Big Data and Regulation in the Insurance Industry[*]

Lawrence S. Powell, PhD

Executive Director

Alabama Center for Insurance Information and Research

University of Alabama

Big Data and Regulation in the Insurance Industry

Executive Summary

The purpose of this study is to assess the need for additional regulation and regulatory resources to address Big Data applications in insurance. I find the current regulatory framework is well-suited to address Big Data concerns. Moreover, with cooperation between insurers and regulators, the full benefits of Big Data can quickly become available to consumers.

While Big Data has the potential to improve insurance markets for consumers and insurers; some regulators and advocates voice concerns that Big Data could harm consumers unless new regulations and regulatory resources are created. Their concerns include opacity and complexity of Big Data methods, potential bias from training fraud detection algorithms on past behavior, and excessive granularity of insurance rates.

I address each of these concerns in more detail in the paper. Briefly, opacity and complexity are irrelevant, because regulators can reject rate filings unless or until insurers and vendors explain them to the regulators' satisfaction. Big Data applications are less biased than any other potential method of claims handling; in fact, Big Data can mitigate problems of bias if they exist in the claims process. Finally, concerns regarding granularity of rates are technically unfounded, and practically very unlikely in a competitive and regulated market.

Consumers stand to benefit from Big Data in at least five general ways. First, Big Data applications can make insurance pricing more accurate. As one example, using data from telematics devices, insurers can charge prices that reflect risk of automobile crashes much more accurately than by using historical loss and demographic data. This type of application has the additional benefit to society of decreasing frequency and severity of losses.

Second, Big Data applications can improve customer satisfaction. They can be used to intervene directly in customer interactions to prevent or solve problems. They can be used to streamline the claims process and maximize speed and accuracy of claim payments. Big Data can also improve the online customer experience and to create new insurance products to meet consumer demands.

Third, Big Data applications can narrow the coverage gap by increasing availability of coverage in areas and exposures that are otherwise difficult to underwrite. Catastrophe models, telematics, and satellite imagery are all examples of this effect.

Fourth, Big Data applications are effective in identifying and mitigating insurance fraud. This presents opportunities to cut the estimated $40 billion annual cost of fraud.

Fifth, there are opportunities to improve insurers' operational efficiency using Big Data. One clear example is to pre-fill insurance applications using public and proprietary data. This makes it easier for consumers to shop for insurance and bolsters competition.

The next conclusion in this study is that the current regulatory system is well-suited to address Big Data. Regulations are comprehensive and appear to serve consumers well. Regulators have the authority and laws needed to protect consumers and maintain a functional market. As evidence, I describe current laws, regulations, and processes. In addition, I note the scarcity of consumer complaints related to concerns voiced in the NAIC Big Data Task Force.

Finally, I evaluate competition in insurance markets. Competition is an important concept, because competitive markets provide important consumer protections. I find that insurance markets display all of the characteristics of competitive markets. They have multiple buyers and sellers with moderate market share.

Insurance markets also demonstrate low to moderate returns. For example, average return on net worth for P&C insurers is lower than that of commercial banks, public utilities, and the Fortune All-Industry Index. Confirmed complaints are quite rare, indicating a reasonable level of consumer satisfaction.

Given the positive results of this comprehensive evaluation, it seems clear that current regulation is adequate. Therefore, creating additional regulations would unnecessarily delay the beneficial effects of Big Data in insurance markets.

Big Data and Regulation in the Insurance Industry

Insurance, by its nature, is a data-driven business. As such, insurers appear to be embracing the age of Big Data with equal parts caution and enthusiasm. While their enthusiasm reflects centuries of experience drawing inference from large datasets; insurers' caution is a product of uncertainty surrounding how consumers and regulators will respond to new analytical constructs.

The National Association of Insurance Commissioners (NAIC) has allocated substantial attention and resources to issues surrounding use of Big Data in the insurance industry. A series of working groups and task forces have formed to discuss these issues, with the most recent being the Big Data (EX) Working Group under the Innovation and Technology (EX) Task Force. At the time this is written, the Working Group is apparently implementing three earlier charges as follows:

1) Review current regulatory frameworks used to oversee insurers' use of consumer and non-insurance data. If appropriate, recommend modifications to model laws/regulations regarding marketing, rating, underwriting and claims, regulation of data vendors and brokers, regulatory reporting requirements, and consumer disclosure requirements.

2) Propose a mechanism to provide resources and allow states to share resources to facilitate states' ability to conduct technical analysis of and data collection related to states' review of complex models used by insurers for underwriting, rating, and claims. Such mechanism shall respect and in no way limit states' regulatory authority.

3) Assess data needs and required tools for regulators to appropriately monitor the marketplace and evaluate underwriting, rating, claims, and marketing practices. This assessment shall include gaining a better understanding of currently available data and tools and recommendations for additional data and tools as appropriate. Based upon this assessment, propose a means to collect, house, and analyze needed data.

The purpose of this manuscript is to assist the NAIC as it evaluates the need for additional regulation or regulatory resources to protect consumers from insurers' use of Big Data. I begin by providing a definition of Big Data to clarify the term and to set the scope for this project. Next, I review uses of Big Data in the insurance industry and discuss how these developments affect consumers. Third, I evaluate the current insurance regulatory system regarding its ability to address Big Data or predictive models. Fourth, I evaluate the degree of competition in insurance markets to estimate the potential for insurers to consistently

profit from inappropriate risk classification practices. Finally, I analyze consumer complaint data to assess the adequacy of current market forces and regulatory interventions.

Evidence is not consistent with the need for additional regulation or regulatory resources. The current regulatory system is well-suited to address issues related to Big Data. Given the potential for Big Data to benefit consumers and society through its use in insurance markets, and the potential for misplaced regulatory efforts to dampen this effect, new regulation is not warranted at this time. In addition, the market is very competitive, showing many participants, moderate profits, and very few complaints. Therefore, allocating additional resources to regulate Big Data would be neither appropriate, nor efficient.

Defining "Big Data"

Big Data has become a catchall phrase used colloquially to represent the capture and analysis of large databases to glean useful results. A singular, comprehensive definition of Big Data is elusive. A common definition of Big Data is the analysis of data demonstrating such volume, velocity, and variation, that standard data-processing equipment is not sufficient for the task.

Given the speed of technical advancement in data-processing hardware and software, the concept of a large database is a moving target; however, much faster increases in production and collection of data suggest this definition will be valid for years to come. Therefore, it is common for Big Data analysis to occur on a server (perhaps in a cloud), rather than on a desktop or laptop computer. For example, as of January 2017, the Dell Computers website advertises a desktop machine with two terabytes of hard drive storage and 32 gigabytes of memory as an advanced analytics computer starting at $3,499.[1] The most common data analysis software packages use the spreadsheet format, such as Microsoft Excel. These programs are typically limited by the maximum number of rows (1,048,576) and columns (16,384), as well as available memory in a computer. In practice, even on an advanced, modern computer, analyzing one gigabyte of data in a spreadsheet does not provide adequate performance or stability.

In comparison, the National Oceanic and Atmospheric Administration (NOAA) currently stores nearly 200 petabytes[2] of data and collects approximately 30 petabytes of additional data per year. To put this in perspective, an average high-speed broadband Internet connection for home or business can download approximately 100 Mbps.[3] At this speed, it

---

[1] http://www.dell.com/us/business/p/precision-t7910-workstation/pd?oc=cup7910w7p_3&model_id=precision-t7910-workstation accessed January 2, 2017

[2] A petabyte is 1,000 terabytes or 1,000,000 gigabytes. According to http://www.whatsabyte.com/, a petabyte could hold approximately 500 billion pages of standard printed text.

[3] See http://www.speedtest.net/awards/us/isp. Mbps are megabits per second. A megabit is one eighth of a megabyte. There are 8 billion megabits in a petabyte.

would take more than seventy-six years to download each additional year of data produced by NOAA. These data are used by insurance companies and vendors to understand, and predict losses from, insured weather perils. Clearly, it would not be feasible to store or analyze these data on an above-average desk top computer, or using common data analysis software.

Velocity is the speed at which data are created and analyzed. Not only must there be sufficient hardware and software to store and analyze large amounts of data, these tools must also be capable of reading and analyzing data quickly as they are created. For example, Big Data techniques are used to monitor audio and video recordings as they are created to alert authorities to the presence of certain threats. In this case, it is imperative that large amounts of data are analyzed in near-real-time. As I describe in more detail below, insurers can apply these techniques to audio data created by customer service telephone calls to improve customer service outcomes.

The last descriptor for Big Data is variation. Variation recognizes that, unlike the perfectly formatted and organized data we learn to use in a statistics class (perhaps downloaded from www.census.gov), many forms of Big Data are neither formatted, nor organized. Examples include audio and video recordings, or the constant flow of "likes," "posts," "tweets," and "snapchats" on social media. Therefore, a large part of Big Data work is to turn this massive barrage of words, numbers, images, sounds, and other descriptors into data from which users may draw inference.

Indeed, it is a marvel of modern technology that data scientists can harness even a small sliver of available data to help private and public entities make more informed decisions. However, this is the future of strategic business endeavors. Big Data represents opportunities to improve solvency through a more thorough understanding of risk.

Insurer use of Big Data

There are a number of ways Big Data can be used in insurance operations. The first and most obvious areas to apply Big Data are underwriting and pricing. In addition, Big Data provides new tools to improve customer experience, reduce the coverage gap, mitigate fraud, and improve efficiency of processes. In the following, I briefly address each in turn.

Pricing and underwriting:

The most intuitive role for Big Data in insurance operations is to more accurately match premiums and losses through pricing and underwriting. An early example of Big Data in insurance pricing is the use of telematics data. Telematics devices collect real-time data about

the driving behavior of policyholders. They record variables including mileage driven, in addition to speed, time, location, braking, cornering, and lane changes. Using this information, a driver's premium is tailored to her own driving behavior. Such rating plans are referred to as pay-how-you-drive (PHYD) plans.

The public-policy benefits of PHYD telematics are enormous. First, PHYD is fully transparent to consumers. In many cases, these devices provide immediate feedback to drivers on their driving behavior. By providing feedback, or even just the knowledge that the insurer is monitoring the driver, automobile crashes, and attendant injuries and fatalities are sure to decline.

Customer satisfaction:

Another objective of insurance companies using Big Data is to improve customer satisfaction. There are two general approaches to improving customer satisfaction. The first is to avoid or quickly resolve service-related problems. The second is to provide benefits likely to align with individual consumers' preferences.

One Big Data approach to mitigating customer service problems is to monitor and react to customer service interactions in real time. Advanced data processing and analytics are capable of detecting or estimating customer satisfaction from data produced during a telephone conversation. Consider a scenario where a consumer has called her insurance company to ask a question or report a claim. Based on the words and tone of the conversation, software detects a problem. The system can automatically escalate the call to a supervisor with the necessary knowledge and authority to resolve the problem.

Increasing coverage:

A common example of Big Data in insurance is the use of computerized catastrophe models for pricing and underwriting windstorm and earthquake perils. Because catastrophic losses are quite rare, it is not possible to estimate their probability distributions from historical data alone. Catastrophe models augment available data on catastrophic losses with dynamic simulation models representing the components of such losses and their interactions. Then they use the more robust model to estimate loss outcomes and probabilities. Without such models, private insurance coverage for the most risky exposures would be much more expensive, or completely unavailable.

Mitigating fraud:

Annual estimates of the cost of insurance fraud in the United States, excluding health insurance, are as high as $40 billion.[4] Because almost all families purchase some type of P&C insurance, and the cost of fraud is paid by consumers, reducing insurance fraud is a worthy goal. Willis Towers Watson (2016) report that 28 percent of insurers were using predictive analytics to identify fraud in late 2015. By 2018, the number is expected to grow to 70 percent.

Big Data methods present enhanced capabilities for fraud detection and mitigation. Two approaches to fraud mitigation that fit squarely within the Big Data description are text analytics and social media analysis. SAS (2016) notes that as much as 80 percent of data in claims files is unstructured text such as typed reports and recorded interviews. Text analytics methods glean useful information from unstructured text. For example, if multiple adjusters interview a large number of claimants in seemingly unrelated claims, and they all provide the same story, word for word, this might suggest they have been coached to give specific statements by a fraud enterprise. Text mining can also be used to identify other "red flags" for organized insurance fraud, such as a large increase in an individual attorney's number of lawsuits.

Another example of Big Data fraud detection is mining the social web to find relationships among people with multiple claims. Such "network link analysis" has been successfully used by insurers to identify organized fraud rings.

Operational efficiency:

Since the beginning of the computer age, technology has been used to make businesses run more efficiently. It is not surprising that Big Data can improve insurer efficiency in ways that benefit consumers. One shining example is the use of Big Data to pre-fill insurance applications. Using a combination of public and proprietary data sources, insurance agents can now complete the majority of personal lines insurance application fields by requesting only one or two items from a prospective customer. With only a driver's license number or Social Security number, and automobile VIN, an insurer can complete an application for automobile insurance. Similarly, with only an address and identification number, a homeowners insurance application can be completed in just a few minutes.

While this is certainly an improvement in insurance company operations, it is also very beneficial to consumers. By streamlining the application process, the search costs for personal lines insurance trend lower and lower. This boosts to level of competition in these markets to better serve consumers, without the need for costly regulation.

---

[4] See https://www.fbi.gov/stats-services/publications/insurance-fraud

Big Data in the Current Regulatory System

In this section, I review the current insurance regulatory system to determine if it is capable of regulating insurer use of Big Data. Specifically, I consider consumer protection issues in pricing and underwriting, marketing, and claims. More generally, the regulatory system provides a host of tools for regulators to ensure fair and appropriate outcomes for insurance consumers. These include rating laws applied in product filings for rates, forms, and underwriting manuals, market conduct examinations, complaint reporting systems, unfair trade practices and unfair claims practices statutes that govern marketing and claims handling, as well as certain Federal laws that require transparency and privacy protection for insurance consumers.

In addition, I point out the importance of regulatory innovations in recent decades, like SERFF and IIPRC, aimed at making insurance markets more competitive and consumer friendly. Such initiatives demonstrate the clear consensus that innovation is a positive development for insurance consumers. If Big Data is treated by regulators as more of a threat than an opportunity, consumers will not realize the potential benefits of Big Data in insurance markets. However, by applying current rules and regulations, regulators can maintain the high level of consumer protection while ushering in a new wave consumer-friendly innovation.

Consumer complaints data provide relevant evidence of how well the combination of market competition and current regulations protect consumers. In this section, I also present data on insurance consumer complaints.

Insurance regulators in every state have systems for receiving and addressing consumer complaints. Regulators catalogue and report data related to consumer complaints to the NAIC. The following analysis uses data retrieved from the NAIC's Consumer Information Source (CIS) website[5] on which consumers are encouraged to review complaint data when selecting an insurance company. The NAIC CIS website offers several caveats and disclaimers with these data. In the Appendix, I discuss and analyze these issues, concluding that, while imperfect, these data are a reasonable approximation of complaints reported by consumers to state insurance regulators.

I collect complaint data from the CIS website for year 2015 in three categories: Number of Complaints by Coverage Type, Reasons Why Complaints Were Submitted, and Final Decisions Regarding Complaints. The database provides 309 variables that represent the

---

[5] https://eapps.naic.org/cis/

number of complaints in various categories and, while it does not offer definitions, many seem intuitive or self-explanatory from their titles.

The NAIC insurance company database includes 3,885[6] company codes for P&C insurers. Of these, 2,905 appear in the online CIS complaints database. More than half (1,598) of these companies do not report any complaints in the database. A more thorough description of these data appears in the Appendix.

Table 1 shows number of complaints by reason category and the amount of premium dollars per complaint in each category.[7] Claim handling is most common reason at 27,962 complaints reported, followed by underwriting (13,168), policyholder service (5,408), and marketing and sales (1,247). Total reasons for complaints is 47,785. Recognizing that direct premium written sums to $572.5 billion, there is on average one complaint reason provided for every $12 million of premium.

Table 1: Complaints by Reason, 2015

Reason Category	# of Complaints	$ Premium per Complaint
Underwriting	13,168	$43,472,765
Marketing and Sales	1,247	$459,061,246
Claim Handling	27,962	$20,472,404
Policyholder Service	5,408	$105,852,325

Sources: NAIC CIS complaint database (see appendix) and NAIC Statistical Compilation of Annual Statement Information for Property/Casualty Insurance Companies in 2015.

Pricing and Underwriting:

Current insurance rating laws require rates and rating factors that are not excessive, inadequate, or unfairly discriminatory. This logical and elegant set of requirements has served the public interest for decades. Through the product filing process, regulators are able to monitor and control price levels and factors used to calculate prices.

A recurring theme in the current discussion about how regulations should or should not change to accommodate Big Data is that somehow Big Data and predictive analytics enable insurers to circumvent important elements of the regulatory process. While these new tools

---

[6] Includes insurers that are no longer active.

[7] Total P&C direct premium written in 2015 was $572.5 billion.

are fascinating and powerful, they do not pose any new threats to consumers that cannot be handled in the current regulatory framework. This is due largely to the fact that the burden of regulatory compliance falls primarily on regulated insurers, not on regulators or vendors.

Suppose, for example, an insurance company submits a rate filing in which it proposes a rating factor developed by a third-party vendor. The regulator reviewing the filing has concerns about the appropriateness of the new rating factor. In this scenario, the regulator is not obligated to approve the filing until the insurance company explains the new factor to the regulator's satisfaction and demonstrates compliance. Note that this does not require direct regulation of vendors, or regulators' review and possession of all data used by insurers and vendors to make rates. It merely requires regulators to reject rate, rule, or form filings that do not demonstrate compliance with existing laws.

On the other hand, Big Data presents opportunities to improve transparency in the rate regulation process. Using Big Data tools, insurers and vendors can demonstrate affirmatively that new rating factors are related to expected losses, do not harm protected groups, and avoid other sub-optimal outcomes.

Deliberations of the NAIC Big Data Working Group have heard concerns that the use of Big Data will further segment prices to the point that each driver or homeowner has her own individual rate. The fear is that this could temper the beneficial risk pooling effect of insurance. While this concern may seem valid on its face, it is both technically incorrect[8], and highly unlikely in a competitive market. This degree of segmentation would produce unstable rates across people and over time, resulting in unhappy consumers. Unhappy consumers are free to buy insurance elsewhere. Moreover, it would be within a regulator's authority to reject rate filings that would cause this type of market disruption.

Finally, complaint data related to underwriting are not consistent with a need for additional regulatory resources. Underwriting issues generate approximately one complaint for every $43.5 million of premium written.

Marketing:

The use of Big Data to market insurance products has the potential to improve customer satisfaction by matching consumers with appropriate insurance coverage. This can also decrease the coverage gap, which threatens the financial solvency of citizens, and presents

---

[8] Segmentation does not restrict benefits of risk pooling. Segmented pricing allows pooling of heterogeneous exposures without creating cross subsidies.

extraordinary risk to the United States Treasury in the form of home mortgage defaults and post-disaster public assistance.

During Working Group meetings, some have voiced concerns about the lack of transparency in Big Data marketing applications; however, public meeting records have not captured a concrete example of how consumers might be harmed by marketing analytics.

To the extent that marketing analytics could create adverse outcomes for consumers, it is highly unlikely that such outcomes would avoid regulatory detection via consumer complaints and market conduct examinations. In addition, the comprehensive nature of Unfair Trade Practices statutes and other consumer protection laws exist to render inappropriate marketing practices illegal, cease said practices, and punish the individuals or entities that perpetrate such crimes.

The NAIC complaint database for 2015 includes only 1,247 complaints filed against P&C insurers related to marketing or sales. The total P&C premium written in 2015 was approximately $572.5 billion. Thus, consumers file one marketing complaint for every $459 million of premium written. In light of this ratio, additional regulation related to marketing practices would struggle to pass any reasonable benefit/cost analysis.

Claims:

Big Data also offers substantial opportunities to improve efficiency and accuracy in the claims process. In a previous section, I provide examples of Big Data applications in fraud detection and mitigation. Big data techniques can also be used to improve the speed and quality of claim outcomes by estimating the complexity of claims and allocating resources efficiently.

One concern voiced to the Working Group is that using Big Data techniques in the claims process could discriminate against protected classes and low-income claimants. This concern asserts an unsubstantiated assumption that algorithms trained on historical data will perpetuate past biases in claims handling.[9] It seems that this concern would be addressed and mitigated, not caused, by Big Data techniques. Claims algorithms are objective and Big Data allows use of the most current data. If past claims data represent the personal biases of claims professionals, it is intuitive that the most recent data would be the least biased. Moreover, Big Data analytics can be used to identify and exclude inappropriate biases if they are found in historical claims data.

Suppose, hypothetically, that an insurer chooses to treat claimants unfairly, market conduct examinations and complaint handling processes will detect such transgressions, and unfair claims practices statutes will hold them accountable. Additionally, the combination of existing laws, regulations, and market characteristics, along with extreme scarcity of

---

[9] I am not aware of evidence supporting this position.

observations consistent with these concerns, suggests very low benefit-to-cost ratio for new regulatory resources. As evidence, among the 120 variables capturing reasons for complaints in the NAIC complaint database, only three fields collect any type of biased treatment. They include Marketing/Sales: Unfair Discrimination (1 complaint in 2015), Underwriting: Unfair Discrimination (7 complaints in 2015), and Underwriting: Redlining (zero complaints in 2015). The NAIC D Committee has not created a complaint category involving discrimination or bias in claims, suggesting the problem is quite rare.

While the number of claim-related complaints is quite small relative to the volume of insurance transactions (27,962 complaints and $572.5 billion premium, or one complaint per $20.5 million premium), the largest share (just over 80 percent) of these complaints involve problems that could be mitigated with Big Data. They include unsatisfactory settlement offers (23 percent), denial of claims (23 percent) and claim handling delays (35 percent).

Competition in Insurance Markets

Regulation is only appropriate when markets do not work as they should. When markets work well, regulation is neither necessary, nor beneficial for anyone. Competition is the primary indicator of market quality. When markets are competitive, consumers are treated fairly and prices reflect cost of production and a fair profit.

By the classic definition, a competitive market is characterized by four traits.[10] First, it includes multiple independent sellers with low to moderate market shares. Second, there are multiple consumers with low to moderate market shares. Third, the product is sufficiently homogeneous or transparent such that consumers can differentiate value across offered prices. Finally, barriers to entry and exit are low, allowing new suppliers to enter the market if prices rise above the fair-market price, or allowing current suppliers to exit the market if they cannot produce the product at the fair-market price.

In addition to these four traits, competitive markets exhibit two more characteristics, as a result of competition. First, profit levels are moderate. Second, consumer satisfaction is reasonable. Insurance markets demonstrate all of these traits and characteristics.

Market characteristics:

Anyone who has watched television or surfed the Internet in the last decade has likely observed the extreme veracity of insurance industry advertising. In 2015, P&C insurers spent just under $6 billion advertising their products to consumers. Thus, among many other slogans and clever characters, we all know that Allstate protects us from "Mayhem," Flo from Progressive can get us a bundle discount, and GEICO's gecko can save us 15 percent in 15 minutes or less. There is little point in advertising when markets are not competitive.

---

[10] Competition is defined as "workable competition" in the sense suggested by Clark (1940).

Beyond the prima facia volume of advertising, there is substantial evidence that the insurance industry's market structure is quite competitive. I analyze 2014 data on entry, exit, market share and participants for six of the largest lines of PC insurance by direct premium earned: private passenger automobile, commercial automobile, homeowners, fire, commercial multiple peril, and workers compensation. Collectively, these lines of insurance represent 71 percent of direct premium earned. In addition they provide a broad cross-section of consumers.

Table 2 presents the maximum, mean, and minimum, of four competition measures across states and lines of business. For example, the concentration measure for private passenger automobile insurance shows the largest (Alaska = 0.18), average, and smallest (California = 0.07) concentration measure reported by any state in 2014.

Table 2: Competition Measures

Attribute	Measure	Private Passenger Auto	Homeowners	Commercial Auto	Commercial Multiple Peril	Fire	Workers Compensation
Concentration	Maximum	0.18	0.20	0.11	0.39	0.18	0.48
	Mean	0.11	0.10	0.05	0.06	0.06	0.12
	Minimum	0.07	0.04	0.03	0.03	0.03	0.03
Number of Sellers	Maximum	57	58	90	81	84	81
	Mean	35	40	63	59	63	53
	Minimum	12	15	32	30	39	29
Market Entry	Maximum	14	19	29	24	29	25
	Mean	6	8	19	15	18	17
	Minimum	1	2	7	9	9	10
Market Exit	Maximum	16	17	28	21	26	22
	Mean	9	7	17	12	17	15
	Minimum	2	1	7	5	10	7

Source: NAIC 2014 Competition Database Report

Information in Table 2 is consistent with competitive markets. In each line of business, there are many sellers with the ability to enter and exit the market. The number of sellers is defined as the number of insurance groups selling a line of insurance that write at least 0.01 percent of the market.[11] Commercial automobile and Fire insurance demonstrate the largest average number of insurance companies per state, both with sixty three. Personal automobile insurance has the lowest average number of sellers per state with a robust thirty five. The minimum number of sellers in each line of business represents Alaska or Hawaii. Both are

---

[11] This is the definition used in the NAIC 2014 Competition Database Report.

relatively small states (by population) and present large cost of market entry given their remote geographic locations relative to other states.

Market entry and market exit are the number of insurance groups that enter and exit each state during the most recent five years. Entry is defined as an insurance group writing at least 0.01 percent of premiums in a state, when the group did not write that much the preceding year. Exit is defined at the same cutoff in the opposite direction. In every state there are both entries and exits during the five-year period, suggesting barriers to entry and exit are reasonable.

The concentration measure is a Herfindahl index. A Herfindahl index is a scale from zero to one, with larger numbers signifying higher concentration. It is calculated as the sum of squared firm market shares.[12] The benefit of using the Herfindahl index over a more simple measure is that the Herfindahl measures number of firms and market share simultaneously.

While this index is a good measure of market concentration, it is not a perfect measure of competition. The theory behind concentration as a competition measure is that, when a market has fewer participants, it is easier for those participants to collude against consumers. However, if a market has only two sellers that are close substitutes, competition between the two sellers can create sufficient competition to function well for consumers. In addition, a seller can gain large market share by being more efficient or providing more value to consumers than their competition. Nonetheless, concentration measures, along with other observations, can provide information on the competitiveness of markets.

The concentration measures reported in Table 2 imply the average state has the functional equivalent of approximately ten to twenty participants with equal market shares. The largest concentration measures are recorded for workers compensation insurance in Oregon (.48), Maine (.43), Rhode Island (.40), Idaho (.39), and Colorado (.37); however, the return on net worth observations for workers compensation coverage in these states are among the lowest in the country, suggesting these insurers are not colluding to increase profit.[13]

---

[12] The equation for the Herfindahl index is $\sum_{i=1}^{n} \left(\frac{Premium_i}{Total\ State\ Premium} \right)^2$, where i counts n firms in a given state and line of business.

[13] The ten-year average return on net worth by state is Oregon 6%, Maine 7%, Rhode Island 7%, Idaho 3%, and Colorado 4%.

Insurance Industry Returns:

Overall, the P&C insurance industry exhibits moderate returns in comparison to other industries. Table 3 presents average annual return on net worth from 2006 through 2015 for P&C insurers, commercial banks, electric and gas utilities, and the Fortune 500 all industry index. At less than 7 percent return on equity (ROE), P&C insurers experienced the lowest return of all industries represented in the table. While consistent with vigorous competition, it is somewhat alarming that the P&C insurance industry displays lower returns than the banking industry, which suffered a severe crisis during the calculation period, and electric and gas utilities, whose prices are set by regulators to avoid excessive returns in the industry's monopoly structure.

Table 3: Average Annual Return on Net Worth by Industry: 2006 – 2015

P&C Insurance	6.6%
Commercial Banks	8.4%
Electric & Gas Utilities	10.0%
Fortune 500	13.8%

Sources: ISO®, a Verisk business® and Fortune, as reported by Insurance Information Institute at:
http://www.iii.org/publications/insurance-fact-book-2016/property-casualty-financial-data/financial-results

While overall average returns are clearly moderate at best, it is natural to be curious about specific lines of insurance and geographic areas in individual years. Performance recorded in any single year, state, and line of business lacks sufficient credibility for analysis; however, it is informative to consider the distribution of returns across all three dimensions. Figure 1 shows the distribution of returns by line of business, state, and year for the six lines of insurance analyzed in Table 2 for the ten-year period from 2005 through 2014.

Figure 1 displays an image quite consistent with market competition. A large majority of return observations represent moderate return on equity. In addition, the shape of the chart is rather symmetrical, implying that the better-than-average and worse-than-average outcomes are approximately offset. These attributes are also apparent in Table 4, which summarizes the distribution presented in Figure 1 by certain percentiles.

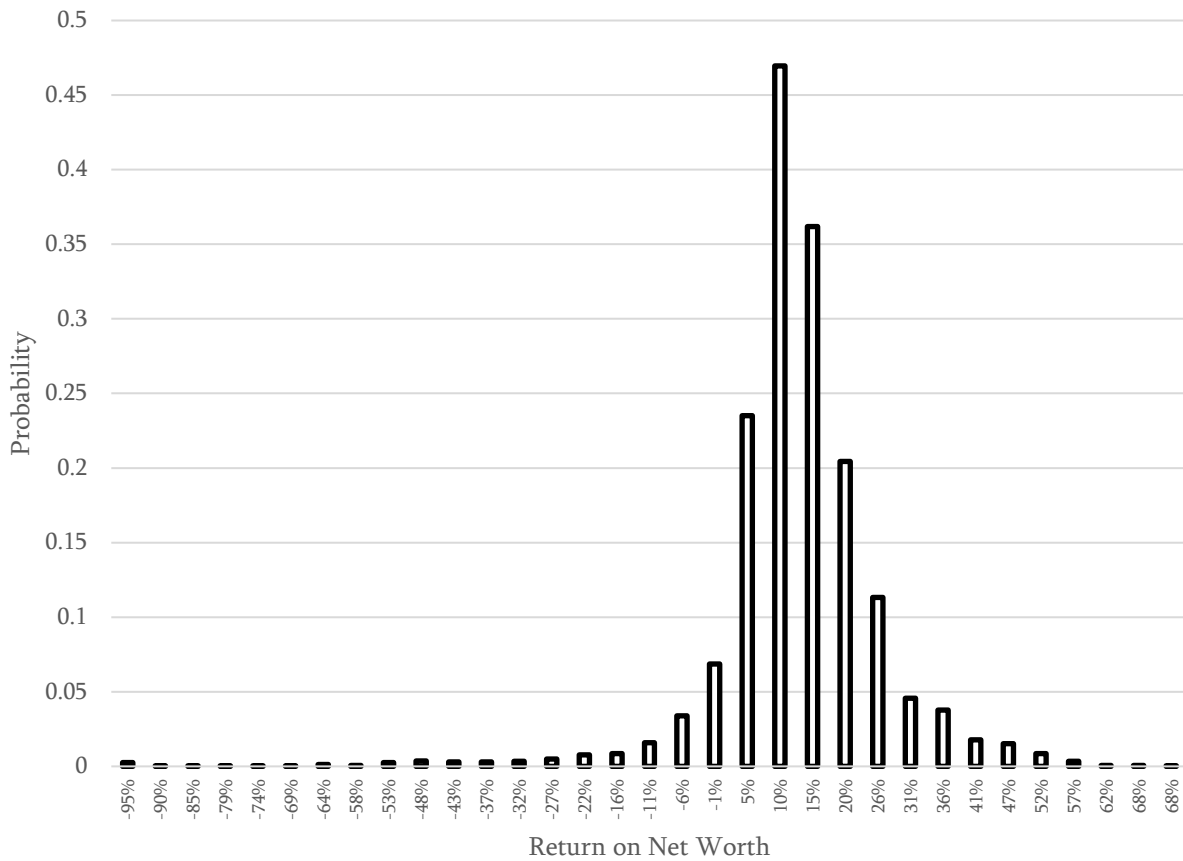Figure 1: Distribution of Insurance ROE by Line of Business, State, and Year, 2005-2014

Table 4: Percentiles of Return on Net Worth

Percentiles	Return on Net Worth
99th	45.5%
90th	23.7%
75th	15.7%
50th	9.5%
25th	4.9%
10th	-0.4%
1st	-39.9%

Source: NAIC Report on Profitability by Line and by State in 2014.

Note: Returns data exclude 40 line/state/year observations of workers compensation returns for the four states with monopolistic state funds (Ohio, North Dakota, Washington, and Wyoming), leaving 3,020 observations for analysis.

Consumer Satisfaction:

I observe an inverse measure consumer satisfaction via complaint data reported to state regulators. As mentioned above, these data are collected from the NAIC CIS website and are further described and discussed in the Appendix. The number of consumer complaints is not meaningful on its own. It must be measured against a control variable. In Table 5, I present complaints per insurance company interaction with consumers and claimants. I estimate the number of interactions using data from NAIC statistical publications on homeowners and automobile insurance (NAIC 2017a; NAIC 2017b), and other sources. The number of interactions includes policy renewals, policy changes (adding and removing houses and cars), and the number of first-party and third-party claimants. To improve exposition and parsimony, I present details of these calculations in the Appendix.

Table 5: Complaints per Consumer Interaction, Automobile and Homeowners

Automobile Insurance Interactions per Complaint

Total Interactions	222,990,917
Auto Insurance Complaints	16,880
Interactions per Complaint	13,210

Homeowners Insurance Interactions per Complaint

Total Interactions =	95,622,537
Homeowners Complaints =	9,951
Interactions per Complaint	9,618

Total Insurance Interactions per Complaint (Home and Auto)

Total Interactions =	318,613,454
Total Complaints =	26,831
Interactions per Complaint	11,878

Sources: NAIC CIS Complaints Database; NAIC 2017(a); NAIC, 2017(b); U.S. Department of Transportation: https://tinyurl.com/zryn7ay; Statista: https://www.statista.com/statistics/199974/us-car-sales-since-1951/; The Car Connection: http://www.thecarconnection.com/news/1090254_the-united-states-of-used-cars ; Insurance Information Institute http://www.iii.org/fact-statistic/homeowners-and-renters-insurance; Realtors' Association of America.

For both homeowners and automobile insurance together, there were 11,878 interactions per complaint. In other words, a complaint is generated in less than 0.01% of consumer interactions with insurers. This is not consistent with market problems caused by Big Data or otherwise.

This review of insurance consumer complaints does not suggest Big Data, or any other aspect, is creating problems that exceed regulators' capabilities. Given the high level of competition, low to moderate returns, and reasonable levels of consumer satisfaction, I find no evidence suggesting a need for additional regulations or regulatory resources.

<u>Summary and Conclusions</u>

Big Data has the potential to improve insurance markets for consumers and insurers. Improved markets would be evidenced by decreasing losses, improved accuracy of insurance rates, lower average cost of insurance, and greater customer satisfaction. However, some regulators and advocates voice concerns that Big Data poses a threat to consumers. The NAIC has sponsored a series of working groups and task forces to examine these potential effects on consumers, and evaluate the need for additional resources to help regulators monitor and evaluate insurance company operations that involve Big Data methods. The purpose of this study is to assist the NAIC in this evaluation.

There are at least five general ways Big Data can improve insurance markets for consumers. First, Big Data applications can make insurance pricing more accurate. As one example, using data from telematics devices, insurers can charge prices that reflect risk of automobile crashes much more accurately than by using historical loss and demographic data. This type of application has the additional benefit to society of decreasing frequency and severity of losses.

Second, Big Data applications can improve customer satisfaction. They can be used to intervene directly in customer interactions to prevent or solve problems. They can be used to streamline the claims process and maximize speed and accuracy of claim payments. Big Data can also improve the online customer experience and to create new insurance products to meet consumer demands.

Third, Big Data applications can narrow the coverage gap by increasing availability of coverage in areas and exposures that are otherwise difficult to underwrite. Catastrophe models, telematics, and satellite imagery are all examples of this effect.

Fourth, Big Data applications such as text analytics and network link analysis are effective in identifying and mitigating insurance fraud. This presents opportunities to cut the estimated $40 billion annual cost of fraud.

Fifth, there are opportunities to improve insurers' operational efficiency using Big Data. One clear example is to pre-fill insurance applications using public and proprietary data. This makes it easier for consumers to shop for insurance and bolsters competition.

The next conclusion in this study is that the current regulatory system is well-suited to address Big Data. Regulations are comprehensive and appear to serve consumers well. Regulators have the authority and laws needed to protect consumers and maintain a functional market. As evidence, I describe current laws, regulations, and processes. In addition, I note the scarcity of consumer complaints related to concerns voiced in the NAIC Big Data Task Force.

Finally, I evaluate competition in insurance markets. Competition is an important concept, because competitive markets provide important consumer protections. I find that insurance markets display all of the characteristics of competitive markets. They have multiple buyers and sellers with moderate market share.

Insurance markets also demonstrate low to moderate returns. For example, average return on net worth for P&C insurers is lower than that of commercial banks, public utilities, and the Fortune All-Industry Index. Confirmed complaints are quite rare, indicating a reasonable level of consumer satisfaction.

Given the positive results of this comprehensive evaluation, it seems clear that current regulation is adequate. Therefore, creating additional regulations would unnecessarily delay the beneficial effects of Big Data in insurance markets.

References:

Clark, J.M., 1940. "Toward a Concept of Workable Competition," *American Economic Review,* Vol. 30, No. 2, pp 241-256

NAIC, 2016. 2015 Insurance Department Resources Report

NAIC, 2017(a). Auto Insurance Database Report 2013/2014

NAIC, 2017(b). Dwelling Fire, Homeowners Owner-Occupied, and Homeowners Tenant and Condominium/Cooperative Unit Owner's Insurance Report: Data for 2014

SAS, 2016. Simplifying Fraud Analytics: 10 Steps to Detect and Prevent Insurance Fraud, http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/simplifying-fraud-analytics-105573.pdf

Willis Towers Watson (2016). "Insights - P&C insurers' big data aspirations for advanced predictive analytics: 2015 Predictive Modeling and Big Data Survey (U.S.)," https://www.towerswatson.com/en-US/Insights/Newsletters/Americas/americas-insights/2016/pc-insurers-big-data-aspirations-for-advanced-predictive-analytics

Appendix: NAIC Complaint Database

In the spirit of ASOP 23,[14] it is important to note the uncertainty inherent in data collected from the NAIC CIS database. NAIC provides the following disclaimer on the websites from which consumers access CIS data. On the introductory page, it states:

> "Consumer Information Source data is voluntarily supplied by state insurance departments and compiled and coded by the NAIC. Not all states provide complaint data to the Consumer Information Source. Please note that this database does not contain a complete record of all complaints filed and should not be used as the sole basis for insurance decisions."

However, retrieving complaint data by state for an insurance company that operates in all states shows complaints reported in every state.[15]

Another peculiarity in these data is how complaints are defined. The summary report of complaints by disposition found on the CIS website, Dispositions Regarding Closed Confirmed Consumer Complaints,[16] states that all complaints reported are "confirmed" complaints. It goes on to define a confirmed complaint as follows:

> A Confirmed complaint is a complaint in which the state department of insurance determines:
> a) The insurer, licensee, producer, or other regulated entity committed any violation of:
>   1. an applicable state insurance law or regulation;
>   2. a federal requirement that the state department of insurance has the authority to enforce; or
>   3. the term/condition of an insurance policy or certificate; or
> b) The complaint and entity's response, considered together, indicate that the entity was in error.

While the definition states that the insurance company was in the wrong in each instance, the report and the database include a field titled "Company Position Upheld," suggesting that the insurer did nothing wrong in between 9 percent and 20 percent of all complaints.

I collect CIS complaint data from the CIS website for years 2014 and 2015 in three categories: Number of Complaints by Coverage Type, Reasons Why Complaints Were Submitted, and

---

[14] Actuarial Standard of Practice (ASOP) #23 deals with the quality of data used by insurance companies.
[15] State Farm Mutual Automobile Insurance Company, NAIC# 25178, displays complaints in fifty states and the District of Columbia.
[16] https://eapps.naic.org/documents/cis_aggregate_complaints_by_disposition.pdf

Final Decisions Regarding Complaints. The database provides 309 variables that represent the number of complaints in various categories and, while it does not offer definitions, many seem intuitive or self-explanatory from their titles. The first field is the total number of complaints received for the company. The "total complaints" number is less than the sum of the other 308 variables for a large fraction of insurance companies that appear in the database. The NAIC Standard Complaint Data form[17] sheds light on this result. It appears that the total number of complaints in the database is the number of actual complaint forms submitted. The form includes check boxes for the 308 variables. In some categories, the complaining entity is allowed to check up to four categories that apply. This results in a maximum of nineteen complaint categories that can be counted for one incident.

The online database for P&C insurers in 2015 includes 40,380 total complaints. When reported by line of business (complainant may choose up to four categories), there are 73,280, an average of 1.8 line categories per complaint. When reported by reason for complaint (complainant may choose up to twelve categories), there are 47,785, an average of 1.2 reason categories per complaint. Finally, when reported by disposition (complainant may choose up to three categories), there are 48,013, also giving an average of 1.2 disposition categories per complaint.

Analysis of Complaints per Interaction

In Table A.1, I present complaints per insurance company interaction with consumers and claimants. I estimate the number of interactions using data from NAIC statistical publications on homeowners and automobile insurance (NAIC 2017a; NAIC 2017b), and other sources. The number of interactions includes policy renewals, policy changes (adding and removing houses and cars), and the number of first-party and third-party claimants.

---

[17] The form is available from http://www.naic.org/documents/library_forms_cds_form_a.pdf

Table A.1: Complaints per Consumer Interaction, Automobile and Homeowners

Automobile Insurance Interactions per Complaint

				Insured Vehicles	Average Cars per Household
Auto Policy Renewals =	105,221,951	=		199,921,707	1.9
			÷	New Vehicles Purchased	Used Vehicles Purchased / Sold x 2
Auto Policy Changes =	87,780,710	=		7,780,710	80,000,000
			+	Insured Vehicles	Percent of Vehicles Filing Claims
Auto Claimants	29,988,256	=		199,921,707	0.15
			×		
Total Interactions	222,990,917				
Auto Insurance Complaints	16,880				
Interactions per Complaint	13,210	=	222,990,917	÷	16,880

Homeowners Insurance Interactions per Complaint

	Homes Insured				
Home Policy Renewals =	82,262,618				
	Houses Sold x 2				
Home Policy Changes =	9,000,000				
				Homes Insured	Percent of Homes Filing Claims
Home Policy Claims =	4,442,181	=		82,262,618	0.054
			×		
Total Interactions =	95,704,799				
Homeowners Complaints =	9,951				
Interactions per Complaint	9,618	=	95,704,799	÷	9,951

Total Insurance Interactions per Complaint (Home and Auto)

Total Interactions =	318,695,716	=	222,990,917	+	95,704,799
Total Complaints =	26,831	=	16,880	+	9,951
Interactions per Complaint	11,878	=	318,695,716	÷	26,831

The first step is to estimate the number of automobile policies that are renewed each year. I start with the most recent available (2013) number of earned exposures from NAIC (2017a).

There were 199,921,707 car years insured that year. Dividing by the average number of cars per household (1.9) yields 105,221,951 policy renewals.

Next I estimate the number of policy changes by the number of cars purchased and sold in the United States. Each time a new vehicle is purchased, someone must either purchase a policy or add a car to their policy. For used vehicles, the buyer adds a vehicle and the seller removes a vehicle, creating two customer interactions. Thecarconnection.com uses Big Data techniques to estimate 40 million used cars change hands each year. Statista.com reports almost 7.8 million new cars sold per year. Of course, there are other ways that people interact with their insurance companies. They add drivers, change limits, ask questions, and a host of other tasks. Therefore, the number of cars changing hands is a very conservative estimate of policy changes and other interactions. The total number of policy changes for adding and subtracting vehicles is 87,780,710.

Finally, I estimate the number of claims filed by applying the claim frequencies reported in NAIC 2017a to the number of insured automobile years. This yields just under 30 million claims, for a total of 222,990,917 interactions. The NAIC CIS Complaint Database includes 16,880 complaints related to automobile insurance in 2014. Dividing interactions by complaints gives 13,210 interactions per complaint.

I follow a similar process for homeowners insurance. The total number of homes insured is from NAIC (2017b). Policy changes is two times the number of homes sold, assuming the buyer and seller usually make changes to their policies. The claim frequency estimate is from the Insurance Information Institute. Total interactions sum to 95,622,537 and 9,951 complaints were reported. The quotient is 9,618 interactions per complaint.

For both homeowners and automobile insurance together, there were 11,878 interactions per complaint. In other words, a complaint is generated in less than 0.01% of consumer interactions with insurers. This is not consistent with market problems caused by Big Data or otherwise.

References

NAIC, 2017(a). Auto Insurance Database Report 2013/2014

NAIC, 2017(b). Dwelling Fire, Homeowners Owner-Occupied, and Homeowners Tenant
and Condominium/Cooperative Unit Owner's Insurance Report: Data for 2014