298

Table 2-6, Probability Transition Matrix: The above matrix estimates the transition probabilities for states east of the Mississippi River.

Number		1	2	3	4	5
106	1	0.085	0.151	0.179	0.170	0.415
244	2	0.053	0.201	0.484	0.176	0.086
1026	3	0.019	0.097	0.719	0.122	0.043
249	4	0.056	0.177	0.526	0.157	0.084
157	5	0.293	0.185	0.210	0.140	0.172

Table 2-7, Probability Transition Matrix: The above matrix estimates the transition probabilities for states west of the Mississippi River.

Tables 2-6 and 2-7 are revealing. The probabilities on the diagonal are fairly high for the East, while several of the off-diagonal probabilities are high for the West. For instance, both the transition from category one to category five and the transition from category five to category one in the West are much higher than the corresponding values in the East. The fact that such extreme movements are more possible in the West suggests a more volatile economy, as one might see in an agricultural system. (As an example, the current economic boom in North Dakota is discussed briefly in chapter 1). Interestingly, the probabilities in the third column in the East (transitions into category three from any category) are noticeably high. This could suggest that economies in the East stabilized more easily over time.

	1	2	3	4	5
East	0.028	0.091	0.731	0.113	0.037
West	0.056	0.134	0.586	0.138	0.086

Table 2-8, Stationary Probabilities: This table shows the long-run proportion of time that eastern and western US states spend in each income category

Table 2-8 describes the stationary distributions of the two regions, lending further support to the claims made above. While the stationary probabilities for the West certainly are not uniform, they are more evenly distributed than those for the East. In the long run, western states probabilistically spend more time in the extreme intervals than do eastern states, while eastern states are more likely to settle in category three. As this analysis shows, using Markov chains to describe smaller subsets of the data can mathematically verify intuitive claims about US economic history in the context of this dataset.

Up to this point, I have used probability transition matrices to show how likely states are to move from category to category when considering all 82 years of data, as well as when examining the first half and second half of the data separately. I chose the halfway point somewhat arbitrarily simply to show how these probabilities can change depending on the data we are interested in considering. However, this change point may not be the best year to examine changes in the probability transition matrix. For this reason, finding out where changes are most likely to occur using a statistical approach is an interesting question to consider. The next chapter explores this process and its implications.

Chapter 3

Finding Change Points Using Maximum Likelihood Estimation

The likelihood function in the previous chapter represents the simplest situation, in which we are not accounting for change points. Because the probabilities of movements between categories will fluctuate most dramatically at these points in time, change points are probably associated with important events in our nation's economic history.

When attempting to identify a change point, we must find the year (in the interval $T=2$ to $T=80$, where T is the number of years since 1929), as well as the values of p and q that maximize the likelihood function. The log likelihood function for this situation is shown in equation (3.1). I have taken out the multinomial coefficients as they have no effect on the maximization of this function. This equation is derived from equation (2.4). For simplicity, I now use p_{j5} and q_{j5}, rather than referring to them as one minus the sums of $\{p_{j1} \ldots p_{j4}\}$ and $\{q_{j1} \ldots q_{j4}\}$, respectively.

$$\log L(\mathrm{T}, p, q) = \sum_{i=1}^{T-1} \sum_{j=1}^{5} \sum_{k=1}^{5} (Y_{ijk} \log p_{jk}) + \sum_{i=T}^{81} \sum_{j=1}^{5} \sum_{k=1}^{5} (Y_{ijk} \log q_{jk}) \tag{3.1}$$

For simplicity, I will define the profile likelihood function based on equation (3.1). The profile likelihood eliminates certain parameters by maximizing over their potential values while retaining the one(s) that we are interested in. Equation (3.2) defines the profile likelihood function in the context of this chapter, where we wish to focus on the change point parameter T.

$$L_{prof}(T) \stackrel{d}{=} max_{p,q}\, L(p, q, T) = L(\hat{p}_T, \hat{q}_T, T) \tag{3.2}$$

where $\hat{p}_T$ and $\hat{q}_T$, respectively, are the maximum likelihood estimators of p and q for a given fixed value of T. This says that the profile likelihood with respect to T is maximized over transition probabilities p and q, where T is the transition year in the interval 1:81 that is most likely to serve as a change point. The values of $\hat{p}_T$ and $\hat{q}_T$ may be calculated using equation (2.6), where the sums indicated by the subscripted dot are taken over 1 to $(T-1)$ and T to 81, respectively. We are primarily concerned with finding the best possible value for *T*, for which we use the maximum of the profile likelihood.

Figure 3-1 plots each year against the value of the profile likelihood function for the corresponding T.

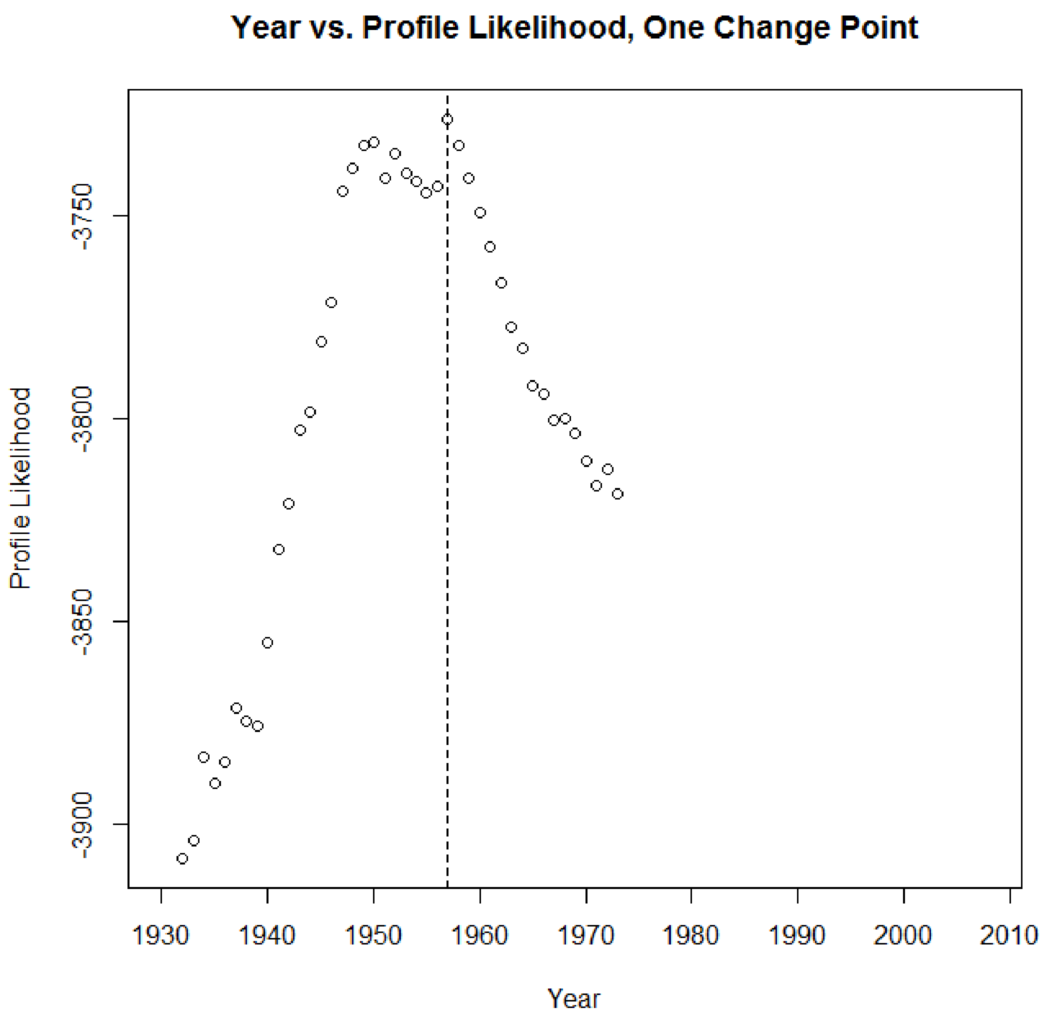

Figure 3-1: This is a plot of *(T+1929)* against $L_{prof}(T)$. The profile likelihood is undefined for values of T greater than and equal to 45, or years from 1976 on.

As we can see, there appears to be a point in time that maximizes the function at the year 1958. Additionally, the Figure 3-1 stops plotting points in 1976. For values of *T* close enough to 2 or 80, the values of Y_{jk} equal zero for some values of *j* and *k*, which means that the corresponding values of $\hat{p}_T$ or $\hat{q}_T$ will also be zero. In terms of equation (3.1), the profile likelihood will contain terms that are zero times the log of zero, which is undefined. By looking back at Figure 2-1, we see that by 1970, some of the transition intervals are not represented in the bar graph for certain years. Figure 3-1 proves that for each year beginning in the mid-1970s and after, there is at least one probability estimate of zero in each year's probability transition matrix.

Economically, this peak in the maximum likelihood function at 1960 suggests that the nature of the US economy changed around this time; specifically, larger swings in per-capita personal income became less possible. As it turns out, this may not be the best way to interpret these results. Historically, the US economy of the 1950s was more or less static, as life had settled down since World War II and many Americans enjoyed a generally comfortable, though not overwhelmingly prosperous, living situation. Shortly thereafter, however, the 1960s brought about positive change through avenues such as President John F. Kennedy's "New Frontier" and President Lyndon B. Johnson's "Great Society." By the end of the decade, the US had seen the creation of Medicare, increased grants to schools, and an overall war on poverty. Therefore, this economic change point should probably be interpreted as more gradual than sharp, as Figure 2-1 suggests. Further, a general change in economic prosperity came in the 1970s with a couple of key economic events. The first was the collapse of the Bretton Woods system, when President Richard Nixon removed the gold backing of the dollar and upset the international currency system. While modern economists would probably agree that the gold standard was a flawed system, this upsetting of the status quo certainly interrupted the prosperity of the previous decade. Additionally, the 1973 oil crisis put additional pressure on economies all over the world when the Organization of Petroleum Exporting Countries issued an embargo that increased the price of oil dramatically. This somewhat dichotomic state of the economy over a period of just a few years would suggest to me that a more realistic change point would have appeared closer to 1970 rather than at 1960.

We can also consider a situation in which we find the two most likely years of change over this span of data. The method is similar to that laid out at the beginning of this chapter, but takes into account not just one maximum value of T, but two maxima, T_1 and T_2. The equation for this is given in (3.3).

$$\log L(T_1, T_2, p, q) = \sum_{i=1}^{T_1-1} \sum_{j=1}^{5} \sum_{k=1}^{5} (Y_{ijk} \log p_{jk}) + \sum_{i=T_1}^{T_2-1} \sum_{j=1}^{5} \sum_{k=1}^{5} (Y_{ijk} \log q_{jk}) + \sum_{i=T_2}^{81} \sum_{j=1}^{5} \sum_{k=1}^{5} (Y_{ijk} \log r_{jk}) \tag{3.3}$$

Figure 3-2 plots each combination of years (for $T_1<T_2<81$) against the value of the profile likelihood function for each pair of years.

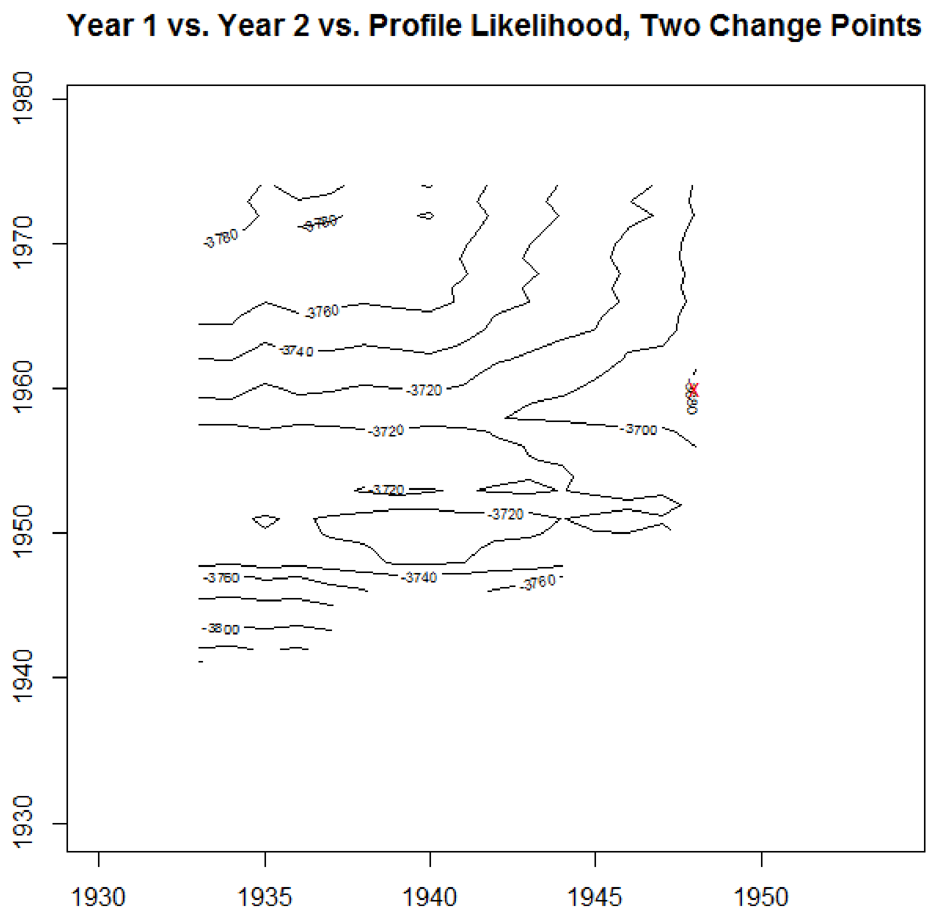

Figure 3-2: This is a plot of *(T_1+1929)*, *(T_2+1929)*, and the profile likelihood. The profile likelihood is maximized at $T_1 = 19$, which corresponds to 1948, and $T_2 = 30$, which corresponds to 1959.

The profile likelihood function for the two change point case is maximized when $T_1 = 19$ and $T_2 = 30$. This corresponds to the years 1948 and 1959 and is plotted on Figure 3-2. Continuing with the previous discussion, it perhaps makes sense for these years to serve as change points when considering the two-year case, as World War II ended in the mid-1940s and 1959 led right into the prosperous 1960s discussed previously.

In principle, we might consider extending the methods in this chapter even further to consider three or more change points. However, beyond three change points, the validity of the estimates becomes an issue. As Figure 3-1 shows, the profile likelihood is undefined whenever any of the 25 counts in the matrices for any of the time periods determined by the change points equals zero. Specifically, Figure 3-1 says that the profile likelihood is undefined for all years after and including 1976. Therefore, we cannot use the method of maximum likelihood estimation laid out in this chapter for more than a few change points. We simply are left with too few data to consider a truly “general” case, and reliable estimates for change points reach a natural limit fairly quickly.

Chapter 4

Discussion and Conclusions

While simple statistical models exist, they often overlook certain nuances when applied to complex economic data. There are simply too many phenomena and too much variability to succinctly sum up why our economy behaves the way it does. However, by using quantitative techniques already applied to a variety of real-world problems, we can at least begin to understand the driving forces behind key economic indicators like growth and income. This paper looked at available data on per-capita personal income over the last 8 decades to uncover trends and identify how our economy has changed over time.

I first quantified each US state's percent change in per-capita personal income from year to year. I then categorized these changes into five intervals, thus creating an overall system for analyzing growth changes. The boxplots and maps in chapter 1 visually communicate a clear trend of economic stabilization; that is, over time, the states' income seems to change less and less from year to year.

After exploring general trends in the data, I created a Markov chain to quantify the transitions. The resulting estimated probability transition matrix (found using the same intervals as I used in the maps and boxplots) summarized the twenty-five probabilities of moving between intervals. The probability of staying in the middle interval from year to year turned out to be the most likely outcome, followed by the other diagonal probabilities, as is consistent with the stabilization hypothesis from the introductory plots. Calculation of the stationary distributions told the same story: in the long run, the US economy is most likely to spend time in the middle transition category. More focused analyses indicate that pre-1970 changes exhibited much more variability than post-1970 changes and that eastern United States have shown more stability than the West both in the probability transition matrix and the stationary distribution, which can perhaps be attributed to the west's more agricultural, and therefore volatile, economy.

It is important to note that the cutoffs for the transition categories, and even the number of categories, were chosen somewhat arbitrarily. Different choices may have led to different specific results; however, given the overall interpretation of Figure 1-1, it seems that the broader results would have been similar. Additionally, I carried out all analyses in this paper under the assumption that all 49 US states in the sample were operating independent Markov chains. This is quite an unrealistic assumption, as economies of states that share geographic proximity probably behave similarly and experience similar trends.

This thesis demonstrates how to calculate change points using maximum likelihood estimation. By looking at how the estimated probability transition matrix changes over different subsets of years, I was able to find the year in which the economy was most likely to have undergone a significant change. The data show that this year was most likely around 1960. Economically, the 1960s were a period of great change, but because this change was rather gradual, I would have expected this change point to appear closer to 1970. I also discussed the possibility of finding multiple change points, noting the limitations of the maximum likelihood method for more than a few change points given the size of this dataset.

Throughout this paper, I used Markov chains to quantify economic trends throughout most of the last century. In other words, I used a specific technique among many that would be appropriate to analyze a simple, though large, dataset. Although the particular model used here includes some fairly strong and probably unrealistic assumption, I have been explicit about what those assumptions are. More sophisticated assumptions would require correspondingly sophisticated analyses, which are beyond the scope of this particular thesis. The Markov chain analysis gives a quantitative way to discuss qualitative trends in the economy, particularly because it enables us to consolidate a rather large dataset into a simple summary. While I have not made any predictive breakthroughs, it is encouraging that my change point analysis identified a particular year when something seems to have shifted under the simplistic model assumed throughout the paper, and that this year has a reasonable economic interpretation. Therefore, I have given examples of some of what can be done with the right amount of statistical background and economic knowledge. Simply knowing what information to consider and what trends to look for is the start to getting the most out of mountains of data.

Appendix A

R Code

```
##Chapter 1

#Read data
x <- read.csv(file="C:/Users/Brynne/desktop/Thesis Data for
.csv",head=TRUE,sep=",")

#Convert to Matrix
y<-as.matrix(x, byrow=F)

#Categorize:
#Change rows in y to analyze transitions over any span of years
#Change cutoffs to fit better with data
tmp <- findInterval(y[,1:82], c(-0.05, -0.02, 0.02, 0.05))

#matrix
data <- matrix(tmp,49,82)+1
#Find transitions
M <- matrix(0,5,5)
for(i in 1:nrow(data)){
for (j in 1:(ncol(data)-1)) {
M[data[i,j], data[i,j+1]] <- 1 + M[data[i,j], data[i,j+1]]
}
}
M
sweep(M, 1, rowSums(M),'/')
par(mfrow=c(1,1), mar=c(6,4,6,2)+.1)

#boxplot of y with trend lines
boxplot(y, xaxt="n")
abline(h=c(-1/20,-1/50,1/50,1/20),col=2)
x=c(2,12,22,32,42,52,62,72,82)
names =
c("1930","1940","1950","1960","1970","1980","1990","2000","2010")
axis(1, at=x, labels=names)
title(main = "Boxplots of Each Transition Year, 1929-2011",  ylab="1-
Year Percentage Change in Per Capita Income", xlab="Years by Decade")

#Stacked Bar Chart
#Create a matrix of total number of states in each category by year
counts <- apply(data, 2, tabulate, 5)
```

```
mp <- barplot(counts, beside=FALSE, main="Proportion of States in Each
Transition Category by Year",col=c("yellow", "orange", "white",
"darkblue", "green"), xaxt="n", yaxt="n", border=NA)
#Custom horizontal axes
x=mp[c(2,12,22,32,42,52,62,72,82)]
names =
c("1930","1940","1950","1960","1970","1980","1990","2000","2010")
axis(1, at=x, labels=names)
axis(3, at=x, labels=names)
#Custom vertical axis
y=c(5,10,15,20,25,30,35,40,45)
names2 = c("5","10","15","20","25","30","35","40","45")
axis(2, at=y, labels=names2)
#Key
legend("right", c("-5% or less", "-5% to -2%","-2% to 2%","2% to 5%",
"5% or greater"), fill=c("yellow", "orange","white","darkblue",
"green"), inset=.2)

#Map first year
cat0 <-
c("Alabama","Arkansas","Delaware","Kentucky","Michigan","Mississippi",
"New Mexico","North Dakota","Oklahoma")
cat1 <-
c("Arizona","Illinois","Indiana","Louisiana","Montana","Ohio","South
Dakota","Tennessee","Texas","Wyoming")
cat2 <-
c("California","Connecticut","Florida","Georgia","Iowa","Kansas",
"Nebraska","New York","North Carolina","Rhode Island","South Carolina",
Utah","Virginia","Washington","West Virginia","Wisconsin")
cat3 <-
c("Colorado","Maryland","Massachusetts","Minnesota","Missouri","New
Jersey","Oregon", "Pennsylvania","Vermont")
cat4 <- c("District of Columbia","Idaho","Maine","Nevada","New
Hampshire")

library(maps)

f <- function() {
      map('state')
      title("Change in Per-Capita Personal Income, 1929-1931", xlab="")
      map('state',cat0,add=T, col="purple", fill=T, density=50)
      map('state',cat1,add=T, col=4, fill=T, density=50, angle=-45)
      map('state',cat2,add=T, col="brown", fill=T, density=50)
      map('state',cat3,add=T, col="green", fill=T, density=50, angle=-
45)
      map('state',cat4,add=T, col="yellow", fill=T, density=50)
      legend("bottomleft", c("-5% or less", "-5% to -2%","-2% to
      2%","2% to 5%", "5% or greater"),
```

```
	fill=c("purple", "4","brown","green", "yellow"), inset=-0.0005)
	map('state',add=T)
	}
	f()

#Map most recent year

cat2 <-
c("Alabama","Arizona","Arkansas","California","Colorado","Connecticut",
"Delaware","Florida","Georgia","Idaho","Illinois","Indiana","Kansas","K
entucky","Louisiana","Maine","Maryland","Massachusetts","Michigan","Min
nesota","Mississippi","Missouri","Montana","Nevada","New
Hampshire","New Jersey","New Mexico","New York","North
Carolina","Ohio","Oklahoma","Oregon","Pennsylvania","Rhode
Island","South Carolina","Tennessee","Texas","Utah","Vermont",
"Virginia","Washington","West Virginia","Wisconsin","Wyoming")
cat3 <- c("Iowa","Nebraska")
cat4 <- c("North Dakota", "South Dakota”)

f <- function() {
	map('state')
		title("Change in Per-Capita Personal Income, 2009-2011",
		xlab="")
		map('state',cat2,add=T, col="brown", fill=T, density=50)
		map('state',cat3,add=T, col="green", fill=T, density=50,
		angle=-45)
		map('state',cat4,add=T, col="yellow", fill=T, density=50)
		legend("bottomleft", c("-5% or less", "-5% to -2%","-2% to
		2%","2% to 5%", "5% or greater"),
		fill=c("purple", "4","brown","green", "yellow"), inset=-
		0.0005)
		map('state',add=T)
		}
	f()

##Chapter 2
#Solve stationary distribution

P <- t(matrix
(c(0.143,0.214,0.071,0.286,0.286,0.007,0.201,0.667,0.104,0.021,
0.003,0.055,0.856,0.080,0.006,0.005,0.102,0.655,0.208,0.030,0.200,0.120
,0.360,0.160,0.160),5,5))
new=diag(rep(1,5))-P
new[,1]=rep(1,5)
pi=solve(t(new),c(1,0,0,0,0))

#Analysis by region
```

```
r <- read.csv(file="C:/Users/Brynne/desktop/Thesis Data by
Region.csv",head=TRUE,sep=",")
mat <- as.matrix(r, byrow=F)
tmp <- findInterval(mat[28:49,], c(-0.05, -0.02, 0.02, 0.05))

data <- matrix(tmp,22,82)+1

M <- matrix(0,5,5)
for(i in 1:nrow(data)){
for (j in 1:(ncol(data)-1)) {
M[data[i,j], data[i,j+1]] <- 1 + M[data[i,j], data[i,j+1]]
}
}
M

tmp <- sweep(M, 1, rowSums(M),'/')
round(.Last.value,3)
rowSums(tmp)


##Chapter 3
#Changepoints
N=81
fn <- function(N) {
      M <- matrix(0,5,5)
      for(i in 1:nrow(data)){
      M[data[i,N], data[i,N+1]] <- 1 + M[data[i,N], data[i,N+1]]
}
return(M)
}

#3D Array
z <-
(abind(fn(1),fn(2),fn(3),fn(4),fn(5),fn(6),fn(7),fn(8),fn(9),fn(10),
fn(11),fn(12),fn(13),fn(14),fn(15),fn(16),fn(17),fn(18),fn(19),fn(20),
fn(21),fn(22),fn(23),fn(24),fn(25),fn(26),fn(27),fn(28),fn(29),fn(30),
fn(31),fn(32),fn(33),fn(34),fn(35),fn(36),fn(37),fn(38),fn(39),fn(40)
fn(41),fn(42),fn(43),fn(44),fn(45),fn(46),fn(47),fn(48),fn(49),fn(50),
fn(51),fn(52),fn(53),fn(54),fn(55),fn(56),fn(57),fn(58),fn(59),fn(60),
fn(61),fn(62),fn(63),fn(64),fn(65),fn(66),fn(67),fn(68),fn(69),fn(70),
fn(71),fn(72),fn(73),fn(74),fn(75),fn(76),fn(77),fn(78),fn(79),fn(80),
fn(81),along=0))

#Log Likelihood Function
#One CP
loglike <- rep(0,80)
for (yr in 2:80) {
      for (a in 1:5) {
```

```
            numer <- colSums(z[1:(yr-1),a, ,drop=F])
            phat <- numer / sum(numer)
            numer2 <- colSums(z[yr:81,a, ,drop=F])
            qhat <- numer2 / sum(numer2)
            loglike[yr] <- loglike[yr]+sum(numer*log(phat)) +
sum(numer2*log(qhat))
            print(phat)
            print(qhat)
      }
}

u <- c(2:80)
v <- loglike[2:80]

plot(u,v, xaxt="n",main="Year vs. Profile Likelihood, One Change
Point", xlab = "Year", ylab = "Profile Likelihood")
x = c(2,12,22,32,42,52,62,72,82)
names =
c("1930","1940","1950","1960","1970","1980","1990","2000","2010")
axis(1, at=x, labels=names)
abline(v=29, lty=2, lwd=1)

#Two CP
loglike <- matrix(0,80,80)
      for (yr1 in 2:79) {
      for (yr2 in (yr1+1):80) {
            for (a in 1:5) {
                  numer <- colSums(z[1:(yr1-1),a, ,drop=F])
                  phat <- numer / sum(numer)
                  numer2 <- colSums(z[yr1:(yr2-1),a, ,drop=F])
                  qhat <- numer2 / sum(numer2)
                  numer3 <- colSums(z[yr2:81,a, ,drop=F])
                  rhat <- numer3 / sum(numer3)
            loglike[yr1,yr2] <- loglike[yr1,yr2] + sum(numer*log(phat))
            + sum(numer2*log(qhat)) + sum(numer3*log(rhat))
            }
      }
      }
loglike[loglike==0] <- -Inf

a <- 1:25
b <- 1:50
contour(1929+a, 1929+b,loglike[a,b],main="Year 1 vs. Year 2 vs. Profile
Likelihood, Two Change Points")
points(1948,1960,col=2,pch="x")

#Comparison code
```

```
pl = function(cutpoints) { # Relies on your existing definition of the
'z' array
cutpoints <- sort(cutpoints)
profLogLik <- 0
for(i in 1:(1+length(cutpoints))) {
      a <- c(1,cutpoints)[i]
      b <- c(cutpoints-1, 81)[i]
      profLogLik <- profLogLik + profLogLikPiece(z[a:b,,,drop=F])
      }
      profLogLik
      }

      profLogLikPiece = function(zz) {
      phat <- apply(zz,c(2,3),sum); phat <- sweep(phat, 1,
      rowSums(phat), '/')
      print(phat)
      sum(sweep(zz, 2:3, log(phat), '*'))
}

uu<-c(1:81)
vv<-c(pl(1),pl(2),pl(3),pl(4),pl(5),pl(6),pl(7),pl(8),pl(9),pl(10),
pl(11),pl(12),pl(13),pl(14),pl(15),pl(16),pl(17),pl(18),pl(19),pl(20),
pl(21),pl(22),pl(23),pl(24),pl(25),pl(26),pl(27),pl(28),pl(29),pl(30),
pl(31),pl(32),pl(33),pl(34),pl(35),pl(36),pl(37),pl(38),pl(39),pl(40),
pl(41),pl(42),pl(43),pl(44),pl(45),pl(46),pl(47),pl(48),pl(49),pl(50),
pl(51),pl(52),pl(53),pl(54),pl(55),pl(56),pl(57),pl(58),pl(59),pl(60),
pl(61),pl(62),pl(63),pl(64),pl(65),pl(66),pl(67),pl(68),pl(69),pl(70),
pl(71),pl(72),pl(73),pl(74),pl(75),pl(76),pl(77),pl(78),pl(79),pl(80),
pl(81))

plot(uu,vv, xaxt="n",main="Year vs. Profile Likelihood, One Change
Point",
xlab = "Year", ylab = "Profile Likelihood")
x = c(2,12,22,32,42,52,62,72,82)
names =
c("1930","1940","1950","1960","1970","1980","1990","2000","2010")
axis(1, at=x, labels=names)
```

REFERENCES

Oldham, Jennifer. "North Dakota Fracking Boom Leaves Oil Hub a Bust: Muni Credit." *Bloomberg News*. N.p., 26 Feb. 2013. Web. 27 Mar. 2013. <http://www.bloomberg.com/news/2013-02-27/north-dakota-fracking-boom-leaves-oil-hub-a-bust-muni-credit.html>.

R Core Team (2012). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ross, Sheldon M. *Introduction to Probability Models*. New York: Academic, 1972. Print.

Silver, Nate. *The Signal and the Noise: Why Most Predictions Fail* - *But Some Don't*. New York: Penguin, 2012. Print.

U.S. Bureau of Economic Analysis (BEA). N.p., n.d. Web. 17 Jan. 2013.

ACADEMIC VITA

Brynne McGarry

222 W. Beaver Ave., Apt. 508, State College, PA 16801

bmm5269@psu.edu

Education

B.S., Applied Statistics, 2013, The Pennsylvania State University, University Park, Pennsylvania

B.S., Economics, 2013, The Pennsylvania State University, University Park, Pennsylvania

Honors and Awards

- Tracy Winfree McCourtney Scholarship, Penn State College of the Liberal Arts, July 2012
- Graduation Student Marshal, Penn State Department of Statistics, May 2013

Association Memberships/Activities

- National Society of Collegiate Scholars

Professional Experience

- Grader, Penn State Department of Statistics, January-May 2013
- Lab Intern, Penn State Department of Statistics, August 2010-May 2013