Analyzing Trends in Loss Frequency and Severity

A Major Qualifying Project Report
Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
In partial fulfillment of the requirements for the
Degree of Bachelor of Science
In Actuarial Mathematics

Submitted on April 30, 2009

Sponsoring Agency: **The Hanover Insurance Group**

Submitted to:
On-Site Liaison: **Alyssa Lopes**, Actuarial Assistant, Personal Lines, The Hanover Group
On-Site Liaison: **Jonathan Blake**, Managing Actuary, Commercial Lines, The Hanover Group
Project Advisor: **Jon Abraham**, WPI

Submitted by:
Ethan Brown, Actuarial Mathematics
Norman Lam, Actuarial Mathematics

Abstract

Hanover Insurance evaluates historical data to analyze trends in loss frequency and severity of claims. The trends are caused by external factors, such as legislative, environmental and economic forces. Trends were analyzed using two different approaches, one correlating the trends from prior data to external factors, and another comparing the impact of events to trends in the data. The analysis mathematically quantified the effect of each external force and isolated factors which were most significant to the trends.

Acknowledgements

We would like to thank Alyssa Lopes and Jonathan Blake from Hanover Insurance for their support and time. We would also like to thank Professor Jon Abraham for his insights and directions while we developed and completed our project. Also, many thanks to our fellow students and friends from WPI, Haiying Liu, Aryeh Shatz, and Mingzhu Zheng, for their feedback and partnership.

Executive Summary

The Hanover Insurance Group is a Worcester-based insurance company offering a variety of insurance products. The company uses its historical data to evaluate trends in its insurance policies using several internal methods.

The goal of this project was to **examine external factors** and compare these with Hanover's historical data, helping predict future loss trends. Steps included:

- Studying historical data for Hanover
- Researching external factors and events that may impact the insurance business
- Comparing trends in external factors and the impact of events to trends in historical data
- Determining which external factors and events correlated best with historical data

Two methods were used for comparing external data to historical data. The first method involved correlating a wide variety of external factors to the frequency, severity, and pure premium of the historical data. After narrowing down the individual forces which correlated best with the historical data, a simple linear programming approach was used to observe if a combination of external factors would correlate better with the historical data.

The second method was to create a scoring method for individual events to compare to the historical data. A timeline was created of significant events over the past twenty years. Then historical data trends were examined to find the points in time where large changes occurred, signifying a possible impact from an event. Events that correlated were given a score based on the magnitude of the occurrence. Once the events were all scored, generalized event types and corresponding scores were defined in order to predict the effect of future events.

The purpose of this project was to define which external factors and events shared trends or impacted the historical data with Hanover. Using this information Hanover would be able to predict future losses and be able to react to any major event that occurs. We were able to define several individual factors that closely correlated to the historical data and create a basic scoring method for major events impacting insurance. Although we did not find a combination of factors to perfectly match the historical data, this project provides Hanover with a basis for predicting future losses and gave the group a better understanding of trends in loss.

Table of Content

Table of Tables

Table of Figures

1. Introduction

The Hanover Insurance Group was created in 1852 and is based in Worcester, Massachusetts. It is a medium-to-large sized company, ranked in the Fortune 1000 and is traded publicly on the New York Stock Exchange under the symbol THG. Hanover is a property and casualty insurance company with a main exposure in Massachusetts, Michigan, New Jersey, and New York. They offer several lines of insurance, ranging from personal to commercial, automotive to homeowners products. This Major Qualifying Project was proposed to us by Hanover to analyze data from their homeowner and automotive personal lines of insurance.

The task for the project was to examine external forces such as legislative, economical, and environmental factors, and compare them to the data that was given in order to explain overall trends and individual events in the data. We used two approaches to accomplish this: correlation and linear programming of the factors, and scaling significance of individual events to explain extreme changes in the data. The first method involved researching several factors across many different subjects over a long period of time, such as change in population, GDP, etc, and comparing the trends of these factors to the trends in the data from The Hanover. In comparing the factors to the data, we looked for very high correlation in order to cull less significant factors. Once we narrowed the factors down to the most significant ones, we used linear programming to attempt to find a perfect mix of factors to explain overall trends in the insurance data. The second method involved researching individual events such as the passing of new laws, inventions, etc, and scaling the impact of the events to explain their significance to the fluctuation in insurance data. Once the scaling was complete, each substantial increase or decrease in the data could be explained by individual events.

2. Background

2. 1 Key Concepts

2.1.1 Correlation Coefficient

A correlation coefficient is a number that represents how well two variables relate. If the number is positive, then the number means that when the first variable increases (or decreases), then the second variable increases (or decreases). If the coefficient is negative, then when the first variable increases (or decreases), then the second variable decreases (or increases). The magnitude of the number also explains how well the variables relate. A higher number means that the two variables correlate well, while a number that is close to zero means that the two variables barely relate.

There are several methods that can be used in order to measure correlation between two variables. The most common is the Pearson product-moment correlation coefficient. This is the model that is used during our calculations.

2.1.2 Linear Programming

Linear Programming is a method for optimizing a linear function usually involving multiple variables. Usually a list of linear equations are submitted and then a mathematical model is made in order to find the best outcome for the set of equations. In this project a simple linear programming model was examined. Only two variables were examined at a time and the analysis of the combination of variables was performed by assigning specific weights to the variables instead of using a model where computer calculation was needed. (Wolfram Research, Inc.)

2.1.3 Event Scaling

Event Scaling is the process of assigning a numeric value to an event on a timeline. These values explain the effect of the events on other data, in this case the historical data for Hanover. The scale can take any form as long as it is consistent. This project uses a scale of 1, 2, and 3 for events. In practice, a negative number would mean the event has a negative effect on the data, and a positive number would have a positive effect.

2.2 Insurance Terms

2.2.1 Rate Making

The process of calculating a premium to charge a customer for an insurance policy is known as rate making. In this process, loss frequency and severity are analyzed in order to predict how much money the company must make to break even, and then adjust the price to make a profit. Pure premium is also reflected in the premium price to account for commissions for insurance salespeople, company expenses, and other miscellaneous expenses. Premium figure that is created through this process reflects a group of policy buyers who share a similar expectation of loss. To create a different premium for each policyholder would be impractical. The data that is examined is normally recorded on a quarterly basis. (McClenahan, 2001)

2.2.2 Exposure

Exposure is the name for the basic rating unit that affects the premium. The unit varies based on the type of coverage that is being provided by the insurance company. For example, a car year is considered one automobile insured for a period of twelve months. A policy covering three cars for a six month term involves 1.5 car years. There are several exposure statistics examined: written exposure, which are the units of exposure from policies that were written during the period;

earned exposure, which are the exposure units that experienced loss during the period; in-force exposures, which are exposure units that experienced loss at a certain point in time. The units of exposure that this project uses in calculations are the earned exposures. (McClenahan, 2001)

2.2.3 Claims

The demand for payment by a policyholder or by an injured third party is considered a claim. The claims are organized by accident date-the date on which the accident occurred, leading to the claim-and by report date-the date on which the insurer is notified of the claim. The claims are recorded as "feature-paid" in the historical data for Hanover, which is used in calculations. (McClenahan, 2001)

2.2.4 Losses

Losses are the amounts paid or to be paid to the claimants under their insurance policy contracts. There are several divisions of losses that are recorded: paid losses, the losses of a period that have been paid to the claimant; case reserve, the amount that is expected to be paid for a claim in the future; accident year-case incurred losses, the sum of paid losses and case reserve for a specific year; ultimate incurred losses, the accident year-case incurred losses plus the losses that have not yet been reported. For this project, paid losses are used in calculations with claims and exposure. (McClenahan, 2001)

2.2.5 Frequency

The amount of claims per exposure unit is called the frequency. The equation for frequency is:

$$\mathbf{Fk = (kC)/E},$$

where Fk is the frequency per k exposure units, k is the scale factor for the frequency, C is the number of claims, and E is the number of exposure units. The frequency used in this project is done on a per-unit basis so it compares to the external data which is also on a per-unit basis, so there is no scale factor. (McClenahan, 2001)

2.2.6 Severity

Severity is the average loss per claim on a policy. It can be calculated using any recorded type of losses, such as paid loss, case reserve, etc. In this case, paid losses are used, and the formula for this calculation is:

$$\mathbf{S = L/C},$$

where S is the severity, L equals paid losses, and C equals the amount of claims for the period. The severity is already calculated on a per-unit basis, so there is no scale factor that needs to be eliminated. (McClenahan, 2001)

2.2.7 Pure Premium

Pure premium is the amount of money needed to pay the amount of losses over the entire exposure. The formula for this quantity is:

$$\mathbf{P = L/E},$$

where P is the pure premium, L is the paid losses, and E is the number of exposure units. The pure premium can also be written as:

$$\mathbf{P = C/E \times L/C},$$

with C equaling claim count, which is the same as:

$$\mathbf{P = F \times S}.$$

Therefore, when frequency is calculated on a per-unit basis, pure premium is the product of frequency and severity. Since the pure premium is based on both frequency and severity, and is more volatile than the other factors, it was not examined as closely as frequency and severity were in this project. (McClenahan, 2001)

2.3 Background of Historical Information

Hanover offers several insurance products, divided in to personal and commercial lines, and insuring a wide range of items, such as businesses, cars, homes, investments, and boats. For this project, we were given information from two different personal lines: homeowner's and automobile insurance. The homeowner's insurance covers a variety of expenses caused by losses, including additional living expenses (renting a hotel while the house is repaired), liability coverage for damaged items, medical payments to others, and inflation. The policy can cover catastrophes like floods if desired, and the company offers coverage to renters and condominium owners in addition to homeowner's. The automobile insurance can also covers a variety of expenses such as collision repairs, medical payments for passengers or other drivers, liabilities for property damage, and several other possible expenses. (The Hanover Insurance Group)

The information given to us was presented in a Microsoft Excel file and was sorted by insurance type, state, and coverage. For the homeowner's insurance, there were twenty-five states where Hanover conducts business. The possible coverage choices were condominium insurance, tenant (renter's) insurance, homeowner's insurance, and "all," which is an aggregate of all of the coverage for the state. Also, for homeowner's, catastrophes could or could not be included in the coverage. For the calculations that were performed, only homeowner's insurance coverage excluding catastrophes were examined because that is the largest source of business for Hanover in the states that were covered.

The automobile insurance contained a larger sampling of data because of the multitudes of coverage offered. Hanover conducts business in twenty-three states for auto insurance, and offers different coverage options in each state. The possible coverage offered by Hanover are bodily injury: the money needed to pay for bodily injury to others; collision, the damage caused to the policyholder’s vehicle in a collision with another car or object; comprehensive, which is collision insurance plus car theft if a new car is stolen within six months of purchase; physical damage, which covers damage caused by the policyholder to property; personal injury protection, which covers medical expenses for the policyholder and/or passengers; and uninsured motorists protection, which pays for damages caused by uninsured or underinsured motorists to the policyholder. In this project, each of the coverage were used because Hanover has differing scales of business for each coverage in each state.

3. Methodology

3.1 Searching for Factors

Before we began to compare external factors to Hanover's historical data, first we brainstormed any factors that we thought could have an effect on the two different insurance types, homeowner's insurance and automotive insurance. The list of possible factors was very broad and covered a wide range of topics. Next we graded each factor on our list for the ability to research information on the topic and how well we thought the information would correlate with Hanover's data. We then focused on searching for data on a few different topics for each insurance separately. While we searched online databases for information, we then expanded our search in order to collect more specific data, i.e. GDP was broken up into GDP-Consumption, GDP-Services, etc. Once we collected data on as many factors from our original list as we could, we expanded our search even more, finding data on several things that did not seem to be relative to insurance. Eventually, we gathered as much data as we thought necessary to begin our correlation comparisons to Hanover's data, and began to examine the information.

3.2 Correlation of Factors

In order to correlate the external data to Hanover's historical data, we used a Pearson correlation coefficient. We selected this method for correlation because the Pearson coefficient because it is widely used to measure the correlation between two variables. The coefficient itself is denoted by the variable "r" and is calculated in the equation:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y}$$

The possible values of r range from -1 to +1, with the strength of the correlation being greatest as the absolute value of r goes to 1. In this study, we decided to eliminate factors that failed to have an r value that was less than -0.7 or greater than 0.7. (Trochim) Was we had established the Pearson coefficient as our guide, we needed to select a time period to compare the data on. Since Hanover's data ranged from 1996 to midway through 2008, we selected a ten-year period from 1998-2007 for evaluation. We also initially decided to select data from Massachusetts, Michigan, New Jersey, and New York from Hanover's data to compare with the external factors we had found because those four states were the main sources of business for Hanover. For the homeowner's data, we only examined the data from the homeowner's line that excluded catastrophes, while for the auto data we originally looked at bodily injury and collision coverage. Eventually once we created a Microsoft Excel model to automate the correlation calculation, we expanded our evaluation to all automotive coverage. We were also able to examine all states and different time period lengths once the Excel model was created.

Once the Excel model was established, we were able to draw conclusions from the data analysis. We initially filtered the information by state since we wanted to examine each state individually. Then we looked at the attributes by filtering out correlations that were within our preferred range mentioned above. Next we examined the resulting correlation for each of the coverage and determined how consistent the correlations were as the period fluctuated. The factors that retained a high correlation for each coverage through the greatest number of varying periods were marked as possible external factors for the state we were examining. We continued our analysis for each state and generated a chart with the resulting factors [Appendix B & C]. For an overall conclusion by state we looked at which factors that were marked for the most coverage for that state and denoted them as the strongest correlation factors.

3.3 Linear Programming

As a supplement to the conclusions from the basic correlation results, we wanted to further examine our results by using simple linear programming of two or more factors. We hoped that by selecting two or more factors that we had found to correlate well with Hanover's data, then combining them with varying weights, would result in a stronger correlation to the historical data. For example, factor A has a correlation of 0.80 to the data, factor B has a correlation of 0.85, and by assigning a weight of 0.7 to factor A and 0.3 to factor B, then summing the weighted factors, the combination would correlate to the data with an r value of 0.90. We used Microsoft Excel to carry out our simple linear programming method.

3.4 Timeline and Event Scaling

We wanted to study if individual events had an effect on the historical data was well as the trends for external factors. First we researched events that could have impacted both homeowner's and auto insurance since 1960, noting events like new legislation, technology, and economic changes. Once the timelines were created, we restricted the time period to examine, settling on 1990 to 2008 because Hanover's data was within that timeframe. Next we took the graphs of frequency and severity for Massachusetts, Michigan, New Jersey, and New York over 1996 through 2008 and highlighted points on the graphs where the trend changed direction. Next we correlated specific events on the timeline to the highlighted points, and assigned a score to each event based on the magnitude of the change in the graph. Once each graph had events correlated and had been scaled, we compared the score of the events for each state and assigned an average score for the events that correlated with multiple states. Noticing similarities between events, we were able to create archetypes of events, such as large economic trends and changes in national interest rates, and assign a score to each archetype. We concluded that these archetypes with scores were the predictors for future changes in trends for frequency and severity.

3.5 Excel Module Usage

For our project, a Microsoft Excel model was build for two reasons. The initial purpose was to create an aid in generating correlations for our small database of external factors to the Hanover's insurance data. The next purpose was to develop a user friendly model for Hanover Insurance which will allow them to include additional external factors or expand on Hanover internal database in order to draw conclusion on new data. This process was done through Microsoft Excel and the usage of macros.

In the model the user is able to select from all the possible internal data from Hanover. The initial option is to select the kind of insurance; Auto or Homeowner. Then a state from a list of state which is filtered depending on whether the selected insurance is sold for that state. Currently Hanover sells their policies in 23 states. The user then selects the coverage that is available for the selected state. Now the user selects the external factors to correlate with the Hanover's data selections above. Lastly, the period which is the number of years to correlate going back from 2008, will need to be determined. Once everything is correctly selected, the "Select" macro will output the correlation data in the "DataOutput" tab for the three attribute, Frequency, Severity, and Pure Premium.

Additionally, by only selecting a type of insurance and a state, the Generation macro will cycle through all the coverage for that state, all the external factors, and all the possible periods. The outcomes of the generated correlations for all the three attributes are stored in the "Results' tab for further analysis. Moreover, for Hanover, this model can easily expand on the number of external factors as more research are done by adding either quarterly or annual data to the Factors tab. Option in the selection will be able to determine that a new factors is added as well as whether it is annual or quarterly data for proper periods selection.

This was the method we use to find the correlation on the external factors to Hanover's data however analysis are to be done in the Results tab to draw conclusions.

4. Auto Insurance Analysis

4.1 Correlation Approach

In this project we worked with two sets of data, the data on Hanover's Homeowner and Auto Insurance and the externals factors. In order to determine which of the correlation would best help in predicting future loss trends, we use the Pearson correlation coefficient. With this method we can measure the strength of the linear relationship between our two sets of data.

4.1.1 Pearson Correlation Coefficient

We used the Pearson Correlation to compute the coefficient "r" which measures the linear association between our two sets of data. For each set of data, the method required the sum, the sum of the squares of each item, the sum of the products of the matched items, and the count of number of items in each set. We then applied those values into the follow formula to determine the r-value. (Trochim)

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y}$$

The r-value was what we used to determine whether an external factor has strong correlation to a set of data from Hanover's data. If the r-value was a positive value, it implied that there was a positive association thus the factor being examined could be consider as a good determinate for Hanover's data. Similarly when the r-value was negative. However we needed to also consider the strength of the r-value. Since we were analyzing two set of arbitrary data, we determined that if a factor had an r-value of greater than 0.7 or less than -0.7 it would a strong enough correlation to be considered.

4.1.2 Time periods

For a factor to be a determinate for future loss trends, it should not only have a strong correlation with current Hanover's data but also Hanover's historical data. With the information provided on Hanover insurances and the availability of the information on external data, we were able to look at a ten year window of historical data for both sets, ranging from 1998 to 2007. If a factor had a strong correlation aggregately throughout that ten year window, we could conclude that the factor might be a possible predictor for future trends.

4.1.3 Coverage

Hanover had provided us with comprehensive data on their Auto Insurance over 23 states. Since Hanover does most of this business in Massachusetts, Michigan, New York and New Jersey, thus for this project, we focused on those states. Each state has between five to six different coverage and we examined three attributes for each of the coverage, Frequency, Severity, and Pure Premium. We wanted to determine how well a given factor would correlate with each of the attribute over the ten year period. However, since Pure Premium is a determined by Frequency and Severity, we excluded Pure Premium in our analysis and conclusion.

4.1.4 Massachusetts Analysis

For MA, there were five different coverage; BI, CM, CO, PD, and PIP. This section will highlight the few factors that were considered a good determinate for each of the coverage.

For BI, the total number of vehicle theft nationally per year (Vehicle_Theft), correlated positively with frequency and negatively with severity, however the correlation was weaker when looking beyond the ninth year where the correlation went below an average of 0.7. Structure, which was a portion of the Gross private domestic investment, correlated negatively for frequency and positively with severity throughout the 40 quarters of data with a peak of 0.99, however when looking at a greater period length, the correlation diminished down to 0.84 for frequency and 0.7 with severity.

For CM, there were many well correlated factors but very few were correlation were consistently high throughout all periods. Only Personal Consumption and Service of GDP correlated consistently with an average of -0.95 throughout almost all of the 40 quarters for both frequency and severity. The factor, population of the United States, was another consistent factor which correlated negatively with frequency yielding an average coefficient of -0.92. As for Severity, Diesel Prices correlated well with an average coefficient of -0.85 throughout almost all quarters.

For CO, the number of vehicle occupants killed in fatal nationally per year (Speed_Vehicle_Occupants), correlated well with frequency peaking at 0.92, but diminishing slowly each year down to 0.81 by end of the tenth years. The factor, Tobacco_Everyday, which is the number of smoker that smoke on a daily bases also correlated well with an average coefficient of 0.95 with frequency however it was only within a short term of 7 years. There were no factors that correlated well with Severity.

For PD, between frequency and severity, there were different factors the yield strong correlation. Both factors, Crime Rate Total and Vehicle Theft, correlated well with frequency and yielded a high average correlation coefficient of 0.95, however both factor had a diminishing correlation when looking at longer period length. On the other hand, severity did not have many factors that had a strong correlated. Only personal consumption expenditures of GDP (P_Consump) and Services of GDP, correlated well with an average coefficient of-0.90 or better, however that was only when looking at first 7 years.

For PIP, there was only one strong factor that correlated well with frequency. The Crime Rate Total correlated strong throughout 9 years with a consistent correlation averaging of 0.93. For severity, there were no factors that had a strong correlation.

4.1.5 Michigan Analysis

For Michigan there are six coverage, BI, CM, CO, CSL, PD, and PIP. Overall there were many factors have correlated very well with frequency but fewer for severity. Section below will present those external factors that correlated the best.

For BI, overall the personal consumption and services of GDP correlated well for both frequency and severity. It correlated consistently with an average coefficient of -0.89 with frequency and 0.80 with severity. However Population yielded an even better correlation with frequency with not only a consistent correlation but also an average coefficient of -0.94 for all ten years. For severity, the fatality rate of 100,000 registered drivers in Michigan (Fatality_Registered), yield an average correlation of 0.90, also for the past 10 years.

For CM, Population correlated extremely well with both frequency and severity. Throughout all ten years Population yielded a consistence correlation with an average of -0.96 for frequency. And for severity, we saw population yield a correlation coefficient average of 0.97 through the past nine years, and the correlation drop significantly for the tenth year. Additionally for severity, the fatality rate for 100,000 registered drivers yielded a stunning average correlation coefficient of 0.98 however it was diminishing as years pass.

For CO, there were many factors that correlated well with frequency however almost none for severity. Only the GDP factors, personal consumption and services, correlated highly for both frequency and severity. Both personal consumption and services correlated slightly below a coefficient average of -0.9 for frequency throughout all 40 quarters. For severity, the two factors, correlation yielded an average coefficient of 0.83 and 0.8.4, respectively, throughout the 40 quarters. Correlations for both were less consistent for the first two years, however as we look at additional periods, we saw a consistent correlation. Additional for frequency, the fatality rate for register driver yield an even higher positive average correlation of 0.94.

For CSL, there were no factors that correlated well with severity. As for frequency, we saw Obese, which was the percentage of population with a BMI that is considered obese, correlated very well with CSL with an average coefficient of 0.95. BMI_OK, which was the percentage of the population with a BMI value that is considered normal, also correlated well with CSAL with an average coefficient of 0.90. Lastly, we also saw population being a well correlated factor of an average coefficient of 0.85 to CSL but we also see a diminishing average as more periods were correlated.

For PD, similar to CSL, there were no factors that correlated well with severity. For frequency, there was wide range of factors that correlated well. The percentage of population that were consider Obese correlated the best, resulting in a average correlation coefficient of -0.91, however for the ninth and tenth year we saw large decline in correlation. Additionally, for the percentage of population with BMI that was considered normal resulted in a more consistent correlation however resulted with a lower average correlation of 0.85 throughout the ten years. Lastly for PIP, the strongest correlation for both frequency and severity was with population. The high correlation coefficient average of -0.93 for frequency and 0.97 for severity for all ten years only tell half of the story. The correlation was actually increased as the period length gets longer. As we included more years into the correlations the coefficient value increased for both frequency and severity. Other factors such as the number of vehicle occupants killed in a fatal crashes nationally per year and percentage of population that are consider obese correlated well, both had an average of around -0.90 correlation with frequency and 0.91 with severity.

4.1.6 New Jersey Analysis

For New Jersey there were six coverage, BI, CM, CO, CSL, PD, and PIP. Overall, many of the coverage fluctuate significantly over the years however below will highlight those factors that correlated despite such variations.

For BI, there were no factors that correlated well beyond the first three years therefore no factors were considered. As for severity, many of the factors showed high correlation; in fact all factors relating to fatality had a very strong and consistent correlation with an average coefficient of 0.93 to 0.97. Additionally, the factors relating to tobacco, particularly, the number of people who smokes daily and number of adults who are smokers exhibit a strong correlation both with an average coefficient of 0.93. Lastly population and GDP properties both yielded a correlation coefficient of 0.88 with BI severity.

For CM, overall there were almost no factors that yield any significant correlation because in 2000 the frequency for CM was exceptionally high. After excluding such extremities, only population and crime rate displayed a decent correlation, both with a coefficient of 0.8 for frequency. For severity even excluding a odd year of 2000, there were no factors that yield any correlation.

For CO, there was not a single factor that it correlated well with except one factor. Since CO's data fluctuated significantly throughout each year, almost no factors could even correlate to it, however interest rate was able to follow such fluctuation closely and yield an average correlation coefficient of 0.81 when looking back for 8 years for CO's frequency.

For CSL, Robbery, which was the number of robberies annually, correlated well with only frequency, with an average coefficient of 0.85. However this factor has a fast decreasing coefficient as the period increased in length thus this factor might only be used as a short term predictor. As for severity, the total number of fatal crashes annually (Speed_Fatal_Crashes), yielded a significant correlation coefficient of 0.74, but again only for a short term of only 6 years.

For PD, again there were not many factors yielding strong correlations. For frequency, Overweight, which is the percentage of the population with a BMI that is considered overweight, showed an average correlation coefficient of 0.84. And the demand for non-highway gasoline also showed strong correlation of a coefficient of 0.9 but only for the first 7 years before PD's frequency turned the other direction.

Lastly for PIP, there are many factors that significantly correlated with the severity, however not many for frequency due to the sudden dip between the year of 2000 and 2001. Excluding the extremities for frequency, the price of gas and diesel gas showed a strong correlation of an average coefficient of 0.84. As for severity, both elements of GDP, personal consumption and services, showed a very high consistent correlation throughout all 40 quarters yielded an average correlation of 0.94 and 0.95, respectively.

4.1.7 New York Analysis

For New York there were six coverage, BI, CM, CO, CSL, PD, and PIP.

For BI, there was a sudden increase in frequency and severity only between the years of 2003 and 2005 which was very different from rest of the years. No factors were able to capture such changes.

For CM, the frequency tended to fluctuate over the years and there was only one factor, the number of vehicle occupants killed in fatal nationally per year (Speed_Vehicle_Occupants), correlated decently well with frequency with a average coefficient of 0.84. As for severity, which fluctuated much more than frequency, correlated well with only one factor. Although not strong correlation, the number of smoker that smoke on only at sometimes correlated consistently with an average coefficient of 0.72.

For CO, the frequency did not correlate well with most of the factors except for two factors. Overweight which correlated well with an average coefficient of 0.87 but had a decreasing correlation with longer period length, and the demand for non-highway gasoline correlated very consistently around the mid -0.95 however after the seventh year it sharply declined. On the other hand, for severity, we saw many factors that were well correlated. One of the more consistent factors was population which a strong correlation throughout all 10 years yielding an average coefficient of 0.97. Other factors that correlated well included robbery, daily and adult tobacco users, and the number of fatal crashes involving registered vehicles, and factors of related to fatalities.

For CSL, both the frequency and severity fluctuated over the years. Many factors correlated when looking at a wider period length. Excluding the short term correlations we saw population and the percentage of population being obese as the two strongest factors that were more consistent when looking at a longer period length. Obese yielded an average coefficient of -0.90 with frequency when looking at least four years of data. And population yielded an average coefficient of -0.9 with frequency when looking at least five year of data. However for severity, which fluctuated much more than frequency only saw the demand of non-highway gasoline correlated consistently with an average coefficient of 0.91.

For PD, although there were no extremities for frequency, only the percentage of the population who uses tobacco some days has a strong correlation. It held an average correlation coefficient of 0.80 for eight years. As for severity, we saw very strong correlation with burglary yielding a consistent high correlation with an average coefficient of 0.96 for nine years. Population and gas prices also correlated well but only when looking at a wider period range. Excluding shorter period range correlations, we saw population yielding an average coefficient of 0.85 when looking at least three years and gas prices yielding an average coefficient of 0.91 when looking at least 16 quarters.

Lastly, for PIP, the frequency was decently correlated with the percentage of the population who were consider overweight and the number non-motorists killed in a fatal crash annually, both yielding an average coefficient of 0.81 and 0.80, respectively. For severity, only the demand for gasoline held a strong correlation with an average coefficient of 0.89 for highway demand and 0.87 for total demand.

4.1.8 Conclusion

Overall for Massachusetts data, it seemed to have a strong correlation with population and GDP factors such as personal consumption and services. Some of the minor factors that correlated with Massachusetts were diesel gas prices and crime rates. For Michigan, we also saw similar conclusion being that there was a strong correlation with population and factors of GDP. Some of Michigan minor factors included, daily tobacco user and percentage of the population that is consider obese. New Jersey, which had the hardest time to correlate with any external factors, resulted with decent correlation with multiple factors; population, number of fatal crashes involving registered vehicles, and percentage of population that are consider obese. Minor factors for New Jersey included GDP factors and percentage of the population that use tobacco daily. Lastly, for New York, the factors that correlated the best varied among the coverage. Factors included percentage of the population considered to be overweight, non-highway gasoline demand, population, factors relating to fatal crashes, and people who uses tobacco daily.

4.2 Timeline Approach

4.2.1 Years Examined

Similarly to the homeowner's insurance analysis, a timeline needed to be created in order to compare events to trends in automotive data. This time, data was gathered on events starting in 1970 and ending in 2008. Data was culled down again to what we considered to be the most relevant events. Four timelines were made from the data, one from 1970-1980, another from 1980-1990, another from 1990-2000, and one last one from 2000-present.

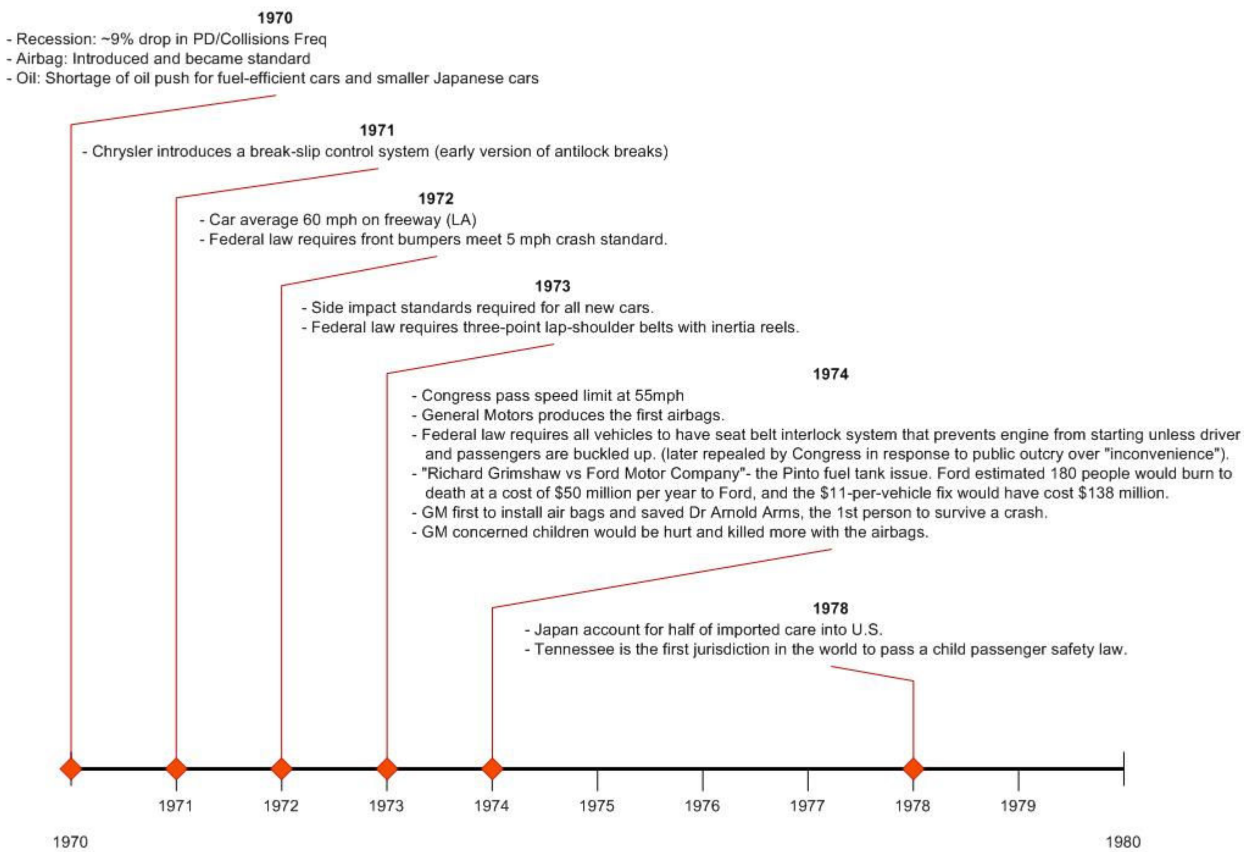

Figure 1: Timeline - Automotive Events 1970-1980

From 1970-1980, several new technologies were introduced and driving laws were passed, helping make cars safer for drivers, passengers, and pedestrians.

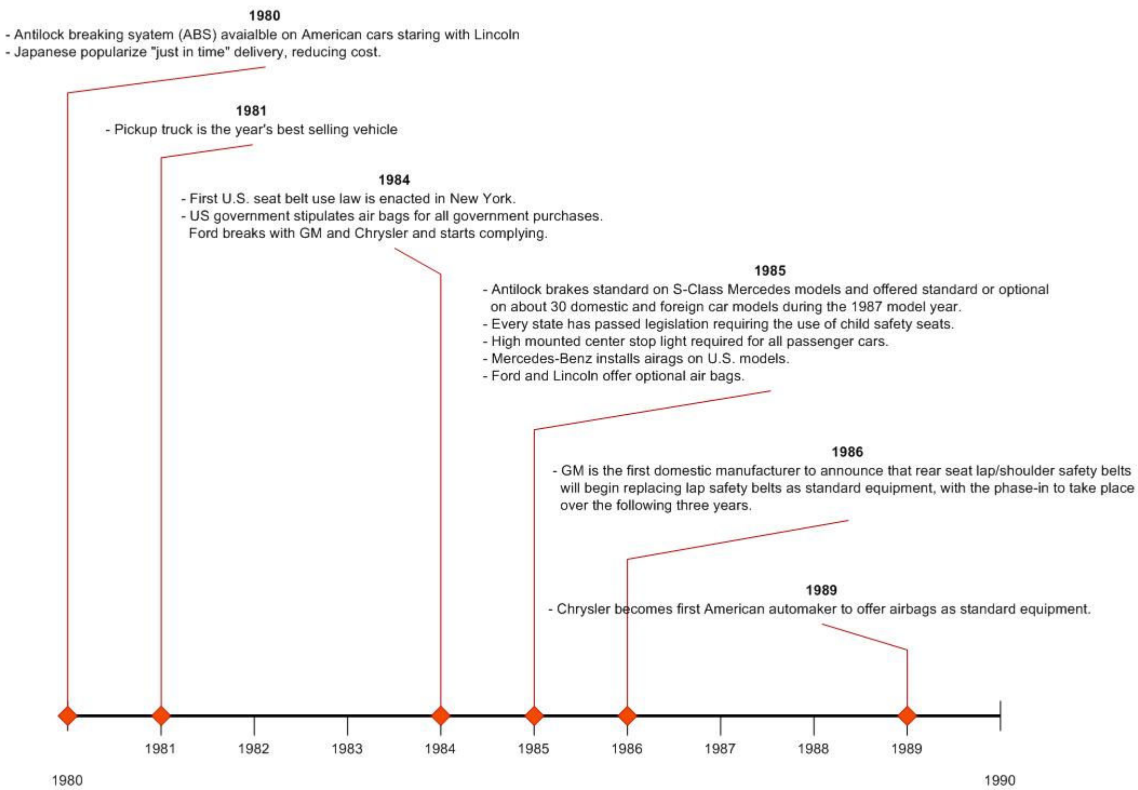

Figure 2: Timeline - Automotive Events 1980-1990

The period from 1980-1990 saw more laws enacted, increasing safety. However, we begin to see more technological advances happen by manufacturers instead of the industry as a whole. Similarly to the homeowner's analysis, events from 1970-1995 were interesting, but largely ignored because the data for Hanover only ranged from 1996-2008.

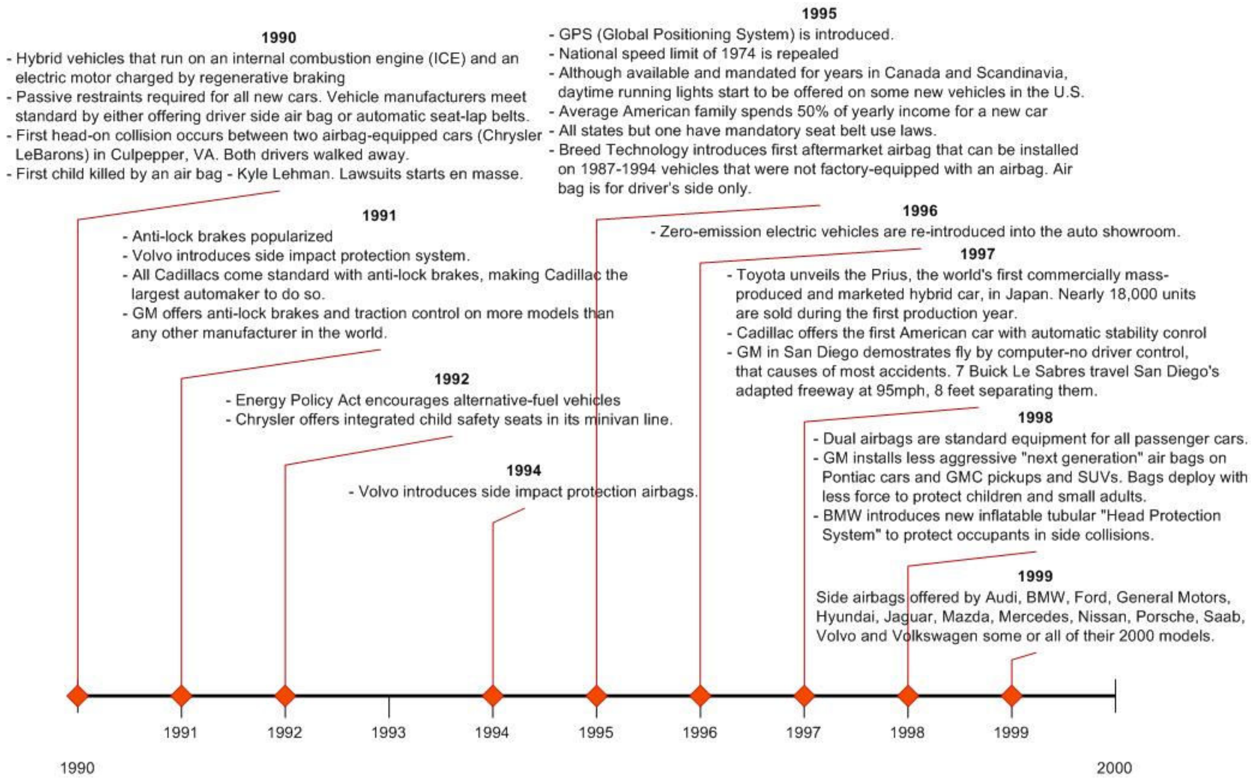

Figure 3: Timeline - Automotive Events: 1990-2000

During the 1990s, there were major breakthroughs in technology almost every year. However, all but one event happened by manufacturer; only two events that affected the entire industry was in 1995 when the national speed limit was repealed and in 1998 when dual airbags became standard equipment for all passenger cars. All laws were enacted prior to 1996 as well.

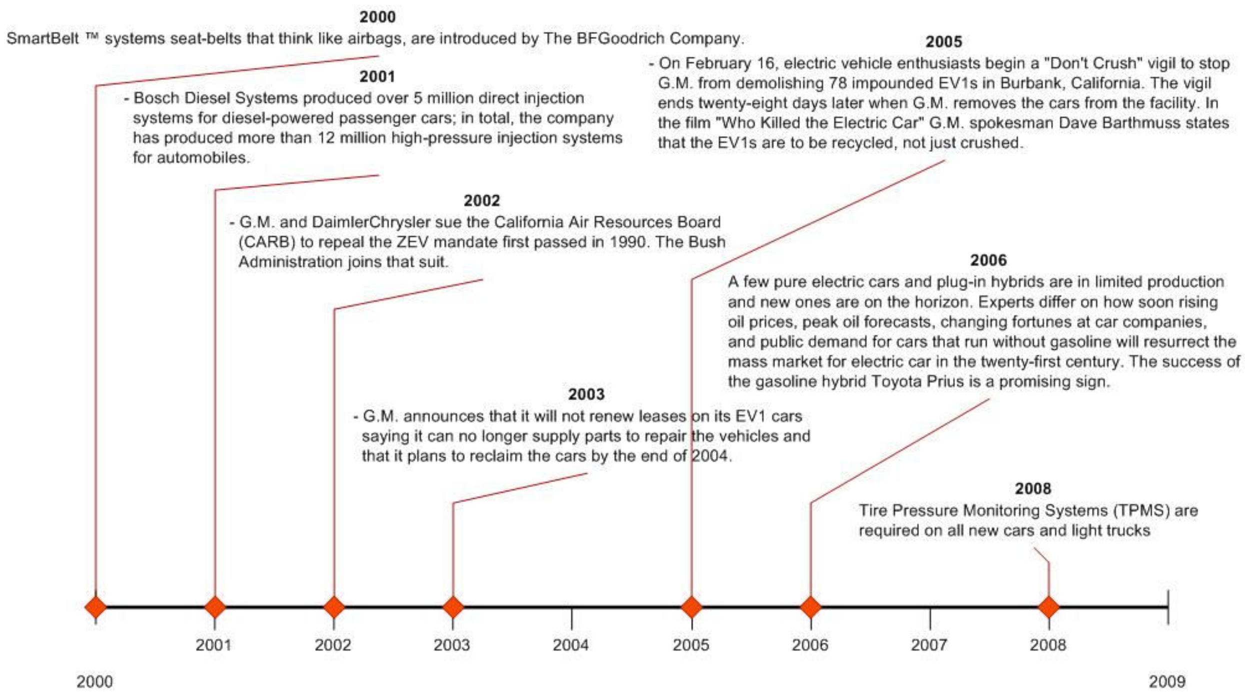

Figure 4: Timeline - Automotive Events 2000-present

Like the 1990s, most of the events that occurred were exclusive to the manufacturers. The only event that affected the entire industry occurred in 2008, when Tire Pressure Monitoring Systems became required on all new cars and light trucks. Because this event occurred in 2008, it is too early to tell if it affected the frequency or severity of Hanover's auto data.

4.2.2 Conclusions

Since there were only three events that occurred during 1996-2008 that could have affected the entire auto industry, it is impossible to create a scoring method for the automotive insurance trends. More data from Hanover creating a longer period of analysis would aid in creating a scoring method, but there are still too few events that affect the entire industry. Therefore, Hanover's auto data is affected more by the trends of external factors than by individual events.

4.3 Overall Conclusions for Automotive Insurance

The frequency and severity for Hanover's automotive data is almost exclusively affected by the trends of external factors. While individual events were examined, it was determined that there was a lack of events affecting the entire auto industry, and auto insurance itself. An event might cause a change in frequency or severity from time to time, but the infrequent events made these changes inconsistent. Overall, external factors that correlated with the auto insurance trends the most were United States population, gross domestic product (GDP) from personal consumption, and GDP from services.

5. Homeowner's Insurance Analysis

5.1 Correlation Approach

For the homeowner's insurance the analysis was very similar to the analysis performed on the auto insurance for Hanover. The specific data for homeowner's insurance was the homeowner's insurance frequency and severity excluding catastrophes. A series of external factors was gathered, many of them the same as the external factors that were examined for auto insurance. In fact, there was only one new data set for an external factor that was not applicable for auto insurance, and that was the median price of homes from 1998-2007. A Pearson correlation coefficient was again used in order to see the degree of correlation between the external factors and Hanover's historical data.

There were a few factors that correlated well over the 10-year term from 1998-2007 for Massachusetts homeowner's frequency and severity. The best factors to correlate with frequency were median home values, US population, and the percentage of people with a body mass index (BMI) that is considered obese. For severity, the US population and the GDP for personal consumption and services, along with the real GDP, correlated the best. It is possible to further analyze the data and trends of homeowner's insurance using an Excel model similar to the model used in the auto insurance. However, once the model was created, we focused on finding trends for automotive insurance because there were more coverage for autos from Hanover and because the scaling method for autos provided no conclusions due to a lack of events affecting the entire industry. For the future, the Excel model used for the auto insurance can be modified to include homeowner's data and provide correlation coefficients for all factors examined on periods of different lengths.

5.2 Timeline Approach

5.2.1 Years Examined

In order to create a timeline, we first had to gather any information that we thought was relevant to homeowner's insurance. Data was gathered on events starting in the 1960s and ending in 2008. After all of the information was pooled together, we culled the group down to the events that we thought were the most relevant and would have the greatest effect, if any, on Hanover's data. Next we made three timelines, one from 1960-1990, another from 1990-2000, and a final one from 2000-present.

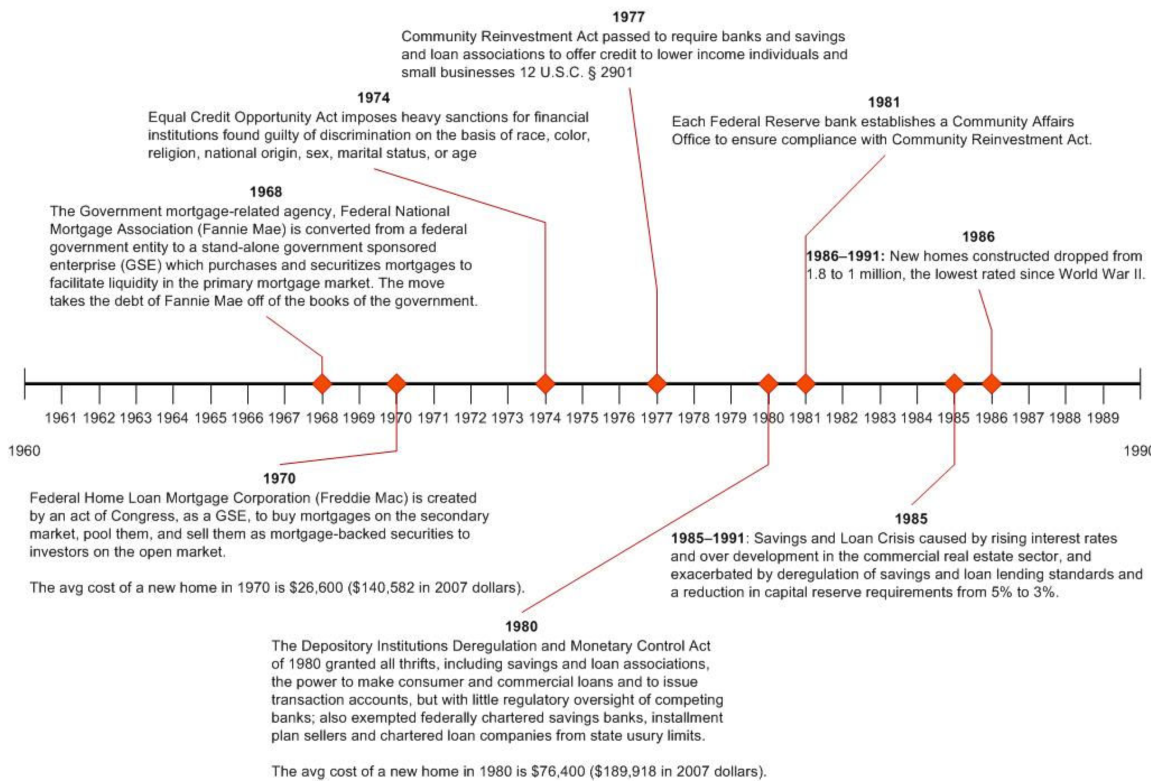

Figure 5: Timeline - Housing Events 1960-1990

The events that occurred from the 1960s through 1995 was largely ignored because Hanover's data only ranged from 1996 through 2008, but it did provide good practice for locating which types of events seemed to occur more often and would have an impact on homeowner's insurance.

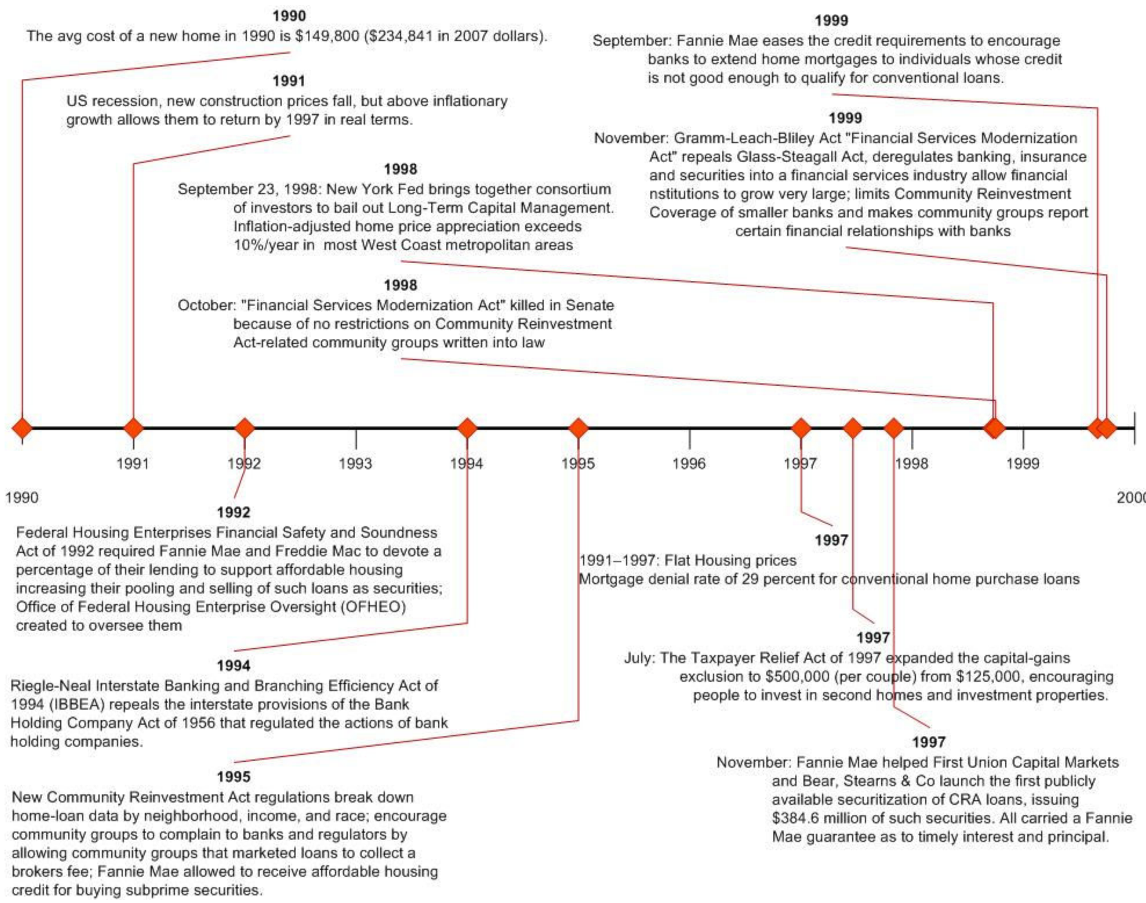

Figure 6: Timeline - Housing Events 1990-2000

The 1990s featured three events that seemed to affect the data from Hanover. Starting in 1991 and lasting through 1997, there was a prolonged period where housing prices in America were flat and there was a mortgage denial rate of 29% for regular home purchases. In September of 1999, Fannie Mae eased credit requirements which meant that more people now qualified for home mortgages, which in turn meant that more people were able to purchase homes. Finally, in November of 1999, Congress passed the Gramm-Leach-Bliley Act, which deregulated banking, insurance, and securities which allowed financial institutions to grow very large.

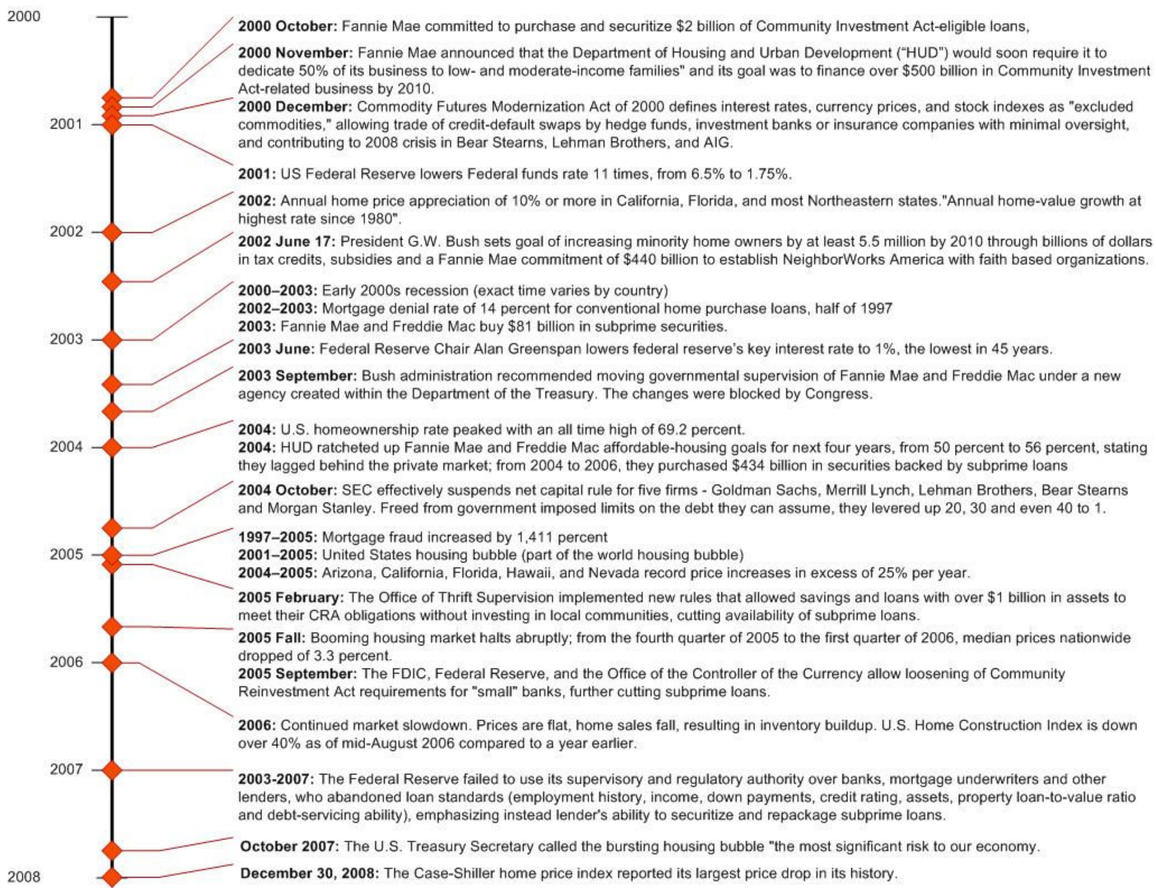

Figure 7: Timeline - Housing Events 2000-Present

In the current decade, four events affected data for homeowner's insurance. In 2001, the US Federal Reserve lowered the Federal funds rate eleven times, from 6.5% to 1.75%. Two years later, the Fed's key interest rate was lowered to 1%, the lowest in 45 years. In 2004, the HUD ratcheted up Fannie Mae and Freddie Mac's affordable-housing goals by six percent. Finally, around 2007 the housing bubble burst and the homeowner's market was characterized by falling house prices, sales, and construction rates.

5.2.2 Trend Matching and Event Scoring

Once the timelines had been created, we needed to compare the sequence of events to the trends in the frequency and severity in Hanover's data to see which events had an impact and how great the impact was. We took the graphs of frequency and severity for the four main states in Hanover's business, Massachusetts, Michigan, New Jersey, and New York, over time, and examined the dates where the trend changed direction. At each point where the trend changed, we attempted to correlate an event on the timeline to show that the event had an impact on the data. At each point where an event correlated, we gave the event a score of one, two, or three to signify the impact of the event. After scoring the events that correlated with all graphs, we averaged the scores to create an overall score for the events. This overall score was used to create a set of archetypical events which accompanying scores which could be applied to future events in order to predict the new event's impact on frequency and severity for Hanover. Following are the graphs of frequency and severity for the four main states with each point that correlated highlighted and matched with its event.

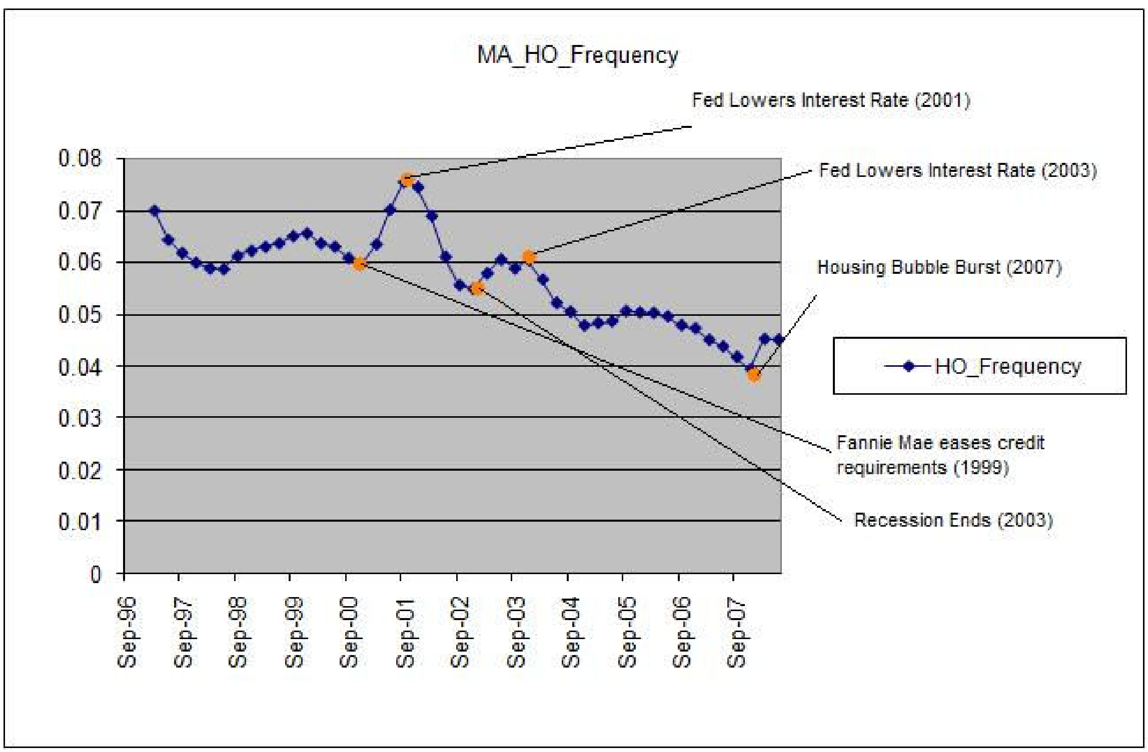

Figure 8: Massachusetts Homeowner's Frequency 1996-2008

The first event that occurred with an effect on frequency was in December of 1999, when the result of Fannie Mae easing credit requirements caused frequency to begin increasing with a score of 2. Next, in 2001 when the Fed lowered the interest rate several times, the frequency began decreasing with a score of 3. At the end of 2002, the economical recession ended, and frequency began increasing for a short while, giving the event a score of 1. In 2003, the interest rate was lowered again and frequency decreased at a score of 2. Finally, in 2007, the housing bubble burst, ending a long period of decline in frequency, giving the event a score of 3.

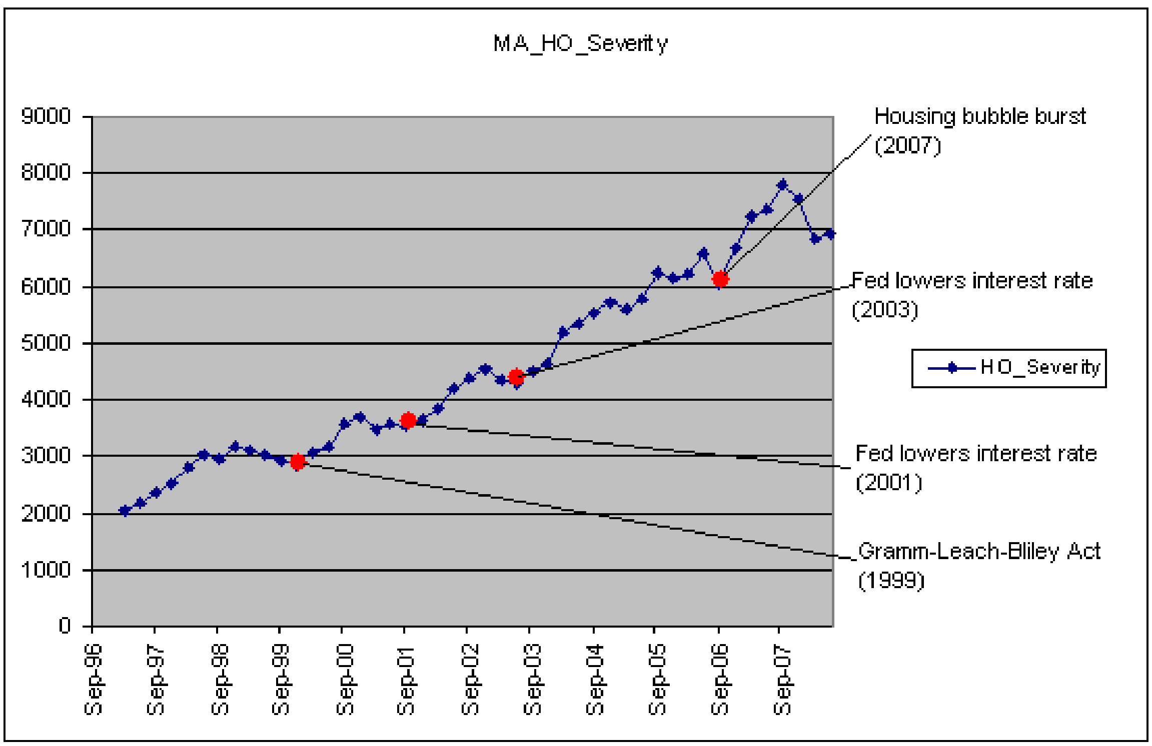

Figure 9: Massachusetts Homeowner's Severity 1996-2008

The Gramm-Leach-Bliley act that deregulated banking, insurance, and securities in 1999 caused the severity to go from decreasing to increasing, with a score of 2. Next in 2001, the Fed lowered the interest rate several times, causing the severity to again shift from decreasing to increasing with a score of 2. The Fed lowered interest rates again in 2003, this time causing severity to go from decreasing to increasing with a score of 3. Finally, in 2007, the housing bubble burst, causing the trend to again increase with a score of 3.

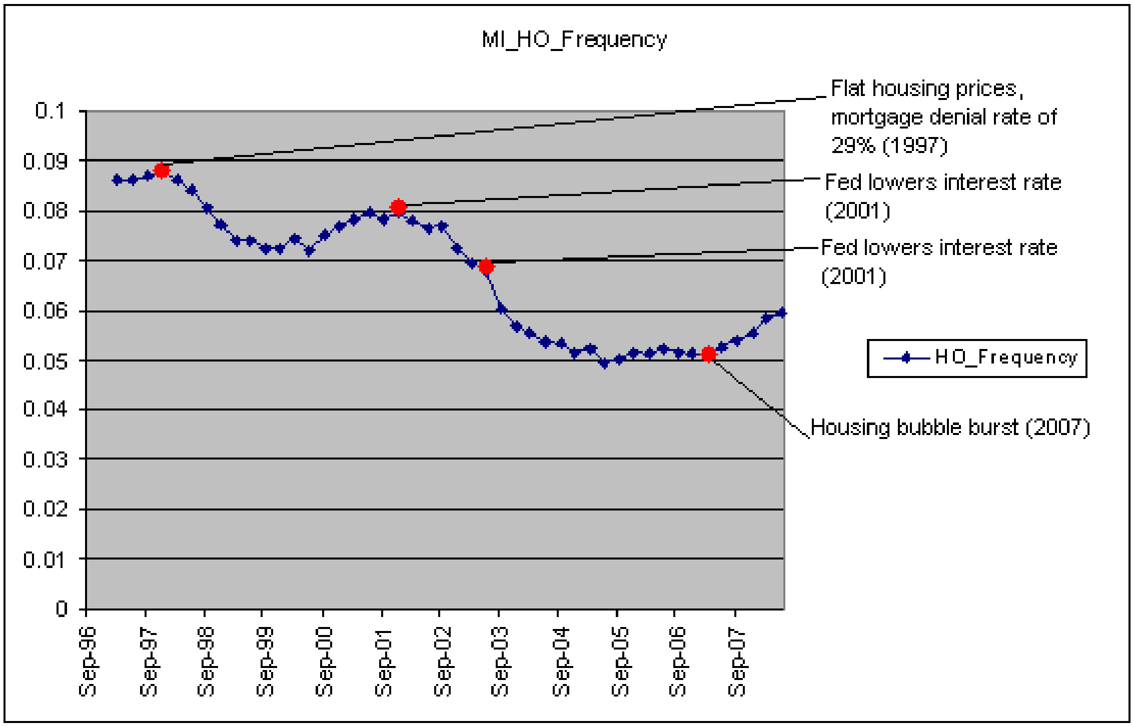

Figure 10: Michigan Homeowner's Frequency 1996-2008

The frequency of Michigan was affected by similar events to Massachusetts, except there seemed to be a short lag in the reaction from Michigan to some events. In 1997, the period of flat housing prices and a mortgage denial rate of 29% ended, causing frequency to drop at a score of 3. Then in 2001, the Fed lowered the interest rate, causing the frequency to descend at a score of 3. After a short increase, the Fed lowered the interest rate again, causing the frequency to decrease at a score of 3. However, the shifts caused by both interest rate changes came a quarter later than the changes that occurred for Massachusetts. Finally, the housing bubble burst in 2007, reversing a long trend of decline in frequency at a score of 3.

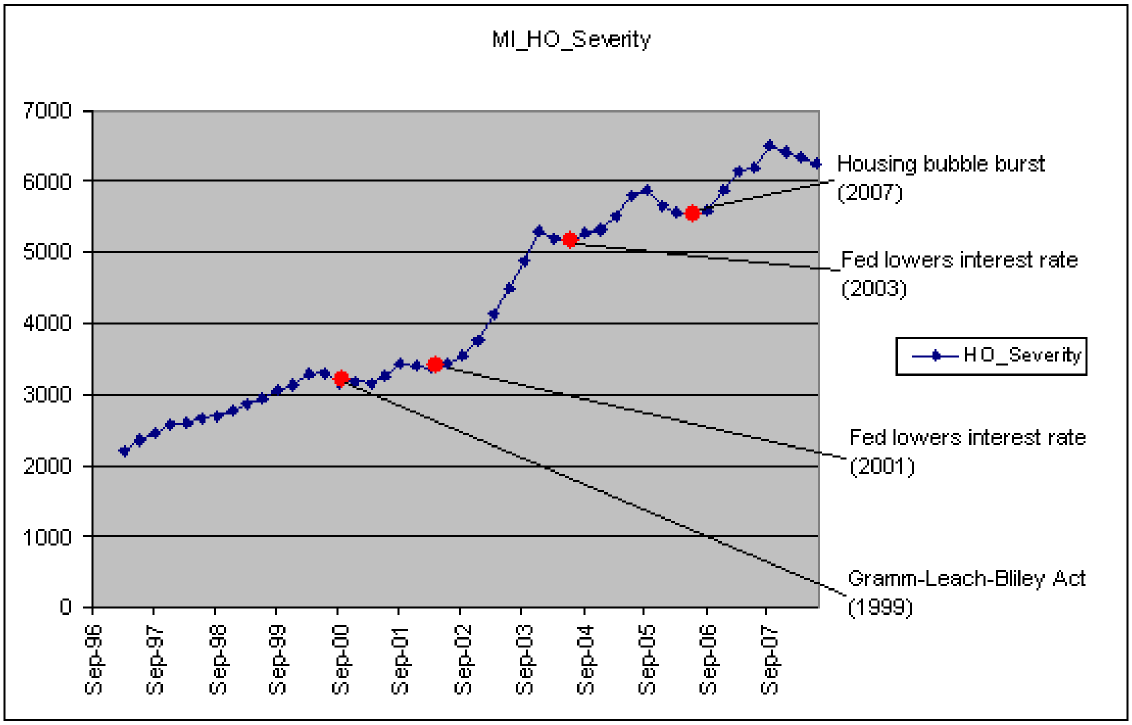

Figure 11: Michigan Homeowner's Severity 1996-2008

The severity of Michigan also featured lag for some events compared to Massachusetts. In 1999, the Gramm-Leach-Bliley Act caused the severity to increase with a score of 1. When the Fed lowered the interest rate in 2001, the severity increased sharply for a score of 3. A short decline was reversed in 2003 when the interest rate was lowered again at a score of 2. Once again, the effect of the change in interest rate was felt after a quarter lag similarly to frequency for Michigan. Finally, the burst housing bubble caused the severity to start to increase again for a score of 2.

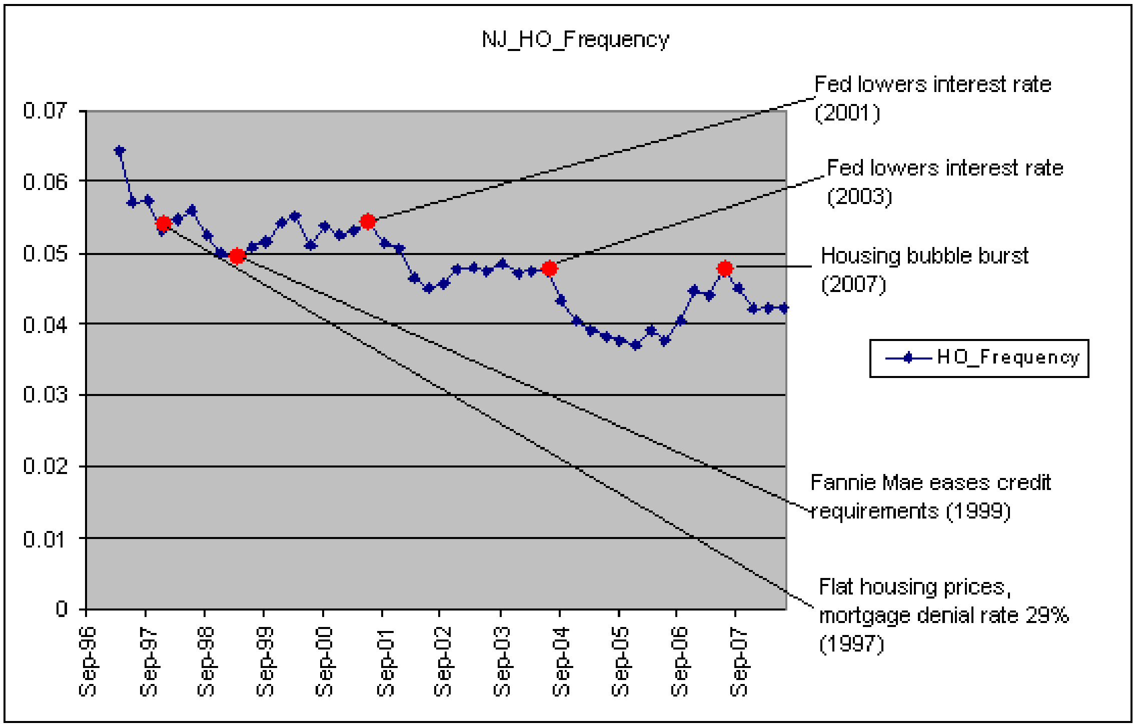

Figure 12: New Jersey Homeowner's Frequency 1996-2008

In 1997, the end of the period of flat housing prices and a mortgage denial rate of 29% caused an end to the declining frequency for a score of 2. Then when Fannie Mae eased credit requirements in 1999, the frequency began to increase again with a score of 2. The lowering of the interest rate by the Fed in both 2001 and 2003 caused the frequency to decrease with a score of 3. Finally, in 2007, when the housing bubble burst, a steep increase in frequency was reversed with a score of 3.

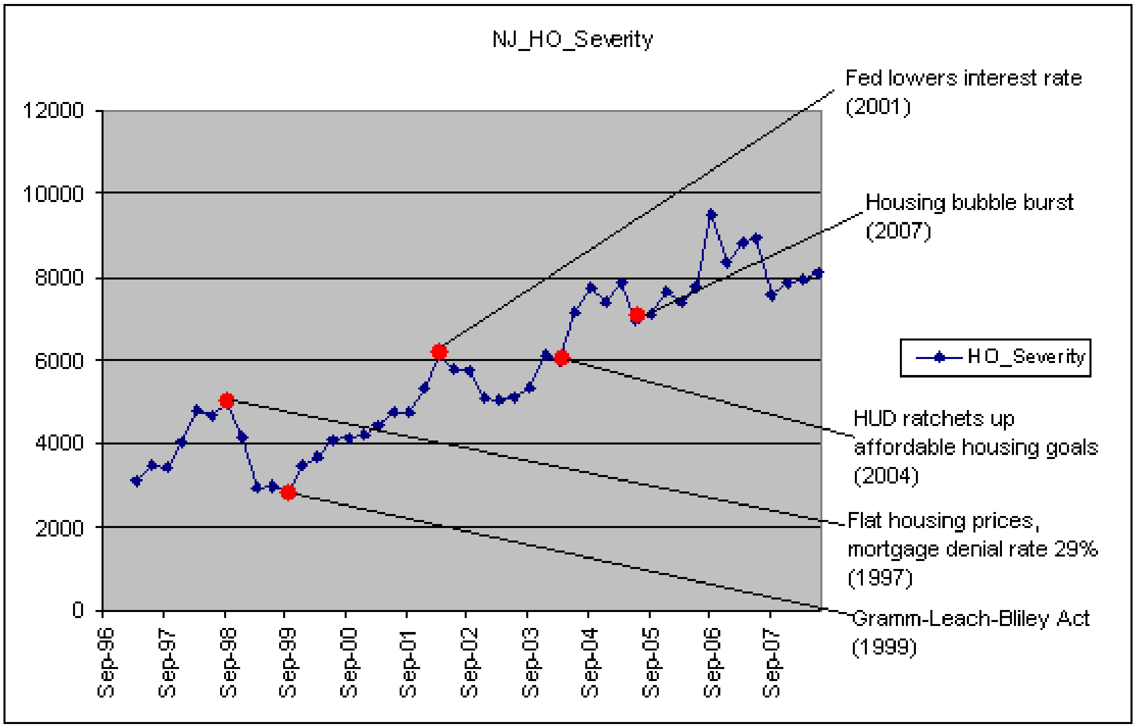

Figure 13: New Jersey Homeowner's Severity 1996-2008

Severity was again affected by similar events to Massachusetts and Michigan. Once the period of flat housing prices and a mortgage denial rate of 29% ended, the severity decreased sharply with a score of 3. In 1999 that trend was reversed when the Gramm-Leach-Bliley Act was released with another score of 3. In 2001, the Fed lowered the interest rate which caused the severity to decrease with a score of 2. When HUD ratcheted up affordable housing goals in 2004, this caused the severity to sharply increase after a short period of decline with a score of 3. Finally, just before the housing bubble began to burst in 2005, the severity began a very volatile increasing trend, for a score of 3.

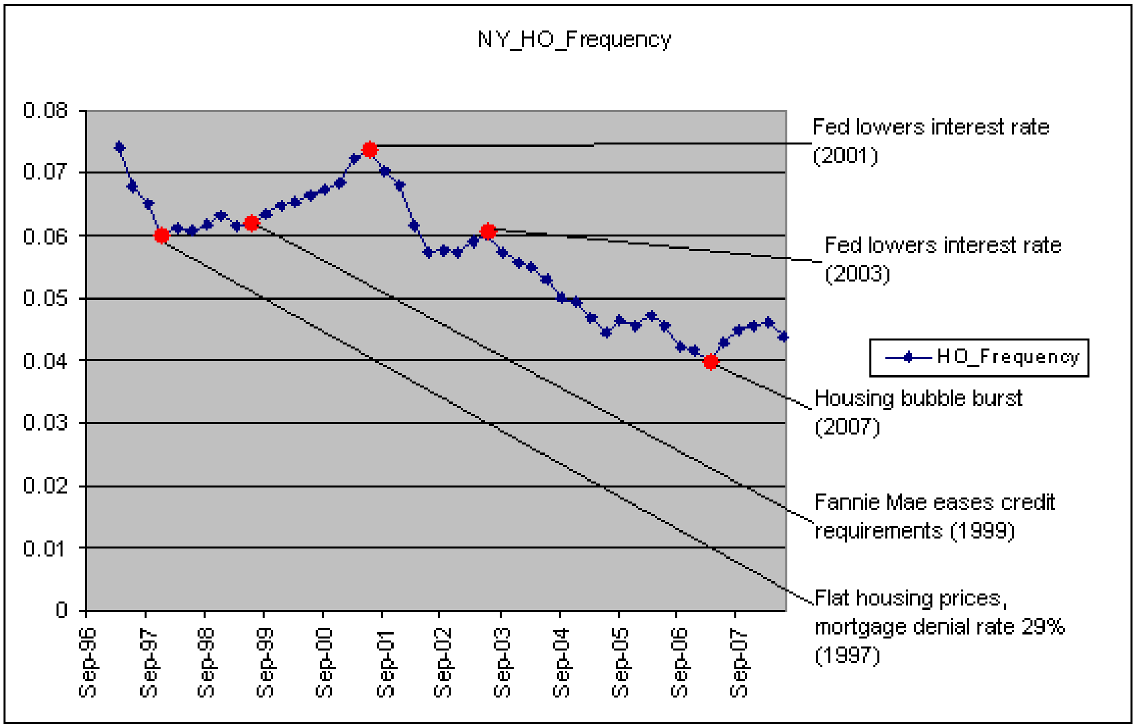

Figure 14: New York Homeowner's Frequency 1996-2008

New York's frequency seemed to be affected by similar events as the rest of the major states for Hanover. In 1997, and sharp decline in frequency was reversed when the period of flat housing prices and a mortgage denial rate of 29% ended, for a score of 2. A short decline in frequency was reversed when Fannie Mae eased credit requirements in 1999 for a score of 1. The lowering of the interest rate in 2001 and 2003 by the Fed caused sharp decreases in frequency after shorter periods of increase, both scoring a 3. Finally, when the housing bubble burst in 2007, the frequency ended a long trend of decline and began increasing for a score of 3.

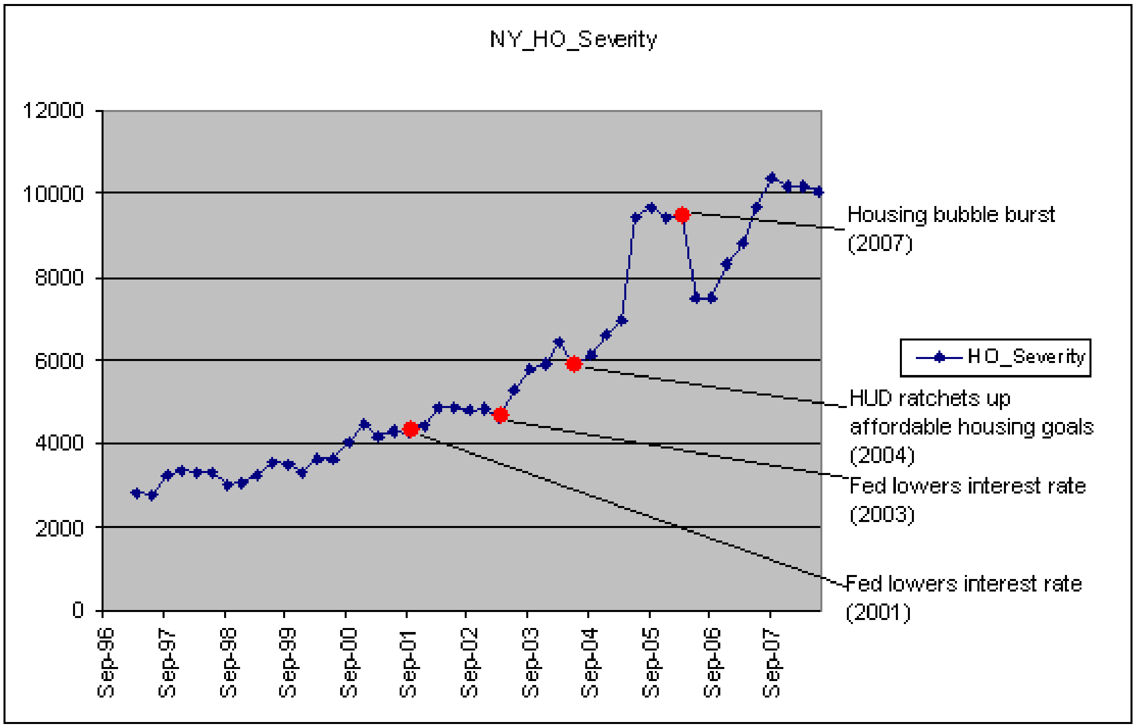

Figure 15: New York Homeowner's Severity 1996-2008

New York's trend in severity was most similar to Michigan's. It was first affected when the Fed lowered the interest rate several times in 2001, for a score of 1, and started a short increasing trend. In 2003, when the Fed again lowered interest rates, a declining trend was reversed again, this time into a sharper and longer increase in severity, for a score of 2. When the HUD ratcheted up affordable housing goals in 2004, a short drop in severity was answered by a steep increase for a score of 3. Finally, just before the housing bubble burst in 2007, New York experienced a sharp drop in severity followed by a sharp increase, for an overall score of 3.

After scoring each of the relevant events, we were able to create an overall score for each event:

Table 1: Overall Score for Relevant Events

Event	Score (Frequency)	Score (Severity)
Flat housing prices/Mortgage denial rate of 29% (1997)	2	2
Gramm-Leach-Bliley Act (1999)	-	1+
Fannie Mae eases credit requirements (1999)	1+	-
Fed lowers interest rate (2001)	3-	2+
Fed lowers interest rate (2003)	3-	3+
HUD ratchets up affordable housing goals	-	3+
Housing bubble burst	3	3+

In this table of overall score, the values that carry a positive or negative sign after the score indicates that the frequency or severity increases or decreases as a result of the event. If the event has no sign along with the score, then the event affects the data for Hanover, but each state reacts differently to the occurrence of the event.

Having created the scoring method for the events on the timeline, an score for generalized events can be created:

Table 2: Overall Score for Generalized Events

Event	Example	Score (Freq)	Score (Severity)
End of prolonged economic trend	Housing bubble burst (2007)	3	3
Federal Reserve behavior	Fed lowers interest rate (2001)	3	3
Fannie Mae/Freddie Mac behavior	Fannie Mae eases credit requirements	1	1
New legislation passed	Gramm-Leach-Bliley Act	-	1

There are several generalized events that occur often enough to affect frequency and severity for Hanover. However, it is extremely difficult to predict if the event will affect frequency or severity positively or negatively. If more events had occurred or if a longer time period of data from Hanover was available, a more specific scoring system could be created. In addition, the scoring system is based on the judgment of the person scoring the events, so one person may assign a completely different score to an event than the scores presented in this project.

5.2.3 Conclusions

There were several events that impacted the frequency and severity of Hanover's historical data over the period from 1996 through 2008. After scoring the events, a more generalized set of events was created as a basis for any future events that could possibly occur. These events included: the end of a prolonged economic trend; any behavior by the Federal Reserve; any behavior by Fannie Mae or Freddie Mac; and any new legislation that is passed. A score was assigned for the events for both frequency and severity. However, a direction of change in frequency or severity for any of the generalized event could not be assigned. This was because there was not enough data to infer any assumptions about a change in direction. In addition, the scores assigned to the data were subject to the discretion of the scorer, and one person may have different judgment from another. Therefore, a scoring method is an interesting and potentially powerful tool for determining future loss trends in frequency and severity, but more data is needed, a system to reduce the impact of the judgment of the scorer, and more research into the method itself can be looked at in the future.

5.3 Overall Conclusions for Homeowner's Insurance

For the homeowner's insurance, both the scoring method and correlation approach held value in helping to predict future loss trends. While a group of factors that affect frequency and severity for all states in Hanover's portfolio were not found, once the Excel file used for auto insurance is modified, results will be attainable. As for the scoring method, the approach works, but is not completely consistent, and is subject to the judgment of the person scoring the data. Also some states react differently to events than other states, so with the current quantity of data available from Hanover, it is virtually impossible to accurately predict the impact of future events on frequency and severity. However, this method can be useful once a longer timeframe of Hanover's data can be compared to the timeline of events. Therefore, in the future both the correlation approach and scoring method should continue to be examined in order to provide better, more accurate predictors for future losses.

6. Conclusions & Recommendations

Overall we found a few concrete conclusions for the homeowner's and auto insurance data from Hanover which will aid in predicting future loss trends. For the automotive insurance, the US population and the GDP of services and consumption correlate well with the historical data for the company, and can be used to predict losses in the future. It was also shown that the frequency and severity for auto insurance were not impacted by external individual events because few events affect the entire auto industry. Therefore, only the correlation approach should be used in comparing external factors to Hanover's data and in predicting future losses for auto insurance.

On the homeowner's side, conclusions were only drawn for the correlation approach for Massachusetts, due to emphasis placed on the auto insurance. With further examination stronger conclusions can be made using the Excel model. The scoring method did prove to be much more useful for homeowner's insurance, providing a list of generalized events with accompanying scores. However, the overall conclusions for scoring were weak because the method is based more on judgment than actual data and because individual states react differently to events. Therefore, more emphasis should be placed upon the correlation approach for homeowner's insurance, but the impact of external events cannot be ignored.

A simple linear programming method was briefly explored in an attempt to find a combination of factors which could provide a more accurate predictor of future losses. However, after initial analysis did not improve upon the correlations of individual external factors to Hanover's data, combined with feedback from our sponsors from Hanover, we discontinued using this method and began to create our Excel model.

While some conclusions were drawn for both automotive and homeowner's insurance, more can be done to improve the accuracy and consistency of the conclusions. The linear programming method is something to be explored in the future, as it is possible that a combination of several external factors perfectly correlate with Hanover's data. The scoring method for homeowner's insurance should also be examined further because with more data the accuracy can be improved. Finally, while we were able to find several external factors that correlated well with Hanover's data, it is possible that there are some external factors which we did not obtain data for that may possibly correlate even better.

Bibliography

- 1980sFlashback. (n.d.). *EconomyPrices*. Retrieved from 1990sFlashback: www.1990sflashback.com/1990/Economy.asp
- Benner, K. (n.d.). *AIG woes could swat swap markets.* Retrieved from CNN Money: money.cnn.com/2008/09/16/news/derivatives_benner.fortune/?postversion=2008091621
- Center for Disease Control and Prevention. (n.d.). *BRFSS Prevalence and Trends Data*. Retrieved from Department of Health and Human Services: http://apps.nccd.cdc.gov/brfss/display.asp
- Center for Disease Control and Prevention. (n.d.). *Four Level Smoking Status*. Retrieved from Department of Health and Human Services: http://apps.nccd.cdc.gov/brfss/display.asp?yr=All+Years&state=US&qkey=4394&grp=0&SUBMIT3=Go
- Center for Disease Control and Prevention. (n.d.). *Obesity and Overweight: Trends: U.S. Obesity Trends | DNPAO | CDC*. Retrieved from Department of Health and Human Services: www.cdc.gov/nccdphp/dnpa/obesity/trend/maps/index.htm
- Davidson, A. (n.d.). *How AIG fell apart.* Retrieved from Reuters: www.reuters.com/article/newsOne/idUSMAR85972720080918
- Econstats. (n.d.). *EconStats - GDP data | GDP COMPONENTS | REAL*. Retrieved from Econstats: http://www.econstats.com/gdp/gdp__q2.htm
- EconStats. (n.d.). *EconStats : BLS unemployment Rate unemployment employment labor force*. Retrieved from EconStats: www.econstats.com/BLS/blsla/blsla_cn25_m1.htm
- Econstats. (n.d.). *US Int Rates - Mortgage Rates | ES*. Retrieved from Econstats: http://econstats.com/r/r_em18.htm
- EconStats. (n.d.). *US Int Rates - US Yield Curve, Treasuries | ES*. Retrieved from EconStats: http://econstats.com/r/r_em1.htm
- Famento, Inc. (n.d.). *History of the Automobile.* Retrieved from xtimeline: www.xtimeline.com/timeline/History-of-the-Automobile

- Fatality Analysis Reporting System. (n.d.). *FARS Encyclopedia: States - Crashes and All Victims*. Retrieved from FARS: www-fars.nhtsa.dot.gov/States/StatesCrashesAndAllVictims.aspx
- Journal, A. B. (n.d.). *Fannie Mae increases CRA options.* Retrieved from BNET: http://findarticles.com/p/articles/mi_hb6632/is_200011/ai_n26424963/?tag=rel.res1
- Labaton, S. (n.d.). *Agreement Reached on Overhaul of U.S. Financial System*. Retrieved from The New York Times Company: partners.nytimes.com/library/financial/102399banks-congress.html
- Leonnig, C. D. (n.d.). *How HUD Mortgage Policy Fed The Crisis*. Retrieved from The Washington Post: www.washingtonpost.com/wp-dyn/content/article/2008/06/09/AR2008060902626.html
- McClenahan, C. L. (2001). Ratemaking. In C. L. McClenahan, *Foundations of Casualty Actuarial Science (Fourth Edition)* (pp. Chapter 3, pp. 75-148). Casualty Actuarial Society.
- Miller, R. H. (n.d.). *AICPA Simplification - PROPFIN*. Retrieved from House of Representative: www.house.gov/natcommirs/aicpa7.htm
- National Academy of Engineering. (n.d.). *Automobile - Timeline.* Retrieved from Greatest Engineering Achievements of the 20th Century: http://www.greatachievements.org/?id=3880
- National Highway Traffic Safety Administration. (n.d.). *Traffic Safety Facts*. Retrieved from NHTSA: www-nrd.nhtsa.dot.gov/pdf/nrd-30/ncsa/tsf2000/2000spdfacts.pdf
- Office of Federal Housing Enterprise Oversight. (n.d.). *House Price Index*. Retrieved from Office of Federal Housing Enterprise Oversight: www.ofheo.gov/hpi.aspx?Nav=62
- P., M. (n.d.). *Automobile Business Timeline, 21th Century.* Retrieved from World Business: www.mapreport.com/subtopics/b/c.html
- PBS. (n.d.). *Timeline: Life & Death of the Electric Car.* Retrieved from NOW | PBS: www.pbs.org/now/shows/223/electric-car-timeline.html
- Ritholtz, B. L. (n.d.). *A Memo Found in the Street* . Retrieved from Barron's: http://online.barrons.com/article/SB122246742997580395.html

- Shiller, R. J. (n.d.). *Understanding Recent Trends in House Prices and Home Ownership.* Retrieved from Federal Reserve Bank of Kansas City: www.kansascityfed.org/publicat/sympos/2007/PDF/2007.08.01.Shiller.pdf
- The Disaster Center. (n.d.). *United States Crime Rates 1960 - 2007*. Retrieved from The Disaster Center: www.disastercenter.com/crime/uscrime.htm
- The Hanover Insurance Group. (n.d.). *Hanover Insurance - Personal Insurance*. Retrieved from Hanover Insurance - Our Policy is Performance: www.hanover.com/thg/personal/index.htm
- Trochim, W. M. (n.d.). *Correlation*. Retrieved from Research Methods Knowledge Base: www.socialresearchmethods.net/kb/statcorr.php
- U.S. Treasury. (2009, February 25). *U.S. Treasury Releases Terms of Capital Assistance Program*. Retrieved February 2009, from http://www.ustreas.gov/press/releases/tg40.htm
- Wikipedia. (n.d.). *Timeline of the United States housing bubble*. Retrieved from Wikipedia: en.wikipedia.org/wiki/Timeline_of_the_United_States_housing_bubble
- Wolfram Research, Inc. (n.d.). *Linear Programming*. Retrieved from Wolfram MathWorld: http://mathworld.wolfram.com/LinearProgramming.html

Appendix A: Final List of External Factors Used to Correlate Data

BMI	
BMI-OK	Percentage of population with a BMI that is considered normal
Overweight	Percentage of population with a BMI that is considered overweight
Obese	Percentage of population with a BMI that is considered obese
Crime Rate	
Property	Rate of property crimes per US population per year
Burglary	Rate of burglaries per US population per year
Robbery	Rate of robberies per US population per year
Total	Rate of total crimes per US population per year
Fatality	
Fatality Rate	Fatality rate per 100,000 drivers
Fatality/Licensed	Fatality rate per 100,000 licensed drivers
Fatality/Registered	Fatality rate per 100,000 registered drivers
Fatality/Miles	Fatality rate per 100 million vehicle miles traveled
Gas Type	
Gas	Price of Gas
Diesel	Price of Diesel Fuel
Miles Traveled	Total miles traveled per year
Gas Consumption	
Total	Total gas consumed per year
Highway	Total gas consumed on highways per year
Non-highway	Total gas consumed on non-highways per year
GDP	
Real GDP	Total GDP
Personal Consumption	GDP of personal consumption expenditures
Services	GDP of services
Structures	GDP of structures
Interest	Quarterly interest rate of US yield curve
Population	Total US population
Speed Related	
Fatal Crashes	Total fatal US crashes by year
Vehicle Occupants	Vehicle occupants killed in fatal US crashes by year
Non-vehicle	Non-motorists killed in fatal US crashes by year

Vehicle and Non-vehicle	Combined occupants and non-occupants killed in fatal US crashes by year
Registered Vehicles	Fatal crashes involving registered vehicles
Licensed Vehicles	Fatal crashes involving licensed drivers
Tobacco	
Adult	Adults who are current smokers
Everyday	People who smoke everyday
Someday	People who only smoke on some days
Former	People who quit smoking
Vehicles	
All Vehicles	All vehicles involved in fatal car crashes
Automobiles	All automobiles involved in fatal car crashes
All Trucks	All trucks involved in fatal car crashes
Light Trucks	All light trucks involved in fatal car crashes
Autos and Light Trucks	All automobiles and light trucks involved in fatal car crashes

Appendix B: Auto Correlation Results

MASSACHUSETTS	Frequency	BI	CM	CO	CSL	PD	PIP	Severity	BI	CM	CO	CSL	PD	PIP	Overall
Overweight														x	1
Obese								x							1
BMI_OK															
Crime_Rate_Total	x			x		x	x								4
Robbery															
Burglary								x	x						2
Vehicle_Theft		x				x									2
Fatality_Rate_MA			x												1
Fatality_Licensed_MA	x		x	x		x									4
Fatality_Registered_MA	x		x	x			x								4
Fatality_Miles_MA	x		x												2
Nonhighway_Gasoline_MilGal_Demand															
Highway_Gasoline_MilGal_Demand															
Total_Gasoline_MilGal_Demand								x							1
Population	x		x					x	x				x		5
Speed_Licensed_Drivers	x							x	x				x		4
Speed_Registered_Vehicles	x							x	x				x		4
Speed_Total_Vehicle_Nonvehicle		x													1
Speed_Nonvehicle															
Speed_Vehicles Occupants	x			x		x									3
Speed_Fatal Crashes															
Tobacco_Former															
Tobacco_Someday															
Tobacco_Everday	x		x	x											3
Tobacco_Adults	x		x												2
Diesel_Price	x	x					x		x	x					5
Miles															
Gas							x								1
P_Consump	x		x				x	x		x			x		6
Services	x		x					x		x			x		5
Structures		x							x						2
Interest_Rate															

MICHIGAN	Frequency	BI	CM	CO	CSL	PD	PIP	Severity	BI	CM	CO	CSL	PD	PIP	Overall
Overweight															
Obese	x	x		x	x	x		x							6
BMI_OK					x	x									2
Crime_Rate_Total			x	x					x						3
Robbery									x						1
Burglary															
Vehicle_Theft															
Fatality_Rate_MI															
Fatality_Licensed_MI										x					1
Fatality_Registered_MI	x	x	x	x		x		x	x	x					8
Fatality_Miles_MI				x				x		x					3
Nonhighway_Gasoline_MilGal_Demand															
Highway_Gasoline_MilGal_Demand															
Total_Gasoline_MilGal_Demand															
Population	x	x	x	x	x		x	x	x	x				x	10
Speed_Licensed_Drivers	x	x	x	x	x	x		x		x				x	9
Speed_Registered_Vehicles	x	x	x	x				x	x	x				x	8
Speed_Total_Vehicle_Nonvehicle															
Speed_Nonvehicle															
Speed_Vehicles Occupants															
Speed_Fatal Crashes															
Tobacco_Former															
Tobacco_Someday															
Tobacco_Everday	x	x	x	x		x		x							6
Tobacco_Adults		x				x		x							3
Diesel_Price	x		x				x	x	x						5
Miles															
Gas	x														1
P_Consump	x	x	x	x			x	x	x	x	x			x	10
Services	x	x	x	x			x	x	x	x	x			x	10
Structures															
Interest_Rate															

NEW JERSEY	Frequency	BI	CM	CO	CSL	PD	PIP	Severity	BI	CM	CO	CSL	PD	PIP	Overall
Overweight						x									1
Obese	x		x				x	x	x					x	6
BMI_OK									x					x	2
Crime_Rate_Total	x		x	x			x							x	5
Robbery					x				x						2
Burglary					x										1
Vehicle_Theft															
Fatality_Rate_NJ									x					x	2
Fatality_Licensed_NJ									x					x	2
Fatality_Registered_NJ								x	x					x	3
Fatality_Miles_NJ								x	x					x	3
Nonhighway_Gasoline_MilGal_Demand						x									1
Highway_Gasoline_MilGal_Demand													x		1
Total_Gasoline_MilGal_Demand															
Population	x		x	x				x	x					x	6
Speed_Licensed_Drivers	x							x	x					x	4
Speed_Registered_Vehicles	x		x				x	x	x					x	6
Speed_Total_Vehicle_Nonvehicle												x			1
Speed_Nonvehicle															
Speed_Vehicles Occupants									x					x	2
Speed_Fatal Crashes												x			1
Tobacco_Former															
Tobacco_Someday															
Tobacco_Everday	x							x	x					x	4
Tobacco_Adults									x					x	2
Diesel_Price							x		x					x	3
Miles															
Gas							x							x	2
P_Consump								x	x				x	x	4
Services								x	x				x	x	4
Structures															
Interest_Rate				x											1

NEW YORK	Frequency	BI	CM	CO	CSL	PD	PIP	Severity	BI	CM	CO	CSL	PD	PIP	Overall
Overweight			x	x			x					x			4
Obese					x			x							2
BMI_OK															
Crime_Rate_Total					x										1
Robbery									x		x				2
Burglary											x		x		2
Vehicle_Theft															
Fatality_Rate_NY											x				1
Fatality_Licensed_NY											x				1
Fatality_Registered_NY	x				x			x			x				4
Fatality_Miles_NY											x				1
Nonhighway_Gasoline_MilGal_Demand	x		x	x			x					x			5
Highway_Gasoline_MilGal_Demand														x	1
Total_Gasoline_MilGal_Demand					x									x	2
Population					x			x			x		x		4
Speed_Licensed_Drivers					x			x			x		x		4
Speed_Registered_Vehicles					x			x			x		x		4
Speed_Total_Vehicle_Nonvehicle															
Speed_Nonvehicle							x								1
Speed_Vehicles Occupants											x				1
Speed_Fatal Crashes			x	x											2
Tobacco_Former							x								1
Tobacco_Someday						x				x					2
Tobacco_Everday	x				x			x			x				4
Tobacco_Adults											x				1
Diesel_Price								x			x				2
Miles															1
Gas															
P_Consump								x			x		x		3
Services								x			x				2
Structures															
Interest_Rate															

Appendix C: Homeowner Correlation Results

Factor	Frequency	Severity
BMI		
BMI OK	0.694487078	-0.895716442
Overweight	0.255724283	0.030556154
Obese	-0.827175718	0.948896945
Crime Rate		
Property	0.784124249	-0.885451544
Burglary	-0.466531021	0.288919791
Robbery	-0.699332564	0.88839827
Total	-0.385512115	0.728264213
GDP		
Real GDP	-0.815658639	0.970411572
Personal Consumption	-0.795697954	0.967810003
Services	-0.78752814	0.964050347
Structures	0.106735043	-0.105755631
Home Value	-0.892153671	-0.517498162
Interest	0.0245	-0.184944788
Mortgage	0.551301878	-0.614328135
Population	-0.83400956	0.975489738
Tobacco		
Adult	0.802736399	-0.937453297
Everyday	0.793419388	-0.953750961
Someday	-0.123593897	0.300768506
Former	0.235060632	-0.022242511