Casualty Actuarial Society E-Forum, Winter 2012 Volume 2

The CAS *E-Forum*, Winter 2012-Volume 2

The Winter 2012-Volume 2 of the CAS *E-Forum* is a cooperative effort between the CAS *E-Forum* Committee and various other CAS committees, task forces, or working parties.

This *E-Forum* includes papers from two call paper programs and two additional papers. One program, conducted by the CAS Committee on Management Data and Information in conjunction with the Insurance Data Management Association (IDMA), was issued for papers on data management, data quality, and data technology topics. The purpose of this program is to develop a source of literature on data topics important to casualty actuaries. The other program was conducted by the CAS Committee on Ratemaking.

The call papers will be presented during the CAS Ratemaking and Product Management Seminar in Philadelphia, March 19-21, 2012.

CAS *E-Forum,* Winter 2012-Volume 2

Table of Contents

E-Forum Committee

Windrie Wong, *Chairperson*

Cara Blank
Mei-Hsuan Chao
Mark A. Florenz
Karl Goring
Dennis L. Lange
Shayan Sen
Rial Simons
Elizabeth A. Smith, *Staff Liaison*
John Sopkowicz
Zongli Sun
Yingjie Zhang

For information on submitting a paper to the *E-Forum*, visit http://www.casact.org/pubs/forum/.

Social Media Analytics:
Data Mining Applied to Insurance Twitter Posts

Roosevelt C. Mosley Jr., FCAS, MAAA

Abstract

The use of social media has grown significantly in recent years. With the growth in its use, there has also been a substantial growth in the amount of information generated by users of social media. Insurers are making significant investments in social media, but many are not systematically analyzing the valuable information that is resulting from their investments.

This paper discusses the application of correlation, clustering, and association analyses to social media. This is demonstrated by analyzing insurance Twitter posts. The results of these analyses help identify keywords and concepts in the social media data, and can facilitate the application of this information by insurers. As insurers analyze this information and apply the results of the analysis in relevant areas, they will be able to proactively address potential market and customer issues more effectively.

Keywords. Social media, analytics, data mining, text mining, clustering, association analysis

1. INTRODUCTION

Regardless of where you look, you can see an explosion in the use of social media. Online communities have developed that focus on both personal and professional lives. Groups have been formed that focus on every potential area of interest, including food, sports, music, parenting, scrapbooking, and actuarial issues. It is estimated that there are over 900 social media sites on the internet. Some of the more popular platforms are Facebook, Twitter, LinkedIn, Google Plus, and YouTube. To help understand the explosion in the use of social media, consider the following statistics which were compiled at www.Banking2020.com [1] in January 2011 and by Danny Brown on his blog at www.dannybrown.me [2].

- People spend over 500 billion minutes per month on Facebook.
- There are 200 million registered Twitter accounts.
- There are more than 70 million users of LinkedIn worldwide.
- YouTube receives more than 2 billion viewers per day.
- Seventy-seven percent of internet users read blogs.

The majority of the population is using social media in some form or another. Given the substantial increase in the use of social media, there is a significant amount of information that is

being generated. As seen in the same sources referenced above, the volume of this content is staggering:

- More than 30 billion pieces of content are shared each month on Facebook.
- Every minute, 24 hours of video is uploaded to YouTube.
- As of December 2010, the average number of tweets sent per day was 110 million.
- There are currently 133 million blogs listed on leading blog directory Technorati.

So not only are people joining and accessing social media sites, but they are also spending time engaging in social media and creating a significant amount of content. As a result of this time spent on social media and the information being generated, businesses have taken notice and are attempting to leverage the power of social media to help them succeed. According to wealthinvest.com [3],

- Two-thirds of comScore's U.S. Top 100 websites and half of comScore's Global Top 100 websites have integrated with Facebook.
- Many businesses now have established Twitter accounts in an attempt to connect with current and potential customers.
- Eighty-eight percent of companies use LinkedIn as a recruitment tool.
- Corporate blogging accounts for 14% of blogs.

The commitment that businesses are making to increase their presence in social media is also being shown in the resources they are committing to this effort. According to eMarketer, U.S. marketers will spend over $3 billion to advertise on social media sites in 2011, which is a 55% increase over what was spent in 2010, and 11% of what they spend on online advertising overall. Also, according to Banking2020.com, 50% of Chief Marketing Officers at Fortune 1000 companies say they have launched a corporate blog because it is a cost of doing business today. So not only is the corporate investment being evidenced by dollars spent but also in the time it takes to create and maintain social media efforts.

Insurance companies have joined in this effort by businesses to use social media. The Customer Respect Group (CRG) produces a monthly newsletter entitled "Social Eyes: The Insurer's View of Social Media." [4] This newsletter focuses on trends and news related to the use of social media by insurance companies. As part of the newsletter, CRG tracks the use of insurer social media sites.

One of the categories that is tracked is the use of Facebook. In the June, 2011 issue of "Social Eyes," CRG tracks 36 insurance company corporate Facebook pages that have a collective total of over three million fans. CRG also tracks other insurance company-related Facebook pages, such as advertising personalities or pages that are targeted toward a specific demographic. These ten Facebook pages have over 5.7 million fans, with the largest being the Facebook page of Flo, the Progressive Girl, who has just over 3 million fans. There is also a section in the newsletter that tracks the Twitter followers and activity of 30 corporate insurance company Twitter identities.

CRG describes the different ways that social media is being used by insurance companies. Insurance companies are using social media for broad purposes, such as communicating general content and promotional advertising, but some are also using the platform to contact and communicate with customers directly. In addition, insurance companies are using social media as a platform for promoting and raising money for charities, i.e., donating to a particular cause for every new fan or for every new Twitter follower.

The use of social media sites has grown significantly, and this fact is being recognized by businesses, including insurance companies. In response, insurance companies and other businesses are investing significant time and resources into establishing and maintaining a social media presence. All of this is feeding into the exponential increase in the amount of data and information that is being generated. This then raises a significant question. What are companies in general, and insurance companies specifically, doing with all of this information? Every Facebook post, every Tweet, every blog entry, every connection with social media generates a new data point, a new bit of information that may be of value to insurance companies. This information might help an insurance company service a policyholder better, connect with a potential new customer, identify a need or concern in the marketplace, or uncover a competitive issue they may be facing. Social media platforms provide opportunities for consumers to share their thoughts with a broader audience, and in understanding how customers are feeling or what they are facing, insurers can better interact with consumers.

How can insurers access and begin to make use of this information? Based on the statistics above related to the amount of social media content that is available, the task can be overwhelming. One approach is for a company to hire a team of people to monitor social media sites for content that might be valuable to a company. This would require the team to read through content, identify information that the company may be able to use, and then to identify the proper channels through which to route that information. While this can and is being done, the obvious challenge is that as

the amount of social media content continues to increase, more human resources will be required to analyze it. This means either increasing the budget for social media efforts, or simply living with the fact that you cannot analyze everything, thus potentially missing valuable information.

The purpose of this paper is to describe, through the use of a specific example, how data mining and analytics can be applied to social media. Data analytics provides insurance companies with a systematic, organized, and powerful way to analyze social media information and extract the valuable information in it without needing to read through every piece of content. Using the power of analytics, key areas of importance can be identified, and these areas can then be investigated further. This can optimize the time spent by focusing the analysis on those areas that are of the greatest potential benefit to the company.

1.1 Research Context

The context of this research will focus on data and text analytics. Since much of the data from the social media sites will be text-based data, the process of preparing and analyzing the data will focus on principles of preparing text data for analysis. The author was unable to find anything in CAS literature that focuses specifically on the analysis of social media. However, a good discussion of the principles of text mining in CAS literature is in a paper written by Louise Francis entitled "Taming Text: An Introduction to Text Mining." [5] Building on these concepts, there are some unique considerations when analyzing text data from social media sites which will be discussed in this paper.

1.2 Objective

The purpose of this paper is to describe, through the use of a specific example, how data mining and text analytics can be applied to social media to identify key themes in the data. Specifically, this paper will describe the analysis of Twitter posts related to the keyword Allstate. Allstate was chosen purely based on the public availability of historical Twitter data. While this example helps to make some of the points and concepts clearer, the purpose of this paper is not to provide a detailed analysis of Twitter activity related to Allstate, but to demonstrate how analytics can specifically be applied to social media information related to a property and casualty insurance company.

1.3 Outline

Section 2 will provide a general background and description of Twitter and will describe the data used in this analysis. Section 3 will provide some general descriptive statistics about the data. Section

4 will discuss the steps that were taken to prepare the data for analysis. Section 5 will describe the analysis of the tweets and also provide some insight into the results of the analysis. Section 6 will outline some of the challenges associated with analyzing social media information. Finally, Section 7 will give applications of these types of analyses for insurance companies.

2. TWITTER BACKGROUND AND DATA DESCRIPTION

Twitter is a social networking site that allows users to send and read short messages of a maximum of 140 characters. Twitter was created in March 2006 and was officially launched in July 2006. The growth of Twitter has been phenomenal, currently having reached over 200 million users and handling over 200 million tweets per day. Users sign up for an account on Twitter, and once they have an account they can begin to "tweet," which is the terminology for sending a message. Users can subscribe to other user's tweets, a process known as "following." These subscribers are known as "followers." By default, tweets that a user sends are visible to everyone; however, users can also choose to send tweets specifically to their followers that will not be visible to the public.

Users on Twitter are identified by a user name, and this user name is proceeded by the "@" symbol. When a user identifies another user in their tweet by their user name, it will be visible to the public, and the user that is referenced will be notified by Twitter that they have been "mentioned." If a user sees a tweet that is interesting and wants to pass the information along, they can "retweet" the post, which is similar to forwarding an email message to a new set of users, in this case their followers. Retweets will generally be identified with an "RT" that is embedded in the message. Lastly, messages can be grouped by topic or type by the use of hashtags (#). A hashtag preceding the topic will allow Twitter users to find tweets related to a particular topic when performing a search.

Twitter also has a location function. If users are tweeting from a mobile device, they can choose to turn on their location, and their latitude and longitude will be captured with the tweet.

Tweets can be related to anything, but much of the content on Twitter is related to several key categories. These categories were outlined in research done by Pear Analytics in 2009 on 2,000 tweets [6]. This study found that tweets were primarily related to six categories:

1. Pointless babble – 40%
2. Conversational – 38%

3. Pass along value – 9%
4. Self-promotion – 6%
5. Spam – 4%
6. News – 4%

While these numbers are related to a study that was done two years ago when Twitter was not as widely used as it is now, the general categorization of tweets likely still holds. As it relates to insurance companies, the areas of interest would be categories 2, 3, and 6, which account for 51% of tweets. Certainly, not all tweets in these categories will be useful to insurers, so the challenge is to determine how to analyze the tweets in such a way that the important information is separated from the information that is not important.

To demonstrate how this can be done, a dataset of insurance company tweets was identified for analysis. Twapperkeeper.com was a web service that tracked and archived Twitter posts based on archives that were set up by users. To track tweets related to a particular topic or user, users could go to this site and establish an archive, and twapperkeeper.com would track and archive those tweets. On July 29, 2010, an archive titled #allstate was established. Tweets for this hashtag were collected from this starting point, with the exception of about a five-week time period (to be explained later). This analysis used tweets through August 12, 2011. [Author's Note: Since the paper was completed, twapperkeeper.com was fully integrated into hootsuite.com. This site allows users to archive social media data based on defined search criteria. However, archives established by other users cannot be accessed unless explicit permission is given by the owner of the archive.]

The data that was captured from twapperkeeper.com includes the following information:

- User: the username that sent the tweet
- Tweet: the content of the tweet
- Timestamp: the date and time the tweet was sent (GMT)
- Tweet ID: Twitter identification number of the tweet
- Geo: latitude and longitude of the user

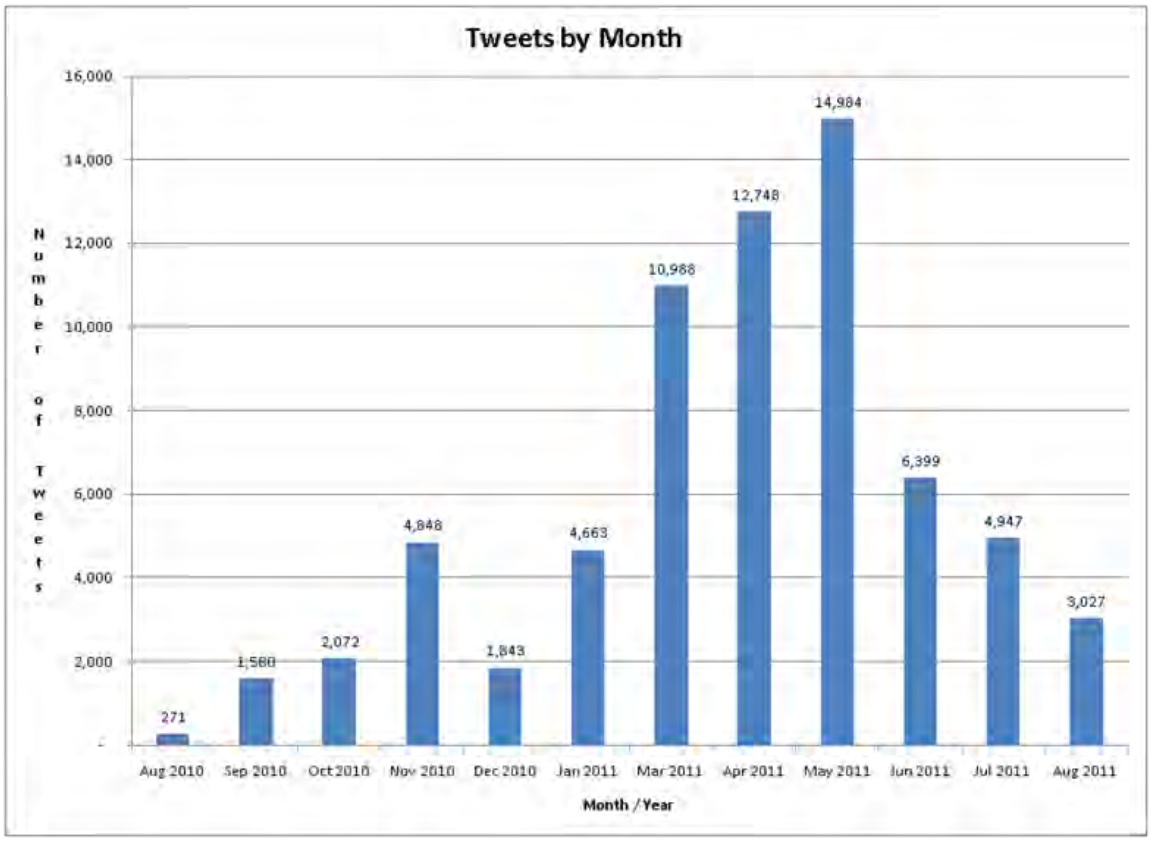
It must also be remembered that this data was captured based on the use of the hashtag Allstate. Therefore, it will not capture every tweet that uses the word Allstate, but rather those tweets where the user specifically identified Allstate as a keyword. Also, twapperkeeper.com makes no guarantees

that they capture all tweets that meet the archive criteria, so there could potentially be tweets with #allstate that were not captured. While this may introduce a bias, the concepts for analyzing the tweets are still valid.

3. GENERAL DESCRIPTIVE STATISTICS

There are a total of 68,370 tweets that were used as part of this analysis. The tweets used began on August 1, 2010 and ended on August 12, 2011. The number of tweets by month is shown below.

Figure 1: Tweets per Month



As can be seen in the figure above, the number of tweets captured per month varied between 1,500 and just under 5,000 through January 2011, at which point the number of tweets increased to 10,000 – 15,000 per month for March through May 2011. June and July settled back to pre-March, 2011 levels. August 2011 only represents 12 days of tweets, so it was not a complete month as of the time this paper was written.

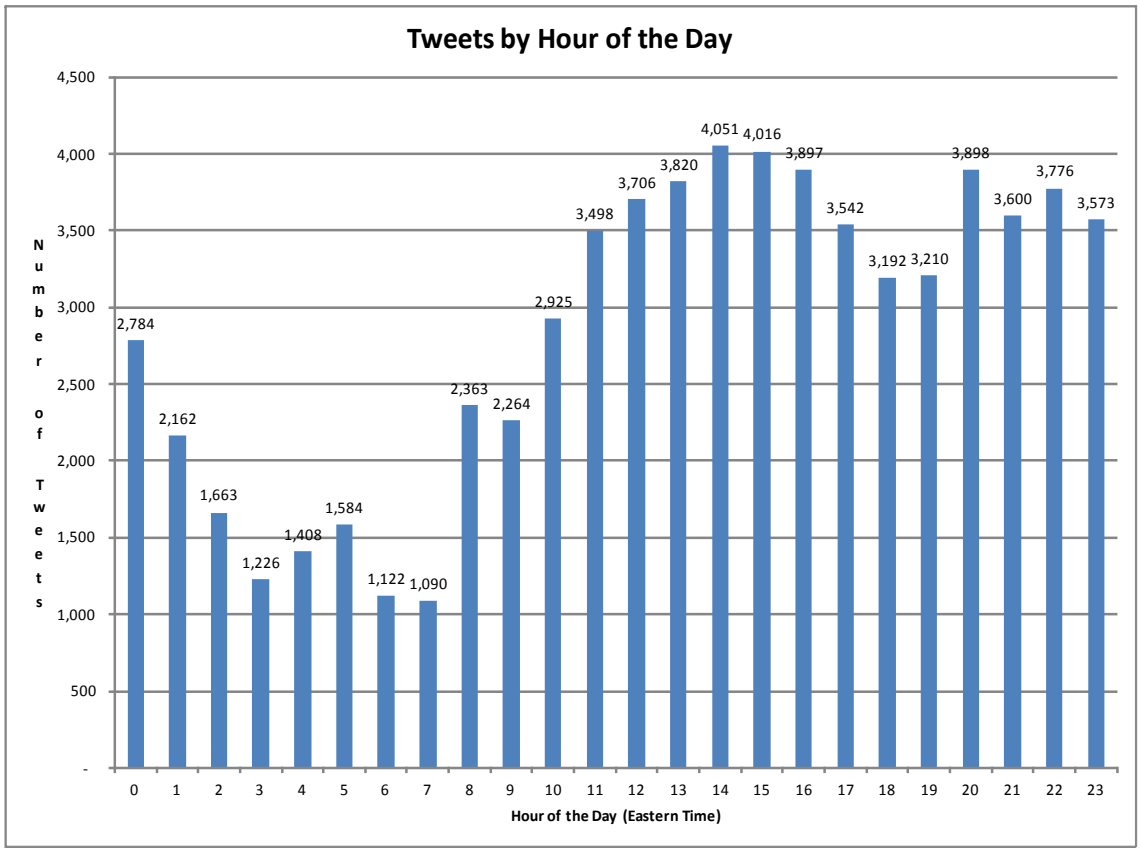
Throughout this paper, as will be described later, we will attempt to uncover concepts being communicated in the data through the use of various data mining techniques. Generally, what the uncovering of these realities does is create more questions which require more investigation to come up with more complete answers. There are several questions which are apparent from Figure 1. First, there seemed to be a significant increase in the number of tweets per month from March to May 2011, followed by a decrease. In addition, as mentioned earlier, there is about a five-week period during which no tweets were captured. We will address the issue related to the increase in tweets later in this paper. Regarding the missing tweets, the author contacted twapperkeeper.com to inquire about the missing data. Early in 2011, there were some issues with the archiving servers, and as a result some of them had to be taken offline for a period of time. The #allstate archive was taken offline during part of January and February, and thus resulted in the five-week period during which no tweets were captured.

What was interesting as well is that there were 40,258 unique users that generated the 68,370 tweets. This equates to an average of only 1.69 tweets per user. The Allstate corporate Twitter identity generated the largest number of tweets at 1,169, which only equated to 1.6% of the total tweets. Over 33,000 of the users made only one comment over the study time period, so the overall conversation was not dominated by just a few users. In fact, the top 100 users only accounted for 13.4% of the tweets, and the top 1,000 users only accounted for 29.2% of the total number of tweets. Even when looking at the traditional 80/20 rule (80% of the involvement comes from 20% of the participants), this particular data falls significantly short of this criteria. The top 20% of the users only account for 52.9% of the tweets.

This underscores one of the realities that underlie social media. The content is really driven by the community rather than specific users. Certainly, there are users that are more active than others, and this activity can be a source of interest for further investigation by a company. But overwhelmingly what social media brings is a sense of the feeling of the community. And if the same sentiment is being expressed by multiple individual users, then it may be something an insurer wants to pay attention to.

In addition to looking at the trends in tweets by month, we also have summarized the tweets by hour of the day to understand the most active times of day for the use of Twitter related to this archive. Figure 2 shows the summaries by hour in Eastern Time.

Figure 2: Number of Tweets by Hour of the Day



As can be seen from the chart above, the most active times for the use of Twitter for this archive were between 11:00 am and midnight Eastern Time. This time period will represent the most active time periods across most of the time zones across the United States. Therefore, this distribution of time periods appears to be reasonable. There is still activity outside of this time period, but the tweet activity is between one-third and two-thirds lower during the late night and early morning hours. This type of information might be helpful to those that are responsible for monitoring, analyzing, and responding to Twitter activity, especially if trends in the type of activity by time of day can be determined.

There are other types of general descriptive statistics that can be calculated based on the data. The goal in generally describing the data is to determine whether the data appears to be reasonable, to determine the applicability of the data for the intended purpose, and to identify any potential gaps or concerns with the data.

4. DATA PROCESSING

The analysis of social media data is heavily dependent on the ability to analyze text data. However, there are some unique considerations in the analysis of social media data that make it different than a normal text mining analysis. One major consideration is that social media data tends to be informal, so issues with misspellings and abbreviations will be a larger challenge. In addition, in the case of Twitter, there are certain symbols that actually do have a meaning and therefore extra care needs to be taken in cleansing the text.

Much of the work required to analyze social media data will be spent obtaining and preparing the data for analysis. This is not a trivial exercise, and the proper approach for a company will depend on the ultimate purpose of the analysis. While the purpose of this paper is not to provide a complete discussion on obtaining social media data, we have listed a few approaches here that the reader can pursue further. As described earlier, the source of the data for this analysis is twapperkeeper.com, which is a service that captures and archives Twitter posts. There are third party applications which can capture social media data from websites, and these applications appear to be both web-based services as well as stand-alone programs. Developers can also create computer programs which monitor and capture information from their social media sites. Programs also exist which scrape information from screens, and this can be applied to monitoring and collecting social media data from websites. Ultimately, companies will need to work with their information technology departments to determine the best approach for collecting and storing social media data.

The first step in analyzing the text data is to remove all the punctuation and symbols. This information generally does not add to the understanding of the text and will make it more difficult to decipher the words that are part of the tweet. This information includes single and double quotation marks, parentheses, punctuation marks, and stray symbols (dollar signs, stars, etc.). In the initial data cleaning, the signs that actually have meaning for Twitter (@, #) were retained.

Once the tweet has been cleansed of punctuation marks and symbols, then the tweet can be parsed into words. This parsing occurs by identifying spaces and using these spaces as the indication of one word ending and another word beginning. The number of potential words will depend on the source of the data. In the case of Twitter, posts are limited to 140 characters, so identifying up to 35 words for this analysis was sufficient. There are other sources of social media such as Facebook where one will need many more words than this to capture all the content.

At this point, the data is structured in a manner shown below:

Table 1: List of Words in Rows

Tweet ID	User	Tweet	Word1	Word2	…	Word35
1	@mosley	Text of tweet	W1	W2	…	W35

Next, we want to determine the frequency of words present in these tweets. To do this, we stack all of the word columns into one column, and then summarize the word frequencies based on this combined column. This will change the data structure from tweets in rows to tweets in columns with one word per row. In stacking these words, care must be taken to maintain the Tweet ID and the word order, since this will be important later in the analysis. The structure of the data once this transformation is made is shown below:

Table 2: List of Words in Columns

Tweet ID	Word Order	Word
1	1	Word1
1	2	Word2
…	…	…
1	35	Word35

This data can then be summarized by word to determine the frequency of words in the tweets. It is here that one will find many different types of words that may not be beneficial to the analysis, words such as "a, an, the, in, I, on, and of." Therefore, these words should be removed at this point so they do not unnecessarily slow down the analysis. The top 10 words in this analysis are shown below.

Table 3: Top 10 Keywords

Word	Word count	Pct of Words
allstate	70815	7.0%
insurance	16868	1.7%
Rt	9292	0.9%
jobs	6093	0.6%
commercials	5502	0.5%
arena	5327	0.5%
good	5132	0.5%
mayhem	5113	0.5%
job	4548	0.5%
like	4007	0.4%

The most prevalent keywords are not surprising. They are related to several different areas that are important to insurance companies. Obviously, the term "insurance" would be expected. "Jobs" appear to be an important topic in insurance company tweets, showing up in total over 10,000 times in the dataset. In addition, there are two words in the top ten related to Allstate advertising, including "commercials" and "Mayhem," which is one of the current advertising campaigns being run by Allstate. Lastly, there are two words related to insurance company slogans present in the top 10, "good" and "like."

In addition to the analysis of the frequency of words present in the tweets, because tweets use hashtags to identify keywords, an analysis of the keywords identified can also be undertaken. Based on the #allstate archive, the top 10 hashtags are:

Table 4: Hashtag Frequency

Hashtag	Count
#allstate	5,959
#jobs	4,328
#job	2,726
#hiring	1,266
#insurance	1,071
#sales	962
#mayhem	498
#tweetmyjobs	315
#tweetajob	283
#coupon	248

There are similar words present in the hashtags as there are in the keyword analysis. Five of the categories are related to employment, including either "jobs" or "hiring." Mayhem is also mentioned here as well. The difference in this list that stands out is the hashtag "coupon." After further investigation of these tweets, they appeared to be coupons offered by either Allstate as they participated in home shows or agencies offering coupons for local merchants.

A simple application of understanding the keywords that are present in tweets would be to set up rules that are triggered by certain keywords. For example, one the keywords present in the data is claims. This word could be a trigger for identifying insureds with questions or concerns about filing claims or the claims process. Another keyword is quote. This could be a trigger to identify potential customers who are looking for information on insurance prices.

There are two issues with the analysis of text data that can be corrected at this point. The first issue is misspelling, and the second issue is different cases or variations of the same word. Again, because of the informal nature of social media, misspellings are common. Also, as can be seen in the top ten words, there can be different tenses or cases of the same word, as with "job" and "jobs." In each of these cases, the desire in the analysis is to reflect the intent of the user. In order to do this, we want to either correct the spelling or make the different tenses consistent. Both of these issues can be addressed using the same approach.

Two techniques were used in this analysis to identify spelling and tense or case differences. One approach is the comparison of two strings by computing the Levenshtein edit distance (LED). The LED is defined as the number of insertions, deletions, or replacements of single characters that are required to convert one string to another. For example, the LED between "job" and "jobs" is 1,

since the strings would be equal by either adding an "s" to the first word or deleting the "s" from the second. Another approach to comparing strings is to calculate the generalized edit distance (GED). The GED between two strings (string1 and string2) is calculated as the minimum cost sequence of operations for constructing string2 from string1. In this calculation, different operations have different penalties associated with them. For example, inserting or deleting a character to create string1 incurs a penalty of 100, but the difference of a blank between the strings only incurs a penalty of 10 points. In the case of the example above, the generalized edit distance between "job" and "jobs" is 100 points, since inserting or deleting an "s" incurs a penalty of 100 points. [Note: The author used SAS™ for this analysis, and the specific SAS functions used were COMPLEV and COMPGED. A description of the COMPGED word operations and the points associated with each can be found at the following URL: http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002206133.htm.]

In the case of both distance calculations, the smaller the distance between two words, the more similar the two words are. The keywords that are identified can be compared to the remainder of the words to determine if there are misspellings, or if there are different tenses or cases of words that are present. Based on an investigation of the distances, a cutoff point can be selected to investigate further whether words can be considered as the same. Once these distances have been calculated, then the list of words can be edited to correct the misspellings and to make word variations consistent.

The last step in the data preparation is to add to Table 1 a set of indicators based on the identified keywords. Based on the keywords identified, an indicator can be added to the table that indicates the presence of a word in a particular tweet. For this analysis, there were 116 keywords identified, and so 116 indicators were added to the dataset that are either 0 or 1 depending on whether the word was present. The structure of the final table with tweets by row is shown below.

Table 5: Example of Structure of Final Table

Tweet ID	User	Tweet	allstate	insurance	commercial	company	…
19	@mosley	Allstate insurance company	1	1	0	1	…

5. TWEET ANALYSIS METHODOLOGY AND RESULTS

In the data exploration phase, keywords are identified in the data and adjustments are made to the data to prepare it for analysis. The purpose of the analysis of social media is to identify patterns and trends that are present in the information which may be of further use to the insurance company. To achieve this goal, we need to identify patterns and combinations of words that will indicate themes and ideas. One step in doing this is a simple correlation analysis, which will identify correlations between pairs of words. There are also two additional types of analyses that will be performed on the data. The first will be a clustering analysis which will group tweets based on their similarities or dissimilarities. The second will be an Association Analysis which analyzes the occurrence of specific words together.

5.1 Correlation Analysis

The correlation statistic used was a Cramer's V statistic for pairs of keyword indicators. Cramer's V indicates the level of association between two nominal variables. To calculate the Cramer's V statistic, assume a 2 x 2 matrix indicating the frequency of the combination of two words.

Table 6: Frequency of Word Combinations

Word 1 / Word 2	**0**	**1**
0	n_{00}	n_{01}
1	n_{10}	n_{11}

The notation n_{ij} represents the frequency of the combination of words in the dataset. For example, n_{00} counts the number of tweets where neither Word 1 or Word 2 were present. $n_{.j}$ is the total frequency for column j, while $n_{i.}$ is the frequency for row i. Given this two-by-two table structure, the formula for Cramer's V can be simplified to:

$$\text{Cramer's V} = \frac{n_{00}\, n_{11} - n_{01}\, n_{10}}{\sqrt{n_{0.}\, n_{1.}\, n_{.0}\, n_{.1}}} \qquad (1)$$

The result is a number between -1 and 1. A value of -1 indicates a perfect negative correlation, a value of 1 indicates a perfectly positive correlation, and 0 indicates no correlation. The top 20

combinations of words with the largest Cramer's V statistics are shown below.

Table 7: Top 20 Cramer's V Statistics

Number	var1	var2	Cramer's V Statistic
1	state	farm	0.861
2	financial	personal	0.803
3	good	hands	0.734
4	agency	purchase	0.663
5	jobs	gravy	0.661
6	esurance	answer	0.612
7	girl	neighbor	0.508
8	youtube	jonas	0.505
9	watch	neighbor	0.489
10	work	neighbor	0.483
11	Love	basketball	0.454
12	Youtube	video	0.452
13	Geico	progressive	0.427
14	Girl	watch	0.420
15	insurance	company	0.413
16	Farm	neighbor	0.405
17	Agent	exclusive	0.394
18	Girl	work	0.387
19	Watch	work	0.366
20	Billion	answer	0.360

There are several categories of correlation that are apparent in this list. There are several correlations which are related to competitors, including State Farm, Progressive, and GEICO. There are also several pairs of words related to employment, including agency purchase, and jobs. There are also pairs of words related to characteristics of the company, which include "good hands," "personal financial," and "insurance company." Another category includes entertainment and other related items, such as "YouTube Jonas," "YouTube video," and "love basketball."

The YouTube and Jonas categories related to a public service campaign that was sponsored by Allstate in which the Jonas brothers participated. "Love and basketball" refers to a movie that Dennis Haysbert had a part in. Haysbert is now a spokesperson for Allstate, which is the connection to this dataset.

While the correlation is a simple approach that will begin to uncover combinations of words, it does not give a complete picture of the words and concepts that may be present in the tweet. The

most obvious limitation is that only pairs of words are compared in this example. Since phrases can be up to 35 words long, understanding only the relationship between pairs of words will miss concepts that are present. In addition, it is difficult to understand how many records are impacted by certain pairs of correlated words because multiple combinations could be present in the same tweet. This problem is addressed by techniques that determine the presence of combinations of words and phrases, which are discussed in the remainder of this section.

5.2 Clustering

The next approach that can be applied to social media analysis is a cluster analysis. The clustering procedure is based on calculating distances between observations and is used to segment databases. Clusters are developed such that observations that are in the same cluster tend to be similar, and objects in different clusters tend to be dissimilar. The clusters developed in this paper use the Ward's Minimum-Variance method. Using this method, the distance between two clusters is calculated as the ANOVA sum of squares between the two clusters summed up over all the variables. The goal of each step of the process is to minimize the within cluster sum of squares.

This method was applied to the 116 keyword indicators which were identified in the data exploration phase. The results of applying this method to the #allstate archive resulted in 47 clusters, or 47 groups of observations that were combined based on their similarities. For each cluster, there are several ways the output can be viewed to attempt to understand what the results show. For each cluster, the percentage of tweets that contain each word is calculated, and from these percentages it can be determined which concepts are predominate within a particular cluster. In addition, the percentage of tweets in a cluster that contain a particular word can be compared to the overall percentage of tweets that contain that word. This will help analysts to see where words are showing up most, even if a keyword does not show up in a large percentage of tweets overall. The percentage of times a word shows up in a particular cluster can also be ranked across all clusters, which will also help the analyst see quickly where words are showing up most frequently.

To determine how important a word is in a cluster, we calculate a ratio called cluster lift.

$$\text{Cluster Lift (word)} = \frac{\text{Percentage of tweets in a cluster that include word}}{\text{Percentage of all tweets that include word}} \quad (2)$$

This calculation provides an indication of how much more likely a word is to be in a particular cluster than it is present in the dataset overall. The cluster lifts are shown in Exhibit 1. This is shown

for a selection of the keywords. We have highlighted cells with a cluster lift greater than 4.0 to show the keywords that are prevalent in each cluster (4.0 was chosen for demonstrative purposes – the proper level of significance for a particular analysis should be determined by the reader). For example, in cluster 18, the word "mayhem" has a cluster lift greater than 15, and the word "guy" has a cluster lift of 5. Therefore, it can be reasonably concluded that tweets in this cluster have something to do with Allstate's commercial personality. In cluster 12, the term "good" has a cluster lift of 12.8, and "hands" has a cluster lift of 19.5. This cluster thus has tweets related to the slogan associated with Allstate.

Exhibit 2 shows the keywords for each cluster that are greater than the 4.0 threshold. Based on the keywords present, the clusters can be grouped into general themes. In reviewing the keywords from the cluster analysis, the tweets were grouped into the themes shown in the table below.

Table 8: Key Themes

Theme	Number of Tweets	Percentage of Tweets
advertising	12,976	18.7%
agency	4,150	6.0%
arena	5,621	8.1%
blank	21,002	30.3%
claims	1,466	2.1%
competition	2,467	3.6%
description	5,499	7.9%
employment	2,327	3.4%
foundation	957	1.4%
news	662	1.0%
other	6,740	9.7%
praise	1,464	2.1%
quotes	1,807	2.6%
roadside	1,232	1.8%

There are several key themes which are present in the analysis. Ignoring the blank and other categories for a moment (we will come back to them), the largest percentage of tweets are related to advertising. More specifically, most of these tweets were discussing the Allstate commercials. While there were some reactions on both sides, the majority of the tweets related to the commercials were positive, with words being used like "funny," "like," and "love." The next largest category was a

category associated with the Allstate arena, referencing upcoming or recent concerts and sporting events there. There were also a significant number of tweets related to employment with Allstate, many describing opportunities as an agent, but some referring to employment opportunities with the company. Many of these showed up in the categories labeled "agency," "employment," and "description." More detailed analyses should be performed on each of the significant themes to determine the types of things, both positive and negative, that are being said about the particular area.

Once these general themes or concepts are extracted from the results, the right areas within the company can be brought in to discuss what to do with the results. For example, given the significant number of tweets regarding the advertising, the marketing department may be able to take feedback from the comments posted by users to improve or build upon an advertising campaign. There was also a theme related to claims. Claim department executives would most likely be interested in hearing this unsolicited feedback that is being provided not to the company, but to a community of millions of users. This information could be used by the claims department in a number of ways, including improving the claims process and addressing complaints about claims handling.

As mentioned, the largest two categories are the missing and the other categories. The cluster with no keywords occurred because there were no keywords in the tweets that met the threshold that was set, which was four times greater than average. For the clusters in the other categories, there were a collection of words that related to a number of different things, but no general theme was identified. This will generally be the case, especially when dealing with social media data. There will be a collection of posts or data elements that a text mining process may have some difficulty classifying. Relating back to the Pear Analytics study, these could be tweets that fall into the pointless babble, self-promotion, or spam categories. There are a couple of ways that these issues can be addressed. The first is to try and do a more detailed analysis on these specific clusters, analyzing the presence of keywords or running a separate cluster analysis on this subset. Another approach is to use a rank of cluster lifts to find the most prevalent words in the cluster, whether or not they meet the standard of four times the average.

When producing this ranking, which is shown in Exhibit 3, it can be seen that several different concepts are present in cluster 21, which was the cluster that had no keywords that met the threshold. The five words that had the largest cluster lifts are, in order: Northbrook, personal, rep, claim, and April. As can be seen, these tweets represent a series of different ideas. Northbrook is the location of the Allstate corporate headquarters, and there also appears to be some references to

claim representatives as well. A combination of investigating some of the tweets as well as applying some of these word and concept analyses to the tweets specifically within that cluster will help uncover more detail in the cluster and potentially extract more information from those tweets.

It is also here that we can begin to see the reason for the increase in tweets in March, 2011. The increase essentially came from several categories of tweets. There was a significant increase during the time period of tweets related to employment and agency opportunities with Allstate, and there was also an increase in the number of tweets related to current and future events at the Allstate arena. There also appeared to be an increase in the number of tweets soliciting customers to receive quotes for insurance during this period.

5.3 Association Analysis

Cluster analyses are helpful in that they create groups of data points that have a relationship with each other, and these relationships can be examined in more detail to discover underlying concepts in the groups. The cluster analysis does have some disadvantages, though. The first is that each data point can belong to one and only one group. For examples like this one where Twitter messages are by definition short, this will not be a huge disadvantage because most tweets will only be focused on one concept. However, for social media outlets without such limitations, like Facebook or a blog, you may have a case where multiple concepts and themes will be present in each data point. In this situation, it will be difficult to assign data points to one and only one group. To address this, an analysis that highlights combinations of words regardless of where they occur can also help an analyst understand the key and important concepts in a body of text data.

One of the techniques that can be applied to highlight combinations of words is association analysis. Association analysis has its background in market basket analysis. It is used in retail environments, such as grocery stores or pharmacies, to identify items that tend to be purchased together. This allows stores to optimize the layout of their store and potentially increase sales by cross-selling or up-selling customers. This analysis determines the likelihood of a combination of items occurring together as well as a confidence around the projection. Ultimately, the association analysis produces a set of if-then rules (if item A is present in a transaction, then item B will be present as well), and the lift associated with the rule.

There are several calculations that are made as part of an association analysis to determine the strength of relationships. The support is a measure of how often items occur together.

$$\text{Support} = \frac{\text{Transactions that contain items A \& B}}{\text{All transactions}} \qquad (3)$$

Confidence measures the strength of the association by measuring how often item B is present when item A is present.

$$\text{Confidence} = \frac{\text{Transactions that contain items A \& B}}{\text{Transactions that contain item A}} \qquad (4)$$

The Expected Confidence is the proportion of items that satisfy the right side of the if-then association rule. This provides the expected presence of Item B if there was no relationship between Items A and B.

$$\text{Expected Confidence} = \frac{\text{Transactions that contain item B}}{\text{All transactions}} \qquad (5)$$

The lift can then be calculated as the ratio of the confidence (4) to the expected confidence (5). The higher the lift, the more the presence of Item B is influenced by the presence of Item A.

Applied to text mining, transactions would be the text field, and the items would be the words themselves. More specifically, in this analysis, a transaction would be a tweet, and the items would be the words in the tweet.

There can be literally thousands of rules generated from an association analysis, and for this example the author calculated 2,000 rules. The lift of the association rules ranged from a high of 140.21 to a low of 11.6. This means that we could calculate more than 2,000 rules and still generate rules that have a significant lift. Exhibit 4 shows an extract of the rules that were generated as a result of this analysis. As indicated above, many of the rules were associated with jobs and employment, so these were excluded those from this extract. The extract shows the statistics described above and a description of the rules. For example, notice Rule Index 281. The left side of this rule is TV, and the right hand of the rule is "Mayhem & ad." The expected confidence is .9%, meaning that Mayhem & ad only appear in only .9% of the tweets. However, when the word TV is present, Mayhem and ad occur in 50.9% of the tweets, as shown in the confidence percentage

column. The lift for this combination is over 55. As a result of reviewing these rules, subjects of tweets can be identified. Tweets that satisfy these rules can be identified and the sentiments that are being expressed related to these topics can be explored further.

The results of an association analysis can also be viewed using a link graph. A link graph displays the rules by using nodes and links. The size of the node varies based on the number of transaction counts (in this case, tweets) that the rule relates to, and the color and thickness of the line varies based on the strength of the relationship. Exhibit 5 shows an example of the entire link graph based on the analysis of the tweets. As you can see from the graph, there is a large mass of nodes and links, and then a number of nodes and links that surround the big mass. The big mass of links in the middle relates to a series of job, employment, agency, and financial opportunities with Allstate or its agencies. There was a significant amount of activity during this time period related to these opportunities, and tweets with the same or similar content were sent many times over the time period. Obviously, more detailed analysis would need to be done on this group to determine the value of the information here.

Exhibit 6 shows the link graph zoomed in on the top left. This zoom represents a collection of words related to "good," which is the most prevalent word in this set of rules. Looking at this group of tweets, there are actually a couple different themes being represented here. The first is the Allstate slogan, "You're in Good Hands." The other concept is the State Farm slogan, "Like a Good Neighbor, State Farm is There." So even though these tweets are related to the hashtag #allstate, there are a number of references to their competitor's slogan in these tweets. Many of these tweets were related to tweets of customers expressing their opinions regarding the different companies, and some were even related to jokes that were circulating around Twitter regarding the two insurance companies.

Exhibit 7 shows a zoom on the right side of the link graph. Here, there are several association rules created related to competitors (State Farm, GEICO, Progressive), and several related to the TV ads (Mayhem, TV, ad). In addition, there were a couple of other rules that highlight specific areas of interest related to Allstate. One was related to the Allstate "X The TXT" campaign, which involved the Jonas Brothers. This was a campaign to highlight the dangers of texting while driving. Many tweets were sent that expressed a positive reaction to this campaign. There was also a rule related to "$1 billion." This was a reference to the purchase price that Allstate paid to acquire Esurance. This news was tweeted about pretty heavily for a short time period when the acquisition was made public.

Again, the purpose of the association analysis is to highlight words and concepts that are coming

through in the data. Once these areas have been highlighted, the company can decide how this information should be disseminated and applied. For example, the mixing of slogans might highlight an opportunity to more clearly define the brand, while the interest in jobs and employment with Allstate might help develop a targeted recruiting campaign.

Although not covered in this paper, as an extension to this analysis, sequencing can also be done. A sequence analysis looks not only at the association of words within phrases, but also analyzes the order of the words within the phrases. This analysis then provides the likelihood of particular word orders and phrases, which can potentially give further insight into the content, especially when re-ordering words may change the meaning of a phrase. The investigation of sequence analysis is left to the reader.

5.4 Geographical Information

The Twitter location information that is available from mobile Twitter applications can save where the user was when the tweet was sent. The user generally has the option to turn location services on or off. In this analysis, 3,918 tweets have location information saved with them, which is 5.7% of the tweets. Associating the tweet with the location can have obvious applications for an insurance company. There could be claim implications, especially if a customer is somehow using social media right after a claim happens. Also, if there are multiple users within a certain geographical area that are tweeting about the same issue (premium quotes, for example), this may identify a concern or an opportunity for the company.

Exhibit 8 shows a map of the locations of the users for the tweets where the location was populated. The top five countries are shown in Table 9. As to be expected, most of the tweets originated from the United States and Canada. There were also a small percentage of tweets that were sent from the United Kingdom and China, and even some from the Gulf of Mexico.

Table 9: Top 5 Countries

Country	Number of Tweets	Percent
USA	3693	94.3%
Canada	58	1.5%
UK	44	1.1%
Gulf of Mexico	25	0.6%
China	22	0.6%

We also focused our attention on specific states that were generating lots of tweets. There were a large percentage of tweets in several states including Virginia, Washington, and Illinois that were related to financial representative and agency opportunities with Allstate. Many of the tweets from Illinois were users talking about and checking in from events that were going on at the Allstate Arena. There were also a few tweets related to roadside assistance help in several states. A large number of the location tweets were related to Foursquare check-ins, which allows users to check in at particular locations, and these check-ins can then be passed through Twitter.

6. CHALLENGES WITH SOCIAL MEDIA ANALYSIS

While the value that can be gained from analyzing social media data is great, there are challenges associated with social media analyses which will require further exploration. One of the first challenges will actually be accessing and collecting this information. As we discussed earlier, there are applications available which will allow companies to begin collecting and analyzing social media data, and companies may also have the ability to build internal programs that do this. The key is to make sure that the data is being collected in a consistent and complete matter, and that it is easily accessible for analysis.

There are also challenges in analyzing social media as it relates to analyzing text data. One of these challenges relates to the context. There are many times when a Facebook post or a tweet is simply a response to another post or tweet. Depending on how the user is responding or what a company may be tracking, the information that is being responded to may or may not be available. To the extent that it is not available, this creates a challenge for the analysis to understand exactly what this information means. If it is available, the challenge becomes connecting the right set of social media data together to be able to understand the broader context of a conversation. This is not a trivial exercise, and there is still emerging work being done to improve the analytics in this respect.

In this analysis, we are assuming that each tweet has equal weight. However, there are reasons that one might want to weight tweets differently. One reason is because users have different numbers of followers, and a tweet from a user with 1,000 followers is likely to be seen by more users that a tweet from a user with 100 followers. So a tweet could be given a higher weight given the influence of the person sending the tweet. Another reason for giving a tweet a higher weight may be the fact that it is a retweet. If it is a comment that other users are agreeing with, it can spread faster

and influence more users. So the number of followers and number of retweets should be considered in the weighting of the tweets.

The focus of this paper has been about determining the subject matter associated with a tweet. Another challenge in terms of the analysis of social media data is understanding customer sentiment. Words in a tweet are simply that, words. They do not carry the normal emotion that is present in a face-to-face conversation, in which case the listener could detect happiness, sadness, sarcasm, etc. There are things that users attempt to do to try and convey different emotions (smiley face, sad face, "lol," TYPING IN ALL CAPS, etc.), but even when a person reads an electronic communication they might get the sentiment wrong. Depending on the forum being analyzed, there are a number of understood rules that communicate different things, such as changing the font color to indicate sarcasm. While this analysis focused on the words being used, to understand sentiment would require a more thorough investigation into the ways that users communicate sentiment, and then attempting to capture those sentiments within the data in a structured way. This is another challenging area where there is still work to be done.

Another issue inherent in social media data arises from the fact that social media data is unfiltered. There are no system edits that ensure the social media data that was captured is accurate, and this may result in false information and statements that are driven by pure emotion rather than fact. This will make the process of sifting fact from fiction a delicate one. However, companies can guard against this by not simply accepting all the things that come out of a social media analysis, but filtering the results through their overall understanding of the business. Also, generally for a trend to show up significantly in a social media analysis, there has to be more than a few users making statements, but multiple users expressing similar sentiments that cause it to rise to the top. Ultimately, it will be very important for the analyst to consider the source of the data in the interpretation of the results.

Typical actuarial predictive modeling analyses are based on historical data that is usually at least three months old and could be at least a year old. Also, typical actuarial predictive analyses are repeated relatively infrequently. Generally, it is at least a year until an analysis is repeated, and could be even longer than that. In order to be able to apply the results of a social media analysis in a timely manner, the analysis will need to occur frequently, and as close to real-time as possible. Trending topics can literally begin in an instant and can become widespread very fast, and if the analysis occurs too long after the topic is trending, it may be too late for the company to do anything useful about it. Therefore, the analysis will need to be automated such that it can be updated quickly and

the results reviewed in a timely manner.

Lastly, the world of social media is not limited to those that use the English language. This will especially be true for companies that have an international presence, and even true for companies that function solely in the United States. The concepts discussed in this paper apply generally to other languages. However, if a company has social media data which includes information in multiple languages, the differences in the languages will necessitate at least a separate initial analysis. This analysis could then be combined later if it is possible to translate the words and sentiments into one language.

7. APPLICATIONS OF SOCIAL MEDIA ANALYTICS

As can be seen in the discussion above, there are a number of things that come out of the analysis of social media. A number of different techniques can be applied to understand the words and combination of words present in each tweet, and how active and popular these particular topics and conversations are. Ultimately, the result of this data-mining exercise will be an identification of trends in the social media conversation, and as a result of the identification of these trends, there are a number of ways that an insurance company can apply this information.

One area of application is in customer service. Some of the tweets in the analysis were related to claims, and depending on the content of the analysis, this information could be used to address potential concerns or questions regarding the claim process. There could also be questions raised related to policy information or provisions in a social media forum, and identifying and proactively addressing these issues gives the insurer an opportunity to provide superior customer service to their policyholders.

Another application of social media analytics would be a better understanding of customer sentiment about the company. One example of this is the customer reaction to company advertising campaigns. As customers provide feedback on advertising and commercials, for example, companies can use this raw, unsolicited feedback to make their marketing programs more effective. Also, if customers are particularly happy or unhappy with a company about a particular issue, it provides the company with an opportunity to attempt to proactively address this issue before it takes on a life of its own.

Social media analytics would also allow the insurance company to gather competitive intelligence from several different perspectives. In this analysis, there were some users that were very vocal

about their preferences of one insurance company over another. While an entire strategy cannot be based on the feelings of a few current or potential customers, if customer sentiment begins to build for or against the company or a competitor, this can be understood and the company can react to it. Also, in this analysis, there were obviously some users who may have been employees or agents of other companies that were soliciting customers. These trends can also be identified and monitored, and could potentially provide insight into competitive issues.

From a broader perspective, the use of social media can be used to identify broader trends in the market that the company may be able to take advantage of. One example of this might be an influx of social media data that suggests more people are looking for quotes or shopping for insurance in a particular area, or identifying concerns with finding affordable insurance in a particular area. Again, this could be brought forward to the right area within a company and proper steps taken to respond to these market trends.

8. CONCLUSIONS

As can be seen all around us, the use of social media has grown significantly, and has transformed the way that people interact. This growth in social media has led to an increase in the amount of information that is being generated, and this information provides insight to companies, including insurers, related to their business. Analytics can be applied to social media to identify key words and phrases that are being expressed, and these findings can be used by insurers to assist in managing their business, and to interact more effectively with current and potential customers.

Acknowledgment

The author acknowledges Gary Wang and Nick Kucera for their review and for comments that improved the paper.

9. REFERENCES

[1.] Banking.com Staff, "Social Media Statistics: By-the-Numbers, January, 2011," Banking.com. 24 Jan. 2011, accessed 22 Aug. 2011, http://www.banking2020.com/2011/01/24/social-media-statistics-by-the-numbers-january-2011-part-ii/>.

[2.] Brown, Danny, "52 Cool Facts About Social Media," Danny Brown. 3 Jul. 2010, accessed 22 Aug. 2011, < http://dannybrown.me/2010/07/03/cool-facts-about-social-media/>.

[3.] Browne, Sean, "Statistics: Social Networks will Receive 11% of Online Ad Spending in 2011," *Wealthvest Marketing*, 20 Jan. 2011, accessed 22 Aug. 2011, < http://www.wealthvest.com/blog/2011/01/20/statistics-social-networks-will-recieve-11-of-online-ad-spending-in-2011/>.

[4.] Customer Respect Group, "Social Eyes: The Insurers' View of Social Media," Volume 1, Number 5. July, 2011.

[5.] Francis, Louise A., "Taming Text: An Introduction to Text Mining," Casualty Actuarial Society *Forum,* Winter 2006, pp. 51–88, http://www.casact.org/pubs/forum/06wforum/06w55.pdf.

[6.] Kelly, Ryan, "Twitter Study Reveals Interesting Results About Usage—40% is 'Pointless Babble,'" Pear Analytics blog post, http://www.pearanalytics.com/blog/2009/twitter-study-reveals-interesting-results-40-percent-pointless-babble/.

[7.] "Twitter Study–August 2009," August 12, 2009, San Antonio, TX: Pear Analytics, http://www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf.

Biography of the Author

Roosevelt C. Mosley, Jr. is a principal with Pinnacle Actuarial Resources, Inc. Roosevelt has 18 years of experience in the property and casualty actuarial field, including over a decade of experience in the application of advanced analytic techniques to insurance companies. Roosevelt is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. Roosevelt's experience in the area of insurance analytics includes rating, underwriting, claims, and marketing.

Tweet Analysis
Cluster Results

Exhibit 1

Cluster	insurance	rt	commercial	arena	good	mayhem	job	like	commercial	hands	car	company	auto	financial	guy	lol	agent
1	0.519	7.879	1.778	0.000	0.741	2.333	0.000	17.635	0.054	0.722	0.627	0.000	0.000	0.000	2.805	2.426	0.216
2	0.300	0.082	0.000	0.000	0.000	0.000	6.631	0.010	0.000	0.043	0.097	0.022	0.036	0.113	0.038	0.014	24.915
3	1.471	0.572	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.142	0.259	5.575	20.240	0.000	0.000	0.000
4	0.002	7.879	0.598	0.000	0.000	0.399	0.325	0.000	0.000	0.252	0.420	0.118	0.123	0.453	0.780	0.000	0.237
5	0.169	1.697	2.118	0.018	0.047	0.149	0.055	1.287	3.412	0.175	0.289	0.000	0.046	0.000	2.702	25.118	0.105
6	0.000	0.000	0.025	0.000	0.000	0.000	0.315	0.000	0.000	0.194	1.097	0.121	0.308	0.000	0.456	0.000	0.050
7	0.313	0.000	1.335	0.063	0.047	1.990	0.081	17.635	0.000	0.641	0.407	0.083	0.015	0.000	2.384	0.343	0.138
8	0.000	0.000	0.000	0.000	0.000	0.000	0.055	0.000	0.000	0.054	0.122	0.056	0.000	0.000	0.222	0.000	0.000
9	5.215	0.408	0.000	0.000	0.222	0.000	0.641	0.057	0.000	0.000	0.714	13.089	22.112	0.000	0.000	0.000	2.951
10	5.196	0.484	0.014	0.000	0.036	0.000	1.555	0.065	0.007	0.014	0.000	19.571	0.008	5.789	0.017	0.019	3.778
11	4.715	0.183	0.037	0.000	0.120	0.045	0.000	0.103	0.055	0.000	2.054	0.118	13.177	0.000	0.000	0.000	0.147
12	0.000	1.447	0.145	0.000	12.888	0.212	0.079	3.092	0.420	19.540	0.221	0.020	0.022	0.007	0.506	2.517	0.144
13	0.036	0.791	0.177	3.909	0.190	0.268	0.411	0.366	0.264	0.000	0.229	0.000	0.459	0.071	0.318	0.087	0.175
14	3.824	0.274	0.174	0.000	0.234	0.000	0.162	0.132	0.208	0.013	22.076	0.524	1.056	0.042	0.345	0.274	0.691
15	3.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	9.967	0.000	10.148	0.000	0.000	0.000	0.000
16	0.040	7.757	0.396	0.000	13.737	5.165	0.307	0.137	0.222	0.151	0.171	0.000	0.000	0.000	7.208	1.266	0.098
17	0.000	0.027	0.000	10.828	0.114	0.000	0.027	0.280	0.000	0.000	0.000	0.000	0.000	0.000	0.095	0.087	0.000
18	0.000	0.510	0.000	0.000	0.073	15.496	0.024	0.052	4.322	0.012	0.354	0.012	0.000	0.037	5.055	0.969	0.075
19	0.263	1.080	0.215	0.000	0.231	1.639	0.095	17.635	19.066	0.141	1.061	0.000	0.000	0.000	3.587	3.079	0.000
20	0.000	6.399	0.000	0.000	13.737	0.000	0.000	17.466	0.000	0.000	17.084	0.000	0.000	0.000	0.073	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	1.721	0.000	0.000	0.001	0.000	0.226	0.000	1.874	0.610	0.000	0.000
22	0.248	2.412	0.434	0.000	12.896	0.000	0.000	12.476	0.000	1.994	8.410	0.000	0.451	0.000	1.406	1.025	0.172
23	2.929	0.614	0.818	0.000	0.190	0.350	0.000	0.532	0.479	0.049	1.997	0.279	3.611	0.000	0.548	0.978	0.414
24	0.356	1.074	0.290	0.000	13.113	0.352	0.000	15.831	0.000	3.553	3.261	0.460	0.000	0.000	0.261	2.854	0.576
25	0.379	0.849	0.110	0.013	0.046	0.000	0.133	0.000	19.066	0.056	0.338	0.000	0.032	0.000	2.517	0.000	0.036
26	0.000	0.641	0.020	0.000	0.000	0.000	0.190	0.028	0.000	0.031	0.880	0.613	5.607	0.033	0.000	0.120	0.081
27	0.000	7.879	0.000	13.135	0.066	0.000	0.038	0.034	0.000	0.000	0.000	0.000	0.000	0.000	0.044	0.192	0.000
28	5.215	0.494	0.231	0.013	0.000	0.000	1.967	0.021	0.009	0.047	0.000	0.000	0.000	0.129	0.415	0.218	1.575
29	0.046	0.669	12.772	0.000	0.048	6.392	0.162	0.490	0.106	0.023	0.181	0.000	0.006	0.000	3.791	1.176	0.015
30	4.087	0.234	0.438	0.000	0.027	0.010	0.497	0.081	0.579	0.064	0.015	0.093	22.112	0.041	0.045	0.133	0.167
31	5.215	0.124	0.000	0.000	0.432	0.000	0.000	0.000	0.037	0.000	17.696	0.159	7.211	0.243	0.000	0.000	0.000
32	0.402	2.072	2.735	0.000	0.346	0.000	0.062	0.750	4.143	0.462	0.139	0.000	0.000	0.000	7.267	5.854	0.000
33	0.029	0.077	0.053	0.073	0.019	6.543	0.000	0.122	0.662	0.000	0.123	0.056	0.031	0.029	0.096	0.105	0.106
34	4.874	0.110	0.060	0.000	0.128	0.000	0.000	0.330	0.000	0.000	20.219	4.819	12.193	0.000	0.000	0.117	0.000
35	0.395	1.441	1.631	0.000	0.057	0.664	0.065	1.641	0.660	0.241	0.164	0.000	0.055	0.000	0.378	3.331	0.000
36	5.215	1.828	0.247	0.000	13.623	0.086	0.876	1.413	0.105	14.628	1.220	0.670	2.993	0.000	0.444	1.110	0.350
37	0.401	1.556	0.164	0.000	0.070	0.119	0.203	0.136	0.244	0.100	0.679	0.000	0.340	0.000	0.235	0.580	0.065
38	0.002	0.000	0.000	13.135	0.082	0.005	0.028	0.267	0.000	0.007	0.062	0.007	0.000	0.000	0.065	0.327	0.000
39	5.215	1.602	5.247	0.000	0.114	15.303	0.000	1.829	2.136	0.081	3.664	0.671	1.009	0.000	4.954	0.834	0.210
40	5.170	0.310	0.056	0.000	0.060	0.068	0.000	0.154	0.000	0.000	0.000	0.177	0.193	0.270	0.000	0.000	0.221
41	0.000	0.000	0.411	0.338	13.737	0.538	0.816	0.976	0.294	0.000	0.369	0.078	0.114	0.027	0.561	0.970	0.326
42	0.208	1.050	12.308	0.000	0.119	3.780	0.069	0.580	0.132	0.000	0.420	0.000	0.000	0.000	2.423	3.303	0.000
43	0.780	0.447	0.000	0.000	0.071	0.000	0.000	0.000	0.000	0.000	0.114	0.104	1.824	2.872	0.000	0.000	0.000
44	1.134	0.690	2.199	0.000	0.166	0.234	0.120	0.586	0.230	0.000	1.067	0.061	0.534	0.000	1.110	0.379	0.230
45	3.187	0.297	0.000	0.026	0.709	0.000	3.853	0.000	0.000	0.000	0.044	4.935	0.044	0.901	0.000	0.050	0.854
46	0.069	3.204	0.106	0.000	0.023	0.077	0.459	0.000	0.000	0.032	0.328	0.033	0.037	0.171	0.228	0.042	0.335
47	0.000	0.000	0.028	0.029	0.000	0.000	0.000	0.039	0.000	0.043	0.146	0.000	1.264	0.045	0.050	0.166	0.445

Tweet Analysis
Cluster Keywords

Exhibit 2

Cluster	Number of Tweets	Keywords
1	352	rt like look please foundation
2	1,823	jobs job agent gravy hiring exclusive start states
3	468	auto financial home esurance quotes online news billion business answer
4	3,416	rt
5	1,453	lol black ***
6	503	going after
7	1,464	like look
8	725	tickets quote may rosemont people save year_w first
9	309	insurance company auto free geico more quotes claim progressive quote business
10	2,702	insurance company financial personal rep agency purchase
11	344	insurance auto free home rates best quote online life save safe drivers
12	2,994	good hands *** always roadside
13	289	time great show night year_w after concert last
14	1,466	car coverage accident
15	268	car auto free geico today rates esurance best great progressive coverage save look bad
16	258	rt good mayhem guy more progressive money team bad last roadside
17	1,446	arena chicago tickets show may rosemont tonight live glee concert
18	1,685	mayhem commercial guy hot hilarious
19	416	like commercial man lmao off night voice look
20	314	rt good like car state farm girl watch work neighbor statefarm
21	21,002	
22	147	good like car state farm girl watch work neighbor
23	796	state geico farm rates progressive life april bad
24	88	good like watch work neighbor bad statefarm
25	2,089	commercial funny lmao voice
26	627	auto home after safe win
27	1,045	rt arena chicago show off tonight live office glee concert first
28	4,150	insurance agency purchase
29	3,416	commercials mayhem love hilarious
30	1,516	insurance auto online life states motorcycle
31	509	insurance car auto state free geico today rates quotes check quote coverage online save drivers
32	635	commercial guy lol love man black *** voice basketball always
33	720	mayhem check youtube video jonas
34	214	insurance car company auto state geico farm today quotes coverage may life accident
35	1,214	man black voice
36	362	insurance good hands may life *** always
37	390	check help news driving team safe drivers foundation
38	2,841	arena tickets show may tonight live glee concert
39	241	insurance commercials mayhem guy
40	229	insurance free quotes coverage life year_w motorcycle
41	777	good team roadside
42	578	commercials funny lmao
43	194	sales esurance more online claims billion answer money
44	331	free geico esurance check progressive save statefarm
45	504	company sales hiring agency service bad
46	605	today foundation
47	455	free help service roadside

Tweet Analysis
Top 5 Keywords by Cluster

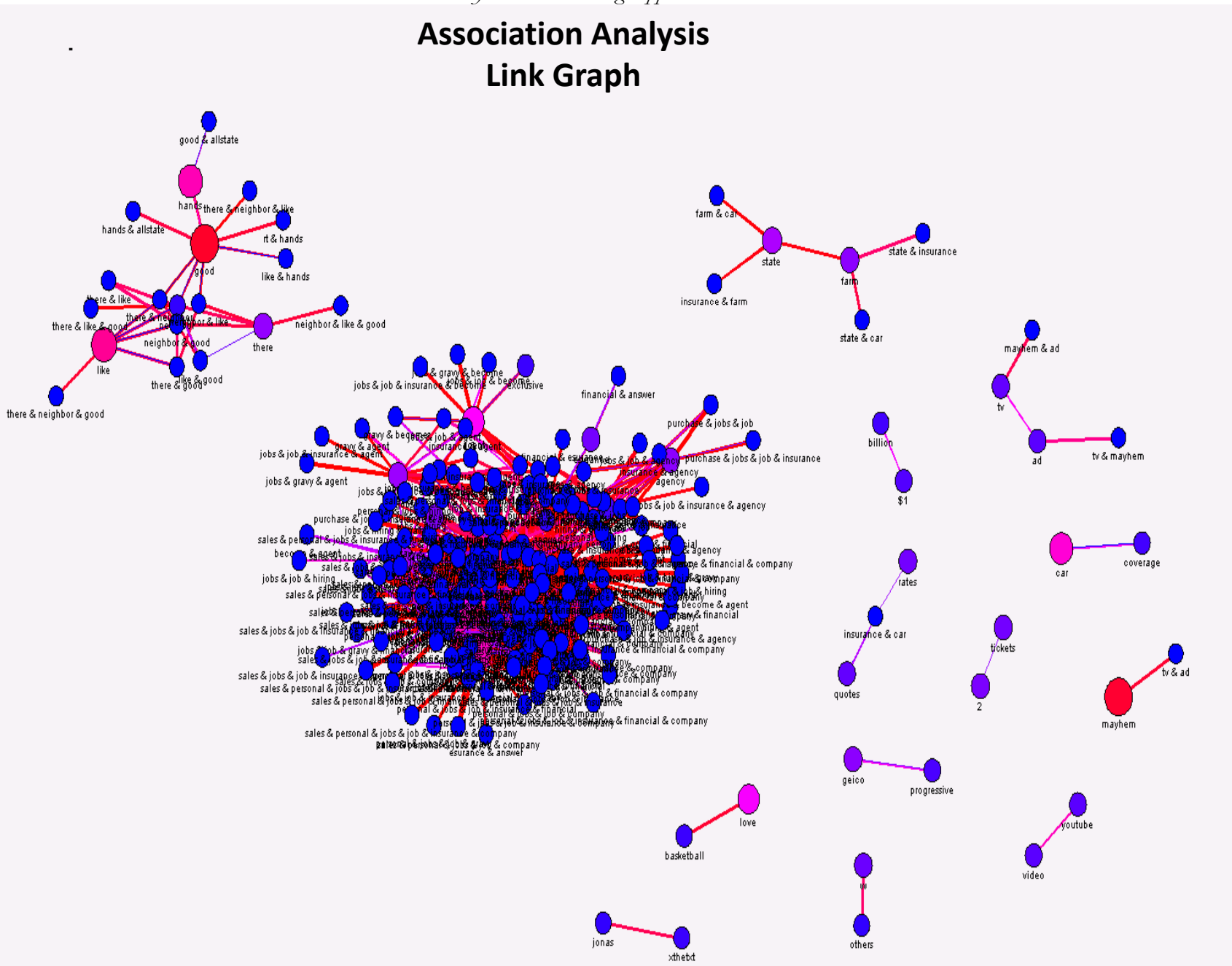
Exhibit 3

	Rank				
Cluster	**1**	**2**	**3**	**4**	**5**
1	foundation	please	like	look	rt
2	exclusive	agent	gravy	start	jobs
3	answer	esurance	billion	financial	home
4	rt	game	never	foundation	check
5	lol	***	black	commercial	still
6	after	going	game	tonight	still
7	like	look	better	please	really
8	people	quote	year_w	tickets	save
9	progressive	quotes	quote	auto	more
10	company	purchase	rep	personal	financial
11	quote	free	auto	life	best
12	hands	good	always	roadside	***
13	last	night	year_w	show	after
14	car	accident	coverage	drivers	insurance
15	rates	progressive	save	today	geico
16	money	progressive	roadside	good	last
17	rosemont	chicago	tickets	arena	glee
18	mayhem	guy	hot	commercial	hilarious
19	commercial	like	look	night	voice
20	neighbor	girl	watch	work	statefarm
21	northbrook	personal	rep	claim	april
22	neighbor	farm	state	work	good
23	farm	state	bad	april	geico
24	statefarm	neighbor	like	good	bad
25	commercial	funny	lmao	voice	hilarious
26	home	win	auto	after	safe
27	chicago	tonight	arena	office	concert
28	purchase	agency	insurance	motorcycle	home
29	commercials	hilarious	mayhem	love	guy
30	auto	states	online	motorcycle	life
31	quotes	rates	drivers	car	free
32	basketball	love	guy	black	***
33	video	youtube	jonas	check	mayhem
34	coverage	farm	geico	state	accident
35	man	voice	black	never	lmao
36	always	hands	good	***	may
37	driving	safe	foundation	team	help
38	glee	arena	concert	tickets	live
39	mayhem	commercials	insurance	guy	best
40	quotes	coverage	year_w	motorcycle	life
41	good	team	roadside	bad	look
42	funny	commercials	lmao	mayhem	lol
43	online	sales	esurance	money	answer
44	geico	progressive	statefarm	save	esurance
45	sales	hiring	bad	agency	company
46	today	foundation	check	great	rt
47	free	help	service	roadside	quote

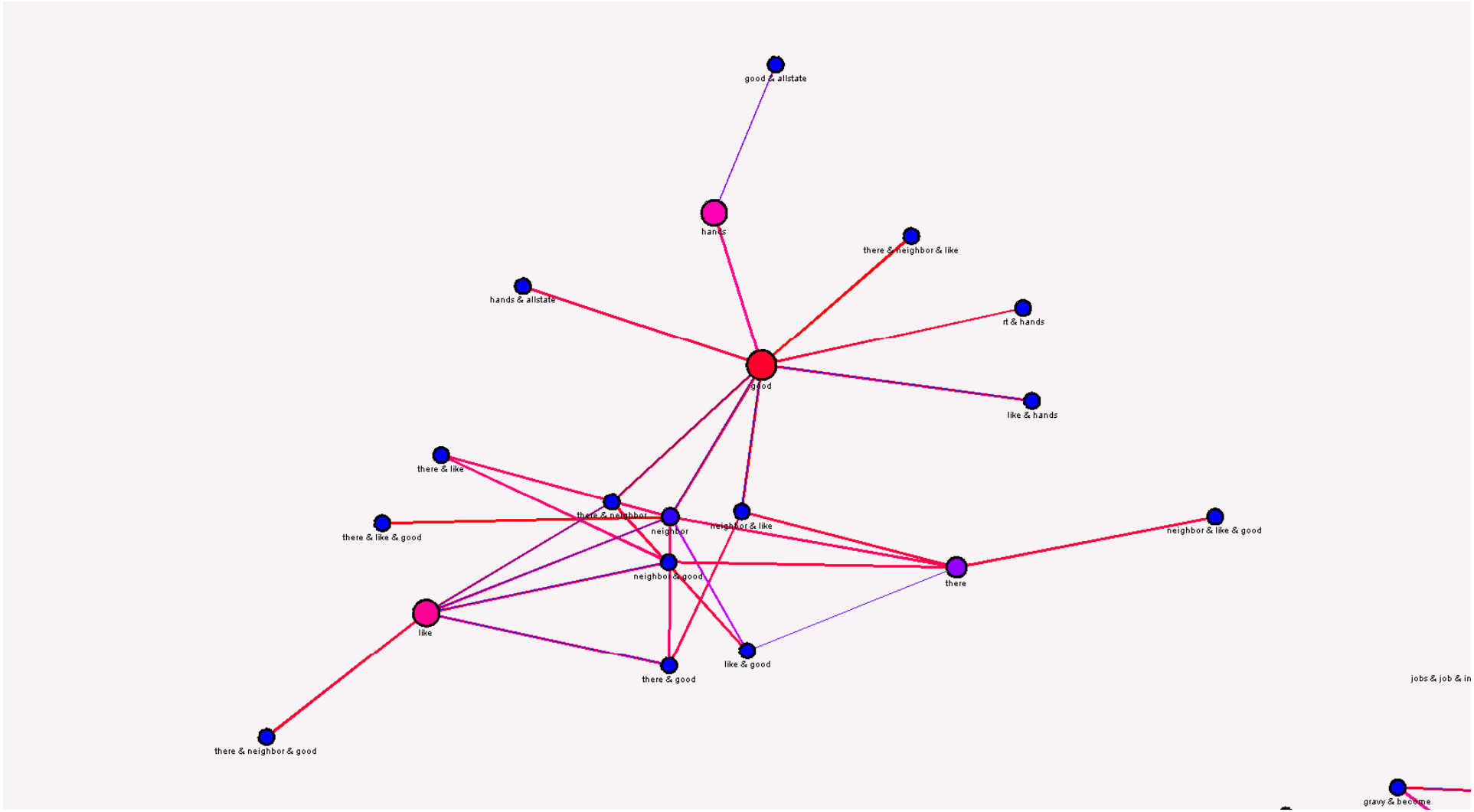
Tweet Analysis
Association Rules

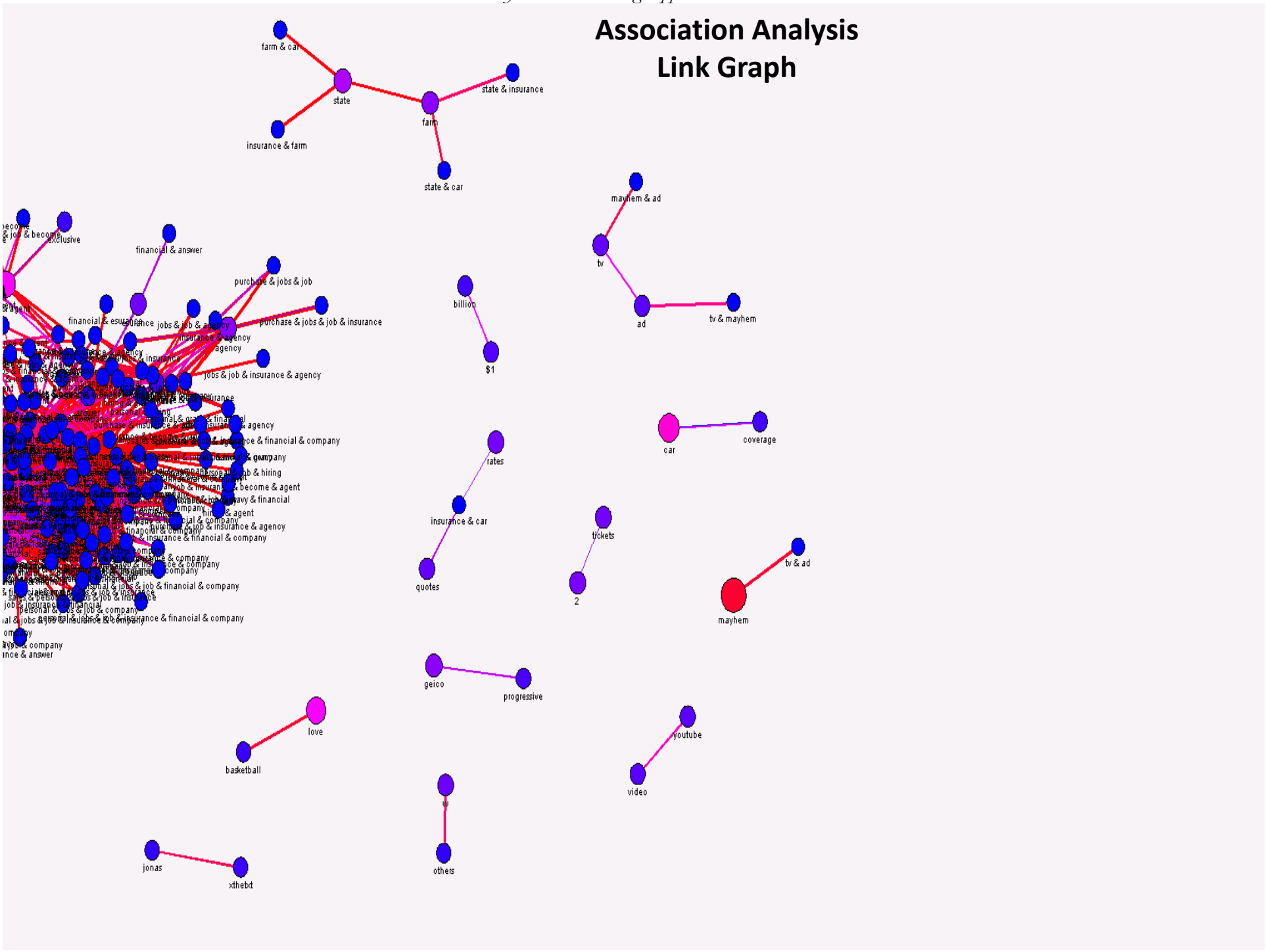
Exhibit 4

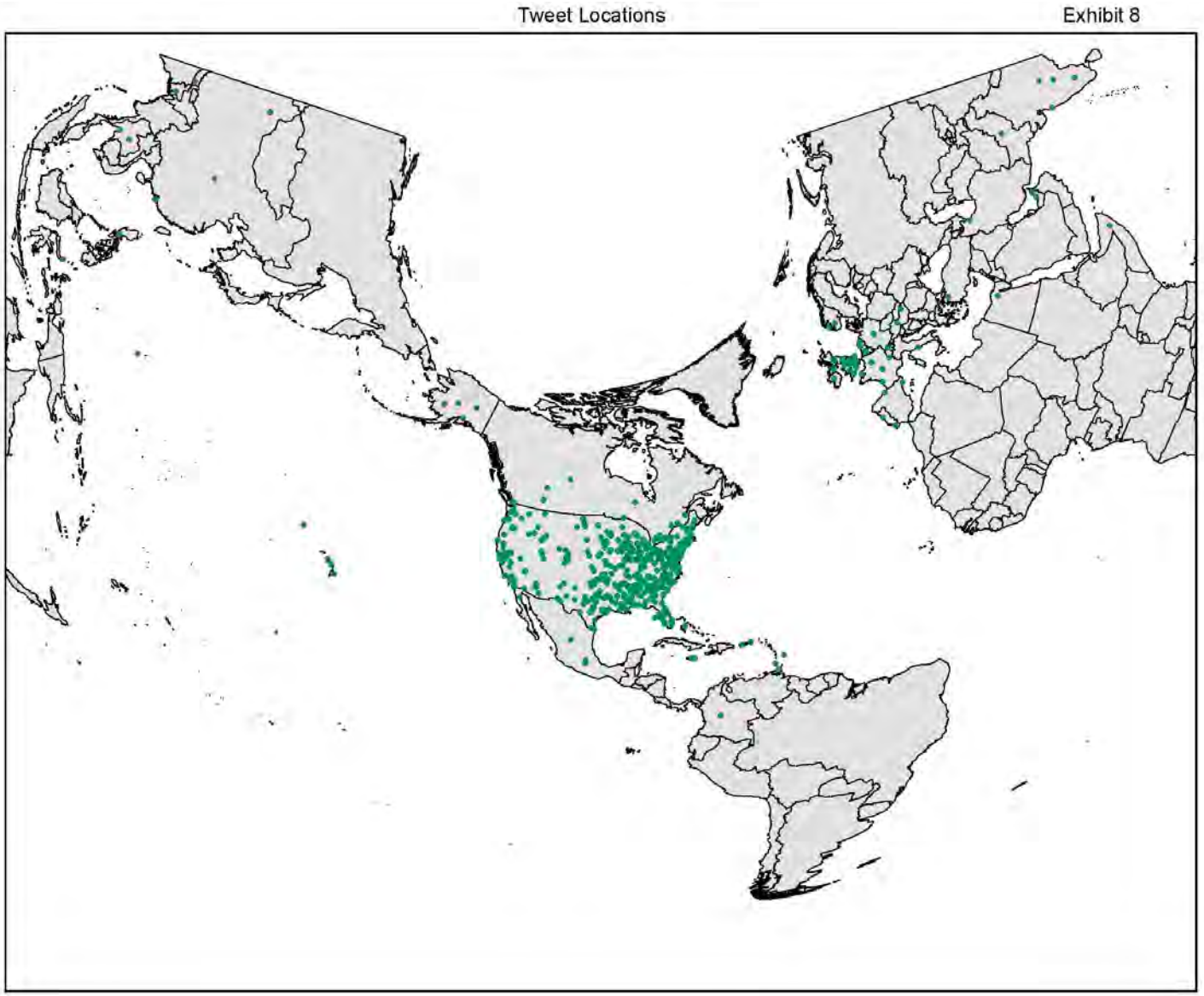
Expected Confidence(%)	Confidence(%)	Support(%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Index
0.632	73.391	0.618	116.070	422	neighbor ==> there & like & good	neighbor	there & like & good	73
0.754	86.831	0.618	115.194	422	neighbor & like ==> there & good	neighbor & like	there & good	78
0.754	81.043	0.682	107.516	466	neighbor ==> there & good	neighbor	there & good	110
0.771	79.174	0.618	102.644	422	neighbor & good ==> there & like	neighbor & good	there & like	134
0.888	91.053	0.760	102.486	519	jonas ==> xthetxt	jonas	xthetxt	135
0.834	85.502	0.760	102.486	519	xthetxt ==> jonas	xthetxt	jonas	136
0.771	75.130	0.632	97.402	432	neighbor ==> there & like	neighbor	there & like	177
1.762	99.167	0.697	56.273	476	financial & answer ==> esurance	financial & answer	esurance	277
0.921	50.911	0.777	55.299	531	tv ==> mayhem & ad	tv	mayhem & ad	281
1.527	84.420	0.777	55.299	531	mayhem & ad ==> tv	mayhem & ad	tv	282
0.700	37.679	0.618	53.855	422	like & good ==> there & neighbor	like & good	there & neighbor	295
1.639	88.285	0.618	53.855	422	there & neighbor ==> like & good	there & neighbor	like & good	296
1.041	52.679	0.777	50.620	531	ad ==> tv & mayhem	ad	tv & mayhem	311
1.762	88.516	0.733	50.229	501	answer ==> esurance	answer	esurance	323
1.639	81.913	0.689	49.968	471	neighbor ==> like & good	neighbor	like & good	331
1.399	65.843	0.897	47.056	613	video ==> youtube	video	youtube	364
2.122	87.639	0.809	41.294	553	state & car ==> farm	state & car	farm	530
0.924	38.138	0.809	41.294	553	farm ==> state & car	farm	state & car	529
2.122	80.923	0.950	38.130	649	state & insurance ==> farm	state & insurance	farm	588
1.174	44.759	0.950	38.130	649	farm ==> state & insurance	farm	state & insurance	587
2.355	89.597	0.618	38.045	422	neighbor & like & good ==> there	neighbor & like & good	there	590
1.527	57.837	0.853	37.886	583	ad ==> tv	ad	tv	595
2.355	88.889	0.632	37.744	432	neighbor & like ==> there	neighbor & like	there	602
2.355	87.430	0.682	37.125	466	neighbor & good ==> there	neighbor & good	there	648
0.826	30.318	0.809	36.727	553	state ==> farm & car	state	farm & car	665
2.670	98.050	0.809	36.727	553	farm & car ==> state	farm & car	state	666
2.122	77.138	2.059	36.346	1,407	state ==> farm	state	farm	673
2.670	97.034	2.059	36.346	1,407	farm ==> state	farm	state	674
2.670	96.434	0.950	36.121	649	insurance & farm ==> state	insurance & farm	state	702
2.355	83.130	0.700	35.299	478	neighbor ==> there	neighbor	there	738
2.128	58.010	0.700	27.258	478	progressive ==> geico	progressive	geico	955
0.973	23.359	0.886	23.999	605	love ==> basketball	love	basketball	1141
3.791	90.977	0.886	23.999	605	basketball ==> love	basketball	love	1142
0.733	14.359	0.697	19.582	476	financial ==> esurance & answer	financial	esurance & answer	1349
2.355	38.571	0.632	16.378	432	like & good ==> there	like & good	there	1493
1.639	26.849	0.632	16.378	432	there ==> like & good	there	like & good	1494
5.676	90.558	0.618	15.954	422	there & neighbor & good ==> like	there & neighbor & good	like	1550
0.682	10.882	0.618	15.954	422	like ==> there & neighbor & good	like	there & neighbor & good	1549
5.676	90.377	0.632	15.922	432	there & neighbor ==> like	there & neighbor	like	1553
0.700	11.140	0.632	15.922	432	like ==> there & neighbor	like	there & neighbor	1554
5.676	88.368	0.689	15.568	471	neighbor & good ==> like	neighbor & good	like	1573
5.676	84.522	0.711	14.891	486	neighbor ==> like	neighbor	like	1613
5.676	83.883	0.632	14.778	432	there & good ==> like	there & good	like	1629
0.754	11.140	0.632	14.778	432	like ==> there & good	like	there & good	1630
0.632	8.470	0.618	13.396	422	good ==> there & neighbor & like	good	there & neighbor & like	1725
0.700	9.354	0.682	13.369	466	good ==> there & neighbor	good	there & neighbor	1728
7.141	91.081	0.777	12.754	531	tv & ad ==> mayhem	tv & ad	mayhem	1766
0.853	10.883	0.777	12.754	531	mayhem ==> tv & ad	mayhem	tv & ad	1765
0.950	11.642	0.849	12.256	580	good ==> like & hands	good	like & hands	1886


Association Analysis
Link Graph
good & allstate
hands
there & neighbor & like
hands & allstate
rt & hands
like & hands
good
there & like
there & like & good
neighbor & like & good
there
neighbor & good
like
like & good
there & good
there & neighbor & good
farm & car
state
state & insurance
farm
insurance & farm
state & car
mayhem & ad
tv
ad
tv & mayhem
billion
$1
financial & answer
purchase & jobs & job
purchase & jobs & job & insurance
jobs & job & insurance & agency
jobs & gravy & agent
car
coverage
rates
insurance & car
quotes
tickets
2
tv & ad
mayhem
geico
progressive
love
basketball
youtube
video
w
others
jonas
xthebxt

Association Analysis
Link Graph




Association Analysis
Link Graph
farm & car
state
state & insurance
farm
insurance & farm
state & car
mayhem & ad
tv
ad
tv & mayhem
billion
$1
financial & answer
purchase & jobs & job
purchase & jobs & job & insurance
jobs & job & insurance & agency
car
coverage
rates
insurance & car
quotes
tickets
2
tv & ad
mayhem
geico
progressive
love
basketball
youtube
video
jonas
xthebxt
others


Tweet Locations
Exhibit 8

Beginner's Roadmap to Working with Driving Behavior Data

Jim Weiss, FCAS, MAAA, CPCU

Jared Smollik, FCAS, MAAA, CPCU

Abstract

Motivation: Usage-based auto insurance has received considerable publicity, but the driving behavior data that fuels these programs has been the subject of limited academic scrutiny. In this paper, we expose the challenges and risks that result from working with this data, and we discuss how actuaries may collect, organize, and analyze driving behaviors for use in insurance applications.

Method: In this paper, we use historical context to identify objectives, challenges, and risks in working with driving behavior data. We identify key tenets of the infrastructure required to support data collection, with a focus on vehicle telematics. We look at sample driving behavior data and how it may be organized into databases for predictive modeling and classification. We conclude with a discussion of sample use cases for the data.

Results: Driving behavior data allows insurers to achieve unique goals in pricing, underwriting, and loss control or mitigation, but it also presents unique challenges and risks.

Conclusions: Sound data management allows insurers to use driving behavior data to achieve organizational goals while rising to the challenges and risks this data presents.

Keywords: Driving behavior data; usage-based insurance; vehicle telematics; black boxes.

1. INTRODUCTION

Usage-based auto insurance (UBI) is defined as "a rating structure that is based, in whole or in part, on the electronic accumulation of data, through a device installed in a motor vehicle, in which an individual's daily driving habits are used to determine a premium rate."[1] The technological framework underlying UBI is referred to as vehicle telematics, that is, "the use of global positioning system (GPS) technology integrated with computers and mobile communications technology in automotive navigation systems."[2] The information generated using telematics—including when, where, how often, and the manner in which a vehicle is operated—is called driving behavior data (DBD). Telematics makes large quantities of detailed DBD accessible to actuaries, who may then use this information as the foundation upon which to build innovative and statistically supportable UBI rating plans.

[1] From Colorado Department of Regulatory Agencies, page 2.
[2] From Telecommunications Systems, Inc., page 4.

The potential uses of DBD are captivating. In *Personal Automobile: Cost Drivers, Pricing, and Public Policy,* Conners and Feldblum observe that "traditional actuarial focus on ratemaking and classification systems has led to an emphasis on pre-accident factors—particularly driver, vehicle, and geographic characteristics—to the virtual exclusion of other factors."[3] They suggest that pre-accident factors relate primarily to claim frequency, while neglected "post-accident factors" such as injury types, medical treatment, and attorney involvement could be more predictive of loss severity. DBD has the potential to bridge the gap between pre- and post-accident factors. For example, vehicles' average speeds may help predict not only accident occurrence but also resulting injury types, and locations of driving may help predict the quality of medical care or likelihood of attorney involvement. DBD may be used to statistically model these and other effects.

Despite its intuitive appeal, UBI uptake has been slow. While over 90% of personal insurers were actively using insureds' credit histories within approximately five years of the earliest adopters, insurers representing only three-fifths of the auto market have implemented UBI programs in the decade-plus since Progressive Casualty Insurance Company (Progressive) launched the first pilot.[4] As a result, little has emerged in the way of academic consensus or a universal model for effectively managing DBD, and individuals new to the UBI space may have difficulty locating resources geared towards the beginner. Strube and Russell developed a model for handling challenges in the modern high-volume transactional processing environment (HVTPE). In it, actuaries identify, procure, and maintain necessary data from various business areas while creating "user-defined" fields to facilitate their own analyses.[5] Many of the lessons learned from HVTPEs are applicable to DBD, but transactional data originates when policies are written, renewed, or endorsed, or when claims are reported or serviced. DBD adds a new dimension by collecting potentially sensitive information at higher frequencies, often remotely, and over an extended period of time. This presents a unique set of data management challenges for the actuarial data manager (ADM). The objective of this paper is to provide a beginner's road map for actuaries and other insurance professionals interested in working with DBD.

The remainder of the paper shall proceed as follows: Section 2 will provide historical background information, identify objectives in working with DBD, and examine key challenges in

[3] From Conners and Feldblum, page 322.
[4] Figures are taken from Sturgeon, page 1 and Towers Watson, paragraph 8. According to Oregon Department of Consumer and Business Services, insurers began using credit history for personal lines of insurance in 1995. Conning survey referenced in Sturgeon was taken in 2001.
[5] See Strube and Russell, page 287.

doing so. Section 3 will focus on one specific challenge: establishing the technological infrastructure required to collect and analyze DBD, i.e., telematics. Section 4 will discuss organizing DBD into databases which may be used for predictive modeling and classification. Section 5 will discuss issues related to utilization of DBD now and in the future. The final section, Section 6, will offer our conclusions.

2. ORIGINS, DESTINATIONS, AND ROADBLOCKS

In his 1929 paper *Notes on Premium and Exposure Bases*, Dorweiler identified critical conditions affecting the "hazard covered by automobile liability insurance, or that cause deviations in this hazard."[6] They were:

- The car—age, condition, etc.
- Highways—road beds, curves, visibility, etc.
- Traffic density
- Laws, regulations, and their enforcement
- Efficiency of driver—age, experience, habits, impairments, etc.
- Mileage
- Speed
- Weather conditions
- Seasonal use of car
- Day and/or night use of car

Most of these data elements, especially the last six, can be considered DBD, but Dorweiler correctly concluded that the lack of necessary "devices and records" (i.e., telematics and the resulting DBD) made the use of such information (i.e., UBI) "impractical under [then] present conditions."[7] As a result, actuaries in the ensuing decades came to rely on what Conners and Feldblum call "proxies for the true ("causative") factors affecting loss costs," such as age, gender, or garaging

[6] Quotes and bullet-list have been reproduced verbatim from Dorweiler, page 337.
[7] Ibid., page 338.

territory.[8] In the 1980s, states such as California and Michigan began to require the use of more "causative"[9] rating variables such as mileage and driver safety record, but even these variables' predictive abilities were contingent upon the veracity of mileage estimates and the credibility of individual driving records.[10] Three distinct waves of innovation helped transform early prophesies of UBI from voices in the wilderness into winds of change.

The first wave was "the computerization of vehicles," which began with the invention of the microprocessor in 1971. Government regulations regarding emissions, safety, and fuel economy in the years that followed effectively required automakers to utilize the new technology to install electronically controlled engines and safety devices in vehicles. Microprocessor-based controls soon took hold of other vehicle systems (e.g., braking). The result was a wealth of DBD describing vehicles' operating conditions, but no efficient way to collect or analyze it. Symptomatic of this was a failed California Senate Bill in 1993 which would have required auto insurance to be purchased by gallon of fuel at the pump.[11] After decades of technological innovation, the highest profile UBI proposal involved collecting DBD with a century-old device (the fuel pump) that Dorweiler had long since rejected as impractical.

The second wave of innovation was "The Computerization of Society." In 1978, the same year the first GPS satellite was launched, Nora and Minc defined *télématique* as the "increasing interconnection between computers and telecommunications."[12] Their vision became a reality in the 1990s with the confluent emergence of satellite-based positioning systems, platforms capable of interfacing with vehicles' on-board computers, and expanded two-way wireless communication abilities. In 1996, President Clinton signed a directive transforming GPS from a primarily military technology into an "international informational utility," and a large market for factory-installed and aftermarket GPS equipment was developed. Federal law effective that year also required all new vehicles to be outfitted with On-Board Diagnostic (OBDII) ports capable of reporting the output from vehicles' various sensors. Finally, the launch of the General Packet Radio Service (GPRS) in 1997 allowed for "always-on" wireless data services. The resulting birth of vehicle telematics

---

[8] From Conners and Feldblum, pages 330-331.
[9] The Actuarial Standards Board's (ASB) Actuarial Standard of Practice (ASOP) No. 12 on Risk Classification states that "it is not necessary for the actuary to establish a cause and effect relationship between the risk characteristic and expected outcome in order to use a specific risk characteristic" (page 4).
[10] See Mahler for more information on the credibility of a single driver.
[11] See "California Lawmakers Reject 'Pay at the Pump' Insurance,", *New York Times*, May 27, 1993.
[12] From Shanken, pages 51-52.

provided more detailed DBD than ever (including locations) and, more importantly, a way to communicate it outside the vehicle using everyday technology.

The third wave of innovation we have named "the automization of auto insurance." One of the first sectors to successfully utilize telematics was commercial trucking, where fleet operators tracked the locations and physical conditions (e.g., fuel level) of their equipment for optimization, preventative maintenance, and other purposes. Progressive brought telematics to the insurance sector with their Autograph pilot in 1998. As part of the voluntary program, the company installed GPS-capable devices in participating insureds' vehicles to record DBD such as the amount of time spent driving at different times of day or located in different "risk zones," all for use in premium determination.[13] Although the effort was allegedly popular with insureds that enjoyed having more control over their premiums, the program was discontinued in 2001 due to technological cost considerations. Undeterred, a number of insurers and organizations serving the industry, including Progressive (who reemerged with another initiative in 2004), have tested or gone to market with different "flavors" of UBI as the third wave continues into the present.

Organizations choosing to work with DBD have various objectives. A primary objective for actuaries is to more effectively segment their employer's book of business. The ASB ASOP No. 12 regarding risk classification observes that "if the variation of expected outcomes within a risk class is too great, adverse selection is likely to occur."[14] Autograph actually functioned more like a monthly utility bill than a risk classification system, retroactively billing insureds based on usage and effectively achieving individual risk segmentation. However, more recent efforts (including Progressive's) have employed risk characteristics such as mileage, time of day, speed, and extreme braking to determine discounts in the future. This is more consistent with the traditional classification approach of "assign[ing] risks to groups based upon the expected cost or benefit of the coverage or services provided."[15] With the potential behavioral variations between drivers of, say, the same age, gender, and credit score, the use of DBD in risk classification has great promise in helping achieve ASB ASOP No. 12's recommended "homogeneity with respect to expected outcomes" in order to avoid adverse selection.

A second objective of DBD is not simply to avoid adverse selection, but rather to actively pursue a more profitable portfolio of insured risks using telematics. Insurers who request DBD from

[13] See Figure 8.7 on page 185 of Cady and McGregor.
[14] From ASB ASOP No. 12, page 5.
[15] Ibid., page 3.

prospective insureds may get fewer applications from individuals who would falsify their garaging location or annual mileage in order to fraudulently lower their premiums, since such information is verifiable using telematics. Furthermore, in 2004, General Motors Acceptance Corporation Insurance Company (GMAC) began offering a discount to insureds for subscribing to OnStar®, a vehicle safety and security service which used telematics equipment factory-installed in General Motors (GM) vehicles. It could be argued that OnStar subscribers were likely to produce fewer or less severe insured losses due to OnStar's possession of their DBD and insureds' resulting access to roadside assistance, emergency response, and stolen vehicle location assistance services. GMAC also offered additional discounts to subscribers for low mileage indicated by their OnStar DBD. Recent announcements from State Farm and the Automobile Association of America Insurance Company (AAA) illustrate similar uses of telematics. Such examples suggest that a sophisticated classification system may not be the only way insurers can use DBD to benefit from positive selection effects.

The GMAC/OnStar case highlights a third objective of DBD, which is to mitigate losses after they occur. Several insurers offer roadside assistance services similar to OnStar which send help to the scene of an event when an emergency phone call is received. Telematics with real-time tracking can enable insurers to respond sooner, such as when an airbag is deployed or a harsh braking event is detected in the DBD. This would be of particular value when a victim is physically unable to make a phone call. Faster response time has the potential to reduce the extent of insured injuries or property damage. Also, DBD's loss reduction capability may not be limited to legitimate claims. The National Highway Transportation Safety Authority (NHTSA) requires that Event Data Recorders (EDRs) voluntarily installed in private passenger and other light vehicles record a minimum set of data elements to facilitate vehicle safety research. EDRs log vehicle data before and after accidents. Such DBD could be used by insurers to analyze for fraudulent claims, so long as it is done in a way that complies with applicable laws governing the use of data from EDRs.

A fourth objective of DBD is to modify risky behavior. Just as fleet managers have long analyzed DBD to recommend safer or more economical driving habits to their drivers, insurers have set up UBI programs that empower individuals to serve as *de facto* fleet managers for their book of insured vehicles. For example, Safeco Insurance's Teen Safe Driver Discount program, introduced in 2008, offers parents weekly reports on participating teens' driving behaviors. Parents are in turn encouraged to review the reports regularly with their teens. In 2009, Liberty Northwest developed a similar tool for commercial fleets called OnBoard Advisor with "accountability features" regarding

fuel, maintenance, and overtime.[16] Both programs award discounts to insureds who exhibit safe driving behaviors. In cases like these, providing insureds with DBD summarized in a meaningful manner may benefit both the insurer and the insured.

Finally, insurance organizations may have objectives not strictly related to their insured books of business. Some may wish to utilize DBD to optimize the performance of their own fleets of auditors' or claims adjusters' vehicles. Others may elect to provide DBD to academia or nonprofits for use in research, such as the recent study that took place in Massachusetts with data supplied by the Commonwealth Automobile Reinsurers. This report concluded that insurers' use of mileage in rating would reduce driving and, hence, greenhouse emissions. The reader will recall that one of the original motivating factors behind the computerization of vehicles was the ability to control emissions, and DBD gives insurers or anyone looking to "go green" a valuable tool to monitor progress. The number of different objectives insurers could have in working with DBD is evident in the many variations on UBI which have gone to market.

While DBD makes various goals achievable, it also presents unique challenges. One such challenge is the "fortress of intellectual property (IP) protection"[17] obtained by Progressive and other innovators in the UBI and telematics spaces. In 1998's State Street Bank decision, the U.S. Court of Appeals of the Federal Circuit ruled that "methods of doing business" could be patented, opening the door for so-called "insurance development labs."[18] One beneficiary was Progressive, who in 1996 (before bringing Autograph to market) initiated the unique step of patenting what they called "the kitchen sink."[19] Their "Motor Vehicle System for Determining a Cost of Insurance" patent protects broad systems for acquiring DBD and methods for using it to produce actuarial classification systems or determine premiums. Subsequent improvements to the invention, including methods of building databases and communicating DBD back to the insured, have also been protected. Similar patents were obtained by other insurers, telematics service providers (TSPs), and non-practicing entities (NPEs, also called "patent trolls").[20]

---

[16] This program later became the subject of litigation. See Bricketto.
[17] "Fortress" characterization is from Bakos and Nowotarski.
[18] "Insurance development lab" characterization is used by Nowotarski in Chartrand article.
[19] "Kitchen sink" characterization was used by Progressive marketing executive Bob McMillan in Hendricks to describe the extensive surveillance capabilities listed in the patent.
[20] A TSP is defined by Langevoort as an organization that "collects, summarizes or stores data from several sources; runs a telematics gateway or backend system; or provides data or services to several clients based on service subscription or demand" (page 11). An NPE is defined by Kauth as an entity that "holds patents, but does not practice the claimed inventions" (paragraph 3).

The jury is still out on the enforceability of UBI- and telematics-related patents. Many hoped the Bilski decision in June 2010[21] would clarify the patentability (or lack thereof) of business processes. Instead, the Supreme Court issued a narrow ruling rejecting the petitioner's application as overly abstract, but not prescribing specific tests for what constitutes a (patentable) business process. That same month, Progressive began suing multiple insurers for patent infringement. Even insurers whose own use of DBD steers clear of the IP fortress risk working with TSPs who may be litigated against. For example, Progressive recently sought to have another insurer's device supplier's patents invalidated for obviousness.[22] That device supplier had also sued Progressive's supplier; it is common for TSPs and/or NPEs to litigate against each other over technology or fleet management techniques. Some analysis suggests that the broadness of patents and their plenitude in this space may adversely impact defensibility.[23] Patents eventually expire, but in the meantime, organizations considering working with DBD must consider the legal and reputation risk involved. If they decide to proceed, actuaries and others should work closely with their companies' legal departments to select reputable TSPs and ensure that uses of the data are respectful of the rights of all parties involved.

A second challenge to working with DBD is consumer privacy. A survey sponsored by the Insurance Services Office (ISO) indicates that approximately 60% of drivers are unlikely to share DBD such as driving locations, speeds, or instances of hard braking and acceleration with their insurers. One reason may be that such data is discoverable and can be subpoenaed. To illustrate, consider that electronic toll collection data has been suggested for use in issuing speeding tickets to individuals who pass between two toll plazas in too short a period of time. DBD presents even greater revenue-generating possibilities for law enforcement. Another privacy threat to insureds is that GPS-detailed data will be sold to third parties for use in location-targeted marketing. Even insurers who abstain from such practices could find their databases hacked, as Epsilon Data Management was hacked in the banking and retail sectors.[24] Finally, insurers who do not clearly explicate the information they collect or how it will be used may face unwanted publicity similar to Apple's following revelations over iPhones' location-tracking capabilities. Such examples illustrate the difficulty even the best-intentioned insurers have in positioning DBD-based initiatives to the public.

---

[21] *Bilski v. Kappos*, 130 S. Ct. 3218, 561 US __, 177 L. Ed. 2d 792 (2010).
[22] See Vanderford.
[23] For example, see M-Cam, Inc., page 2.
[24] See Svensson.

Defenses exist which help make UBI offerings more acceptable to all parties involved from a privacy perspective. Insurers have been able to address some concerns simply by not collecting GPS information. Beyond that, programs using DBD should at a minimum be opt-in instead of opt-out, especially since most require the installation of hardware in insured vehicles. Duri et al. propose a broader framework in which various privacy options, including different use cases and degrees of DBD collection, are selected from upon enrollment, with financial incentives for greater sharing (i.e., discounts). Under this structure, TSPs would be used to manage privacy preferences and share data with the insurer and insured as appropriate. Custom software modules in vehicles would further reduce the transfer of sensitive information by performing necessary calculations on-board. Insureds themselves would not have the ability to hack and manipulate DBD for financial gain.[25] While not bound to this structure, insurers should continually evaluate the adequacy and appropriateness of their own privacy policies.

Laws and regulations regarding privacy further impact insurers' ability to utilize DBD. For example, California prohibits use of location data for most insurance purposes and was the first of many states to implement EDR statutes which require vehicle owners' consent before retrieving accident reconstruction information. Such laws potentially limit insurers' ability to investigate fraudulent claims using DBD. More generally, DBD may be linked with, and has the potential to reveal, personally-identifiable information (PII), a protected class of data which places the insurer under the purview of a multitude of additional legislation.[26] At least one Department of Transportation report concedes "today's patchwork of privacy legislation from federal, state and local governments make it impossible to identify a lowest common denominator for privacy regulations."[27] Nevertheless, it is the responsibility of insurers to make sure UBI programs are mindful and compliant with this evolving body of law to the extent possible.

Perhaps the greatest challenge to successfully utilizing DBD is establishing the required infrastructure. The *Practitioner's Guide to Generalized Linear Models* asserts that credible modeling results are generally achieved using 100,000 or more vehicle-years of data.[28] Given IP and privacy challenges, insurers may find it difficult to enroll enough insureds in UBI to amass so much DBD over a short period of time. Before even thinking about that, however, insurers must first

[25] See Duri et al. for more information on the framework discussed throughout this paragraph.
[26] Hughes Telematics, Inc. asserts such laws "could include the Federal Trade Commission Act, the Fair Credit Reporting Act, the Gramm-Leach Bliley Act, as well as various state laws and related regulations" (page 10).
[27] From U.S. Department of Transportation (NG-911 Transition Issues Report), page 6.
[28] See Anderson et al., page 40.

appropriate costly hardware to collect, transmit, and store the data; software and algorithms to process it; and personnel (e.g., actuaries) to implement and oversee the endeavor. The next section describes some of those challenges.

3. INFRASTRUCTURE

Implementing any sort of UBI program requires resources that many insurance companies do not have. Some of these resources are physical, such as the devices used to collect the data or the computers used to store and process it, and can be purchased, often at significant expense. Others, such as the knowledge and skills required to work with and derive value from the data, must be learned or outsourced. For these reasons, no UBI program should be considered a small endeavor, and insurers must have a clear vision for the form that the program will ultimately take. Such a vision will offer guidance with regard to what types of devices are required to collect, transmit, store, and process DBD. Additionally, many insurers may choose to begin the journey toward a full-fledged UBI program with a pilot to determine its viability and to minimize up-front costs while resources and knowledge are accumulated.

There are a wide variety of devices available on the market to collect DBD, and selecting the most appropriate device or devices for a UBI program depends on evaluating numerous factors. Any robust program will need a comprehensive set of rating variables, such as some of those listed in Section 2, to be as granular as possible. Devices capable of providing the data elements to feed such a program need to have a robust feature set and extensive capabilities. Such devices are currently expected to include GPS technology, accelerometers for each axis of motion, and the ability to read data from the vehicles' own sensors via the OBDII port. Devices like this currently cost between $100 and $200 each, which is down from $300 to $400 just a few years ago for less capable devices and, as expected for almost any form of technology, prices are always dropping. In the future, devices may include additional accelerometers or gyroscopes to determine angular acceleration, in addition to the rectilinear acceleration currently being measured, and vehicle orientation. Programs that only need to rely on a narrower set of rating variables require less capable and less expensive devices for implementation.

Aside from the technical capabilities of a device for collecting DBD, another consideration in device selection is data transmission. Generally, a device will buffer behavior and event data for some period of time and, depending on the type of UBI program, may need to be transmitted

periodically to the administrator or service provider for the program. Transmission can be manual or automatic, wired or wireless. For some programs, ongoing data collection may not be required and the easiest way to transmit the data may be to return the device to the program administrator. The data can be retrieved from the device and the device redeployed in another vehicle, which reduces the cost of purchasing devices to be used in as many vehicles as desired. However, more robust programs do require ongoing data collection and mailing devices back and forth is not a viable option. Any such program requires a device capable of data transmission from a remote location.

In order to keep transmission costs down, some devices allow a user to plug a data cable into the device so that the DBD can be transferred to a computer and subsequently transmitted to the program administrator. While inexpensive, manual transmission has a few drawbacks. Users may find manual transmission to be inconvenient and time consuming, making the UBI program, already considered by some to be a hard sell, less attractive to consumers. Furthermore, if the user forgets or neglects to transmit data periodically and in a timely manner, data may be lost if the device buffer does not have enough capacity to store the data collected between transmissions. Users may also choose intentionally not to report data if they believe, for whatever reason, that it will reflect poorly on their driving behavior and somehow negatively impact them. Another consideration is that not all areas, especially some rural locations, have ready access to high-speed Internet. The DBD files may be quite large for the most devices and some users may not be able to transmit such files on their own. For these reasons, automatic transmission of data is often preferred.

Devices that transmit data automatically are generally outfitted with a cellular or Wireless Fidelity (Wi-Fi) radio. The main benefit of cellular radios is that they can transmit data almost anywhere at any time. The transmission trigger could be a predetermined time or period, such as 300 hours since the last transmission, or based on the amount of data collected, for example it could be programmed to transmit whenever the data buffer is 80% full. Whatever the trigger, it should be selected so that data loss is minimized if there is a lag between when the device is programmed to transmit the data and when it actually can, which may happen if the device falls out of range of the cellular network for a brief period after transmission is triggered. However, ubiquity of transmission comes at a cost: data charges for cellular transmission can be expensive. Cellular data plans currently cost between $10 and $20 per month and thus make up a significant portion of plan administration.[29]

[29] For example, see Bird, paragraph 4.

An alternative to cellular transmission is Wi-Fi. If a device is outfitted with a Wi-Fi radio, the cost of transmitting data via cellular network can be eliminated. Devices with Wi-Fi radios are most useful when the vehicle is regularly garaged at locations where Wi-Fi routers can be installed for purposes of data transmission. For example, if a user returns home most nights, the device could be set up to connect to a router within range of the garage and transmit whenever that connection is made. Similarly, an entire fleet of vehicles that is regularly garaged overnight at a particular location could be outfitted with devices that transmit whenever in range of a router at the garage. Or, if a business has multiple garaging locations, a router could be installed at each so that no matter the location at which vehicles are being garaged, they can transmit data as long as they have a network connection.

Devices with Bluetooth technology are beginning to appear on the market. Instead of transmitting data directly to the program administrator or service provider, these devices connect to other Bluetooth-capable devices, such as smart phones, for transmission via whatever mechanism the other device is able to utilize. This is usually Wi-Fi or cellular, and in any case it offloads the cost of transmission to the user. If the user already has a data plan for his or her cellular phone, there is no need to purchase a dedicated cellular plan for the vehicle's device, but the user must be aware of the amount of data being transmitted so as not to incur additional data overage fees.

Once the type of device required by a UBI program of the desired scope is determined, it is recommended that the program administrator evaluate and test various devices from different manufacturers and service providers to find the one that best meets the program's needs. Alternatively, the robustness and viability of any program can be increased if it is flexible enough to be able to use data collected by any device that meets a minimum set of required criteria. It is easier to outfit large fleets in a short amount of time when devices can be purchased from multiple manufacturers. In cases where devices have already been bought for and installed within vehicles, a particular UBI program may be more attractive if it doesn't require a different device.

As far as the infrastructure for any UBI program goes, the devices installed in vehicles, while crucial, only represent part of the picture. In order to be useful, the data collected and transmitted must be stored until it can be processed and this requires additional hardware and software that may not already be available.

Although it depends on how comprehensive a database is desired and the particular specifications for each data element collected, such as precision and frequency, a typical device will transmit between three and twenty megabytes of data per month. Devices installed in commercial vehicles

tend to report amounts on the higher end of this spectrum because the vehicles are being used more. This is generally due to the nature of the business and the fact that commercial vehicles may be used by different drivers during multiple shifts throughout the day. So, depending on the number of vehicles participating in the UBI program, the data could add up rather quickly. As mentioned earlier, credible modeling results are generally achieved using at least 100,000 vehicle-years of data.[30] At the stated rate of data collection, that amounts to between four and twenty-five terabytes, at a minimum, for credible results. There are currently over 250 million vehicles (2,500 times the minimum credibility standard) registered for use in the United States.[31] While no one is suggesting that every vehicle will be outfitted with telematics devices for collecting DBD in the near future, as UBI programs gain acceptance from consumers and insurance producers alike, the sizes of these databases are expected to grow, as are the costs of storing the data. Every enterprise must also have a plan for backing up data in case of loss. This data redundancy also adds significantly to the cost of maintaining a UBI program.

While storage capacity is one consideration for the database, a company must also have the right software to process and work with it. Some database and statistical packages are not able to adequately handle tens or hundreds, or more, terabytes of data. Consider software packages that can utilize efficient algorithms and parallel processing to speed calculations and analysis. Additionally, some software is constrained by Random Access Memory (RAM) in how large the working database can be. Working with this software requires programming expertise to access the entirety of the data. Instead, consider packages that do not have such constraints on database size.

In addition to working with the data to develop UBI programs, many vendors are providing policyholder-facing software to add value and incentive to customers. Many personal auto policyholders are looking for, or even expect to see, real-time information about where their cars have been driven and whether and when any usage events, such as speeding, hard braking or turning sharply, occurred during a trip. They may also want to be able to locate their vehicle when stolen, which benefits both the policyholder and the insurer because the quicker a vehicle is located and reacquired, the less expensive the claim is likely to be and the more satisfied the insured is. Parents may also prefer to use this information to track teen drivers. Commercial auto policyholders are already using information like this for fleet management purposes, including routing and fuel consumption, as well as ensuring that fleet drivers are following various laws and regulations or fleet

[30] See footnote 27 of this paper.
[31] From U.S. Department of Transportation (Bureau of Transportation Statistics), Table 11-1.

safety guidelines. Thus, the software and hardware infrastructure required for providing these services to policyholders should be considered when determining the features and costs of a competitive UBI program.

Finally, as with any company's database, the administrator of a UBI program must make security a top priority. In recent years, company databases have been hacked for e-mail addresses and passwords, and cases such as these have become more prevalent. A few higher-profile cases have resulted in release of names, addresses, and even credit card numbers. Many people would, and should, be wary of databases containing multiple terabytes of data with so much private, identifiable behavior information. Secure access controls must be implemented so that only the people who should be able to see the data are able to. In addition, any information transmitted to or from the policyholder should be done so in a secure manner with data encryption. A data retention policy must also be put into place to minimize data loss, should it occur. When data is no longer needed, it should be deleted from the working database and all backup copies.

In the next section we turn from discussing the infrastructure necessary to collect and use the data to actually working with it.

4. NAVIGATING THE DATA

The driving behavior data elements collected to support UBI vary depending on the technology used to collect them. Telematics devices typically record time-stamped location and vehicle identification information with every observation. Some devices further identify the driver of the vehicle through tags, PIN codes, or fingerprinting. The American Association of Equipment Management Professionals' telematics data standard focuses on location and three additional data points: distance traveled, driving time, and fuel consumption. The organization finds that these four data elements support 80% of their constituency's reporting needs.[32] Speed, acceleration, deceleration, and braking are less standard elements but are also commonly used in UBI. Such values may be estimated if not collected through telematics. For example, average speed may be estimated using time and mileage, and average acceleration may be estimated using time-velocity formulae. The accuracy of the estimates depends on the frequency at which observations are recorded. More detailed DBD such as seat belt usage, turn signaling, and on-board entertainment (e.g., radio stations) have been suggested for use within patents but are not typical of UBI initiatives to date.

[32] See Bennink, page 1.

When choosing which data elements to collect, actuaries should balance the desire for greater predictive power with the cost of collecting more data.

Table 4.1 displays sample DBD. This table is illustrative and does not represent the data collectible from any particular telematics device. In the example, nine data elements are recorded at sixty-second intervals beginning and ending when the ignition is switched on and off, respectively. Such a series of records is referred to as a "journey." The first and second columns of Table 4.1 show that the driver's journey on April 8, 2012, begins at 8:45 AM and concludes at 8:53 AM.[33] The GPS coordinates in the third and fourth columns indicate that he changes heading at least twice: in the fourth observation, he ceases heading Due North, and in the seventh, he begins heading Due West. Columns five, six, and seven show that the driver's cumulative time, mileage, and fuel consumption, respectively, increase with each observation except the second, when his vehicle has been powered on but has not traveled any mileage. The speeds and accelerations in columns eight and nine show that the driver experiences different driving conditions during each leg of his journey. Because insurance policies are typically written or renewed every six to twelve months, and not every minute or journey, the information in Table 4.1 does not immediately lend itself to use in rating or underwriting.

Table 4.1 — Sample DBD for Single Journey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Obs.	Date	UTC	Degrees Latitude	Degrees Longitude	Seconds	Miles	Gallons	Miles per hour	g-force
1	4/8/2012	14:45:30	-27.117	-109.367	0:00:00	0.000	0.000	0	0.000
2	4/8/2012	14:46:30	-27.118	-109.364	0:01:00	0.000	0.050	0	0.000
3	4/8/2012	14:47:30	-27.117	-109.371	0:02:00	0.080	0.055	20	0.030
4	4/8/2012	14:48:30	-27.150	-109.379	0:03:00	0.552	0.060	50	0.060
5	4/8/2012	14:49:30	-27.109	-109.388	0:04:00	1.499	0.078	65	-0.100
6	4/8/2012	14:50:30	-27.098	-109.398	0:05:00	2.538	0.091	59	0.010
7	4/8/2012	14:51:30	-27.093	-109.410	0:06:00	3.234	0.105	40	-0.070
8	4/8/2012	14:52:30	-27.084	-109.409	0:07:00	3.916	0.124	35	-0.015
9	4/8/2012	14:53:30	-27.076	-109.409	0:08:00	4.445	0.143	0	0.000

[33] Pacific/Easter Time is six hours behind Coordinated Universal Time (UTC) when Daylight Saving Time is not in effect.

Table 4.1 (Continued)

Columns (1) and (2) represent the date and time of each observation in UTC.

Columns (3) and (4) represent the GPS coordinates of each observation in decimal notation.

Column (5) represents the cumulative amount of time elapsed at each observation in seconds.

Column (6) represents the cumulative distance traveled at each observation in miles.

Column (7) represents the cumulative amount of fuel consumed at each observation in gallons.

Column (8) represents the speed at each observation in miles per hour.

Column (9) represents the amount of acceleration or deceleration at each observation in g-force.

The first step in making sense of DBD is to identify when different aspects of vehicle operation occur. When a vehicle is powered off, it is referred to as "parked" or "garaged." The device in our example does not record observations while parked, but some devices do record periodic observations at such times. A vehicle may also be powered on but not in motion (e.g., the second observation of Table 4.1), in which case it is considered "idle." Time spent parking or idling may be of value to fleet managers looking to optimize drivers' fuel consumption or actuaries looking to price the theft peril. Of greater interest, however, are events and conditions which occur when a vehicle is powered on and in motion, i.e., "driving." Periodic observations are excellent at capturing driving conditions as in Table 4.1 but may not identify events which occur in a split second, such as entering a turn at high speed or slamming on the brakes. These maneuvers may directly cause or prevent accidents. As a result, many devices record additional observations when one or more conditions exceeds thresholds or "triggers" prescribed by the actuary, TSP, or the two working in conjunction. For example, braking events may be identified when certain gravitational forces (g-force) are registered and/or decelerations occur over a prescribed duration (e.g., Kantowitz defined braking events by a 0.2 g-force deceleration rate over five seconds). Resulting observations may indicate that a given type of event occurred and/or the set of conditions that triggered its recording. **Table 4.2** contains sample events that were recorded by the telematics device during the same journey that resulted in the periodic observations displayed in Table 4.1.

Table 4.2 — Sample Events

Obs.	(1) Date	(2) UTC	(3) Event Description
2.1	4/8/2012	14:47:00	Shift into Drive
6.1	4/8/2012	14:51:00	45-degree turn at high speed
6.2	4/8/2012	14:51:15	-0.4 g-force threshold broken
8.1	4/8/2012	14:53:15	Shift into Park

Once driving and analogous concepts have been identified, they may be quantified. If the cumulative totals in columns five, six, and seven of Table 4.1 and the totality of events in Table 4.2 were the only values of interest, actuaries could save considerable time and expense by accepting just one observation per journey or policy period. To wit, MileMeter offers coverage by the mile, which is verified only at renewal using photographs of vehicles' odometers.[34] Alternatively, some TSPs offer scores based on driving ability which may be used in a manner similar to credit. Actuaries interested in deeper analysis of DBD would prefer to associate events and "incremental" amounts of driving with unique sets of conditions that produce them. This is analogous to associating exposures, losses, and claims with the risks from which they emanate.

Table 4.3 combines the events from Table 4.2 with the periodic observations from Table 4.1 to produce "incremental" driving, parking, and idling totals. For the sake of simplicity, Table 4.3 displays event counts rather than full event detail. (Note that observations denoting transitions between parking, idling, and driving are duplicated to enable proper allocation of incremental totals between vehicle states.) Incremental totals are calculated in two steps: First, as shown in columns (11) through (13), differences are taken between the cumulative totals in columns five through seven of each observation compared to its predecessor record. Next, two-point rolling averages are taken of the differences to obtain the incremental totals in columns (14) through (16). Averages are used because approximately half of the driving between any two observations can be more closely associated with the prior. For example, column (12) shows that 0.472 miles were traveled between observations three and four. Approximately half of this mileage occurred more closely to observation three. Similarly, half of the 0.947 miles traveled between observations four and five may be more closely associated with observation four. Therefore, the incremental mileage shown in column (15) for observation four is estimated as $(0.472 + 0.947) \div 2 = 0.710$. Rolling averages may

[34] MileMeter is described in Furchgott as being "at the non-tech extreme" of UBI (paragraph 9).

be unnecessary if observations are recorded at sufficient frequency that conditions are unlikely to differ significantly between consecutive observations. Determining incremental totals for each observation allows the user to aggregate driving totals over different sets of conditions.

Table 4.3 — Periodic Observations with Events and Incremental Totals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Obs.	Date	UTC	Degrees Latitude	Degrees Longitude	Seconds	Mileage	Gallons	Miles per hour	g-force
	4/8/2012	14:45:30	-27.117	-109.367	0:00:00	0.000	0.000	0	0.000
1	4/8/2012	14:45:30	-27.117	-109.367	0:00:00	0.000	0.000	0	0.000
2	4/8/2012	14:46:30	-27.118	-109.364	0:01:00	0.000	0.050	0	0.000
2.1	4/8/2012	14:47:00	-27.118	-109.364	0:01:30	0.000	0.053	0	0.000
2.1	4/8/2012	14:47:00	-27.118	-109.364	0:01:30	0.000	0.053	0	0.000
3	4/8/2012	14:47:30	-27.117	-109.371	0:02:00	0.080	0.055	20	0.030
4	4/8/2012	14:48:30	-27.150	-109.379	0:03:00	0.552	0.060	50	0.060
5	4/8/2012	14:49:30	-27.109	-109.388	0:04:00	1.499	0.078	65	-0.100
6	4/8/2012	14:50:30	-27.098	-109.398	0:05:00	2.538	0.091	59	0.010
6.1	4/8/2012	14:51:00	-27.096	-109.404	0:05:30	2.886	0.098	53	-0.050
6.2	4/8/2012	14:51:15	-27.094	-109.407	0:05:45	3.060	0.102	45	-0.430
7	4/8/2012	14:51:30	-27.093	-109.410	0:06:00	3.234	0.105	40	-0.070
8	4/8/2012	14:52:30	-27.084	-109.409	0:07:00	3.916	0.124	35	-0.015
8.1	4/8/2012	14:53:15	-27.076	-109.409	0:07:45	4.445	0.138	0	0.000
8.1	4/8/2012	14:53:15	-27.076	-109.409	0:07:45	4.445	0.138	0	0.000
9	4/8/2012	14:53:30	-27.076	-109.409	0:08:00	4.445	0.143	0	0.000
	4/8/2012	14:53:30	-27.076	-109.409	0:08:00	4.445	0.143	0	0.000

Table 4.3 (Continued)

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
		Differences between Obs.			Incremental Totals				
Obs.	Status	Seconds	Mileage	Gallons	Seconds	Mileage	Gallons	Turns	Brakes
	Parked								
1	Idling	0:00:00	0.000	0.000	0:00:30	0.000	0.025	0	0
2	Idling	0:01:00	0.000	0.050	0:00:45	0.000	0.026	0	0
2.1	Idling	0:00:30	0.000	0.003	0:00:15	0.000	0.001	0	0
2.1	Driving	0:00:00	0.000	0.000	0:00:15	0.040	0.001	0	0
3	Driving	0:00:30	0.080	0.003	0:00:45	0.276	0.004	0	0
4	Driving	0:01:00	0.472	0.005	0:01:00	0.710	0.012	0	0
5	Driving	0:01:00	0.947	0.018	0:01:00	0.993	0.016	0	0
6	Driving	0:01:00	1.039	0.013	0:00:45	0.694	0.010	0	0
6.1	Driving	0:00:30	0.348	0.007	0:00:23	0.261	0.005	1	0
6.2	Driving	0:00:15	0.174	0.004	0:00:15	0.174	0.004	0	1
7	Driving	0:00:15	0.174	0.003	0:00:38	0.428	0.011	0	0
8	Driving	0:01:00	0.682	0.019	0:00:53	0.606	0.017	0	0
8.1	Driving	0:00:45	0.529	0.014	0:00:23	0.265	0.007	0	0
8.1	Idling	0:00:00	0.000	0.000	0:00:08	0.000	0.002	0	0
9	Idling	0:00:15	0.000	0.005	0:00:08	0.000	0.002	0	0
	Parked								

Columns (1) through (9) are described in the notes to Table 4.1.

Column (10) indicates whether the vehicle is parked, idling, or driving at each observation.

Column (11) represents the amount of time elapsed (Column (5)) between the previous observation and the current observation in seconds.

Column (12) represents the distance traveled (Column (6)) between the previous observation and the current observation in miles.

Column (13) represents the amount of fuel consumed (Column (7)) between the previous observation and the current observation in gallons.

Column (14) estimates the amount of time associated with each observation, which is calculated as the average of the differential times in Column (11) for the current and subsequent observations.

Table 4.3 (Continued)

Column (15) estimates the distance traveled associated with each observation, which is calculated as the average of the differential distances in Column (12) for the current and subsequent observations.

Column (16) estimates the fuel consumption associated with each observation, which is calculated as the average of the differential fuel amounts in Column (13) for the current and subsequent observations.

Column (17) represents the number of turns taken at high speed (as identified in Table 4.2) that occurred at each observation.

Column (18) represents the number of heavy braking incidents (as identified in Table 4.2) that occurred at each observation.

Raw DBD describes a relatively limited number of conditions, but resources are available to paint a richer portrait. A simple yet valuable distinction to the actuary might be on-roading versus off-roading. However, GPS coordinates are often imprecise (e.g., due to interference from buildings), so a large percentage of driving may incorrectly appear to be "off-road." In response, algorithms are available to "snap" coordinates to their nearest road segment. Map vendor data may then be used to identify the type of road (e.g., freeways, arterials, or local roads) or speed limit for each segment. Real-time traffic services are also available. GPS and time stamps may even be used to access temperatures, precipitation, or other weather information. Studies into these effects are promising, but a dearth of weather stations in the vicinities of driving may limit the success of such efforts. Due to the costs of licensing third-party data and different levels of refinement between raw telematics and third party databases, actuaries may find linking such information to the DBD to be difficult. Simpler calculations may also be performed. Distances between driving locations and journey origins may be a useful check on traditional rating variables such as radius of operation, and may be calculated using publicly available algorithms. Alternatively, ISO proposed a simple, patent-pending metric to describe the riskiness of driving locations in terms of the average garaging loss costs of encompassing territories. In summary, analysis of DBD need not be limited to raw information collected using telematics.

Maintaining a quality database involves more than simply processing incoming DBD. ASB ASOP No. 23 advises that the actuary should identify "data values that are materially questionable or relationships that are materially inconsistent."[35] An abnormally low number of observations may indicate that an insured has installed his device incorrectly or hacked into his own data stream. Frequent "gaps" in the data or large numbers of events may be signs of technological defect.

[35] From ASOP No. 23, page 4.

Insurance claims which do not have corresponding DBD events (e.g., intense g-force) may represent fraud. Such issues should be pursued with the insured, TSP, or claims adjuster as appropriate. Automated checks and corrective measures may also be applied to the data. Observed speeds or distances may be compared to those implied by incremental mileage-to-time ratios or GPS coordinates, respectively. Missing or unreasonable values may be estimated using data from adjacent observations. For example, when a vehicle shifts from idling into drive, its location may be estimated from the previous observation, since which time it has not moved. In contrast, fuel consumption could be estimated using interpolation between the previous and subsequent records' fuel consumption, since fuel is burned while idling. The actuary will not have opportunity to review every record of DBD but should exercise "professional judgment" in determining whether the data is of "sufficient quality to perform [the] analysis."[36]

The aggregate database of DBD may contain observations for several highly similar journeys. It may not be necessary, or desirable from a privacy perspective, to separately identify each one. Dates may be expressed simply as weekdays, weekends, or holidays, while times may be categorized (e.g., early morning, Ante Meridiem (AM) peak, etc.) by the expected degree of traffic congestion. GPS coordinates, which many companies do not even collect for privacy or other reasons, may be discarded after they are used to determine road type, weather, or other qualitative and quantitative information. Conditions such as speed and acceleration may not be retained to the exact miles per hour (MPH) or hundredth of a g-force,, since riskier cases have been separately identified as events. Cumulative totals are somewhat redundant to incremental totals and may be expressed as broader ranges. A one-way analysis of the DBD may be used to identify logical groupings of the data. **Table 4.4** presents Table 4.3 in a slightly different format which better preserves privacy. This approach is consistent with the Law of Large Numbers: "As the volume of similar, independent exposure units increases, the observed experience will approach the "true" experience."[37]

[36] Ibid., page 5.
[37] From Werner and Modlin, page 216.

Table 4.4 — De-identified DBD

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
					Cumulative Totals		
Obs.	Date	Time	Road	Weather	Minutes	Miles	Gallons
	Holiday	AM Peak	Local	Partially Cloudy	0	0	0
1	Holiday	AM Peak	Local	Partially Cloudy	(0 , 5]	0	(0 , 0.25]
2	Holiday	AM Peak	Local	Partially Cloudy	(0 , 5]	0	(0 , 0.25]
2.1	Holiday	AM Peak	Local	Partially Cloudy	(0 , 5]	0	(0 , 0.25]
2.1	Holiday	AM Peak	Local	Partially Cloudy	(0 , 5]	(0 , 5]	(0 , 0.25]
3	Holiday	AM Peak	Local	Partially Cloudy	(0 , 5]	(0 , 5]	(0 , 0.25]
4	Holiday	AM Peak	Freeway	Partially Cloudy	(0 , 5]	(0 , 5]	(0 , 0.25]
5	Holiday	AM Peak	Freeway	Partially Cloudy	(0 , 5]	(0 , 5]	(0 , 0.25]
6	Holiday	AM Peak	Freeway	Partially Cloudy	(0 , 5]	(0 , 5]	(0 , 0.25]
6.1	Holiday	AM Peak	Freeway	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]
6.2	Holiday	AM Peak	Arterial	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]
7	Holiday	AM Peak	Arterial	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]
8	Holiday	AM Peak	Arterial	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]
8.1	Holiday	AM Peak	Arterial	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]
8.1	Holiday	AM Peak	Arterial	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]
9	Holiday	AM Peak	Arterial	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]
	Holiday	AM Peak	Arterial	Partially Cloudy	(5, 10]	(0 , 5]	(0 , 0.25]

Column (1) indicates whether each observation is recorded on a Weekday, Weekend, or Holiday.

Column (2) indicates the time of day during which each observation is recorded. "AM Peak" is defined by Jun as occurring between the hours of 6:00 AM and 9:00 AM, Local Time.

Column (3) represents the type of road being traveled at each observation. This is obtained by linking the GPS coordinates in Columns (3) and (4) of Table 4.3 to a map database, similar to Jun.

Column (4) represents the weather conditions at each observation. This is obtained by linking the date, time, and GPS coordinates from Columns (1) through (4) of Table 4.3 to a weather database.

Column (5) expresses Column (5) of Table 4.3 in ranges of five minutes.

Column (6) expresses Column (6) of Table 4.3 in ranges of five miles.

Column (7) expresses Column (7) of Table 4.3 in ranges of 0.025 gallons.

Table 4.4 (Continued)

	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
		Over/Under		Incremental Totals				
Obs.	Speed	Speed Limit	Status	Seconds	Miles	Gallons	Turns	Brakes
	-	-	Parked					
1	-	-	Idling	0:00:30	0.000	0.025	0	0
2	-	-	Idling	0:00:45	0.000	0.026	0	0
2.1	-	-	Idling	0:00:15	0.000	0.001	0	0
2.1	[0 , 2.5]	[-17.5 , -12.5)	Driving	0:00:15	0.040	0.001	0	0
3	(17.5 , 22.5]	[- 2.5 , 2.5]	Driving	0:00:45	0.276	0.004	0	0
4	(47.5 , 52.5]	[-17.5 , -12.5)	Driving	0:01:00	0.710	0.012	0	0
5	(62.5 , 67.5]	[- 2.5 , 2.5]	Driving	0:01:00	0.993	0.016	0	0
6	(57.5 , 62.5]	[- 7.5 , -2.5)	Driving	0:00:45	0.694	0.010	0	0
6.1	(52.5 , 57.5]	[- 12.5 , -7.5)	Driving	0:00:23	0.261	0.005	1	0
6.2	(42.5 , 47.5]	(2.5 , 7.5]	Driving	0:00:15	0.174	0.004	0	1
7	(27.5 , 32.5]	[- 2.5 , 2.5]	Driving	0:00:38	0.428	0.011	0	0
8	(32.5 , 37.5]	[- 7.5 , -2.5)	Driving	0:00:53	0.606	0.017	0	0
8.1	[0 , 2.5]	(-∞, -27.5)	Driving	0:00:23	0.265	0.007	0	0
8.1	-	-	Idling	0:00:08	0.000	0.002	0	0
9	-	-	Idling	0:00:08	0.000	0.002	0	0
	-	-	Parked					

Column (8) expresses the speed from Column (8) of Table 4.3 in ranges of 5 MPH. Speed is not generally applicable when the vehicle is parked or idling.

Column (9) compares the speed from Column (8) of Table 4.3 to the speed limit and expresses the difference in ranges of 5 MPH. Speed limits are obtained by linking GPS coordinates from Columns (3) and (4) of Table 4.3 to a traffic laws database.

Column (10) replicates Column (10) of Table 4.3.

Columns (11) through (15) replicate Columns (14) through (18) of Table 4.3, respectively.

The final step in constructing the database is to associate DBD with traditional insurance data (e.g., premiums and losses). One way to do this is to assign every claim to the observation which immediately precedes it, e.g., one with intense g-force. This may prove challenging because claims are not typically recorded in second-by-second detail, and DBD may become irretrievable if

telematics devices are damaged in accidents. Another limitation of this approach is that such observations may only be as frequent (or infrequent) as accidents themselves. Alternatively, actuaries may associate each claim with all the DBD observations that occur during the policy period. This assumes that every moment a vehicle is operated has some impact on its loss propensity. Either of these two approaches supports frequency and severity modeling, because every observation indicates whether or not a claim occurs and, if so, for what amount. If the actuary prefers a pure premium approach, he or she may pro-rate losses to DBD observations based on time, mileage, or fuel consumption, all of which were noted by Dorweiler as potential exposure media for auto. He or she may also consider pro-rating premiums and building a loss ratio model, which would effectively control for the effect of existing rating variables. A more sophisticated approach to account for those effects would be to associate exposure information with the DBD and analyze for interactions. For example, less experienced vehicle operators who partake in risky behaviors may present greater risk than more experienced ones with similar behavioral patterns. Merging traditional insurance data with DBD enables the actuary to take the next step of analyzing the effects of driving behavior on loss experience. The next section looks at how actuaries may act upon this information.

5. APPLICATIONS

There are many potential applications for DBD, including pricing, scoring, underwriting, claims, and others. For actuaries, the endgame is certainly pricing. Actuaries generally already have an informed estimate of the proper aggregate cost of insurance for a group of risks based on traditional methods, but DBD has the potential to make rating individual risks more granular and accurate than ever before. New rating variables can be integrated into old rating schemes as additional variables or by replacing traditional variables. These new rating variables may serve as more accurate measures of the same rating criteria or they may substitute something not previously measurable by using a conventional proxy. For example, in commercial auto the "radius of operations" rating variable is used as an estimate of how much a vehicle is operated since vehicles that tend to be driven further from the garaging location tend to be driven more than vehicles that stay within a smaller radius. This variable can be replaced with a more accurate and verifiable measure of the actual mileage driven and an indicator of how far the vehicle strays from its garaging location based on GPS data. These can help differentiate not only between vehicles used close to and far from the garaging location, but, among these, which are used more frequently than others. Ultimately, total mileage

may prove to be more indicative of loss propensity and may supersede distance from the garaging location.

As another example, measurements based on variables that are more closely related with the aggressiveness of the driver, such as speeding and hard braking, could replace traditional proxies like driver age. Younger drivers may tend to be more aggressive, but when actuaries know what aggressive driving looks like in the data, they can apply surcharges only to those drivers exhibiting such behaviors, whether they are youthful or not, instead of an entire class of drivers comprised of operators with varying driving habits. Also, replacing proxies with variables that may be perceived as more causative, while not necessary under ASB ASOP No. 12, should be well received by consumers, who appreciate transparency in rating and desire to know what they can do to affect premiums. This is also true for regulators, who value fairness and accuracy in rating. DBD may also be used to create an entirely new, independent UBI rating plan without reliance on traditional plans or variables.

Other applications for DBD include scoring and underwriting. Scoring is useful for condensing multiple pieces of information into a single number. A credit score, for example, combines information about an individual's payment and delinquency history, credit utilization, length of credit history, and type of credit into a single number ranging from 300 and 850 (for the Fair Isaac Corporation (FICO) score in the United States) to represent the creditworthiness, or likelihood of default, of that individual.[38] At a glance, a financial institution can make a decision about whether or not to extend a loan to an individual, and if so at what rate, based on the expected risk of that person defaulting on the loan. A similar driving behavior score could be developed to summarize the plethora of information contained in DBD and other linked databases. And, similar to the use of credit scoring in personal insurance since the 1990s, a driving behavior score could be used as a rating variable instead of building a model to account for all of the components that make up the score. Actuaries should be wary, however, of the fact that a score, a single number, while indicative of overall risk, does not tell as complete a story as the individual components. And if the underlying composition of the score differs from company to company, scores may not be portable for use in rating or even to indicate how risky one driver is when compared to others or with respect to the producer's own underwriting criteria.

[38] See Gutner or visit FICO.com.

Regarding the use of DBD in underwriting, behavior data, as individual components or as scores, may be used to make decisions about whether or not to cover a risk or which tier, premium, standard, substandard, etc., to place a risk in for pricing. It may take some time for UBI programs and models to gain acceptance from consumers and regulators for pricing, but in the meantime there is no reason that the information cannot be used in underwriting to help place insureds with similar risks for traditional rating.

Another underwriting application of DBD is in verification of rating parameters to prevent premium leakage. Insurers require policyholders to complete an application form before providing coverage and often upon renewal as well. When insureds complete the application form honestly and accurately, the premium charged represents the best estimate of expected loss for that insured. Insureds may on occasion complete the application forms inaccurately if they don't know the exact answer to a question or if they think it will help reduce the premium. Many questions on an application form, such as annual mileage driven, garaging address, business/pleasure use, and distance to work, whether a vehicle is parked in a garage or on the street overnight, etc. could be verified with the use of DBD to ensure that risks are being charged a fair premium.

Some speculate that DBD could be used to determine how many different people regularly use a vehicle, based on how different individual driving habits may be, to verify that the number of named insureds on a policy is correct. Driving behavior data could notify an insurer of potential mid-term policy changes, such as a change of address, before the policyholder contacts the insurer. Applications like this start to push the boundaries of what DBD and UBI are capable of providing to insurers. In order to realize the potential, actuaries, statisticians, and other data scientists will need to carefully and thoroughly analyze behavior data to transform it into valuable information. As was seen with the use of credit for insurance, the first movers may gain expertise that solidifies their position for years to come.

Outside of rating and underwriting, driving behavior data could be used for claims. For example, if a device installed in a vehicle is capable of cellular transmission and is monitoring the vehicle's acceleration and deceleration, it could detect and immediately report when there is a large change in deceleration, indicative of a collision. This is called first notice of loss (FNOL). With FNOL, policyholders can get the help they need more quickly than if they have to seek it themselves. In severe accidents, where people may be incapacitated, automatic notification of an emergency could be the difference between life and death. FNOL has the potential to increase satisfaction among policyholders while decreasing short- and long-term costs for both bodily injury and property

damage. The faster a claimant is diagnosed and treated for injuries, the less expensive payments tend to be. Similarly, insurers spend millions of dollars per year in vehicle storages costs as a result of delayed notification of claims and the average payout for claims reported more than a day after an accident is 20% higher than those reported within a day.[39]

Device data could also be used for claim adjustment and fraud prevention. It has been estimated that over 40% of reported bodily injury claims result from fraudulent or exaggerated injuries and that about 20% of every claim payment is attributed to soft fraud.[40] If the devices could be used to help recreate the conditions of an accident, insurers could be given more accurate assessments based on soft-tissue and kinematics studies to determine the authenticity of a claim and to help make estimates of claim payments. Just knowing which claims to deny or investigate helps ensure that resources are being used as effectively as possible.

DBD has non-insurance uses as well. Some applications include making improvements to infrastructure safety and better vehicle traffic planning. Claim and GPS data could be used to determine safety levels of roads and intersections. Additional lights or signs could be installed to help prevent accidents. Congested areas and roads around them could be improved to add new lanes or change traffic patterns to help traffic flow more smoothly and consistently. It is arguable that repurposing DBD for applications like these also has an effect on insurance, as improving safety and trip quality may ultimately reduce accidents due to road and driving conditions or driver attitudes, which reduces overall claim payments. However, before data can be repurposed, there are a few regulatory hurdles that must be considered.

Not all states have passed laws related to vehicle telematics, but those that have generally declared that data from the devices is owned by the owner of the vehicle, not the insurer. Therefore, use of the data for any purpose depends on agreement with the policyholder. This can be easily accomplished with the policy itself as long as the policyholder is made aware of the implications that come with installing a telematics device in a vehicle and transmitting data to the insurer. Alternatively, while many states have decided that telematics data can be used for rating, especially when regulators feel that DBD is indicative of causal behaviors and promotes fairness in rating, some that explicitly allow DBD to be used for rating may prefer that it only be used for premium discounts. Still, other states, like California, have restrictions on which data can be used. California does not allow the use of GPS location data for rating, but encourages the use of mileage. For now,

[39] From Diamond Management & Technology Consultants analysis cited in Blumer.
[40] From Brockett, pages 245-246.

policyholders that agree to transmit data to an insurer do so because they believe they will get a benefit from doing so, such as premium discounts and other real-time services, and it should not be difficult to find a market for insurers to begin testing or rolling out UBI programs.

6. CONCLUSIONS

It has long been known that DBD allows insurers to achieve unique goals in pricing, underwriting, and loss control or mitigation. However, such data also presents unique challenges and risks. One way to address these roadblocks is through a sound data management strategy. In making the decision to work with DBD, actuaries should be mindful of privacy and intellectual property rights. Once they have committed to UBI, infrastructure should be designed to cost-effectively capture and transmit required information to the insurer and/or insured, with appropriate security measures taken along the way. Telematics is the data collection mechanism that best supports these objectives. As DBD is collected, it should be organized, enriched with third-party information, and de-identified or consolidated to better enable traditional actuarial functions. Automated checks and professional judgment should be used to ensure data is of sufficient quality for analyses. The resulting database serves as the starting point for various UBI applications. Regulatory and technological breakthroughs can only increase DBD's prevalence in an industry that has thus far been reluctant to fully embrace its possibilities.

Acknowledgments

The authors acknowledge John Baldan, Vinay Deshmukhe, Christopher Sirota, Isaac Wash, and Dorothy Ziegelbauer of Insurance Services Office, Inc., for their insights into working with driving behavior data.

7. BIBLIOGRAPHY

[1.] Abdel-Aty, Mohammed et al., "Relating Crash Occurrence to Freeway Loop Data, Weather Conditions, and Geometric Factors," October 1, 2005, http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_TE/FDOT_BD548_04.pdf.

[2.] Actuarial Standards Board, *Actuarial Standard of Practice #12: Risk Classification* (Revised Edition), December 2005, http://www.actuarialstandardsboard.org/pdf/asops/asop012_101.pdf.

[3.] Actuarial Standards Board, *Actuarial Standard of Practice #23: Data Quality* (Revised Edition), December 2004, http://www.actuarialstandardsboard.org/pdf/asops/asop023_097.pdf.

[4.] Anderson, Duncan et al., "A Practitioner's Guide to Generalized Linear Models (Third Edition)," CAS *Discussion Paper Program: Applying and Evaluating Generalized Linear Models*, February, 2007, http://www.casact.org/pubs/dpp/dpp04/04dpp1.pdf.

[5.] Automotive Service Association, Mechanical Division Operations Committee, "Telematics: Past, Present, and Future," May 2008, http://www.asashop.org/news/asaresources/ASAtelematics_0508.pdf.

[6.] Bakos, Tom, and Mark Nowotarski, "Progressive Builds Fortress of Patent Protection," *Insurance IP Bulletin*, Volume 2004, Number 3, October 15, 2004, http://www.bakosenterprises.com/IP/B-10152004/pwcomplete.html.

[7.] Barkouk, Laiss, "Weather Telematics Leverages Truck Fleets to Generate Weather Data," *GPS Business News*, June 7, 2011, http://www.gpsbusinessnews.com/Weather-Telematics-Leverages-Truck-Fleets-to-Generate-Weather-Data_a3070.html.

[8.] Bennink, Curt, "Standard Makes Telematics Practical," *Equipment Today,* July 2011, http://findarticles.com/p/articles/mi_hb5751/is_20110701/ai_n57858839/.

[9.] Bird, Colin, "State Farm Testing OnStar Competitor," Kicking Tires: The Blog for Car Buyers, August 12, 2011, http://blogs.cars.com/kickingtires/2011/08/state-farm-testing-an-onstar-competitor.html.

[10.] Blumer, Fred, "Insurance Telematics: More than Just New Underwriting Criteria," *Insurance & Technology*, November 18, 2010, http://www.insurancetech.com/blogs/228300120.

[11.] Bricketto, Martin, "Allstate, Liberty Face Progressive Patent Suit," *Law360*, January 13, 2011, http://www.law360.com/articles/219920.

[12.] Brinkhoff, Thomas, "Requirements of Traffic Telematics to Spatial Databases," *Proceedings 6th Annual Symposium on Large Spatial Databases*, Hong Kong: Springer, 1999, pp. 365-369, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.4932&rep=rep1&type=pdf.

[13.] Brockett, Patricia et al., "Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud," *Journal of Risk and Insurance*, Vol. 65, No.2, pp. 245-274, 1998, http://www.derrig.com/research/UsingKohonen'sSelf-OrganizingFeatureMap.pdf.

[14.] Cady, Glee Harrah and Pat McGregor, "Connecting the World," *Protect Your Digital Privacy: Survival Skills for the Information Age*, Que 2002, pp. 148-189, http://books.google.com/books?id=WXVc9I_9nlsC&printsec=frontcover&dq=Protect+your+digital+privacy:+survival+skills+for+the+information+a&hl=en&ei=AxtATunmO8-30AH9tPTIAw&sa=X&oi=book_result&ct=result&resnum=1&ved=0CCkQ6AEwAA#v=onepage&q=Protect%20your%20digital%20privacy%3A%20survival%20skills%20for%20the%20information%20a&f=false.

[15.] California Department of Insurance, "About Us: Provisions of Proposition 103 Affecting the Rate Regulation Division," not dated, http://www.insurance.ca.gov/0500-about-us/0500-organization/0400-rate-regulation/prop-103.cfm,

[16.] "California Lawmakers Reject 'Pay at the Pump' Insurance," *New York Times*, May 27, 1993, http://www.nytimes.com/1993/05/27/us/california-lawmakers-reject-pay-at-the-pump-insurance.html.

[17.] Chartrand, Sabra, "Patents; Insurance Protection for Terrorism, Divorces, Frivolous Lawsuits, and Excessive Gambling Losses," *New York Times*, June 30, 2003, http://www.nytimes.com/2003/06/30/technology/30PATE.html?ei=5007&en=01b31db998d0242c&ex=1372392000&partner=USERLAND&pagewanted=print&position=.

[18.] Colorado Department of Regulatory Agencies, Division of Insurance, *Amended Regulation 5-2-12: Concerning Automobile Insurance Consumer Protections*, August 31, 2010, http://www.dora.state.co.us/Insurance/regs/F5-2-12_083110.pdf.

[19.] Commercial Carrier Journal, "Qualcomm Touts Insurance Telematics, Driver Safety Solution," February 14, 2011, http://www.ccjdigital.com/qualcomm-touts-insurance-telematics-driver-safety-solution/.

[20.] Conners, John B. and Sholom Feldblum, "Personal Automobile: Cost Drivers, Pricing, and Public Policy," CAS *Forum* Winter 1997, pp. 317-341, http://www.casact.org/pubs/forum/97wforum/97wf317.pdf.

[21.] Costlow, Terry, "Privacy Issues May Limit Linkup Between Telematics, Insurance," *Automotive Engineering International*, December 3, 2010, http://www.sae.org/mags/aei/9058.

[22.] Craig, Stephen C., "Telematics-Based Insurance Programs: Dreams and Realities," *Verisk Insurance Solutions Auto Newsletter*, June/July 2011, Issue 1, http://www.verisk.com/insurance/auto-insurance/telematics.html.

[23.] Delaware Division of Motor Vehicles, "OBD II FAQs," October 5, 2010, http://www.dmv.de.gov/services/vehicle_services/faqs/ve_faqs_obdi.shtml.

[24.] Digi International, "Digi Launches Industry's First Fleet Management Telematics Device Family Featuring Five Different Wireless Technologies," October 5, 2009, http://www.digi.com/news/pressrelease?prid=625.

[25.] Dorweiler, Paul, "Notes on Exposure and Premium Bases," CAS *Proceedings* 1929, Volume XVI, Number 33, pp. 319-343, http://www.casact.org/pubs/proceed/proceed29/29319.pdf.

[26.] DriveCam, Inc., "DriveCam Files Patent Infringement Lawsuit Against SmartDrive," Reuters, May 16, 2011, http://www.reuters.com/article/2011/05/06/idUS229826+06-May-2011+BW20110506.
[27.] Duri, Sastri, et al., "Framework for Security and Privacy in Automotive Telematics," *Proceedings of the 2nd International Workshop on Mobile Commerce*, New York: ACM, 2002, http://www.cc.gatech.edu/projects/disl/courses/8803/backup/readinglist_files/p25-duri.pdf.
[28.] Ealey, Lance and Glenn Mercer, "Telematics: Where the Radio Meets the Road," *McKinsey Quarterly*, May 1999, http://mkqpreview2.qdweb.net/Telematics_Where_the_radio_meets_the_road_344.
[29.] Elliott, Peter, and Barry Jennings, "Data Protection Issues with Intelligent Transport Systems, Vehicle Telematics, and Road Pricing," January 7, 2009, http://www.twobirds.com/English/News/Articles/Pages/Data_protection_intelligent_transport_systems_010709.Aspx.
[30.] Ethier, Sheridan, and Randy Martin, "Fast Booting Techniques May Meet Automotive Infotainment/Telematics Activation Needs," *EE Times*, November 17, 2006, http://www.eetimes.com/design/automotive-design/4011112/Fast-booting-techniques-meet-automotive-infotainment-telematics-activation-needs.
[31.] Ferreira Jr., Joseph and Erik Minikel, "Pay-As-You-Drive Auto Insurance in Massachusetts," November, 2010, http://mit.edu/jf/www/payd/PAYD_CLF_Study_Nov2010.pdf.
[32.] Finnegan, Daniel and Christopher Sirota, "Is Vehicle Data Recording Auto Insurance's Future?," June 1, 2005, https://www.qualityplanning.com/media/1185/iso.qpc.vehicle%20data%20recording.v5links.pdf.
[33.] Fletcher, Lauren, and Grace Suizo, "Telematics Use in Work Truck Fleets," Work Truck, March 2011, http://www.worktruckonline.com/Article/Print/Story/2011/03/Telematics-Use-in-Work-Truck-Fleets.aspx.
[34.] Frost and Sullivan, "North America: Slow Takeoff for Telematics-enabled Usage-based Insurance," *GPS Wire News*, June 17, 2011, http://gpswire.net/?p=2290.
[35.] Furchgott, Roy, "More Consumers Are Letting Insurers Monitor Their Mileage," *New York Times*, December 24, 2010, http://www.nytimes.com/2010/12/26/automobiles/26INSURE.html.
[36.] Garth, Denise, "Telematics Data: The Next Level," *IDM Quarterly*, Volume 7, Issue 1, Winter 2010, pp. 1-3, http://www.innovation-group.com/files/documents/Public%20Files/News_Articles/Telematics_Article-_IDM_Quarterly_2010_Winter.pdf.
[37.] GMAC Insurance, "GMAC Insurance and OnStar Create Innovative Insurance Products," January 28, 2004, Ally, http://media.ally.com/index.php?s=43&item=161.
[38.] Gutner, Toddi, "Anatomy of a Credit Score," *Business Week*, November 28, 2005, http://www.businessweek.com/magazine/content/05_48/b3961124.htm.
[39.] Hales, Mike, et al., "Telematics: Reinventing Auto Insurance Part II," *Insurance & Technology*, September 8, 2010, http://www.insurancetech.com/blogs/227400412.
[40.] Harris, Jim, "Event Data Recorders - State Statutes and Legal Considerations," *Accident Reconstruction Journal*, Vol. 18, No. 1, January/February, 2008, http://www.harristechnical.com/downloads/Harris_EDR_article.pdf.
[41.] Helft, Miguel, "Jobs Says Apple Made Mistakes with iPhone Data," *New York Times*, April 27, 2011, http://www.nytimes.com/2011/04/28/technology/28apple.html?_r=2.
[42.] Hendel, John, "Telematics and UBI: The Regulatory Opportunities," *Telematics Update*, September 5, 2011, http://social.telematicsupdate.com/insurance-telematics/telematics-and-ubi-regulatory-opportunities.
[43.] Hendricks, Evan, "Insurer's Patent Targets Driver's Every Move," *The Privacy Times*, November 4, 1999, http://www.privacytimes.com/NewWebstories/home_priv_11_16.htm.
[44.] Hughes Telematics, Inc., "Amendment No. 1 to Form S-1 Registration Statement under the Securities Act of 1933," April 6, 2010, http://www.faqs.org/sec-filings/100407/HUGHES-Telematics-Inc_S-1.A/.
[45.] Hughes Telematics, Inc., "Insurance and Telematics: More Than Just Better Underwriting," not dated, http://www.hughestelematics.com/pp/whitepapers/insurancewp.php.
[46.] Hunkins, Dave, "Emergence of Consumer Solutions in Vehicle Telematics," December 29, 2003, http://www.cs.clemson.edu/~johnmc/courses/cpsc875/projects/emerge.pdf.
[47.] InCode Telecom Group, "Telematics: How Economic and Technological Forces Will Shape the Industry in the U.S.," May 2001, https://confluence.engin.umich.edu/download/attachments/1605717/Telematics_Position_Paper_v11.pdf.
[48.] Insurance Institute for Highway Safety, "Q&A: Event Data Recorders," *Research: Event Data Recorders*, November 2010, http://www.iihs.org/research/qanda/edr.html.

[49.] Ippisch, Tobias, "Telematics Data in Motor Insurance: Creating Value by Understanding the Impact of Accidents on Vehicle Use," Lulu Enterprises 2010, http://www1.unisg.ch/www/edis.nsf/SysLkpByIdentifier/3829/$FILE/dis3829.pdf.

[50.] Iqbal, Muhammad Usman and Samsung Lim, "Location Privacy in Automotive Telematics," August 25, 2009, http://www.gmat.unsw.edu.au/snap/publications/usman&lim2007a.pdf.

[51.] Jennings, Trip, "Internet Upgrade Headed to Northern New Mexico," *The New Mexican*, August 15, 2011, http://www.santafenewmexican.com/Local%20News/Internet-upgrade-headed-to-region.

[52.] Jones, Steve, "Introduction to Road Data - Part One," *Directions Magazine*, August 25, 2010, http://www.directionsmag.com/articles/introduction-to-road-data-part-1/130105.

[53.] Jun, Jungwook, "Potential Crash Exposure Metrics Based on GPS-Observed Driving Behavior Activity Metrics," Savannah: Georgia Institute of Technology, 2006, http://commuteatlanta.ce.gatech.edu/Resources/jun_jungwook_200612_phd.pdf.

[54.] Kaften, Cheryl, "UPS Delivers Fuel-Per-Package Savings That Equal 63.5 Million Miles Not Driven," *Green Technology World*, August 1, 2011, http://green.tmcnet.com/topics/green/articles/203060-ups-delivers-fuel-per-package-savings-that-equal.htm.

[55.] Kantowitz, Barry H., and Sandeep Premkumar, "Safe Vehicles Using Adaptive Interface Technology (Task 6b): Identify Demand Levels of Telematic Tasks," November 2004, http://www.volpe.dot.gov/hf/roadway/saveit/docs/dec04/finalrep_6b.pdf.

[56.] Kauth, Joel, "Keeping the Patent Trolls at Bay - Defending Your Location-Based Patents and Your Business," *Directions Magazine*, April 18, 2011, http://kppb.com/kppb/index.php?option=com_content&view=article&id=81&catid=3&Item.

[57.] Kramer, Jan, "Bundling Telecommunications Services: Competitive Strategies for Converging Markets," December 13, 2007, http://www.im.uni-karlsruhe.de/Upload/Publications/3d6fe37c-fd87-4ba8-aa55-da0f9d519225.pdf.

[58.] Langevoort, Harry, "Tracking and Tracing in a Complex Environment," September 11, 2007, http://www-05.ibm.com/nl/events/presentations/tracking_and_tracing_in_a_complex_enviroment.pdf.

[59.] Lasky, Michael S., "Real-Time Traffic Info Gets You Past Jams," *PC World*, December 22, 2006, http://www.pcworld.com/article/128294/realtime_traffic_info_gets_you_past_jams.html.

[60.] Lewis, Peter, "Car Insurance Patents Threaten Consumer Choice," *Insure.com*, July 10, 2009, http://www.insure.com/car-insurance/patents.html.

[61.] Liberty Northwest, "Liberty Northwest® Launches Onboard Advisor™ for Commercial Fleets," Business Wire, March 3, 2009, http://www.businesswire.com/news/home/20090303006147/en/Liberty-Northwest%C2%AE-Launches-Onboard-Advisor%E2%84%A2-Commercial-Fleets.

[62.] Madison, Bill, "Tell-Tale Telematics," *Claims Advisor*, January 11, 2011, http://claimsadvisor.com/articles/tell-tale-telematics/2/.

[63.] Magney, Phil, "The Impact of AT&T's New Data Plan on Telematics," IHS iSuppli Market Research, June 28, 2010, http://www.isuppli.com/Automotive-Infotainment-and-Telematics/MarketWatch/Pages/The-Impact-of-ATTs-New-Data-Plan-on-Telematics.aspx.

[64.] M-Cam, Inc., "Intellectual Property Analysis of Progressive's U.S. Patent No. 7,124,088," May 16, 2011, http://www.m-cam.com/sites/www.m-cam.com/files/20110516%20-%20Progressive%20v%20Allstate%20et%20al.pdf.

[65.] Mahler, Howard C., "The Credibility of a Single Private Passenger Driver," CAS *Proceedings* 1991, Volume LXXVIII, Number 148, pp. 146-162, http://www.casact.org/pubs/proceed/proceed91/91146.pdf.

[66.] Melnitzer, Julius, "Bilski Ruling from U.S. Supreme Court," *Financial Post*, June 29, 2010, http://business.financialpost.com/2010/06/29/bilski-ruling-from-us-supreme-court/.

[67.] Mestdagh, Wilfred, "Snap to a Road," *MP2K Magazine*, February 12, 2006, http://www.mp2kmag.com/a130--snap.calcxy.gps.mappoint.html.

[68.] Michigan Department of Licensing and Regulation, Insurance Bureau, "Part III: Rating," *A Year of Change: The Essential Insurance Act of 1981*, June 4, 1982, pp. 13-15, http://www.michigan.gov/documents/dleg/The_Essential_Insurance_Act_in_1981_272421_7.pdf.

[69.] Mobile TeleSystems OJSC, "Telematics (Corporate)", not dated, http://www.corp.mtsgsm.com/upload/contents/301/telematika.pdf/.

[70.] Muckell, Jonathan, et al., "Towards an Intelligent Brokerage Platform: Mining Backhaul Opportunities in Telematics Data," Journal of the Transportation Research Board, Issue 2097, September 1, 2009, pp. 1-8, http://www.jmuckell.org/Publications/TRB_88th%20Meeting%20GE%20GRC%2009-3144.pdf.

[71.] Mueller, Scott, "Microprocessors from 1971 to the Present," *Upgrading and Repairing PCs* (17th Edition), Que 2006, http://www.informit.com/articles/article.aspx?p=482324&seqNum=2.

[72.] Muermann, Alexander, and Daniela Straka, "Asymmetric Information in Auto Insurance: New Evidence from Telematics Data," December 2010, http://www.skinance.com/Papers/2011/Muermann.pdf.

[73.] Neleman, Paul, "What Is First Notice of Loss and Why Is It Important to Policy Holders and Insurers?" November 22, 2010, http://ezinearticles.com/?What-Is-First-Notice-Of-Loss-And-Why-Is-It-Important-To-Both-Policy-Holders-And-Insurers?&id=5386395.

[74.] Nora, Simon and Alain Minc, *The Computerization of Society*, The MIT Press, 1980.

[75.] O' Connor, Amy, "Safe Driving Habits Pay Off for Consumers with Usage-Based Insurance Programs," *Insurance Journal*, August 1, 2011, http://www.insurancejournal.com/magazines/mag-features/2011/08/01/208734.htm.

[76.] O' Connor, Brendan, "Comparison of Data Analysis Packages: R, Matlab, SciPy, Excel, SAS, SPSS, Stata," February 23, 2009, http://brenocon.com/blog/2009/02/comparison-of-data-analysis-packages-r-matlab-scipy-excel-sas-spss-stata/.

[77.] Oregon Department of Consumer and Business Services, "Credit Scoring in Insurance - Questions and Answers," November 2007, http://insurance.oregon.gov/FAQs/credit_scoring-qa.pdf.

[78.] Palmer, W. Scott, "Auto 'Black Box' Data: Industry Update," October 31, 2003, http://www.injurysciences.com/Documents/EDRUpdateArticle.pdf.

[79.] Parnell, Karen, and Beng Ceng, "Telematics Digital Convergence—How to Cope with Emerging Standards and Protocols," May 27, 2003, http://www.xilinx.com/support/documentation/white_papers/wp194.pdf.

[80.] PBS&J, *ITS Orange Book*™ - *Smart Highways*, Issue 1, 2005, http://northamerica.atkinsglobal.com/Unpublished/PBSJ_Orange_Books/PDF/OrangeBook_Issue2_screen.pdf.

[81.] Penton Media, "Terion Launches Wireless Service for Trucking," *Drivers: Business News for the Professional Trucker*, November 1, 1999, http://driversmag.com/ar/fleet_terion_launches_wireless/.

[82.] Progressive Corporation, "Innovative Auto Insurance Discount Program to be Available to 5,000 Minnesotans," August 8, 2004, http://www.progressive.com/newsroom/2004/August/Tripsense.aspx.

[83.] Progressive Corporation, "Progressive Testing New Product in Texas That Features Revolutionary Auto Insurance Rating Method," The Free Library, October 27, 1999, http://www.thefreelibrary.com/Progressive+Testing+New+Product+in+Texas+That+Features+Revolutionary...-a056960005.

[84.] Property Casualty Insurers Association of America, "The Predictive Value of Credit-based Insurance Scores," February 11, 2009, http://www.pciaa.net/web/sitehome.nsf/lcpublic/402/$file/pcicreditwp021109.pdf.

[85.] Safeco, "Safeco Offers All Parents of Teen Drivers Peace-of-Mind," May 27, 2008, http://www.safeconews.com/pressrelease.php?p_id=169.

[86.] Shanken, Edward A., "From Cybernetics to Telematics: The Art, Theory, and Pedagogy of Roy Ascott," *Telematic Embrace: Visionary Theories of Art, Technology, and Consciousness*, Berkeley and Los Angeles: Universal of California Press, 2003, pp. 1-96, http://books.google.com/books?id=zN85LrAoDwUC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false,

[87.] Simpson, Andrew, "Progressive Goes National with Usage-Based 'Snapshot' Car Insurance Program," *Insurance Journal*, March 14, 2011, http://www.insurancejournal.com/news/national/2011/03/14/190132.htm#.

[88.] ST Microelectronics, "ST Offers Enhanced 3-Axis Digital Gyroscope," *Sensors Magazine*, April 8, 2010, http://www.sensorsmag.com/machine-manufacturing/news/st-offers-enhanced-3-axis-digital-gyroscope-7014.

[89.] State Environmental Resource Center, "Pay-As-You-Drive Auto Insurance Policy Issues Package," May 10, 2004, http://www.serconline.org/payd/background.html.

[90.] Strube, Joseph, and Bryant Russell, "Actuarial Data Management in a High-Volume Transaction Processing Environment," CAS *Forum* Winter 2005, pp. 274-314, http://www.casact.org/pubs/forum/05wforum/05wf274.pdf.

[91.] Sturgeon, Julie, "Bad Credit Hurts in Many Ways," Bankrate.com, July 21, 2006, http://www.bankrate.com/finance/debt/bad-credit-hurts-in-many-ways-1.aspx.

[92.] Svensson, Peter, "Possible E-mail Theft from Epsilon Slams Banks, Retailers," *USA Today*, April 4, 2011, http://www.usatoday.com/money/industries/technology/2011-04-04-e-mail-theft-phishing.htm.
[93.] Swallow, Kevin, "Kinishi Launches Fingerprint ID Telematics for Drivers," February 26, 2010, http://www.roadtransport.com/Articles/2010/02/26/135661/Kinishi-launches-fingerprint-ID-telematics-for-drivers.htm.
[94.] Telecommunications Systems, Inc., *Form 10-K: Annual Report Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934*, March 3, 2009, http://secwatch.com/tsys/10k/annual-report/2009/3/3/2003794/print.
[95.] Towers Watson, "Towers Watson DriveAbility(SM): UBI Taking Hold Among Auto Insurers," *Market Watch*, June 22, 2011, http://www.marketwatch.com/story/towers-watson-driveabilitysm-ubi-taking-hold-among-auto-insurers-2011-06-22.
[96.] U.S. Department of Transportation, Federal Motor Carrier Safety Administration, Office of Analysis, Research, and Technology, "100 Car Naturalistic Study," March 2006, http://www.fmcsa.dot.gov/facts-research/research-technology/report/100-car-naturalistic-study/100-car-naturalistic-study.pdf.
[97.] U.S. Department of Transportation, Intelligent Transportation Systems, Next Generation 9-1-1 (NG9-1-1) System Initiative, "NG9-1-1 Transition Issues Report," February 2008, http://www.its.dot.gov/ng911/pdf/NG911_TransitionIssuesReport_FINAL_v1.0.pdf.
[98.] U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, "Table 1-11: Number of U.S. Aircraft, Vehicles, Vessels, and Other Conveyances," National Transportation Statistics, 2011, http://www.bts.gov/publications/national_transportation_statistics/html/table_01_11.html.
[99.] Vanderford, Richard, "Progressive Sues HTI Over Car Monitoring Patents," *Law360*, September 16, 2010, http://www.law360.com/articles/194102/progressive-sues-hti-over-car-monitoring-patents,
[100.] Veness, Chris, "Calculate Distance, Bearing, and More Between Latitude/Longitude Points," not dated, http://www.movable-type.co.uk/scripts/latlong.html.
[101.] Verisk Analytics, "ISO Equips Fleet with Telematics Devices to Develop Analytical Methods for Personal and Commercial Auto Insurance," September 21, 2010, http://www.verisk.com/Press-Releases/2010/ISO-Equips-Fleet-with-Telematics-Devices.html,
[102.] Verisk Analytics, "Will Consumers Agree to Install Telematics Devices in Their Cars?," October 2010, http://www.verisk.com/dloads/applied-informatix/verisk-telematics.pdf.
[103.] Walsh, James, "The Best Insurance—Data Backup," Wealth Mountains, not dated, http://www.wealthmountains.com/articles/Article/The-Best-Insurance---Data-Backup/2189.
[104.] Werner, Geoff, and Claudine Modlin, *Basic Ratemaking* (Version 4), Casualty Actuarial Society 2010, http://www.casact.org/pubs/Werner_Modlin_Ratemaking.pdf.
[105.] Ykovv, Liane, "AAA to Offer Customers In-Drive Emergency Response Telematics," August 15, 2011, http://reviews.cnet.com/8301-13746_7-20092469-48/aaa-to-offer-customers-in-drive-emergency-response-telematics/.
[106.] Ziegler, Chris, "2G, 3G, 4G, and Everything in Between: An Engadget Wireless Primer," January 17, 2011, http://www.engadget.com/2011/01/17/2g-3g-4g-and-everything-in-between-an-engadget-wireless-prim/.

Abbreviations and notations

AAA, Automobile Association of America Insurance Company
ADM, Actuarial Data Manager
AM, Ante Meridiem
ASB, Actuarial Standards Board
ASOP, Actuarial Standard of Practice
DBD, Driving Behavior Data
EDR, Event Data Recorder
FICO, Fair Isaac Corporation
FNOL, First Notice of Loss
G-Force, Gravitational Force
GM General Motors
GMAC, General Motors Acceptance Corporation Insurance Company

GPS, Global Positioning System
GPRS, General Packet Radio Service
HVTPE, High-Volume Transactional Processing Environment
IP, Intellectual Property
ISO, Insurance Services Office, Inc.
MPH, Miles Per Hour
NHTSA, National Highway Transportation Safety Administration
NPE, Non-Practicing Entity
OBDII, On-Board Diagnostics II
RAM, Random Access Memory
TSP, Telematics Services Provider
UBI, Usage-Based Automobile Insurance
UTC, Coordinated Universal Time
Wi-Fi, Wireless Fidelity

Biographies of the Authors

Jim Weiss is an Assistant Manager in the Personal Automobile Actuarial Division at Insurance Services Office, Inc.. He works with ISO's Applied Informatix™ business unit in developing analytical tools to support telematics-based initiatives, and moderated and co-presented a session entitled "Effectively Utilizing Vehicle Telematics Data" at the Casualty Actuarial Society Special Interest Seminar on "Cutting Edge Tools for Pricing and Underwriting" in October 2011. Jim is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and a Chartered Property Casualty Underwriter.

Jared Smollik is a Manager in the Increased Limits & Rating Plans Division at Insurance Services Office, Inc.. He is responsible for increased limits for commercial and personal auto, as well as the development of increased limits and other rating factors for the management protection and e-commerce programs. He is also responsible for ISO's Enterprise Risk Management Service for Insurers and has been involved with ISO's telematics research and product development since 2005. Jared is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries and a Chartered Property Casualty Underwriter.

How Individuals Purchase Insurance: Going Beyond Expected Utility Theory

Marc-André Desrosiers, Ph.D. Candidate, FCAS, MBA, BA

Abstract: For insurers to be successful in the long run, they need to put forward an attractive value proposition for insureds and price (sufficiently) for it. To facilitate this, it is best that insurers deepen their understanding of consumer behavior in situations that involve risk. This paper is intended as a survey of the developing economic literature concerning how individuals make choices in the face of risk. It will be shown and illustrated that there are primitives that drive the choice behavior of individuals faced with risk. With understanding of these primitives in mind, a practicing actuary should be able to rationalize observed portfolio profitability or unprofitability, as well as anticipate the effect of some pricing and product changes on the profitability of the affected portfolios.

Keywords: Behavioral Economics; Expected Utility Theory; Committed Consumption; Loss Aversion; Prospect Theory; Consolation Hypothesis

1. INTRODUCTION: THE NEED TO GO BEYOND EXPECTED UTILITY THEORY

To ensure their long-term survival, insurers must assure themselves that they continuously strive to respond to evolving demands of insurance consumers and their related stakeholders. For insurers, this focus on satisfying customer needs and wants is doubly important as insureds are the main source of capital for insurance companies, being the biggest contributor of debt financing (reserves) and a major contributor of shareholder equity (underwriting profit). A sign of potential for long-term survival of the insurer is sustainable and sustained above average profitable growth. Profitable growth can be thought of as the result of a process where, over time, the insurer puts forward an attractive value proposition and also charges enough for it to be profitable. The attractive value proposition leads to the insurer being able to sustain organic growth, or, because an insurer with an attractive value proposition can better leverage a portfolio being acquired, growth by acquisition. This double focus on sustainable growth and pricing leads insurers to consider insurance consumer behavior, so as to make the value proposition as relevant as possible and to capture, in premiums, as much of the consumer surplus as possible.

Over the last few decades, behavioral economists, both theorists and empirical researchers, have made significant headway into the problem of explaining how individuals make purchasing decisions in insurance and, more generally, in risky situations. Their theories take us beyond expected utility theory and move us into a realm of theories that better reflect the context of our daily lives and of our psychology.

Having a theory, or a consistent set of theories, of consumer behavior is important for the actuary as, without it, predictions of the effect of a supply policy change would become, at best, guess work that could easily lead the insurer astray. If the actuary does not rely on a set of behavioral primitives,[1] the actuary will be left with treating all choices with regards to insurance as entirely context dependent. If every individual's decision making with regard to insurance choice were entirely context dependent, it would be impossible to prepare forecasts of the actual effect of, for example, a rate change or the imposition of a minimum deductible.

This paper is intended to provide actuarial practitioners with a survey of recent developments in behavioral economics. In Section 2, we will start by reviewing the traditional argument for why individuals value insurance: the transfer of sizable risk and the associated prospective pricing. In Section 3, we will then show that the traditional expected utility framework cannot explain risk aversion for modest risk. In Section 4, a first alternative theory will then be explored: consumption commitments make individuals more risk averse over moderately sized downside risk gambles than would have been predicted by the expected utility framework. In Section 5, a second alternative set of theories will then be explored. All theories center on the understanding of loss aversion. We will need to explain violations of asset integration, the different nature of preferences over gains and losses, decision weights and other probability distortions, diminishing sensitivity to gains and losses, and reference dependence. In Section 6, we will then explore a third theory to explain risk aversion that goes beyond that predicted by expected utility theory: an individual's insurance choices are modified by whether or not he likes the "objects" being insured. The heart of that theory is the consolation hypothesis that implies that insurance indemnification is also seen as a consolation for the loss of the appreciated "object." In Section 7, we will also share the results of an empirical study that has found that individual insurance choices are highly correlated across similar coverages. Finally, in the Appendix, we will share some thoughts about how to conduct private research relating to insurance consumer behavior.

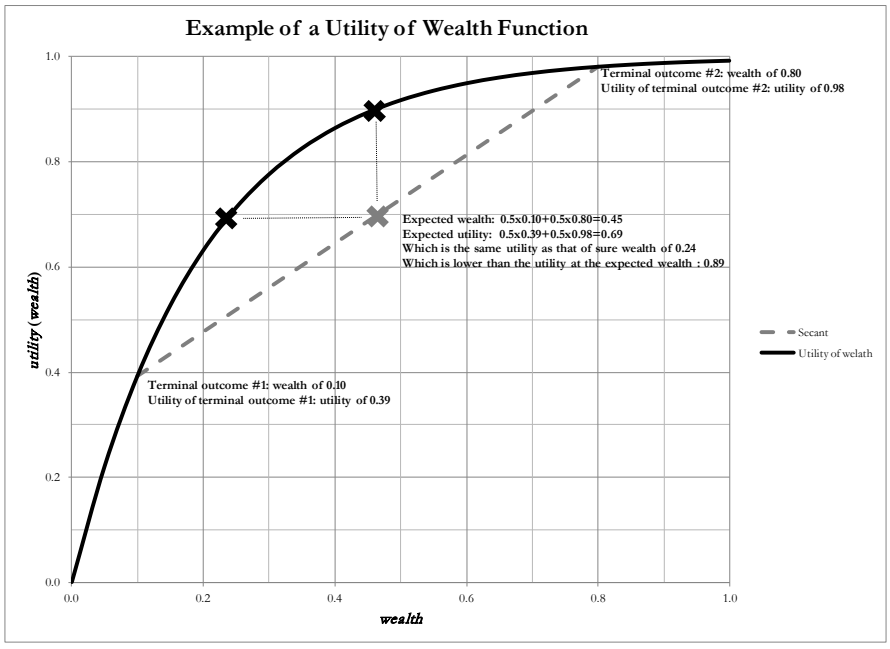
2. RISK TRANSFER, PROSPECTIVE PRICING, AND EXPECTED UTILITY

The most basic argument for why insurance is valuable to individuals comes from the risk transfer associated with prospective rating, assuming that individuals have decreasing marginal utility of wealth and maximize their expected utility.

Let's create a simplified scenario. Just as in the presentation of utility theory in *Actuarial Mathematics* (Bowers, et al. 1997), let's consider the case where an insured is facing a 1% probability

[1] By behavioral primitives, we mean context invariant components of behaviors, especially those elements relating to preferences and perceptions in risky situations.

of having an accident that will cost him $1,000,000 or a 99% probability of suffering no loss at all. The expected outcome for the insured is to suffer a $10,000 loss; that is, the $10,000 expected loss is calculated as 0.01 x $1,000,000 + 0.99 x $0. From experience, we know that many individuals are willing to pay more than $10,000 for this coverage. It must be the case that they are considering more than the expected value to make their decision. For example, they may be entertaining the consequences of facing the $1,000,000 loss and thinking of it as catastrophic. We can suppose that individuals have preferences over wealth that can be expressed as utilities. Generally, we demand preferences to be complete,[2] transitive,[3] and consistent.[4] For preferences relating to financial outcomes, we can always treat one wealth as having a utility of zero and another as having a utility of one, as utility-of-wealth functions are defined uniquely up to a linear transformation $u^*(w) = au(w) + b$, for $a > 0$. To discover the exact shape of the utility function, we can arbitrarily set $u(w_0) = 0$ and $u(w_1) = 1$, for some w_0 and w_1. As we generally think of people as preferring more wealth to less, w_1 should be greater than w_0. We can then ask people to provide us with the certain wealth w^* that would make them indifferent between having that wealth for sure and facing a $p\%/\ (1-p\%)$ chance of having w_0 or w_1, expressed as: $u(w^*) = p \cdot u(w_0) + (1-p) \cdot u(w_1)$. With enough of these points (by varying *p*), we can construct the utility of wealth function for the person. It may look like this.



[2] Completeness relates to the requirement that people can always say if they prefer or are indifferent to two outcomes they are presented with.

[3] Transitivity relates to the requirement that a person cannot have the following set of preferences: (1) preferring outcome *x* to outcome *y*, (2) preferring outcome *y* to outcome *z*, and (3) preferring outcome *z* to outcome *x*.

[4] In identical situations, a person cannot prefer an outcome *x* certain times and outcome *y* other times.

In the example above, the person displays decreasing marginal utility of wealth. We will call that person risk averse, because the person has a willingness-to-pay to eliminate the downside risk. Using Jensen's inequality, it can be generally proven that a person who displays decreasing marginal utility of wealth will always have a positive willingness-to-pay for insurance and a negative willingness-to-pay for gambling. We can think of the willingness-to-pay for insurance, or risk premium, as the vertical distance between the expected utility of wealth line (solid line) and the expected wealth line (dashed line). If we measure the risk premium this way, we are measuring it in units of utility. If the risk premium is measured as the horizontal distance between the expected wealth line and the utility of wealth line, then it is measured in units of wealth.

This model helps rationalize the demand for risks that create catastrophic financial consequences, such as fire insurance. The following are necessary to determine the willingness-to-pay for insurance in this model: (1) initial wealth (a richer person should be more risk tolerant), (2) frequency (the higher the probability of loss, the greater the risk premium), (3) severity (the higher the severity of loss, the greater the risk premium), (4) risk aversion, i.e., the concavity of the utility of wealth function (the more risk averse the person, the greater the risk premium). A consequence of this model is that the upper layer of coverage should always be valued more by the insured than the lower layers of coverage. It is important to note that the coverage can only be valuable to the insured under the condition that the insured can replace the risky prospect of a loss with a certain prospect of paying a premium for insurance. This can be thought of as prospective pricing: when the premium for insurance is determined prior to the period of time when the insured is exposed to loss. In this case, the coverage should be purchased if the price is less than the expected value of the coverage plus the risk premium.

If pricing was retrospective, that is, the insurer charged the total amount of the claims that occurred during the period (possibly taking into account the time value of money) at the end of the period, the insurance coverage would become, at best, a contingent financing agreement, and the risk transfer features of the insurance contract would be greatly diluted.

3. WHY EXPECTED UTILITY THEORY CANNOT EXPLAIN RISK AVERSION FOR MODERATE SIZED RISK

Unfortunately, expected utility theory,[5] with the associated decreasing marginal utility of wealth,[6]

[5] The usual tenets of expected utility theory include (1) linearity in the probabilities (that is, expected utility is the probability weighted average of the utility of the different possible outcomes), (2) asset integration (that is, the considered utilities are the utilities of terminal wealth), and (3) risk aversion (in this context, this refers to decreasing marginal utility of wealth).

[6] Coined risk aversion in this model.

cannot provide a good explanation for why individuals purchase moderate insurance. In fact, it can explain even less why insureds purchase small scale insurance. Suppose that the only reason that individuals purchased insurance is because pricing is prospective, individuals attempt to maximize their utility and individuals have decreasing marginal utility of wealth; then, the risk premium (the amount over zero profit they are willing to pay for coverage) they should be willing to pay for a small risk should be proportionately small compared to the risk premium they should be willing to pay for a substantial risk. In other words, we expect that the marginal loss ratio for the higher layers to be lower than that of the lower layers of coverage, as the higher layers of coverage are more valuable to the insured. Yet, Sydnor (2010) finds that insureds are willing to pay about $100 to decrease their deductible from $1,000 to $500, when the expected loss cost in the layer is about $20. The proportional risk premium in the layer is about 80% ((100 – 20)/100) and this is certainly much greater than the proportional risk premium people are willing to pay for more substantial coverage,[7] considering that the Personal Lines insurance market is roughly competitive.[8]

As Rabin (2000) puts it, "[a]ny utility of wealth function that doesn't predict absurdly severe risk aversion over very large stakes predicts negligible risk aversion over modest stakes." Given the empirical findings that individuals are willing to pay substantial risk premium for modestly sized insurance coverage, there must be other forces than decreasing marginal utility of wealth at work to explain why individuals find value in their insurance coverage.

4. CONSUMPTION COMMITMENTS CHANGE OUR DEMAND FOR INSURANCE

Chetty and Szeidl (2007) have demonstrated that, when individuals have committed consumptions[9] that are costly to adjust, they will exhibit risk aversion for moderate stakes downside risk that goes beyond that predicted by decreasing marginal utility of wealth. With committed consumption and costly adjustments in committed consumption, moderately sized adverse shock to income is absorbed only in non-committed consumption and the welfare loss associated with a moderately sized adverse shock to income is greater than in a setting where all consumption can be costlessly adjusted.[10]

Before going further, let's provide examples of committed consumption and adjustment expenses

[7] By this, we mean the upper layers of coverage.
[8] While there may be supply policy reasons for why the pricing the insurer is offering has the peculiarity that the marginal loss ratio for upper layers is higher than that of lower layers, it does not remove the mystery, from the point of view of expected utility theory, that individuals should actually pay a higher proportional risk premium in the lower layers than in the higher layers. Supply policy reasons for why an insurer may offer such varying prices by layer may have to do with risk signalling through deductible choice, competitive pressures, etc.
[9] They define "committed consumptions" as goods that involve transaction costs that are infrequently adjusted.
[10] Like in expected utility theory.

related to committed consumption. A simple way to classify committed consumption is consumption that is infrequently adjusted downward. Examples of committed consumption would then include housing, cars, furniture and, to a lesser extent, leasing arrangements and health care spending.[11] If we further analyze the case of housing, the costs associated with committed consumption may be better exemplified. Moving is associated with large transaction costs, such as broker fees, monetary and utility costs of moving, and potential capital loss associated with re-sale.[12] We can also think about non-committed consumption as consumption the level of which changes most in adverse income shocks. With that view in mind, it makes sense to treat food and entertainment consumption as non-committed.

Chetty and Szeidl also find that borrowing constraints make individuals appear even more risk averse over moderate downside risk. With borrowing constraints, adjustments of non-committed consumption have to be even more dramatic in moderately sized adverse shocks. To contrast, a person with access to credit may use credit to smooth over a dry spell by using credit as a temporary source of funds.

In many ways, Chetty and Szeidl provide us with a way to explain the results generally associated with good (bad) credit scores in Personal Lines insurance. Suppose that an individual has consumption commitments and has decreasing marginal utility of wealth; then, other things being equal, the individual should take more precautions to avoid positions in which they will not be able to satisfy those commitments. Therefore, that individual is more likely to take more precautions to avoid a loss, assuming that the loss is not fully insured. Also, other things being equal, the individual is more likely to value insurance coverage more than an uncommitted individual or an individual that is not risk averse. This leads to the double prediction that individuals that deal better with their commitments (commonly exemplified by their having a better credit score) should have a lower loss cost (holding constant other risk characteristics) and a lower loss ratio (as, other things being equal, the risk premium they are willing to pay should be higher). While having commitments does not automatically lead to people handling them well, those individuals that started out as more risk averse before they took on commitments should see their risk aversion magnified because of their commitments, be extra motivated to avoid losses, and have increased willingness to pay for insurance.

Another relevant prediction of Chetty and Szeidl's theory is that: "commitments create a force toward providing more insurance for short-term, moderate-stake shocks relative to long-term welfare programs" (Chetty and Szeidl 2007, 861). The P&C analog of that prediction is that the

[11] More generally, [non-small] contracts that include penalties for early termination.
[12] In a soft market, at least.

relative willingness to pay for catastrophic coverage (like fire coverage) may be lower than the willingness to pay for more modest coverage (like crime coverage).

5. LOSS AVERSION AS A DRIVING FORCE FOR BUYING SMALL SCALE INSURANCE

The consumption commitment theory for why individuals display higher than expected[13] willingness to pay for insurance was built without ever appealing to the particular psychology of individuals. With Prospect Theory, developed in *Prospect Theory: An Analysis of Decisions under Risk* (Kahneman and Tversky, 1979), it becomes necessary to delve into the psychology of gain/loss perception. Prospect Theory is the first theory that provided an explanation of the rationality of small scale insurance purchasing at a premium for individuals.

5.1 "A Bird in the Hand Is Worth Two in the Bush"

One of the underlying assumptions of expected utility theory is that individuals consider risks by examining their net wealth given the risk; that is, they integrate the risk to their assets. This is commonly called asset integration. One of the first psychological phenomena that Kahneman and Tversky demonstrated was that people don't always consider gambles using asset integration. The flip side of that statement is that individuals very often consider gambles using a gain/loss perspective. For the moment, let's not address the issue of how the zero point is determined in the individual's psychology[14] and focus on the differential treatment of gains and losses in terms of the individual's preferences. With many, easily repeatable experiments, they have been able to identify that individuals are about twice[15] as sensitive to losses as they are to gains: the expression "a bird in the hand is worth two in the bush" captures that phenomenon well.

The preceding phenomenon could help to explain why individuals would be willing to pay to eliminate their deductibles. Like Johnson, et al. (1993) note, insureds can react quite strongly to mandatory increases in their deductibles. They explain this phenomenon as follows: insureds perceive the deductible payment as a loss to which they are quite sensitive. The product design that they propose to help insureds avoid feeling the loss associated with deductible payments and yet avoid moral hazard is to offer rebates; that is, they propose to incorporate the deductible charge in the premium while at the same time offering insurance rebates to insureds that remain claim-free.[16]

[13] Under expected utility theory.
[14] The treatment of the determination of the zero point is left to Section 5.4.
[15] 2.25 in many calibrations.
[16] Said otherwise, the design they propose is to offer clients a policy with no deductible. Compared to usual policies that incorporate a deductible, these policies will be surcharged. To maintain incentive compatibility and to keep the client

Unfortunately, there has been little research to attempt to explain why it is that individuals are loss averse. Clearly insurance product design and pricing can exploit the phenomenon, but, until such time as the effect is understood, it will make it difficult for insurance practioners to fully exploit the opportunities associated with loss aversion.

5.2 Decision Weights as Opposed to Probabilities

While in expected utility theory the expected utility is computed using probability weights, Kahneman and Tversky have found that people generally overweight very low probabilities. In essence, they have found that the subjective value of gain/loss prospects are not weighted together using probability weights, but rather with decision weights that are greater than probability weights when the probabilities involved are small, and less than probability weights when the probabilities involved are large. When individuals are considering loss prospects with a low probability of occurrence,[17] they tend to attach a decision weight that is greater than the probability weight, and thus exhibit a willingness to pay a risk premium for insurance. This can help explain why the risk premium associated with coverages with very low frequency can be higher than the risk premium associated with coverages that have much greater frequency (e.g., individual theft insurance vs. auto collision coverage).

Yet, it is important to distinguish between decision weights and probability mis-estimation. One example of a heuristic that individuals apply, and presumably actuaries too, leading them to incorrect probability assessments, is when the representativeness heuristic leads to misconception of chance biases such as the gambler's fallacy.[18] The insurance equivalent of the gambler's fallacy is when an insured, or an actuary, effectively assumes that an insurable event will occur because the last insurable event was a long time ago. If there is evidence that the processes leading to insurable events are memoryless (like the claim count being appropriately modeled by a Poisson or an over-dispersed Poisson random variable), the time since the last insurable event doesn't influence when the next event will occur. In this case, the heuristic can be thought of as the person starting from an

---

prudent before and after losses, the policy would offer a discount for those insureds that remain claim-free. One design for the discount that the authors mention involves providing a refund to clients that didn't make a claim, a form of retrospective rating. If the discount is carried forward to the next contract (like a claim-free discount), presumably the discount would be greater for the policy without the deductible than it is for the policies that have a deductible.

[17] A low probability of occurrence generally corresponds to probabilities less than 10%.

[18] The following Wikipedia source conveniently describes what I refer to as the gambler's fallacy. "The gambler's fallacy (...) is the belief that if deviations from expected behavior are observed in repeated independent trials of some random process, future deviations in the opposite direction are then more likely. (...) The gambler's fallacy implicitly involves an assertion of negative correlation between trials of the random process, and, therefore, involves a denial of the ex-changeability of outcomes of the random process. In other words, one implicitly assigns a higher chance of occurrence to an event even though from the point of view of nature or the experiment, all such events are equally probable (or distributed in a known way)." (http://en.wikipedia.org/wiki/Gambler%27s_Fallacy).

estimate of the long-run frequency of insurable events, and thinking about when the next event needs to occur for the observed claiming process to follow its long-term average. If the time between insurable events is exponentially distributed, then the waiting time is not influenced by the time since the last event, and the reasoning leads the person to the wrong conclusion. The person is putting too much emphasis on the representativeness of the long-term average. Other examples of common mis-estimation of probabilities have also been documented and the footnote below indicates a source of information on that topic. [19]

For coverages where individuals may have some difficulty in forming an accurate assessment of their level of risk (e.g., fire peril, frill-type coverages, etc.), probability mis-estimation could play a significant role in explaining why individuals are willing to pay a risk premium. In the case of rare events, the insurer is in a much better position to evaluate the likelihood of losses than the insured is. For example, Gallagher (2010) finds that the flood insurance take-up rate materially increases after a flood occurs in a community. Assuming that the long-term probabilities of flood occurrence are relatively constant,[20] the availability bias[21] could push up their risk premium and induce them to purchase insurance. (The biases arising out of the availability heuristics may be difficult to isolate from another psychological effect that we'll explore in Section 6: individuals tend to have a higher willingness to pay for insurance for objects they like than for objects they don't care for.)

5.3 Gambling and Diminishing Sensitivity to Losses

According to Prospect Theory, the utility of gains increases at a decreasing rate and similarly for losses; in effect, the theory postulates that people become gradually less sensitive to gains or losses as they become greater in absolute value. The theory assumes that our perception of gains/losses functions like our senses in that it becomes more difficult to distinguish between values as the magnitude of those values increases.

With this assumption, we can attempt to explain the "long-shot bias" in end-of-the-day betting when a gambler has been losing overall in a day. Suppose that an individual's reference wealth is morning wealth. As the individual becomes less sensitive to losses as they grow bigger, the

[19] In *Judgment under Uncertainty: Heuristics and Biases* (Tversky and Kahneman 1974), the authors explore three families of heuristics that tend to induce individuals to mis-estimate probabilities:

(1) the representativeness heuristic, exemplified by the following biases:
    (a) insensitivity to prior probability of outcome, (b) insensitivity to sample size, (c) misconception of chance, e.g., gambler's fallacy, (d) insensitivity to predictability, (e) illusion of validity, and (f) misconception of regression [towards the mean];
(2) availability heuristic, exemplified by the following biases:
    (a) biases due to the retrievability of instances, (b) biases of imaginability, and (c) illusory correlation; and
(3) adjustment and anchoring heuristic, exemplified by the following biases:
    (a) insufficient adjustment, (b) biases in the evaluation of conjunctive and disjunctive events, and (c) anchoring in the assessment of subjective probability distribution.

[20] Thus, even if people undertook Bayesian updating of their probability assessment, their probability assessments could not change so much.

[21] An occurrence that they can easily retrieve and imagine.

possibility of finishing the day "in the black" outweighs the potential cost of finishing the day deeper "in the red."[22] The flipside of this predicted phenomenon is that the theory should also predict risk seeking for small-scale downside risk. This is a prediction that insureds may not take up coverages that are sold at a loss for the insurer. This phenomenon may help rationalize why people who suddenly became less wealthy may forego purchasing insurance in some situations. If a person already suffered a disastrous loss, then paying an insurance premium to protect against a not-so-much-more-catastrophic situation may be unattractive. However, as we will see next, further reflection about how the reference point is set is required to rationalize small scale insurance purchasing.

5.4 Not All Money Spent Is Perceived as a Loss

In *Reference Dependent Risk Attitudes* (Koszegi and Rabin 2007), a theory is developed to help us predict the reference point against which gains and losses are measured. The authors propose that different types of situations command different ways of setting the reference point for individuals.

In the case of a surprise choice (that is, a choice that an individual needs to make but where the individual did not anticipate having to make that choice), a person will appear risk neutral if the risk is small and the person is already endowed with a significant amount of risk. Said otherwise, if an individual is given the possibility of insuring on a small scale when the person is already facing significant risk, the person will not be willing to pay a risk premium to insure, if the choice of purchasing insurance comes as a surprise. For example, suppose an individual went to a ski resort, rented skis and was offered the possibility to insure the skis against responsible damages when the person wasn't aware that choice was going to be offered. The theory predicts that the willingness to pay a risk premium for insurance would be next to nil. In effect, what is happening is that the person does not have the time to form plans that influence the formation of the reference point and thus is led to use the current wealth as the reference point. When that happens and the person is already endowed with risk and when the marginal risk is small, the risk is dwarfed by the existing risk, and, for the person to be attracted to purchase insurance, the insurance has to be fair or favorable.

When the person envisions purchasing insurance for a risk, even if endowed with risk, the person should be willing to pay a higher risk premium than when the insurance purchasing option comes as a surprise. The individual has the chance to anticipate the gain/loss "sensation" that will arise if the individual purchases (or not) insurance contingent on a loss happening. In that anticipation, the individual has the chance to consider the case where a loss happens and insurance was not

[22] The expression "in the red" refers to the individual having suffered a net loss; while, the expression "in the black" refers to the individual making a net gain.

purchased, using the initial wealth diminished by the premium as a reference point. In that situation, the prospective insured should be comparing the "sensation" of gain that he could experience when a loss doesn't happen but he paid an insurance premium to the "sensation" of loss he would experience when a loss event occurs but when he didn't insure. Given the greater sensitiveness of potential insureds to losses than to gains, this induces the potential insured to be willing to pay a risk premium to insure. A way to think about the above prediction is that the willingness to pay for small-scale insurance should be greater when the individual can foresee the availability of the coverage. This aspect of the theory can help rationalize cell phone insurance, which is admittedly small scale.[23]

People can appear even more risk averse when insureds evaluate starting from a situation where they have coverage, whether or not they would prefer to not be covered. In that case, an insured is left to compare whether he prefers a world where he isn't covered regardless of a loss happening and a world where he is covered. The narrowing of the considered scenarios with insurance makes the individual prefer to be covered over a greater range of prices: that is, the willingness to pay for insurance appears greater in that case. Thus, the title of this sub-section: "not all money spent is perceived as a loss," and this could help rationalize why existing customers are more willing to accept higher prices than customers who are actively shopping.

6. INCREASED WILLINGNESS TO PAY TO INSURE AN OBJECT WE LIKE

Hsee and Kunreuther (2000) further explored the psychology underlying insurance purchasing. What they found is that there are other factors, other than *monetary* factors, that affect the way we purchase insurance and make a claim when we suffer a loss. For one, they find that individuals are more likely to make a claim if they feel that the party that insures them has wronged them. This can be thought of as the reprisal motive for claiming. This phenomenon should normally not affect insurers, as they are generally not responsible for causing the insurable event, but it may lead to claim inflation if clients are dissatisfied with the service of the insurer, prior to or during the claim settlement. Interestingly, in the claim filing process, the way the coverage is framed in the insured's mind affects the willingness to claim. "If the money is construed as compensation for the lost object, then his willingness to collect the money will depend on his affection for the object. If the money is construed as unrelated to the lost object, then his willingness to collect the money will be independent of his affection" (Hsee and Kunreuther 2000, 146). They provide an example where the compensation for an appreciated object takes the form of a payment from the insurer or a discount on an unrelated object: people were less likely to make a claim if the compensation was a discount

[23] At the very least compared to homeowners or car insurance.

on an unrelated object.

More importantly, the authors found that, when individuals experience affection for the objects insured, controlling for actual and perceived market value, the risk premium that individuals are willing to pay to insure the objects increases. Now, insurers are unlikely to be able to influence the affection an insured has for the insured goods, but the insurer may be able to anticipate the attachment an insured has for different goods insured and adjust pricing so as to capitalize on the increased willingness to pay for insurance. For example, it might be that people that have red cars care more for their cars than people that have blue cars. By collecting the car color information, the insurer could adjust pricing accordingly.

Returning to the case of increased flood insurance purchasing after a flood, the affection theory for increased willingness to pay for insurance can also help explain why flood insurance take-up increases after a disaster: "this occurs because people who have just experienced a disaster know what it feels like to have lost things they love and want to avoid some of the pain by being protected in the future." (Hsee and Kunreuther 2000, 154). Another way to think about this is to suppose that we place insureds in one of two situations: (1) they face a purely financial risk with very low probability of occurrence such that, when an event occurs, insureds are reminded that the risk exists and suddenly become more willing to pay for coverage, and (2) they face flooding risk. The presumption is that the insureds in the second category will be more willing to pay for coverage and to make a claim when a flooding event occurs, because they care more about their house and belongings than about a purely financial interest. A related prediction is that individuals are more likely to file a claim for losses just above a deductible for insurable interests to which they feel a connection.

Another example of coverage choice that can be best rationalized using the consolation hypothesis is the purchase of large-scale life insurance on young children, as young children are not income earners whose salaries need to be replaced at their death. The consolation hypothesis may also be exploited in P&C insurance marketing. For example, an advertising campaign may thoughtfully illustrate situations where people are surrounded by goods and people they care for, and are reminded about how their life would be if an insurer were not there to help them get back on their feet. This can be contrasted with a marketing campaign that does not directly or indirectly appeal to the affection to the insured objects, such as a marketing campaign that would only be aimed at establishing name recognition and brand identity.

7. HOW THE WILLINGNESS TO PAY FOR A COVERAGE IS CORRELATED WITH OTHER COVERAGES

Before concluding, we would like to share the results of a study that finds that willingness to pay for insurance tends to be correlated across coverages. While the study focused on private health and disability insurance and pension choices of individuals, there is no a priori reason to believe that the effect is not present in other insurance markets. In particular, "[the study] find[s] (…) that one's choices in other insurance domains are substantially more predictive of one's choice in a given insurance domain than either one's detailed demographic characteristics or one's claims experience in that domain" (Einav, et al. 2010, 1)

The findings of this study are consistent with the idea that there exists some behavioral primitive that allows for the forecasting of insured behavior in a given context, using information developed in another context. It is important to note that the finding is not that people's behaviors in insurance choices are entirely context insensitive, but that there seems to be some context invariant component to insurance choice behavior.

Unfortunately, at this point in time, an equivalent study has not been conducted for P&C Personal Lines insurance. That being said, the finding provides insights about thoughtful elements that could be included in a well-designed insurer database. If the result also applies in the P&C world, then the risk premium that people are willing to pay for their auto insurance should be correlated with the risk premium they are willing to pay for their homeowners' or renters' policies. Insurers will not be able to measure and capitalize on that phenomenon unless they build client databases that allow them to connect the auto policy with the other policies of the client.

8. CONCLUSION

We've explored recent developments in behavioral economics and attempted to show that these developments could be instrumental in assisting the actuary in forming more reliable theories about how insurance consumers could react to changes in the supply policy. We have also attempted to demonstrate how a solid understanding of behavioral economics could help the actuary rationalize observed insured behavior, such as in portfolio profitability/unprofitability, claim reporting patterns, reaction to marketing campaigns, etc. Clearly, private research will need to take place to allow actuaries in the industry to refine the basic theories presented here, as well as to develop new ones. It is best not to underestimate the R&D challenge of deepening the actuarial understanding of consumer behavior, but it is also important to keep in mind that actuaries have access to highly relevant data that is generally not available to public researchers: field actuaries have a sizable

competitive advantage over their academic counterparts to explore the behavior of consumers but, more importantly, to operationalize these findings into product design, marketing campaigns, strategic pricing, etc.

Acknowledgements

The author wishes to thank Prof. Justin Sydnor for his assistance in this project, as well as the feedback and efforts of the Call Paper Committee members.

APPENDIX: TYPES OF DATA AND DATA ELEMENTS

In this appendix, we will share some thoughts about how to conduct private research relating to insurance consumer behavior. We will share some considerations relating to potentially useful data elements for the actuary to gather.

- When the actuary is attempting to understand consumer behavior, what sources of information can the actuary attempt to access? The information can be quantitative or qualitative. The information can come directly from clients or be obtained indirectly (e.g., through brokers, agents, or underwriters). The data can be a sample or cover the entire population. The data can be generated in-house or it can be outsourced. The possibilities generate a grid of possibilities all of which have advantages and disadvantages.

  For example, the actuary can gather the comments of a regional branch manager that regularly discusses on-going issues with brokers. While the gathered information may not be as reliable as when gathered from consumers, information obtained in this manner tends to be obtained at low marginal cost. Another example is when a direct writer directly surveys its clients. The campaign could become quite expensive, but this would allow the insurer to obtain direct feedback from its clients. In this example, the survey campaign could be done in-house or be outsourced to a marketing firm.

Fortunately, the actuary can begin analyzing consumer behavior using information which is generally readily available. One potential starting point for a consumer behavior analysis can be common statistics like the retention ratio, the new business ratio, closing ratios and quote activity. When analyzing these ratios, it is important for the actuary to take into account the effect of confounding factors. An example of this is an increase in quote activity can arise due to a competitor's rate becoming less attractive or because the insurer has undertaken a marketing campaign to attract quotes. When analyzing these ratios, controlling for seasonality and trends is crucial for the identification of action-reaction effect of a supply policy change, as the ratios of interest could be changing without the supply policy changing. A key family of confounding factors that arise when examining retention relates to nonsupply-related reasons for why the insured hasn't renewed, e.g., they have ceased to exist, they do not have an insurable interest anymore, they lost access to their agent/broker, the product/service/experience does not meet their needs/expectations, etc.

A series of questions could guide the actuary in deciding which variables in the available data could be used for further exploration. In no particular order these questions are:

- Who's the client? Who decides? Who pays? Who influences the client?
    - Is it the head of a household? Is it a property manager?

  The person who decides may not be the person who has an insurable interest.
- What is the customer's level of risk aversion?
    - Is the customer willing to assume more of the risk to reduce the premium involved (e.g., high deductible policies)?
    - It is worthwhile to keep in mind the different variations upon the theme of risk aversion that we have explored in Sections 1 through 6.
    - Is the client more likely to take risk (as can be suspected by known information)? Does the client have a history of accidents and violations? Are there coverages required by the client that entail higher risk-taking behaviors (like insuring a motorcycle)?
- Is the customer "naturally" price sensitive?
    - For example, is it the case that the insured is already near bankruptcy and attempting to save every penny on insurance purchasing?
    - Is the industry of the insured in danger (e.g., small farms)?
    - Are economic conditions unfavorable to the insured?
    - Is the client showing signs of financial difficulty to the insurer? Are there many reinstatements or mid-term transactions? Did the client miss payments with the insurer already?
    - Is the client selecting coverages/options in a pattern that indicated higher price-sensitivity (like purchasing reduced limits, increased spreading of payments, etc.)?
- What are the insurance alternatives available to the customer? What are the substitutes to insuring with you available to the client?
    - Does the insured have a history of [relevant] claims?
- Is the decision emotional? Automatic? Rational?
    - Will the customer move for a small premium increase?
    - Is the customer unwilling to shop around unless a problem arises?
    - Will the customer explore thoroughly all the available alternatives at each renewal?
    - What is the dollar amount in play? What is the relative amount in play compared to the customer's revenues?
    - Are there related contracts in play also?
- How valuable are services, extra protection, etc. to the customer? Is the comparison of value between your products/services/experiences and those of alternatives difficult to do for the client?

- Are there signs that the client sees great lifetime value in its relationship with the insurer? How long has the client been with the insurer? What are the costs for the client to switch insurers?
- How much money is the client already spending with you (in $ or in %)?
- Does your pricing appear fair to the client?
    - How does your price compare across time? Across insurance alternatives? Across clients?

9. REFERENCES

[1.] Baker, Ronald, *Pricing on Purpose: Creating and Capturing Value,* Wiley, 2006.

[2.] "The Economics of Insurance," Chap. 1 in *Actuarial Mathematics,* by N.L. Bowers Jr., et al., Schaumburg, Illinois: The Society of Actuaries, 1997.

[3.] Chetty, Raj, and Adam Szeidl, "Consumption Commitments and Risk Preferences," *The Quarterly Journal of Economics* 122, no. 2 (2007): 831-877.

[4.] Cutler, David M., Amy Finkelstein, and Kathleen McGarry, "Preference Heterogeneity in Insurance Markets: Explaining a Puzzle," *American Economic Review Papers and Proceedings* 98, no. 2 (2008): 157-162.

[5.] Einav, Liran, et al., "How General are Risk Preferences? Choices Under Uncertainty in Different Domains," NBER Working Paper # 15686, 2010.

[6.] Gallagher, Justin, "Learning about an Infrequent Event: Evidence from Flood Insurance Take-up in the US," Department of Economics, University of California at Berkeley, Nov. 14, 2010, p. 47, http://bellarmine2.lmu.edu/economics/papers/gallagher%20jmp.pdf.

[7.] Hsee, Christopher K., and Howard C. Kunreuther, "The Affection Effect in Insurance Decisions," *Journal of Risk and Uncertainty* 20, no. 2 (2000): 141-159.

[8.] Johnson, Eric, et al., "Framing, Probability Distortions, and Insurance," *Journal of Risk and Uncertainty* 7 (1993): 35-51.

[9.] Kahneman, Daniel, and Amos Tversky, "Prospect Theory: An Analysis of Decisions under Risk," *Econometrica* 47, no. 2 (1979): 263-292.

[10.] Koszegi, Botond, and Matthew Rabin, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics* 121, no. 4 (2006): 1133-1165.

[11.] Koszegi, Botond, and Matthew Rabin, "Reference Dependent Risk Attitudes," *American Economic Review* 97, no. 4 (2007): 1047-1073.

[12.] Peter, J. Paul, and James H. Donnelly, "Preface" in *Marketing Management,* McGraw-Hill, 2006.

[13.] Rabin, Matthew, "Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion," Chap. 11 in *Choices, Values, and Frames*, by Daniel Kahneman and Amos Tversky, New York: Cambridge, 2000.

[14.] Sydnor, Justin, "(Over)insuring Modest Risks," *American Economic Journal: Applied Economics* 2 (2010): 177-199.

[15.] Tversky, Amos, and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (1974): 251-284.

Biography of the Author

Marc-André Desrosiers is a Ph.D. candidate at UW-Madison in the Actuarial Science, Risk Management and Insurance program. He also completed his MBA at University of Calgary, after receiving his FCAS. Marc-André has studied Actuarial Mathematics and Philosophy at Concordia University, Montréal. The author also keeps contact with the industry as he is currently working as an external consultant for Intact Financial Corporation Actuarial Commercial Lines department. He is interested in pricing optimization, behavioral economics, customer behavior, and demand modeling. He can be joined at mdesrosiers@bus.wisc.edu.

OCI OK

By Tom Herget, FSA, MAAA, CERA

The use of Other Comprehensive Income (OCI) is receiving its moment in the sun. It is being considered for housing some of the earnings volatility in the International Accounting Standards Board (IASB) current approach to measuring insurance contracts' performance.

The American Academy of Actuaries' Insurance Accounting Task Force[1] recently prepared a white paper to help IASB members understand just what could belong in OCI.[2]

Following are some of the concepts that were raised to help answer this question.

ACCOUNTING BASICS

Any accounting system has fundamental relationships between assets, liabilities, and net worth. No matter what rules or principles exist for an accounting basis, the balance sheet item for net worth is the difference between assets and liabilities.

The income statement is a measure of performance for an accounting period. The change in net worth reflects the excess of income over expenses for the period. The change in net worth that results from this performance is called Comprehensive Income (CI).

CI comprises two components, Profit or Loss (PL) and Other Comprehensive Income (OCI). A search of accounting literature has not revealed principles for assigning elements to either PL or OCI. Items included in OCI are events that rule makers have decided should not be in PL. Some of the events may be characterized as unusual, non-recurring, or items outside the control of management.

Once a contract has expired, earnings under any accounting basis will be the same. No matter what the IFRS, U.S. GAAP, Estonian, or whatever reserving rules are, the change in liabilities will be canceled once the policy obligation is extinguished. All the accruals sum to zero; the only thing left is cash.

At the point where the policy exits the company's inventory, the PL and CI must be equal. Thus,

[1] The American Academy of Actuaries is a 17,000-member professional association whose mission is to serve the public and the U.S. actuarial profession. The Academy assists public policymakers on all levels by providing leadership, objective expertise, and actuarial advice on risk and financial security issues. The Academy also sets qualification, practice, and professionalism standards for actuaries in the United States.

[2] http://www.actuary.org/pdf/finreport/OCI_response_111219.pdf

the last OCI entry reverses all the prior OCI entries so they sum to zero.

AUTHORITATIVE ACCOUNTING LITERATURE

Search results for information related to OCI in three popular accounting systems (U.S. GAAP, IFRS, and U.S. statutory) revealed:

US GAAP: A September 30, 2010 letter from Ernst & Young, LLP states, "There are no clear underlying principles for the recognition of OCI items or for the reclassification of such items through net income."

IFRS: A June 2010 Ernst & Young, LLP industry newsletter reads, "A number of respondents to the exposure draft requested that the IASB also address the issue of the lack of clear underlying principles for the recognition of OCI items (as well as for the reclassification of such items to profit or loss) within IFRS."

U.S. Statutory: Instructions for preparing the U.S. statutory statement include a description of its OCI provision: "The purpose of the capital & surplus account is to delineate certain charges and credits not included in operations such as net capital gains and items pertaining to prior years…"

The conclusion is that under these three accounting bases, there is no articulation of comprehensive principles for recognizing items in OCI.

USES OF PROFIT OR LOSS

PL is used by company management, by authorities, and by investors.

Company Management

All insurance products are developed using models of future cash flows. The models produce results that display returns to policyholders, employees/agents and to the company itself. The returns to the company itself are the PL. In pricing, the anticipated PL should be set neither too low (not enough return) nor too high (likely uncompetitive and unsalable). The actual PL as it emerges is compared to the expected PL to evaluate the success of the product. Under current accounting standards, it is rare, if at all, that items captured in OCI play a part in the product pricing process, since they are typically non-recurring items.

PL can also play an important role in the determination of executive and employee bonus and incentive compensation. This helps align management actions with shareholder interests. OCI may or may not be a component of incentive plans.

Finally, PL is used to trumpet performance results. Each quarter, in print and through earnings

conference calls, PL is the focal point of performance discussions. OCI is usually mentioned and discussed separately in such communications.

Regulatory Authorities

Insurance regulators tend to look at balance sheet adequacy on a current basis before looking at income. However, a string of successive negative CIs would cause alarm.

Insurance taxation bodies have a keen interest in the PL as that serves as the basis for taxable income. In the United States, impacts of management-elected changes are often captured and re-spread into PL or OCI over a specified number of years, according to regulatory policy.

Investors

Generally, the item that attracts investors' attention most is the PL. That seems to be the basis on which management is judged. PL is the numerator of a common benchmark, earnings per share. When a share value is expressed as a multiple of earnings per share, it is the PL that is used as the benchmark although sometimes additional adjustments are made by an analyst.

RECOGNIZING EVENTS IN EITHER PL OR OCI

One could ask should OCI even exist. A case can be made that it doesn't really matter how something gets to the bottom line.

If OCI should exist, performance impacts could be allocated to PL or OCI by either a blanket assignment or through principles.

Blanket Assignment

Authoritative literature could merely state what measures do not belong in PL and should run through OCI. This would involve subjective determinations and could be the result of convenience, simplicity, or political compromises. Any accounting authority can make such a list; without principles, there would be no way to evaluate its propriety.

Principles

There are many viewpoints as to what can constitute "regular" or "normal" earnings (PL) in insurance, especially since there is so much unknown and so much variation around the unknown in insurance products and the investments and capital that support them.

Following are several possible principles. This presentation starts with a clean sheet of paper, incognizant of rule makers' existing preferences or pronouncements. There may be more than one right answer. Any answer may also not be practical. Also, it might be that no single principle is

adequate; a combination may be needed. The purpose of this offering is to discuss different viewpoints.

Here are some possibilities and perspectives, offering advantages and disadvantages, on the following candidates for principles that could be used to distinguish OCI from PL:

- Warranted vs. unwarranted volatility
- Actions within vs. outside of management control
- Ordinary (usual) vs. extraordinary (unusual) events
- Regular results vs. those due to changes in methodologies or assumptions
- Current year results vs. prior (or future) period adjustments

1. Warranted vs. unwarranted volatility: The challenge is to develop a consensus viewpoint among participants as to what type of volatility would be considered unwarranted. There is a common perspective that volatility imposed by accounting conventions that doesn't reflect the underlying economics of the business can be viewed as unwarranted. This is called "accounting mismatch."

One example of unwarranted volatility would be the component of CI created by the fact that assets and liabilities are measured at discount rates that are determined on an inconsistent basis. Another possibility is the fact that one side of the balance sheet may be unlocked (e.g., at fair value) while the other side may be locked in (e.g., at amortized cost).

Advantages:

- Accounting mismatch is objective and it is relatively easy to identify.

Disadvantages:

- It might involve two perpetual independent valuations at reporting time (companies are already doing this with available for sale securities and shadow DAC, shadow VOBA calculations).

2. Actions within vs. outside management control: This is also a challenge to define. Conceptually, management is responsible for every action and inaction of its company. Further, the purpose of insurance is to deal with risks (for the most part) outside of the policyholders' control.

Possible examples of items outside of management's control would be introduction of a new catastrophe model that now dictates more capital is needed. Another candidate would be the use of market interest rates in determining the value of liabilities. A third possibility would be the

introduction of legislation that is disruptive to the current business plan. Some people maintain investment results are outside of management's control.

Advantages:

- This helps measure management performance by removing items that are beyond their control.

Disadvantages:

- Management is responsible for everything; why exempt certain items?
- It may be difficult to ascertain what is or is not within management's control.
- There might be a bias in classifying favorable events to be within management's control and unfavorable ones outside of their control, thus inviting manipulation.

3. Ordinary (usual) vs. extraordinary (unusual) results: Here again, defining extraordinary will be a challenge. To an individual, the arrival of a hurricane may be a life-changing extraordinary event. But to an insurer, this would be a regular component of day-to-day business. A major catastrophe, for which benefits are payable under the terms of a contract is not an extraordinary, external event. Nor would major medical or technological breakthroughs that dramatically reduce the cost of existing coverage be considered an extraordinary event. For this purpose, a determining criterion might be whether there is a provision for such events in the pricing of the product.

Possible considerations for extraordinary events might be a court case that establishes retroactive liabilities in contracts where no such exposure was anticipated (asbestos). Another possibility is the collapse of a counterparty (a reinsurer or a hedge provider), the value of whose promises is now dramatically diminished. Additional candidates are the transfer of a large loss portfolio and the acquisition or sale of a block of business or company.

Advantages:

- This helps provide a better trend line of normal operations.
- This may help management make difficult decisions if there is a separate place in CI to report their impacts.

Disadvantages:

- It is difficult, and getting more difficult, to define the dividing line between the ordinary and the extraordinary.
- There may be a tendency to classify adverse events as extraordinary and favorable events as ordinary, thus inviting manipulation.

4. Regular results vs. those due to changes in methodologies or assumptions: Insurers will frequently review methodologies in light of emerging developments and environments. Companies will often introduce new models or upgrade existing ones. These can be perceived as presenting a better indication of the future. Use of new methodologies can be viewed as a refinement rather than a correction.

Often, assumptions need to be changed. If an event occurs during the current period that dictates a prior assumption is no longer valid, the assumption should be changed. OCI could be used to report the impact of the assumption change. However, applying outdated, prior assumptions to the current period's inventory is a meaningless, if not incorrect, determination.

Advantages:

- This helps provide a better secular performance trend line.
- It may help management make difficult decisions if there is a separate place in CI to report their impacts.

Disadvantages:

- Since insurers should be changing their evaluation of the future regularly, why classify activities as extraordinary when they are part of normal operations?
- To measure the impact of the assumption change, the company would need to quantify by using old assumptions at a new date or new (but premature) assumptions at the old date. Neither would reflect a valid representation of the balance sheet at that time.

5. Current year results vs. prior (or future) period adjustments: Assumptions need to be changed periodically. Sometimes what had appeared to be an aberration is confirmed as a trend. This is a normal situation for the evaluation of mortality and sometimes voluntary terminations. Introduction of a new assumption is appropriate. With the benefit of hindsight, one could say that the change should have been implemented several periods earlier. One use of OCI would be to report the prior period effects in OCI and only the current period in PL.

In the same way, changing an assumption brings into the current year, adjustments to the results of all future years. Using OCI to remove these effects from the current year PL would also improve the usefulness of PL.

Sometimes, a mistake may have been made. Thousands of keystrokes are used to generate an image of a liability or an asset. When these human errors are detected, their impact could be recorded in OCI.

Advantages:

- Items that have prior period impacts can usually be clearly identified as well as quantified.
- This helps provide a better trend line of normal operations.
- This eliminates opportunities for management to manage earnings.

Disadvantages:

- This might become painful for management to constantly address.
- Pointing one's eyes towards a mistake in a prior report might become a source of litigation.
- Changes in estimates could be used to manipulate the emergence of profit through PL; e.g., an insurance liability could be strengthened through OCI in order to improve future PL.

CONCLUSION

Since OCI concepts are being considered as a solution to reducing volatility in the insurance contracts IFRS, it would be a very appropriate time for the accounting industry to consider articulating the principles behind distinguishing between elements of PL and OCI. If the accounting industry desires to provide lists of what should be included in OCI, professionals can submit possible lists. If the accounting industry prefers to develop principles behind what belongs in OCI vs. PL, the actuarial profession would be willing and able to assist, expanding on (or adding to) the five candidates presented. Personally, this author feels that addressing the warranted versus unwarranted volatility (which reveals the accounting mismatch) offers the most information to the user. It is possible that some combination of the above principles offers the most valuable information to a user. The quantification of the impacts of unusual or extraordinary events could always be made in disclosures and not necessarily be assigned to OCI.

Acronyms for Actuaries

By Tom Herget, Chris Kogut, and Anna Wetterhus

Actuaries interact with a dizzying array of programs and institutions from around the world -- and the many acronyms they use as shorthand—every day. But there are just too many acronyms for anyone to keep straight, and new ones are being added all the time. Members of the American Academy of Actuaries' Solvency Committee have put together an acronym reference chart. This handy guide can be folded up to fit in your wallet and quickly reviewed so you can fearlessly enter an elevator.

This chart can be downloaded from the Society of Actuaries Web Site at www.soa.org/fr-acronyms and is a link in the CAS E-Forum, Winter 2012-Volume 2.

Would you care to nominate new candidates? Would you like to enhance an existing description? The authors are committed to keeping this resource current. E-mail any of the three authors with your suggested text additions or changes.

Acronym link:
http://www.casact.org/pubs/forum/12wforumpt2/Herget-Acronyms-Spreadsheet.xlsx