Advancements in the Measurement of Underwriting Risk and Diversification

By

Kenny Wunder

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Business)

at the

UNIVERSITY OF WISCONSIN – MADISON

2019

Date of final oral examination: 7/24/2019

The dissertation is approved by the following members of the Final Oral Committee:

Joan T. Schmit, Professor, Business (Co-Advisor)

J. Tyler Leverty, Associate Professor, Business (Co-Advisor)

Justin R. Sydnor, Associate Professor, Business

J. Michael Collins, Associate Professor, Consumer Science

Dedicated to My Teachers

ACKNOWLEDGEMENTS

To complete a dissertation is to conclude a chapter of life. I have been fortunate to have encountered and been supported by the most fantastic collection of individuals along my path to this point, and no words that follow can convey the depth of gratitude and love I feel for you all. Consider this section a list of IOUs, redeemable any number of times and with no expiration date.

I'll begin with those who came before me in the program, leading by example, and showing me the friendship and support that define the spirit of our department. Marc, though you've been physically away from UW for years, you are still present in so many ways. Every incoming student knows who you are and sees you as the model for what we can accomplish. Please don't stop sending Christmas cards to the office. It's not weird, despite what you think. Gee, thank you for establishing the now-accepted (nay, encouraged) practice of walking up and down the hallways of the 4th floor while contemplating challenging research questions. Doug, I'm only now beginning to understand just how ahead of the curve you were, and I'm looking forward to learning more from you in the future.

Anastasia, Kyeonghee, and Junhao, you have been with me throughout my time in the program, and your contribution to my success cannot be overstated. Anastasia, thank you for helping me through Hansen's econometrics, for laying out a plan of attack for a dissertation based on measuring something previously difficult to define, and for the many pictures and consistent emails you've sent since you graduated. Kyeonghee, you've been an absolute force from the first day I met you. With unwavering persistence and an off-the-charts work ethic, you have established a reputation here that will not soon be forgotten. Junhao, I've seen you go through the wringer time and time again, yet it never seems to faze you. Your willingness to do things like walk around the block over and over, discussing the seeming contradictions arising in insurer behavior, and your constant friendship have meant the world to me. Also, thank you for running marathons.

From my perspective, the spirit of the program has never been as strong as it is today. Youngsun, thank you for introducing me to the subtleties of chalk selection and for sharing your favorites with me. Seulkee, your determination to master a ridiculous language is more than admirable. Thank you for walking with me to get coffee, even though you don't drink coffee. Xiaochen, you're the only one in the office who gets as excited as I do about coding, algorithm selection, and working in Linux, and I'm grateful for our time discussing those things. Nii, thank you for making actuarial science look so good and remaining positive in the face of sleep deprivation. Vaibhav, you've taught me as much about appreciating life, different cultures, and food than anyone. We wrote our first paper together, and I'm already looking forward to our next. You're stuck with me, pal. Chenyuan, you are undoubtedly one of the most brilliant and inspiring people I have ever known. There is no upper bound to what you will accomplish in the field and in life. You will do great things. I'm so happy that you are now the unofficial office manager, as I need not worry about anyone or anything falling through the cracks on your watch. Lisa, aside from being a constant source of knowledge and insight, you've become one of my closest friends. You've inspired me both in and out of the office. I don't have many people in my life that would run 40 miles through the woods (at night, no less) just to ensure I didn't quit. Your ability to endure suffering with a smile is unmatched. We will have many adventures together. Thank you for everything.

There are others outside of Grainger 4260 who I consider irreplaceable. Jodi and Joni, you have played a key role in my near-daily routine, and I will strive to carry your relentless positivity and encouragement into all of my future endeavors. Jodi, I promise to vote in every election. Joni, I'll likely never be cold again, and my desk will never be without a candy bowl. Jed, Peng, and Margie, while I didn't seek you out as often as I should have, your doors were always open to me, and knowing that you would be there if I needed you has added years to my life. Dani, thank you for being interested in the coolest topics, for believing that I was capable of understanding your ideas, and for helping prepare me for my next role. I'm thankful that our time at UW overlapped. Anita and Hessam, you have taught me about managing an enormous class, dealing with diverse students, and being prepared to change direction in curriculum when it needs to happen. I will always have a selection of footwear available to anyone visiting me at home. That's just genius. Richard, you have the kind of reputation with the students that I hope to attain. Thanks for trusting me to handle things in the classroom and showing me what it means to be a real gamer. Paul H., thank you for showing me a different way to teach and a different way to learn from our students. Kirk, thank you for teaching me the value of a heart rate monitor, allowing me to visit and talk with your students, sharing with me your thoughts on what makes a successful actuarial program, making me feel loved, and slowing down your ride, so I could keep up with you for 56 miles. I want to appreciate life as you do. Don, thank you for our endless discussions on running and jazz, for being excited about my accomplishments, and for never kicking me out of your office, though I know you had work to do (and I had no official reason to be there). Belle and Patrick, you keep us all going. There is not one of us whose time at UW is not made better by your consistent work on our behalf. I apologize for being late to send you almost everything ever, and I'm thankful that you never once made me feel bad about missing a deadline.

To my committee members, I owe special thanks. Michael, your willingness to accept the task of being on my committee, your insightful and helpful questions, and your ability to identify places for me to improve are remarkable. I'm thankful for your work on my behalf. Justin, with a single, simple question, you can change the direction of anyone's thinking. You are brilliant in every sense of the word. You were encouraging when I was feeling my lowest, and you've been supportive every step of this process. Thank you. Ty, your office door has always been open to me and your patience seems to know no limits. You help me connect raw, abstract ideas to something that may actually matter to others, and you get as excited as I do about new approaches to old concepts. Your no-nonsense feedback always helped add a bit of humor to the truth of my shortcomings. "Stop rambling on here and make your point." Perhaps I will take your advice.

Joan, you've played many roles in my life over the past 8 years, and you've taught me something valuable in each of them. You've taught me the value of communicating and in working to do it better. You've taught me the value of listening not only to what is said, but what is intended. You've taught me the value of patience, both with myself and with others. You've taught me that the way I interpret the world may not be (and likely is not) the way it actually exists. You've taught me that the small things we do or say can have a significant influence on the world around us, but and our fear of how that influence may manifest can have a significant influence on us. While I've only given you a paragraph here, in the biography of my life, you’d warrant a chapter. You are the person I most admire, and the person I will strive to be.

I’ve done my best to organize this acknowledgements section in a logical way. Yet from writing this dissertation, I’ve learned that there are always a few points that are difficult to place. I’m calling this my “looking back and forward” paragraph. Sean, we’ve been friends forever, and you are at least partially responsible for everything that I’ve accomplished over the past two decades. I own you hundreds of gallons of choco-charm, at least 2.5 tons of pinwheels, and credit for recognizing long before I did that polos make great athletic wear. Thank you for always being in the background. Renea, you are the original inspiration for my desire to be an educator. You believed in me from the beginning, though you had no evidence to support that belief. From getting me math books I couldn’t afford and helping me work through them to walking me through challenging concepts in your office for hours and hours to accepting my calculus exams in colored pencil, I will never forget the things you did for me. I will work to bring to my students’ lives the humor and encouragement you have brought to mine. Looking forward, I could not be more excited. Lars, George, and Bill, we are going to be great friends. Thank you for the roles you have played in getting me pointed in the right direction. I owe you big time.

Finally, my family. Mom and Willie, I couldn’t have made it without your support, and I’ll continue to strive to make you proud of me. Only good things are on our horizon! Cathy, I have never known a stronger person, and you’re a daily inspiration to me. Mary Beth, thank you for reminding me that some people just have fabulous in their blood. Ezra, I don’t even need to thank you or ramble on about why I love you, and that’s one of the reasons I do. Megan, you are the reason for it all. You are the love of my life and the only person I could not live without. We did it.

ABSTRACT

In this work, we propose the use of a risk-based measure of diversification, the Diversification Ratio (DR), to assess insurer underwriting diversification in the U.S. Property-Liability (P-L) insurance industry. We compare DR outcomes to previously used diversification measures for the 10 largest P-L insurers, and we demonstrate that evaluation of the effect of diversification on firm performance differs greatly depending on the measure of diversification used.

Contents

1 Introduction

The concept of diversification is rooted in portfolio theory, wherein diversification occurs when the total risk in a portfolio of assets or liabilities is lower than the sum of the individual risks in the portfolio. While the business units within a firm can be considered as a portfolio of financial assets, three primary challenges exist in applying the portfolio theory framework to measure corporate diversification. The first challenge lies in the segmenting of a firm's activities into asset classes, with each class considered to be of a homogeneous asset type. The other two challenges lie in assessing the risk within each "asset class" or business unit and in determining the correlation among the units. While the literature includes efforts to address segmentation, we do not know of business diversification measures incorporating risk or correlation explicitly. The purpose of this research is to construct a measure of diversification in the U.S. Property-Liability (P-L) insurance industry that considers risk and correlation explicitly.

In this work, we use the detailed financial statements of P-L insurers to construct state-line level underwriting portfolios, which allows us to consider the product and geographic activities of insurers' underwriting business simultaneously. We demonstrate the value of such segmentation by highlighting the extensive variation in insurer loss ratios at the line of business, geographic, and state-line levels. We use this observed variation to determine the risk within each state-line and the correlation between each possible pair of state-lines, which allows us to apply the portfolio theory framework directly to measure the extent of diversification present in insurers' underwriting portfolios. Specifically, we measure the total risk in each insurer's underwriting portfolio and compare this risk to the risk that would be present in the portfolio if all segments were perfectly positively correlated, i.e., if there were no diversification occurring. This measure has been called the Diversification Ratio (DR) in the financial literature.

We then compare DR to HHI for the 10 largest P-L insurers in 2014 and identify two insurers, Liberty Mutual and Hartford, that appear equally diversified when diversification is measured using HHI, but are differentially diversified when we use DR as the measure of diversification. Using DR, we are able to identify the differences between these two insurers' portfolios that drive the difference in diversification.

Finally, we use the measure to investigate the relationship between diversification and firm performance. This relationship has been studied fairly extensively in the insurance industry. We rely upon the general structure of this literature to document how a risk- and correlation-based measure of diversification may yield different insights than prior measures of diversification. Like Villalonga (2004), we find that how diversification is measured matters. In our case, we find that increased diversification as measured by DR is associated with reduced return variability and increased risk adjusted returns, but appears not to be statistically related to overall return, itself.

2 Literature Review

From the earliest literature on portfolio theory, which begins with Markowitz (1952) and is set within the market of financial assets, we understand that diversification depends on two key elements, risk and correlation. In the decades since this work, many studies have applied diversification to areas beyond financial assets. One line of inquiry that has garnered extensive attention is understanding the effect of business diversification, that is, within-firm diversification of business activities across industries or segments, on firm performance.[1] Many of the results have suggested the existence of a diversification discount.[2] One possible explanation for this is

[1] King (1975), Berger and Ofek (1995), Lamont and Polk (2002), Elango et al. (2008), Liebenberg and Sommer (2008), Morris et al. (2017), and others.

[2] Lang and Stulz (1994), Berger and Ofek (1995), Servaes (1996), Martin and Sayrak (2003), Laeven and Levine (2007), Goetz et al. (2013), and others.

that non-diversified firms benefit from greater operational efficiencies than do diversified firms.[3] Additionally, by limiting the scope and direction of growth, firms may maximize their comparative advantage by focusing on core competencies (Berger et al., 2000), leaving them less diversified. Regardless of the varying explanations for the existence of the discount, one feature remains constant throughout this literature: any economic model used to investigate the drivers and/or effects of diversification contains a measure of diversification. As a result, the importance of measuring diversification as accurately as possible cannot be overstated.

The extent to which diversification can be properly measured is a product of both the data available and the type of measure chosen. Villalonga (2004) uses a detailed dataset allowing a more granular firm segmentation than is used in most prior studies and finds a diversification premium for firms diversifying within an industry. The results of that study indicate that the diversification discount may be an artifact of data limitation and the way that limitation restricts both firm segmentation and the measurement of the relationships among the segments, implying that the diversification measure, itself, is material to the outcomes observed.

Beyond financial asset markets, in which the data required to determine the necessary components of diversification, namely risk and correlation, are available, data limitations have resulted in the construction of a number of alternative measures. The simplest measure of business diversification used in the literature is a count of business units, with such units having been defined by their industry categorization, the products they produce, and/or the geographic regions within which they operate.[4] While counting measures offer a sense of breadth, they do not allow the recognition of the relative sizes of the units, the extent of similarity across them, or the

[3] Amit and Livnat (1988), Chatterjee and Wernerfelt (1991), Hoyt and Trieschmann (1991), Lang and Stulz (1994), Markides and Williamson (1997), and others.

[4] For examples in the insurance literature, see Lamm-Tennant and Starks (1993) and Liebenberg and Sommer (2008).

relationships among them. As an example, consider two firms, each producing only two products. Firm 1 derives 50% of its revenue from each product equally, while Firm 2 derives 99% of its sales from Product 1 and 1% from Product 2. Under a count of products, these two firms are equivalently diversified (that is, each has a "count of 2"). Yet, because we observe the proportions in each of the products, Firm 2 appears more closely related to a single product firm.

Basic proportional measures improve upon count-based measures by considering the relative sizes, or weights, of the business units within the firm.[5] While the inclusion of proportions allows us to differentiate between the two firms in the previous example, the inability to identify similarities/differences between the units and the relationships among them persist. Additionally, a list of proportions is not useful in economic models, wherein we seek to employ a single measure in order to control for the extent of firm diversification or measure the association between diversification and other variables. The exception to this lies in the unique case of only two business segments, in which case a single proportion can define the entirety of the list of proportions. In the insurance literature, examples of this bi-segmentation include categorization based on personal and commercial lines, short-tailed and long-tailed lines, and property-liability and life-health.

More advanced proportional measures, such as Herfindahl-Hirschman Index (HHI) based measures, combine the list of segment proportions into a single measure of diversification, making broader application in economic models possible. As such, HHI is the dominant measure of

---

[5] In the existing literature, these proportions have been based on revenue, cost, profit, etc. In the insurance literature, these categories have been defined as business written in particular states or particular lines, as well as broader categories such as personal and commercial lines, long-tailed and short-tailed lines, etc. Basic proportional measures are employed by Berger et al. (2000), Choi and Weiss (2005), Chen et al. (2008), Liebenberg and Sommer (2008), Berry-Stolzle, Hoyt, and Wende (2013), Morris et al. (2017), and others.

diversification used in the literature.[6] HHI is defined as the sum of the squared proportions across categories within a population. That is,

$$HHI = \sum_{i=1}^{n} {p_i}^2 \tag{1}$$

where p_i is the proportion in category i. From this equation, we can see that HHI directly answers the following question: given a population spread across a set of categories, what is the probability that choosing two individuals at random (allowing replacement) from the population results in both individuals belonging to the same category?

Beyond yielding a single diversification metric, HHI provides two additional benefits over basic proportional measures. In the context of business diversification, the squaring of the proportions in Equation 1 emphasizes the effect of a firm's uneven distribution of business activities, increasing the sensitivity of the measure to disproportional allocations. HHI also has an intuitive interpretation that allows a relatively straightforward comparison across a set of firms. A firm with an HHI of n is as equivalently diversified as is a firm consisting of $1/n$ equal sized business units.

Despite the superiority of HHI over counting and basic proportional measures, one significant issue remains: HHI is a measure of concentration, and concentration is not the same as diversification. Missing from HHI, is the variability of outcomes within each of the business units (the risk) and the measurable relationships among them (the correlation). This becomes evident upon recognizing that the use of HHI or another strictly proportional measure will yield the same outcome, whether the business units in which a firm operates are similar or dissimilar, volatile or

[6] For examples, see McCullough and Hoyt (2002), Chen et al. (2008), Liebenberg and Sommer (2008), Berry-Stolzle et al. (2012), Che and Liebenberg (2017), and Morris et al. (2017). Elango et al. (2008) uses the Shannon Entropy Measure, which is similar to HHI in construction and interpretation.

stable, related or independent. A measure of diversification must consider all of these features. If we are missing even a single one, it is not possible to distinguish between the sum of the individual risks and the risk in the portfolio.

3 Diversification Ratio

From portfolio theory, we understand that the total risk resulting from a collection of individual risk sources ("the risk of the collection," or portfolio risk) can be no greater than the sum of the individual risks ("the collection of the risks"). The difference between the two is an indication of the degree of diversification occurring within the collection.

To illustrate, the variance (or "risk") of a portfolio, ${\sigma_p}^2$, can be determined as

$$ {\sigma_p}^2 = \sum_{i=1}^{n} {w_i}^2\, {\sigma_i}^2 + 2\sum_{i=1}^{n}\sum_{j<i}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij} \tag{2} $$

where w_i is the proportion of the portfolio held in asset or liability i, σ_i is the standard deviation of the returns on asset or liability i, and ρ_{ij} is the correlation between the returns on assets or liabilities i and j. Therefore, the determination of portfolio risk requires an assessment of the standard deviation in the outcomes for each underlying component in the portfolio, the weight assigned to each of those components, and the correlations among the component outcomes. The resulting risk of the collection is equal to the collection of risks only in the case of perfectly positive correlation among all risk sources in the portfolio. In any other case, the risk of the collection is strictly less than the collection of risks. This reduction in risk is attributable to diversification, and it is through the quantification of this reduction that we can measure diversification.

In financial settings, diversification occurs through the pooling of assets having unknown future values. Importantly, diversification reduces variability in the expected value of the pooled assets; i.e. the portfolio. In insurance underwriting, where the ultimate costs associated with a policy are not known at the time the policy is issued, insurers achieve diversification by pooling the liabilities arising from policies written across products (lines of business) and geographic areas (states).[7] Though we are working with liabilities rather than assets, the risk-reducing consequence of diversification is the same.

Our approach to measuring diversification, is to measure the relationship between a collection of risks and the risk of the collection. As diversification increases, so will the difference in these risk measures. Specifically, we employ Equation 2 to measure the risk in an insurer's underwriting portfolio, and we compare it to the sum of the risks of the portfolio's component parts. As described earlier, if diversification is occurring in the portfolio, the resulting net risk in the portfolio will be lower than the collection of individual risks. This comparison is known as the *Diversification Ratio* (DR),[8] and it is a measure of how much risk reduction has been achieved through the firm's choice of component parts (state-lines in our analyses). Specifically, we can construct DR for each insurer ($\propto$) in each year (t) as

$$DR_{\propto t} = \frac{\sum_i w_{it}\sigma_{it}}{\sigma_p} \tag{3}$$

where w_{it} is the insurer's proportion of premiums written in state-line i in year t, σ_{it} is the standard deviation of loss ratios in state-line i from time t_{-5} to t_{-1}, and σ_p is as defined in Equation 2.[9]

---

[7] The fundamental pooling insurers undertake is through sales to a relatively large, homogeneous, independent set of policyholders. In our analysis, we are assuming this part of diversification and looking further to the product line and geographic diversification beyond the law of large numbers.

[8] DR has been used in the literature, although not often. See Choueifaty and Coignard (2008).

[9] Villalonga (2004) identifies the importance of correctly segmenting the business portfolio. We segment insurer underwriting portfolios on the state-line level, the smallest component distinguishable in our data. A complete description of "state-line" will follow in Section 5.

The numerator of Equation 3 is the weighted average risk of the collection of risks. It is worth noting that determining this weighted average is equivalent to measuring the portfolio risk using Equation 2 (as standard deviation, rather than variance) in the case of perfect positive correlation, a situation already identified as providing no diversification. That is,

$$\left(\sum_{i=1}^{n} {w_{it}}^2 \, {\sigma_{it}}^2 + 2\sum_{i=1}^{n}\sum_{j<i}^{n} w_{it}w_{jt}\sigma_{it}\sigma_{jt} \cdot 1\right)^{1/2} = \sum_{i} w_{it}\sigma_{it}. \tag{4}$$

The mathematical definition of DR offered in Equation 3 provides three insights into the conceptual interpretation of the measurement. First, a portfolio consisting of a single component or a multi-component portfolio in which all component outcomes are perfectly positively correlated with each other will have a DR equal to 1. Second, because the portfolio risk cannot exceed the weighted average risk in the portfolio, the lower bound of DR is also equal to 1. Third, DR increases as diversification increases. That is, an insurer with a relatively large DR is deriving relatively more risk reduction due to the unique construction of its underwriting portfolio. Specifically, the risk in an underwriting portfolio with a DR of n is lower than the average risk in the portfolio by a factor of n.

Similarly, we can interpret DR in a way that parallels the interpretation of HHI presented earlier: an insurer with a DR of n is as diversified as is an insurer writing in n^2 independent, equal-variance lines in equal proportions. This interpretation provides the same conceptual baseline we get from HHI, yet it includes risk and correlation.

4 Data

To calculate DR, we use data from the National Association of Insurance Commissioners (NAIC) database for property and liability insurers from 1991 to 2014. The database contains underwriting and financial information for all U.S. P-L insurers. This dataset is extensive, containing roughly 10,000 variables for each insurer in each year. We rely on premium and loss data at the state and line of business level, which we collect from the Exhibit of Premiums and Losses (Statutory Page 14) of the annual NAIC statement. We consider stock and mutual insurers, and our dataset includes all group and unaffiliated individual single insurers. We consider all 50 U.S. states and the District of Columbia (51 "states") across 23 lines of business.[10]

In the construction of underwriting portfolios, we consider only insurer-year-state-line observations with non-negative losses and direct written premiums greater than $100,000. Furthermore, because we use industry wide data to determine the risk within and correlation among state-line loss ratios outcomes, we need to ensure that the observed relationships between state-line underwriting outcomes indeed reflect the underlying relationships, rather than firm idiosyncratic behaviors or erroneous data. In order to do this, we fit a lognormal distribution to the observed loss ratios within each state-line-year, and we remove observations more than 2.33 standard deviations from the mean (0.5% in each tail). If a catastrophic event occurs in a particular state-line in a particular year, it likely affects multiple insurers, and although it would result in large loss ratios for participating insurers, these observations remain in our measure. If a particular firm makes a mistake or restructures its business in such a way that an abnormally large or small loss ratio is reported in a given state-line-year, our approach results in the exclusion of the loss ratio observation from the data.[11] The resulting measures, therefore, are associated with the market and not influenced extensively by idiosyncratic firm outcomes.

[10] The set of 23 lines is: Accident and Health, Aircraft, Boiler and Machinery, Burglary and Theft, Commercial Auto, Commercial Multiple Peril, Credit, Earthquake, Farmowners, Fidelity, Financial Guaranty, Fire and Allied lines, Homeowners', Inland Marine, Medical Professional Liability, Mortgage Guaranty, Ocean Marine, Other Liability, Personal Auto, Products Liability, Surety, Warranty, and Workers' Compensation.

[11] In total, we removed 46,130 state-line level loss ratios from the initial 853,478 observations, which is approximately 5.40%.

5 Extent of observed variation

As indicated earlier, existing diversification measures in the insurance literature rely on either line of business level or state level distribution of insurer activity. Studies often include two measures within an economic model, each constructed from an alternative segmentation of business units, such as geographic and line-of-business Herfindahl. The measure proposed here utilizes a more granular segmentation of insurer portfolios to limit the loss of relevant information resulting from the ex post combination of separate state and line level diversification measures. We demonstrate the value of our portfolio segmentation decision by describing the sources and extent of variation within our data and illustrate the loss of relevant detail that results from less granular segmentation.

One of the key metrics used in the assessment of insurance outcomes is the loss ratio. The loss ratio is equal to the insurer's losses (and loss adjustment expenses) attributable to a set of policies divided by the value of premiums collected when the same set of contracts was issued.[12] Loss ratios can be relatively stable or relatively unstable over time across lines of business and geographic regions. Because insurers are required by regulation to hold enough capital to pay losses resulting from the policies written with some level of confidence, greater stability in an insurer's firm-level loss ratios may allow the insurer to allocate capital across its business activities more effectively.

[12] Specifically, we use direct losses incurred divided by direct premiums written. Losses incurred includes both losses paid as well as those losses an insurer expects to pay on the policies written. While long-tailed lines have a relatively greater chance that the losses ultimately paid will differ from expectations, using losses incurred allows us to compare short and long-tailed lines on relatively equal footing, whereas paid losses would add a level of difference. Direct premiums written does not include consideration of reinsurance decisions. This means that our measure of diversification is determined prior to the purchase of reinsurance. Importantly, both the numerator and denominator of DR are made in the absence of reinsurance.

5.1 Line of business level variation

Insurers have opportunities to sell a wide variety of products. These products range from common personal contracts to pay for damage arising out of ownership of homes and vehicles to specialized commercial contracts on events related to product liability lawsuits, loss of income from property damage, compromised customer data, and many others. Importantly, these products yield differing levels of variability in loss outcomes.

Consider Figure 1, which contains histograms of insurer-year-line level loss ratios for personal auto (top) and homeowners' (bottom), two lines of business that are often grouped together in a single business unit as *personal lines*. These histograms contain all insurer-year-line level loss ratio observations that are less than or equal to 1 and occur within the full time window of the data.[13] While loss ratio observations exist beyond these boundaries, Figure 1 provides insight into the distribution of loss ratios within the usual range of observation for the two lines. The distribution of loss ratios for personal auto contains 8,124 observations with a mean of 0.60 and a standard deviation of 0.152, while the distribution for homeowners' contains 8,686 observations with a mean of 0.545 and a standard deviation of 0.210. Above each histogram is a boxplot of the same loss ratio data. Comparison of these two distributions using the 2-sample Kolmogorov-Smirnov (KS) test results in a KS statistic of 0.2211 and a p-value of 0.0000, indicating that these distributions are statistically different.[14]

The extent to which these differences exist across other lines of business can be seen in Figure 2, which is a bloxplot representation of loss ratios for all 23 lines of business. While loss ratios

[13] The occurrence of loss ratios larger than 1 are observed in our data. The purpose of Figure 1 is to illustrate the extensive differences in LR outcome variability across lines. Approximately 96% (90%) of personal auto (homeowners') loss ratios fall in this range.

[14] The loss ratio distributions constructed from all observations (i.e. not only those with a loss ratio less than or equal to 1) for the two lines are also statistically different from one another (KS statistic = 0.1889, p-value = 0.0000).

greater than 2 are observed in our data, the y-axis of Figure 2 is bounded for illustrative purposes. Approximately 98.2% of all line of business level loss ratios observed over our time period fall within this range. While prior studies often group lines of business together into broader categories, nearly all lines are statistically different from each other.[15]

5.2 State level variation within a line of business

Differences in the geographic distribution of insurers' business activities provide another substantial source of variation in underwriting outcomes. Each state in the U.S. has a unique set of regulations that affects insurance products. Each state further offers variation in laws, physical hazards, and business practices. Fortunately, these variations can be accounted for within NAIC data, which provides loss and premium data on a state basis within each line of business. The need to account for state differences within each line of business follows the rationale just described to account for line of business differences.

Figure 3 shows the distribution of loss ratios across three states (Florida, Georgia, and California) in homeowners' (top), a "personal" insurance product, and commercial multiple peril (bottom), a "commercial" insurance product, within the same 5 year time period, 2009-2013. These distributions appear quite different across states within each line of business.

To illustrate the extent of the differences across states within homeowners' (HO), we begin by fitting a distribution to the set of observed loss ratios in Florida *within* HO over this time period. To do this, we use maximum likelihood estimation (MLE) to estimate the parameters of each of

---

[15] The difference between distributions is determined using the KS test. A p-value of less than 0.1 means that the distributions are statistically different. We observe only 3 pairs of lines that are not statistically different from one another at the 0.1 level: warranty/ocean marine (p-value = 0.3260), fidelity/mortgage guaranty (p-value = 0.1110), and other liability/mortgage guaranty (p-value = 0.1488). Each of these 3 pairs contains at least one *financial* (non-standard) insurance line.

98 distributions that yield the best fit to the empirical distribution of Florida-HO loss ratios.[16] From the resulting set of distributions (and MLE determined parameters), we choose the distribution that minimizes the sum of squared errors. The observed loss ratio distribution and the best-fitting distribution for Florida-HO (a Burr distribution) are shown in the top panel of Figure 4. The empirical distribution of Georgia-HO loss ratios is shown along with the same best-fitting distribution from Florida in the bottom panel of Figure 4. A chi-squared goodness-of-fit test reveals that the best fitting distribution for Florida does not statistically fit the observed outcomes in Georgia. Repeating this process, we perform the chi-squared test against the set of observed loss ratios for each of the 51 states. The resulting χ^2 statistic and p-value for each test are provided in Table 1. From these results, we see that the distribution fits the observed loss ratios in Florida, the set used to fit the distribution initially. However, at a significance level of 10%, the best fitting distribution for Florida does not statistically fit 44 out of the 50 remaining states.[17]

To account for variation in outcomes across both lines of business and states, we construct insurer portfolios in a way that considers loss ratios on a state-line basis. In this way, we distinguish among insurers writing business in many states from those writing in only a few states. Additionally, we are able to consider *which* lines of business insurers are writing in *which* states.

To see the value in this, consider an insurer writing an equal amount of business in two states, say Florida (FL) and Georgia (GA), and an equal amount in two lines, say homeowners' (HO) and commercial multiple peril (CMP). This insurer could be writing both lines in each of the states, yielding 4 state-lines: HO in FL, HO in GA, CMP in FL, and CMP in GA. However, this insurer could also be writing only HO in FL and only CMP in GA. This scenario generates multiple

[16] This is the full set of continuous distributions available in the scipy.stats Python module as of 4 July 2019. The full set of distributions can be found at https://docs.scipy.org/doc/scipy/reference/stats.html.
[17] We also perform the comparison between Florida-HO and all other states in homeowners' using the KS test. The results indicate that *every* state-level loss ratio distribution within HO is statistically different from Florida-HO.

potential state-line portfolios, each having a different level of risk and diversification, even if we hold line of business and state concentrations constant. Considering line and geographic diversification separately does not allow differentiation of these alternative state-line portfolios.

Another benefit to working on the state-line level is that we are able to observe variation in an insurer's level of diversification within a line of business. For a single line insurer, line-level analyses would reveal that there is no diversification taking place in the insurer's portfolio. Yet, if this insurer is operating in multiple states within that line, the use of a state-line based portfolio allows us to identify that the insurer is, in fact, diversified. It also allows us to observe changes in this level of diversification over time, even if the insurer is a single line insurer throughout the entire timeframe. Simply stated, state-line considerations allow us to recognize that an insurer writing only homeowners' insurance in Georgia and Florida is diversified geographically, though it only writes in one line of business, and to measure the degree of diversification that results from operating in these two states.

6 Illustrated Example

To illustrate the implication and value of DR, we construct the measure using a sample insurer, Colonial Insurance Group in 2014. Table 2 contains the premium and loss data reported by Colonial in its 2014 annual financial statement. As described in Section 3, the underwriting portfolio is constructed from Colonial's business written at the state-line level. In 2014, Colonial wrote in one state, Texas, and in two lines, homeowners' (HO) and allied lines (AL). In addition to the statement of premiums and losses, Table 2 also contains the state-line level loss ratios (right column) for the two state-lines: TX-HO and TX-AL.

We begin by measuring the underwriting portfolio risk for Colonial using Equation 2. In this simple case of a two state-line portfolio, Equation 2 reduces to

$$\sigma_p{}^2 = w_A{}^2\sigma_A{}^2 + w_H{}^2\sigma_H{}^2 + 2\, w_A w_H \sigma_A \sigma_H \rho_{AH} \tag{5}$$

where $\sigma_p{}^2$ is the variance in the portfolio; w_A, w_H are the weights on TX-AL and TX-HO in Colonial's underwriting portfolio, respectively; σ_A, σ_H are the standard deviations of loss ratios based on industry-wide observations in the TX-AL and TX-HO state-lines; and ρ_{AH} is the correlation between loss ratio outcomes in TX-AL and TX-HO.

The state-line weights in Colonial's portfolio are determined as the proportion of total direct premiums written attributed to each state-line. From Table 2, total premiums are \$6.50 M, of which \$3.84 M comes from TX-AL and \$2.66 M from TX-HO. The state-line weights in the portfolio are

$$\begin{gathered} w_{\boldsymbol{A}} = \frac{\$\,3.84\,M}{\$\,6.50\,M} = 0.59 \\ \text{and} \\ w_{\boldsymbol{H}} = \frac{\$\,2.66\,M}{\$\,6.50\,M} = 0.41. \end{gathered} \tag{6}$$

In order to determine the standard deviation (risk) in each of the state-lines, we return to the industry-wide data. By considering the underwriting results from all insurers writing business in each of the state-lines, we are able to determine the underwriting risk in both TX-AL and TX-HO. As described in Equation 3, these state-line level risk measures, σ_A, σ_H, are determined using data from the 5 prior years, t_{-5} to t_{-1}. In this example, we use all observed loss ratios in each state-line from years 2009 to 2013.

We begin with the TX-AL state-line. The full dataset contains 483 insurer-year-state-line observations for TX-AL between year 2009 and 2013 (inclusive). In Table 3, we provide a sample of the observed underwriting results from this state-line within this timeframe. Figure 5 (top panel) is the histogram constructed from all 483 loss ratio observations. Using these loss ratios, we determine the required state-line level standard deviation via

$$\sigma_A = \sqrt{\frac{\sum\left(x_{A,i} - \bar{x}_A\right)^2}{N-1}}, \tag{7}$$

where $x_{A,i}$ refers to each individual loss ratio observation (i) occurring in state-line TX-AL, and $\bar{x}_A$ is the average of the loss ratios observed in the state-line across the 5-year window. For TX-AL in year 2014, we determine $\sigma_A = 0.6632$.

Using the same process to determine σ_H, we gather all 237 TX-HO loss ratio observations occurring between 2009 and 2013. Table 4 is a sample of these observations, and Figure 5 (middle panel) is a histogram of these state-line level loss ratios. Using Equation 7 for TX-HO, we determine $\sigma_H = 0.3082$.

We now turn our attention to the only missing value in Equation 5, ρ_{AH}, the correlation coefficient. From the bottom panel of Figure 5, in which the two loss distributions are overlaid, we can see that TX-HO has a higher probability of smaller loss ratios and a shorter tail (captured in part by the smaller standard deviation) than does TX-AL. Again, using the KS test, we conclude that these distributions are indeed statistically different (p-value = 0.0008). Correlation, however, is not about the similarity or difference between distributions. Rather, correlation reflects the similarity of *draws* from these distributions. The correlation of draws may fall anywhere between -1 and +1, even if the distributions are identical.

In order to determine the correlation, we return to the data. Using the same time period, 2009-2013, we observe each occurrence in which any insurer writes business in **both** state-lines in the same year.[18] Each of these loss ratio pairs represents an outcome (a draw), faced by the same insurer in the same year, from each of the two state-lines. There are 171 such loss ratio pairs in the data. A sample of these observations is given in Table 5. We then determine ρ_{AH} as the correlation between the resulting vectors (one for TX-AL, and a second for TX-HO) of loss ratio results using

$$\rho_{AH} = \frac{\sum(x_{A,i} - \bar{x}_A)(x_{H,i} - \bar{x}_H)}{(N-1)\sigma_A\sigma_H}. \tag{8}$$

Applying Equation 8 to the full set of loss ratio pairs sampled in Table 5, we find ρ_{AH} to be 0.1083.[19] We now have everything necessary to determine the underwriting portfolio risk for Colonial in 2014 using Equation 5. The result, σ_P = 0.4241, captures the risk of the collection. In order to measure DR for Colonial via Equation 3, we also need to find the average risk (the collection of the risks) in the portfolio. Specifically, applying Equation 4, we determine

$$\begin{aligned} &\sum_i w_{it}\sigma_{it} \\ &\quad = w_A\sigma_A + w_H\sigma_H \\ &\quad = (0.59 \times 0.6632) + (0.41 \times 0.3082) \\ &\quad = 0.5178. \end{aligned} \tag{9}$$

Since we understand diversification to be the relationship between the risk of the collection and the collection of risks, the application of Equation 3 yields the desired measure of diversification. In the case of Colonial Insurance Group in 2014, DR = 1.2199. This means that Colonial's diversification is equivalent to writing business in $1.2199^2 \approx 1.5$ independent, equal variance state-lines in equal proportions.

[18] Because variation in loss ratios captures variation across regulatory, demographic, and economic dimensions in addition to variation in the scale and magnitude of the loss events themselves (e.g. hurricane, flood), we seek a time window short enough to capture changes in these features. Yet, we require a long enough time window to provide the resulting insurer outcome data to determine the risk and correlation. We have carried out the analyses at 3, 5, 7, 9, 12, and 15-year windows as well as using all of our data to determine the underlying correlations. We use the 5-year window here for the reasons above.

[19] A general analysis of state-line / state-line correlations is presented in Section 6.

For comparison, the line level and state (geographic) level Herfindahl measures can be found using the portfolio weights alone. For the line level HHI, application of Equation 1 yields

$$HHI_{line\ level} = \sum_{i=1}^{n} p_i{}^2 = p_A{}^2 + p_H{}^2 = 0.59^2 + 0.41^2 = 0.5389. \tag{10}$$

The interpretation of this outcome is that Colonial's distribution of business activities is equivalent to writing in 0.5389^{-1} ≈ 1.85 lines of business. Similarly, at the geographic level,

$$HHI_{state\ level} = \sum_{i=1}^{n} p_i{}^2 = p_{TX}{}^2 = 1.00^2 = 1.00, \tag{11}$$

which indicates that Colonial operates in only a single state.

7 General Calculations

The example described in Section 6 serves to illustrate the framework necessary to measure diversification in insurance underwriting. The simple two-state-line portfolio requires the construction of only two individual state-line risk measures and a single correlation metric. For large, diffuse insurers, it is not uncommon to find portfolios with 800 or more state-line components. The measure of portfolio risk for an 800 state-line insurer requires, of course, 800 individual measures of state-line risk, but it also requires the determination of nearly 320,000 correlation values, because we must include a measurement of the correlation between each unique pair of state-lines in the portfolio.

7.1 A matrix approach to portfolio variance

The portfolio risk equation, Equation 2, can be written in matrix form as

$$\sigma_p{}^2 = \mathbf{W'VW} \tag{12}$$

where σ_p is the portfolio risk, $\mathbf{W}$ is a vector of portfolio weights, with one element for each portfolio component, and $\mathbf{V}$ is a variance-covariance matrix. The variance-covariance matrix can be written as

$$\mathbf{V} = \mathbf{S} \times \mathbf{P} \times \mathbf{S} = \begin{bmatrix} \sigma_1 & \cdots & & 0 \\ \vdots & \sigma_2 & 0 & \vdots \\ & 0 & \ddots & \\ 0 & \cdots & & \sigma_n \end{bmatrix} \times \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & & \ddots & \rho_{2n} \\ \vdots & & & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \sigma_1 & \cdots & & 0 \\ \vdots & \sigma_2 & 0 & \vdots \\ & 0 & \ddots & \\ 0 & \cdots & & \sigma_n \end{bmatrix} \tag{13}$$

where $\mathbf{S}$ is a diagonalized matrix of standard deviations and $\mathbf{P}$ is a correlation matrix. If we allow $\vec{S}$ to be the vector of standard deviations on the diagonal of $\mathbf{S}$, then we can rewrite the weighted average risk in the portfolio (Equation 4) as

$$\sum_i w_{it}\sigma_{it} = \mathbf{W}'\vec{S}, \tag{14}$$

and determine the Diversification Ratio (Equation 3) as

$$\mathrm{DR} = \frac{\mathbf{W}'\vec{S}}{\sqrt{\mathbf{W}'\mathbf{V}\mathbf{W}}}. \tag{15}$$

For any insurer-year portfolio, $\mathbf{W}$ is a vector of length 1173, with each element being the proportion of the insurer's annual business written in a particular state-line. For example, an insurer operating only in a single state within a single line of business, the weight vector will contain 1172 zeros and a single 1.

In order to apply Equation 15, we need two additional components: $\mathbf{S}$ (the underlying risk in the state-lines) and $\mathbf{P}$ (the correlation among the underwriting results across state-lines). We apply a three step methodology to determine these components: the creation of a data structure, the population of that data structure, and finally the determination of $\mathbf{S}$ and $\mathbf{P}$.

7.2 Creating the data structure

The initial data structure takes the form of a table, wherein each state-line is assigned both a row and a column.[20] The elements of this table are lists of numbers or lists of pairs of numbers that are initially empty but will be populated and used to determine **S** and **P**. The type of list found in each cell of the table is determined by the position of the cell within the table; either the cell is on the diagonal (where the column state-line is the same as the row state-line) or the cell is off-diagonal (where the column state-line differs from the row state-line).

Figure 6 illustrates this concept using a reduced table. Each state-line is assigned both a row and a column. Therefore, there exists in the table a row corresponding to each of TX-AL and TX-HO and a column corresponding to each of TX-AL and TX-HO.[21] The diagonal elements correspond to the intersection of the row and column for a given state-line. Two examples of diagonal elements are labeled with 1's in Figure 6. One such element is the cell in the table where the row labeled TX-AL intersects the column labeled TX-AL. The off-diagonal elements correspond to the intersection of the row/column for a given state-line and the column/row of a different state-line. One such element is the intersection of TX-AL and TX-HO (labeled as "2" in Figure 6).[22]

The lists found on the diagonal of this structure will contain all observed loss ratios from the data resulting from any insurer writing business in the corresponding state-line in any given year as well as the year in which this loss ratio was observed. This is illustrated in Figure 6 in the call-out labeled *Diagonal*. The lists off the diagonal will contain pairs of loss ratios, with one loss ratio

---

[20] We have 23 lines of business and 51 "states," which includes the 50 U.S. states and Washington D.C. Because there exists one row and one column for each state-line combination, this table is square, and each side has dimension 23×51 = 1,173.

[21] As in the example presented in Section 6, TX-AL refers to the state-line "Texas – Allied Lines" and TX-HO refers to the state-line "Texas – Homeowners."

[22] Because of symmetry, there also exists an identical element at the intersection of TX-HO and TX-AL.

appearing for the row state-line and another for the column state-line, and the year in which this pair is observed. This second list is labeled *Off-Diagonal* in Figure 6.

7.3 Populating the data structure

Population of the data structure is achieved through the sequential analysis of each insurer-year portfolio. To describe this process, we begin with Colonial Insurance Group in 2014, which is the portfolio used in Section 6. Colonial's underwriting portfolio and resulting loss ratios at the state-line level are provided in Table 2 and the bottom right of Figure 7..

For each state-line in Colonial's portfolio, the state-line level loss ratio and the year in which this result is observed are recorded in the appropriate diagonal element of the data structure shown in Figure 6. For Colonial in 2014, we observe the loss ratio for each of the two state-lines, TX-AL and TX-HO. Each of these enter the data structure on the diagonal as shown in Figure 7. To populate the off-diagonal elements in the data structure, we first construct a set containing all loss ratio pairs for Colonial in 2014, where a loss ratio pair for insurer $\propto$ in year *t* is determined by

$$(LR_i, LR_j)_{\propto t}\ , \qquad (16)$$

where LR_i is the loss ratio in state-line *i* and LR_j is the loss ratio in state-line *j* (for $j \neq i$). The full set of loss ratio pairs resulting from Colonial's 2014 portfolio contains two elements: (LR_{TX-AL}, LR_{TX-HO}) and (LR_{TX-HO}, LR_{TX-AL}). Each of the loss ratio pairs, as well as the year in which the pair was observed for the insurer is recorded in the off-diagonal of the data structure.[23] This is shown in Figure 7. The processing of this insurer-year portfolio is now complete, and we move on to the next insurer-year portfolio.

[23] Due to the symmetric nature of the problem, each loss ratio pair appears in the data twice: once in the form of (A, B) and a second time in the form of (B, A). In practice, it is more computationally efficient to record each loss ratio pair only once and take advantage of the inherent symmetry.

We select as the next portfolio to process that of Wellington in 2011. Wellington wrote in 3 state-lines in 2011: TX-AL, TX-HO, and TX-OL.[24] Figure 8 (bottom right) shows the state-line level loss ratios reported by Wellington in 2011. The resulting state-line level loss ratios for TX-AL and TX-HO are added to the data structure, augmenting the lists populated during the processing of Colonial. Because Wellington additionally writes in TX-OL, the corresponding loss ratio observed in this portfolio is recorded in the previously empty list found in the data structure where the TX-OL *row* intersects the TX-OL *column* (not shown).

Next, we construct all loss ratio pairs for Wellington in 2011 as described in Equation 16. This results in 6 loss ratio pairs, of which 3 are unique: (LR_{TX-AL}, LR_{TX-HO}), (LR_{TX-AL}, LR_{TX-OL}), and (LR_{TX-HO}, LR_{TX-OL}). The first of these pairs, TX-AL and TX-HO, is added to the off-diagonal list in the data structure containing the same state-line loss ratio pair for Colonial along with the year of observation. The other two loss ratio pairs are recorded in the proper off-diagonal lists within the data structure. This ends the consideration of the second insurer-year portfolio, and the process is repeated for all remaining insurer-year portfolios occurring in the data.

Before moving on to a description of how this data structure is used, there are several things to note. First, processing a portfolio with N state-lines will result in the addition of observed loss ratios to N diagonal elements of the data structure. Second, each of the state-line loss ratios is recorded along with the year in which the observation occurred in the data. Third, the number of off-diagonal loss ratio pairs generated when processing a portfolio does not increase linearly with the number of state-lines in the portfolio. Specifically, an insurer-year portfolio with N state-lines will generate

$$\binom{N}{2} = \frac{N!}{2 * (N-2)!} \tag{17}$$

[24] TX-OL refers to the line other liability (OL) written in the state of Texas (TX).

unique loss ratio pairs if $N \geq 2$ and 0 otherwise. Finally, to give a sense of scope to the magnitude of data contained in the completed data structure, processing our full data set yields approximately 807,000 insurer-year-state-line loss ratio observations (diagonal) and approximately 328 million unique loss ratio pairs (off-diagonal).

7.3 The determination and description of key inputs

In order to carry out the measures of underwriting portfolio risk and diversification we must determine the variance-covariance matrix ($\mathbf{V}$) as described in Equation 13. This requires both the diagonalized vector of standard deviations for all state-lines ($\mathbf{S}$) and the correlation matrix ($\mathbf{P}$).

7.3.1 State-line standard deviations, S

The determination of **S** relies on three key observations. First, **S** is a square matrix with 1173 rows and columns, one for each state-line appearing in the data. That is, the dimensionality of **S** is equal to that of the data structure created and populated in this section, and each element of **S** is mapped to exactly one list of loss ratios or loss ratio pairs in the data structure table. Second, each off-diagonal element in **S** is equal to 0, and each diagonal element (where the row corresponding to state-line *i* intersects the column corresponding to state-line *i*) in **S** will contain the standard deviation of loss ratios for state-line *i*. Finally, as was the case in the Colonial example from Section 6, **S** is based on industry wide observations (those recorded in the data structure) occurring only in the 5 years prior to the year in which the risk and diversification are measured for the insurer. This means that **S** is unlikely to be constant over time, and for each year t in the data, beginning in Year 6 (1996 in our data), we find $\mathbf{S}_t$ using only observations from the 5 previous years.

For each state-line *i*, we begin by identifying the list element in the filled data structure that corresponds to the intersection of the row and column for state-line *i*. As described previously (and sampled in Table 3) this element is a list of loss ratios occurring in state-line *i* along with the year in which each loss ratio was observed. For each year, we measure the standard deviation in the empirical distribution of loss ratios occurring within the 5-year time window leading up to year *t*. Specifically, the element in $\mathbf{S}_t$ corresponding to state-line *i*, is the standard deviation of all loss ratios occurring in state-line *i* from years t_{-5} to t_{-1}, inclusive.

In Figure 9, we provide an example of the variation in observed state-line loss ratios. This figure summarizes the loss ratios observed between 2009-2013 in each state in the homeowners' line of business. The states are presented in increasing order of mean loss ratio within homeowners', and the standard deviation of each of these distributions will fall on the diagonal of $\mathbf{S}_{2014}$. A similar collection of state-line loss ratio distributions exists for each of the 23 lines of business, and for each of the 1173 state-line distributions, we can determine the risk in the state-line as the standard deviation of loss ratios.

In Figure 10, we provide the distribution of state-line risk measures within each line of business in 2014 for a subset of lines (non-financial lines). Each of the boxplots in Figure 10 is constructed using 51 risk measures, one for each of the 51 states within the line, with each risk measure determined as the standard deviation in state-line level loss ratios experienced across all insurers between 2009 and 2013. The lines are presented in order of increasing average risk, with the average being taken across states. This figure differs from Figure 3 in that it illustrates the average risk across states within the line, while Figure 3 describes the distribution of insurer loss ratios within the lines without regard to geographic region. Larger standard deviation within these distributions implies that the underwriting experience across states within the line differs to a greater degree. We summarize these distributions for the full set of 23 lines in Table 6 for the years 1996 and 2014 (shown in Figure 10). For lines in which the variation in loss ratios across states is large, the collection of states within the line in which an insurer operates becomes increasingly relevant. The inclusion of correlations amplifies this importance.

7.3.2 State-line/state-line correlations, P

In order to determine the correlation matrix $\mathbf{P}$, we use a similar approach to that used to find $\mathbf{S}$. As is the case with $\mathbf{S}$, $\mathbf{P}$ is a square matrix with dimensionality of 1173 and varies from year to year. For each year t, we determine $\mathbf{P}_t$ using only observations from years t_{-5} to t_{-1}. However, unlike $\mathbf{S}_t$, the diagonal elements of $\mathbf{P}_t$ are all equal to 1 and the off-diagonal elements are between -1 and +1. Each off-diagonal element of $\mathbf{P}_t$ can be mapped to exactly one list element in the populated data structure. Specifically, element i,j in $\mathbf{P}_t$ corresponds to the list of loss ratio pairs in the data structure found at the intersection of the row for state-line i and the column for state-line j, with $i \neq j$ (e.g. Table 5). The value of element i,j in $\mathbf{P}_t$ is determined as the correlation between the two lists of loss ratios found in the corresponding data structure element that occur from years t_{-5} to t_{-1}.[25]

In order to provide a more meaningful descriptive analysis of the large matrix (1173×1173) of resulting correlations for a single year, we first reorganize the correlation matrix as shown in Figure 11.[26] The rows and columns are ordered by state within each line of business. From the full matrix, we take the intersection of personal auto (PA) and homeowners' (HO) as a sample matrix. This area is shaded in Figure 11. This matrix has a row for every state in HO and a column for every state in PA. Hence, it contains the full set of state-line/state-line (SLSL) correlations in which one line is HO and the other is PA. Through a description of this sample matrix, we will demonstrate

---

[25] Here we use the Pearson product-moment correlation under assumptions of normality. Alternative approaches to determining correlation may be more appropriate given the skewness of loss ratio distributions for some state-lines (Bishara and Hittner, 2015) and represent an opportunity for future work.

[26] In any of the calculations, the ordering of the rows and columns of $\mathbf{P}_t$ is arbitrary. We require only that $\mathbf{P}_t$ is symmetric and the state-line ordering is consistent with that of $\mathbf{S}_t$ and the vector of portfolio weights, $\mathbf{W}_{it}$. Here, we use this reorganization purely for descriptive purposes.

the information contained within the full SLSL correlation matrix and provide the context necessary to summarize and describe resulting DR outcomes in Section 8.

There are three important things to note regarding this sample matrix. First, this is not a correlation matrix. Rather, it is a collection of the observed SLSL correlations that exist between PA and HO that we have organized to facilitate meaningful description. Second, each element in the sample matrix is the correlation in loss ratios between one state in HO and another state (or the same state) in PA. For example, the matrix contains the correlations for (HO in California × PA in California) as well as (HO in California × PA in Wisconsin). Finally, the matrix does *not* contain cross-state correlations within a line of business. That is, (HO in California × HO in Wisconsin) is not represented. The matrix contains 2601 (51×51) elements.

Using the 2014 correlation matrix, $\mathbf{P}_{2014}$, we create the sample matrix just described. We provide a boxplot of the correlations found in the sample in Figure 12 (first panel). Each point represents the correlation between a state in HO and a state in PA, determined using all observed SLSL loss ratio pairs (as in Equation 16) from 2009-2013. The summary statistics for this distribution are provided in Table 7 in the panel labeled “Raw SLSL Correlations”

Each correlation value represented in the figure is determined empirically using Equation 8 and the set of observed pairs occurring within the timeframe. In order to consider the credibility of each of these resulting correlations, it is important to account for the number of loss ratio pairs used to determine each correlation value. Correlations that are determined using relatively few SLSL pairs may not be as credible as those resulting from many data points. The second panel in Figure 12 shows the number of observed pairs used to determine each correlation found in the boxplot. Each resulting correlation is represented by a horizontal line, and the length of the line is proportional to the number of observations used to determine the correlation. A summary of these observation counts is presented in Table 7 in the panel labeled “Credibility.”

In order to account for the impact of potentially limited data when describing the set of correlations, we weight each of the resulting correlations by the number of observations (SLSL loss ratio pairs) that were used in its determination. Panel 3 of Figure 12 is a boxplot created from these weighted correlations. The summary statistics are provided in Table 7 in the panel labeled “Weighted SLSL Correlations.” From this table, we see that the largest positive (negative) correlation in observed insurer loss ratios between HO and PA within a state-line is 0.596 (-0.407). We also see that 119 (40) loss ratio pairs were used to determine this correlation value, eliminating concern that this seemingly large correlation results from a lack of observations with the data.

7.3.3 Correlations over time

Now that we have identified the features relevant in describing a SLSL correlation matrix and used them to document the resulting correlations between states in HO and states in PA, we can observe how the correlations change over time. Temporal changes in correlation may result from significantly large loss events that are limited to a subset of lines and/or states, state-level changes in regulatory or legal environments, or changing demographics over time. That is, the underlying relationships among state-line outcomes change over time and adopting year-specific correlations allows us to capture such changes within the measure.

Using the same pair of lines, HO and PA, we construct a boxplot of weighted SLSL correlations (as in Figure 12 – Panel 3) for even years, beginning in 1996. This chart is provided as Figure 13. We provide the full set of summary statistics (as described in Table 7) over 4-year intervals in Table 8.

The set of SLSL correlations between HO and PA in Figure 13 reveals a slow upward trend from 1996-2006 followed by an extended period of relative stability. This summary viewpoint, however, conceals many relevant details about the relationships between specific pairs of state-lines. For example, some SLSL relationships result in systematically larger or smaller correlations over time, and some are more or less volatile over time than others. Figure 14 is constructed using the same data as Figure 13 without the boxing, adding instead the tracking over time of 4 specific SLSL pairs.[27] In this figure, we observe that, among the SLSL pairs: (PA in AZ × HO in NC) has the most stable correlation over time (in red), (PA in OR × HO in SD) has the most variable correlation over time (in blue), (PA in HI × HO in ND) has the smallest (most negative) average correlation over time (in green), and (PA in IA × HO in IA) has the largest (most positive) average correlation over time (in orange). When taken together, Figures 12 and 13 show that, even though the average of the SLSL correlations between PA and HO may be both close to zero and stable over time, each observed SLSL combination may be consistently small or large and may be relatively stable or volatile over time. These features illustrate the importance of measuring diversification at the state-line level. That is, binning all observed loss ratios at the line of business level before determining risk and correlation or averaging across all SLSL observations within a pair of lines results in a loss of relevant nuance only available when working on the state-line level.

The fact that the largest average correlation between a state in PA and a state in HO over the entire timeframe occurs when both lines are found in the same state is not surprising. Table 9 provides a list of the 20 largest average (over time) SLSL correlations within PA × HO. Of these 20, 12 result when both lines are offered in the same state (shaded in the table). It is worth noting that of the 2601 resulting correlations, only 51 of these are from within-state line pairing. In

[27] To further clarify, the SLSL pairs represented here are those in which one state is found in the personal auto line of business and the other in homeowners'.

contrast, the 8 relatively high interstate correlations appearing in Table 9 are unexpected and indicate opportunities for future study.

We can apply these analyses further to investigate the relationships among state-line underwriting results for other pairs of lines. In Figure 15, we provide the distributions of observation-weighted SLSL correlations wherein one line is personal auto and the other comes from each of a set of other lines. From this figure, we observe that while personal auto and homeowners' has the highest average correlation between interstate interline state-line combinations, the correlations existing between states in homeowners' and states in workers' compensation differ extensively. This implies that, for insurers writing both lines, the particular set of states in which the insurer operates within either line becomes more influential on the insurer's underwriting risk and diversification. The rightmost boxplot in Figure 15 summarizes the SLSL correlations wherein both lines of business are personal auto. We see that, relative to the correlations among state-line outcomes across lines, the within-personal-auto state-level outcomes have a much larger mean correlation. This implies that an insurer's geographic spread of business within personal auto is less effective at reducing diversification than is entering other lines of business, in general.

Figure 16 is similar to Figure 14, but it illustrates the SLSL relationships over time between PA and workers' compensation (WC). If we consider the 20 largest average (over time) SLSL correlations within PA × WC (similar to Table 9), a different pattern emerges. Table 10 shows that, of the 20 largest average SLSL correlations, 9 of them occur when the WC state is Wyoming. This observation, however, is more likely a consequence of limited empirical data than an insight into the underlying relationship between WC in Wyoming and states in PA. Figure 17 shows the number of insurer-year observations over the relevant time period (2009-2013) for each state in

WC. The four leftmost states in the chart (shaded) are the only states in the U.S. WC market with a monopolistic state fund. As a result, fewer loss ratio observations occur in the data. Importantly, while this may lead to somewhat spurious risk and correlation values for these state-lines, the reason for the limited data (few insurers operating in the state-lines) also implies that these values are used rarely when measuring risk and diversification for insurers.

8 Comparing diversification outcomes

8.1 Summarizing DR

With all component parts available, we are able to calculate DR using Equation 15 for all insurer-year portfolios from 1996 to 2014.[28] Using 1996 and 2014 as sample years, we summarize DR outcomes, and we compare DR and HHI for the 10 largest insurers by premium volume in 2014 to illustrate differences between the measures. We conclude this section by using DR to analyze an insurer portfolio in order to identify the insurer's underwriting activities that have the greatest influence on its diversification outcome.

In 1996 (2014), the mean industry-wide DR is 1.673 (1.757), the median is 1.426 (1.461), and the standard deviation is 0.773 (0.859). In Table 11, we provide a detailed breakdown summarizing DR across all insurers at 6-year intervals (intermediate years not shown to conserve space) from 1996 through 2014, the first and last years observed in the dataset. For each year, the table contains summary statistics for DR at the line of business level. These values are determined using DR for every insurer writing business in the line during the reported year.

[28] Because we require 5 years of data to determine state-line risk and correlation, and our first year of observation is 1991, the first year in which we can determine DR is 1996.

The interpretation of DR presented at the end of Section 3 adds perspective to the summary of DR outcomes presented in Table 11. In 1996, the industry mean DR is 1.673 and the average DR across only those insurers writing homeowners' is 2.101. We interpret DR in terms of the equivalent number of independent, equal-variance, equally weighted state-lines that would yield the same DR outcome. Specifically, an insurer with a DR of 1.673 (2.101) is as equivalently diversified as is an insurer writing in 2.80 (4.41) independent, equal-variance lines in equal proportions. That is, the average homeowners' insurer is approximately 57% more diversified than is the average insurer in 1996. The least diversified insurers are those writing in a single state-line, yielding a DR of 1. In contrast, Travelers Insurance Group in 2013 has the largest DR observed in our data, 5.179. This is equivalent to writing equal business in 26.82 independent, equally risky state-lines. While Travelers' DR in 2013 is 3 times larger than the mean DR for all firms in 1996, the interpretation presented here shows that this comparison yields a 13 fold difference in the number of independent, equal-variance, equally weighted state-lines.

8.2 Diversification of the 10 largest insurers

8.2.1 DR and HHI

Greater insight into how DR differs across firms can be gained through a comparison of large insurers. We provide Herfindahl and DR measures for the ten largest U.S. P-L insurers in 2014 in Table 12.[29] The table is organized in order of decreasing diversification, as measured by the line of business Herfindahl.[30] We plot these ten insurers' Herfindahl measures against their DR measures in Figure 18. Insurers become less line diversified moving from left to right in the figure, and diversification as measured by DR increases from bottom to top. The general relationship

[29] Largest as determined by total direct premiums written in 2014.
[30] Higher concentration is indicated by a lower Herfindahl.

between the two measures is observable as a downward slope from left to right in the figure. That is, DR tends to increase as HHI decreases, as one would expect. This relationship extends beyond the set of largest insurers. The Pearson correlation coefficient for DR and HHI is -0.778, and it is significant. Similarly, the correlation between DR and geographic HHI is -0.683, while it is 0.305 between geographic and line-level HHIs. Each of these correlations has a p-value of 0.0000, indicating that they are all statistically significant.

8.2.2 Concentration plots

We introduce in Figure 19 a set of concentration plots as a way to visualize and compare the underwriting portfolios for three insurers falling across the diversification spectrum in 2014: Travelers' (left), State Farm (center), and Progressive (right).[31] In these plots, the y-axis is a set of 23 bins, each corresponding to one line of business in the data. The x-axis goes from 0 to 1 and is a measure of proportion. These concentration plots provide a visual representation of concentration across three dimensions. First, the horizontal dotted lines separate the 23 lines of business into three categories. Personal lines are grouped together at the top, commercial lines form the middle grouping, and financial lines are binned at the bottom. Within each of these three categories, the proportion of the insurer's direct premiums attributed to lines within the category is plotted in pink. Similarly, the proportion of total direct premiums attributable to each line is displayed in blue. Finally, within each line of business, the proportion of the line level direct premiums for each of the 51 states is plotted as a black line.

From the concentration plots in Figure 19, we see that Progressive is the most concentrated of the three. Almost all of Progressive's business is in personal auto. In Table 12, we see that

[31] In 2014, Travelers is the most diversified firm in our data, Progressive is the least diversified, and State Farm lies near the middle.

Progressive is the most concentrated and least diversified of the Top 10. Within the personal auto line of business, however, Progressive is fairly well spread across geographies, with a few observable exceptions. The large proportion of Progressive's business that comes from personal auto implies that the correlations at play within Progressive's portfolio are those describing the relationship between states *within* the personal auto line (summarized in the rightmost boxplot in Figure 15). That is, concentration in personal auto, even in the case of low geographic concentration, leads to the inclusion of large correlation values among the state-lines in the portfolio. State Farm has a smaller Herfindahl than Progressive (0.4759 versus 0.7787), and this difference is evident in the concentration plot. From the pink bars, we can also discern that both insurers are personal lines insurers. At the other extreme, Travelers is the least concentrated of the Top 10 and is a commercial lines insurer. We can see from the blue bars that Travelers does no more than 20% of its business in any single line.

8.2.3 Comparing similarly concentrated but differentially diversified insurers

Within Figure 18, a particularly interesting relationship exists between Liberty Mutual (NAIC code: 111) and Hartford (NAIC code: 91). Both insurers have similar diversification as measured by line-level HHI. Liberty Mutual has an HHI of 0.187 and Hartford has an HHI of 0.194. Interpretation of these outcomes indicates that Liberty Mutual (Hartford) is as diversified as is an insurer writing business in 5.34 (5.14) lines in equal proportions, a difference of about 3.8%. Their DRs, however, differ to a much larger extent. Liberty Mutual has a DR of 4.713, while Hartford has a DR of 3.910. Interpretation of these insurers' DR measures as described in the beginning of this section reveals that Hartford is as diversified as is an insurer writing 15.29 independent, equally risky state-lines in equal proportions, while for Liberty Mutual it is 22.22 state-lines. That is, Liberty Mutual is 45% more diversified than is Hartford in 2014. What is it about Hartford's specific distribution of business activities that leads to this striking difference in diversification?

The concentration plots for Liberty Mutual and Hartford are provided in Figure 20. We can measure the line of business Herfindahl for each insurer using Equation 1. This measure requires only the proportion of the insurer's business in each line, and we can determine it using the data required to plot the blue bars in the concentration plots. Interestingly, the set of 8 lines with visible blue bars for Hartford is the same set visible for Liberty Mutual. So large is each insurer's proportion of business in these lines that each insurer's Herfindahl can be approximated using this subset of lines. Table 13 contains the details of each insurer's proportional distribution of business across these 8 lines and the squares of those proportions. Summation of these squared proportions (Equation 1) reveals the similarity in line of business Herfindahl measures. Therefore, it must be the case that the difference in diversification is not due to line of business concentration, but rather due to *which lines of business* and *which states within those lines* the insurers operate.

In order to identify the features of Hartford's underwriting portfolio that contribute to a lower (than Liberty Mutual's) diversification, despite its equivalent line of business concentration, we consider how risk, correlation, and concentration interact within Hartford's portfolio. We can observe within Figure 10 a summary of relative state-line level risk measures across states within each of the 8 lines of business that make up the bulk of Hartford's (and Liberty Mutual's) portfolio. From the figure, we see that workers' compensation and commercial multiple peril have greater average risk across states than do personal auto and homeowners'. Figure 10 also shows that personal auto has lower variation in state-line risk across states within the line than do homeowners', workers' compensation, and commercial multiple peril. The greater the variation in state-line risk, the more relevant is the geographic distribution of the insurer's activities within the line. That is, the more different are the state-level outcomes within a line, the more differential is the potential impact of their inclusion on the risk and diversification of an insurer's book of business. Figure 20 also allows us to observe the geographic (state-level) concentration within each line of business for each insurer.

While both insurers are similarly geographically concentrated in most lines, Hartford's geographic distribution of its workers' compensation business has a few dominant state-level sources of revenue. We provide a detailed magnification of the workers' compensation portion of Hartford's concentration plot in Figure 21. We see that Hartford provides workers' compensation insurance in almost all states. We observe that Hartford's workers' compensation business is most concentrated in California (CA), New York (NY), and New Jersey (NJ). Together the workers' compensation premiums coming from these 3 states make up just over 38% of Hartford's workers' compensation business and over 11.5% of Hartford's total business. To understand the full effect of this geographic concentration *within* workers' compensation on Hartford's total diversification, we must consider the risk in these state-lines. We must also consider the correlation among their underwriting results, both within the workers' compensation line and in other lines.

The variation in state-line level loss ratios within workers' compensation is observable in Figure 22. From the figure, we observe that CA, NY, and NJ have the 7th, 5th, and 15th highest risk among all 51 geographies, respectively. This fact is not inherently detrimental to diversification. Diversification refers to the extent that the risk in the portfolio is reduced relative to the risk that would exist in the portfolio if the loss experiences were perfectly positively correlated across all state-lines. As we will show, diversification suffers most when two heavily weighted state-lines are high risk and are highly correlated.

The correlation coefficients magnify the risk and weight combinations within the state-line portfolio. In order to observe how these features culminate in Hartford's relatively lower diversification, we return to the construction of the diversification measure. We combine Equation 2 and Equation 3 to show DR in an expanded form.

$$DR = \frac{\sum_i w_i \sigma_i}{\left(\sum_{i=1}^{n} w_i{}^2 \sigma_i{}^2 + 2 \sum_{i=1}^{n} \sum_{j<i}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij}\right)^{1/2}} \tag{18}$$

The numerator of Equation 18 is determined entirely from the weights on and risk within each state-line. This same product of weight and risk is also present in the denominator. That is, the combination of the relatively large risk in CA, NY, and NJ within workers' compensation and the relatively large weights on these state-lines plays into both the numerator and the denominator. As described in Section 3, the numerator and denominator are equivalent in the case of perfect positive correlation among all state-line loss ratio experience. Holding weights and risk constant, DR is larger when the correlations are smaller (closer to -1). Furthermore, the impact of correlation on diversification increases as the product of the weights on and risk within the two state-lines increases. That is, ρ_{ij} becomes more important in determining diversification as $w_i w_j \sigma_i \sigma_j$ increases. This results from the observation that diversification decreases as the denominator in Equation 18 increases. Identifying the state-line combinations within an insurer's portfolio that most contribute to increasing the denominator of Equation 18 is equivalent to identifying the state-line combinations in the insurer's business portfolio that most contribute to a reduction in diversification. The state-line combinations within an insurer's portfolio that play a relatively large role in reducing diversification may stem from relatively large weights, state-line risk, or SLSL correlations. If we hold state-line risk and SLSL correlation constant for all pairs of state-lines, the allocation of weights alone (i.e. the concentration of underwriting activity) drives diversification. Diversification depends on all of these features.

Hartford's underwriting portfolio in 2014 consists of 478 state-lines. Application of Equation 18 to Hartford's portfolio results in the summation of 228,484 factors in the denominator. Of these, 478 do not include SLSL correlation and are also present in the numerator. The remaining 228,006 summed factors in the denominator include correlation, risk measures, and weights. We determine the values for each of these factors and present the 10 largest in Table 14. In all 10 of the SLSL combinations in Table 14, at least one state-line is found in workers' compensation. In 9 of the combinations, both state-lines are in workers' compensation, and workers' compensation in California is present in 8 of the 10. The table also contains the values for each of the variables contributing to the SLSL denominator component. The disproportional weight on the CA-WC state-line within Hartford's workers' compensation business certainly contributes to the size of each of these factors, but this concentration component is only part of the reason Hartford's diversification is relatively lower. Also relevant is the (relatively) large risk in the CA-WC state-line observed in Figure 22. Finally, from Equation 18, we can see that the largest components in the denominator result when the weights are placed on riskier lines **and** those lines are relatively highly correlated. All three combinations of state pairs generated from CA, NY, and NJ are present in this list. From Table 14, we can conclude that the biggest driver of Hartford's lower diversification is not only its disproportional allocation of business across states within workers' compensation, but that the larger weights are placed on states with greater variability in loss ratio outcomes and that the outcomes in these states are relatively highly correlated.

9 Applying the Diversification Ratio

9.1 Diversification and firm performance: the basic model

A common model structure used to study the relationship between diversification and firm performance takes the form

$$PERF_{it} = \propto \; + \boldsymbol{\beta} DIV_{it} + \boldsymbol{\Gamma} X_{it}, \tag{19}$$

where $PERF_{it}$ is a measure of firm i's performance in year t, DIV_{it} is a vector of diversification measures, and X_{it} is a vector of firm or industry characteristics.[32] As an example, Liebenberg and Sommer (2008) (L&S) investigate the effect of operating in multiple lines of business on insurer performance, finding evidence in support of the diversification discount for insurers operating in more than a single line of business, relative to those operating in only a single line.

Their model is as follows:

$$\begin{aligned} PERF_{it} = \propto + \boldsymbol{\beta}\,(&MULTLINE_{it} + GEODIV_{it} + PCTLH_{it} + \boldsymbol{LINE_{it}} + \boldsymbol{STATE_{it}}) \\ &+ \boldsymbol{\Gamma}\,(SIZE_{it} + CAPASSET_{it} + SDROA_{it} + WCONC_{it} \\ &\qquad + MUTUAL_{it} + \; PUBLIC_{it} + \; GROUP_{it}) + \tau_t + \varepsilon_{it}, \end{aligned} \tag{20}$$

where $PERF$ is measured as return on assets (ROA); the diversification measures used include: $MULTLINE_{it}$, (the primary variable of interest) which is an indicator variable that equals one if insurer *i* operates in more than 1 line of business in year *t* and zero if it operates in only one line; $GEODIV_{it}$, defined as 1 - geographic HHI (which allows $GEODIV$ to increase as concentration decreases); $PCTLH_{it}$, the proportion of business written in Life-Health; and $\boldsymbol{LINE}$ and $\boldsymbol{STATE}$, vectors of indicator variables for the same set of 23 lines and 51 states used to determine DR. Because safer insurers can command higher prices, Equation 20 includes 2 covariates to control

[32] This form is used in: Amit and Livnat (1988), Wernerfelt and Montgomery (1988), Markides and Williamson (1997), Elango et al. (2008), Liebenberg and Sommer (2008), Goetz et al. (2013), and others.

for the influence of insolvency risk on insurer returns, $SIZE_{it}$ (firm size as log of total assets) and $CAPASSET_{it}$ (policyholder surplus divided by total assets) in the vector of firm-level characteristics.[33] Other firm-level characteristics in Equation 20 include: $SDROA_{it}$ (the 5-year standard deviation in ROA); $WCONC_{it}$ (insurer i's proportion of business written in each line multiplied by the industry concentration in the line), which is a measure of insurer i's participation in competitive lines, controls for an insurer's ability to charge higher premiums in less-competitive lines of business, potentially increasing profitability; $MUTUAL_{it}$ (an indicator equal to one if the insurer is organized as a mutual and zero otherwise); $PUBLIC_{it}$ (an indicator equal to one if the insurer is publicly traded and zero otherwise); and $GROUP_{it}$ (an indicator equal to one if the insurer is affiliated with a group). The model also includes a year fixed effect and an error term.

9.2 Modeling firm performance using DR

Because our goal is to observe the effect of diversification on insurer performance, and because our primary variable of interest, DR, is a continuous measure of diversification that accounts for both geographic and line-level activities, our model is relatively simpler and takes the form:

$$PERF_{it} = \propto\ + \boldsymbol{\beta}\, DIV_{it} + \boldsymbol{\Gamma}\,(SIZE_{it} + CAPASSET_{it} + UPR_{it}) + \tau_t + \delta_i + \varepsilon_{it}. \qquad (21)$$

---

[33] Additionally, large insurers potentially derive the variability reducing consequences of within state-line pooling across policyholders more so than do smaller insurers. The diversification across policyholders within a state-line is not captured by the diversification measure presented here, as the number of policies issued by an insurer is not provided in the insurer's financial statements. The inclusion of the within-state-line level pooling effect could be achieved if the data were available through an insurer specific modification of state-line risk elements. Insurers that offer relatively more policies in a state-line would see a relatively larger reduction in the risk in that state-line. The determination of the risk present in the state-line for a given pool size could be determined through the same process employed here, but the observed loss ratios used to determine the state-line variation would be limited to insurers with the same "level" of policy count. The correlation in loss ratios resulting across state-lines is unlikely to be affected by the pool size.

where $PERF_{it}$ is a measure of insurer performance, DIV_{it} is DR_{it},[34] UPR_{it} is underwriting portfolio risk (σ_p in Equation 2), $SIZE_{it}$ and $CAPASSET_{it}$ are as in L&S, τ_t is a year fixed effect, δ_i is a firm fixed effect, and ε_{it} is an error term. Using DR as the measure of diversification in Equation 20 reduces the number of descriptive variables needed in the model. Specifically, $MULTLINE$, $GEODIV$, $\boldsymbol{LINE}$, and $\boldsymbol{STATE}$ are subsumed by DR. Because $WCONC$ is constructed as a modified line-level proportional measure, we do not consider it in our models explicitly, though all results are robust to its inclusion. We do not have the data to include $PCTLH$, so it is absent from our model. Because we are able to measure the underwriting portfolio risk of the insurer directly, we include UPR in place of the $SDROA$ firm characteristic variable. This allows us to implement a second model using the same independent variables as the first, but use $SDROA$ as a dependent variable in order to observe the relationship between diversification and variability in insurer performance.[35] Because we are interested in controlling for, but not necessarily estimating, the effects of largely time-invariant firm characteristics (i.e., $MUTUAL_{it}$, $PUBLIC_{it}$, and $GROUP_{it}$), we include firm and year fixed effects in lieu of these variables.

9.3 Regression analyses

We present the results of our analyses in a series of panels, with each panel containing the regression results for each of 3 performance metrics ($PERF_{it}$) for insurer *i* in year *t*: ROA_{it}, return on assets; $SDROA_{it}$, 5-year standard deviation of ROA; and $RAROA_{it}$, risk-adjusted ROA, which is ROA divided by UPR. The panels differ in the diversification measure or combination of measures (DIV_{it}) used in the models. The diversification measures we consider are: $SPREAD_{it}$, determined as 1 - line of business HHI so that diversification increases as line-level concentration

---

[34] We conduct the analyses with multiple diversification measures for robustness and comparison. Our focus, however, is on DR.

[35] SDROA is the standard deviation in ROA from time t_{-5} to t_{-1}.

decreases;[36] $GEODIV_{it}$, as described above; and DR_{it}, the Diversification Ratio.[37] Reported errors are robust standard errors clustered at the firm level. The summary statistics for our variables are provided in Table 16.

Panel 1 (Table 17) contains the regression results for Equation 20, using all 3 performance metrics and $SPREAD$ as the measure of diversification. The coefficient on $SPREAD$ is negative and statistically significant in Model (1). This implies that as diversification (as measured by SPREAD) increases, firm performance decreases. This finding is in line with L&S and prior studies.[38] That is, these results support a diversification discount. The results of Models (2) and (3) indicate no significant association between diversification as measured by HHI and $SDROA$ or $RAROA$.

We then repeat these 3 regressions using $GEODIV$ as the measure of diversification and report the results in Panel 2 (Table 17). The results are similar to those shown in Panel 1. In Model (4), the coefficient on $GEODIV$ is negative and statistically significant, indicating that as geographic diversification increases, firm ROA decreases. In Model (5) the coefficient on $GEODIV$ is not statistically significant, but the coefficient is positive and statistically significant (at the 10% level) in Model (6), wherein firm performance is measured using $RAROA$. Because insurers may diversify geographically and across lines of business, we include both HHI-based diversification measures, $SPREAD$ and $GEODIV$, simultaneously in Panel 3 (Table 17). The coefficients on the diversification variables change little when both measures are included simultaneously.

[36] Line-level HHI is not included in L&S, because their interest lies in comparing diversified firms to non-diversified firms, rather than evaluating diversification on a spectrum. Because HHI is a common measure of diversification used in the literature, and our goal is to investigate the effect of using DR relative to prior measures, we also include line-level HHI in some model specifications in order to observe differences in outcomes.

[37] The largest variance inflation factor (VIF) among the full set of all independent variables used in any model is 6.09, indicating that multicollinearity among variables is not a concern in our data.

[38] Lang and Stulz (1994), Berger and Ofek (1995), Servaes (1996), Martin and Sayrak (2003), Laeven and Levine (2007), Liebenberg and Sommer (2008), Goetz et al. (2013), and others.

Collectively, Panels 1 through 3 suggest a negative relationship between diversification (when measured using line-level or state-level HHI, or the combination of the two) and ROA but find no association between diversification and the variability in ROA. We also find no association between line-level HHI and RAROA, and a weakly positive significant relationship between geographic HHI and RAROA.

We replace the HHI based measures with DR in Panel 4 (Table 17). Using DR as the measure of diversification, we find no association between diversification and insurer ROA, and we find a negative and statistically significant relationship between diversification and SDROA. We consider this result to be intuitive. Pooling (simple diversification) does not reduce expected loss, yet results in reduced average variability. As shown in Model (12), increased diversification is associated with an increase in RAROA.

Panel 5 (Table 17) shows the regression results when using the full set of diversification variables. Inclusion of both HHI measures alongside DR does not change the sign or statistical significance of the coefficient on DR in any of the models. However, including DR does result in a positive and significant coefficient on $GEODIV$ in Model (14) and a negative and significant coefficient on $SPREAD$ in Model (15).

To appreciate the economic value of these results, consider the median insurer with a DR of 1.474. This insurer is as diversified as is an insurer writing in 2.173 independent, equal-variance state-lines in equal proportions. The addition of one more such state-line to the median insurer's underwriting portfolio would result in a DR of 1.781, a change of 0.307.[39] The coefficient on DR in Model (11) tells us that this change is associated with a reduction in *SDROA* of -0.00086. To

[39] With an initial DR of 1.474, the median insurer is essentially writing in $1.474^2 = 2.173$ independent, equal-variance state-lines in equal proportions. Here, we consider the change in DR resulting from the insurer's participation in one additional independent, equal-variance state line to the extent that an equal proportion of the insurer's business is written in that state-line. Such a change in for the median insurer would yield a DR = $\sqrt{(2.173+1)}$ = 1.781.

generate an equivalent change in SDROA using firm size, *SIZE* would need to increase by 0.064. For the median firm, this is equivalent to growth in total assets of approximately $4.3 million, or 6.6%. The comparison of *DR* and *SIZE* is a natural one, because the negative relationship between insurer size and variability in ROA may also be the result of increased diversification.

We summarize the variation in firm-level DRs over time by firm size in Table 18. From the table, we see that smaller insurers are on average less diversified and have less variation in their diversification levels over time. Table 18 also includes the number of firm-year observations within each decile (based on *SIZE*) having a DR of 1, indicating participation in only a single state-line, as we expect single state-line insurers to be less likely to have varying DRs over time. Because standard errors may be inflated when many observations in the sample have DRs that do not change over time, we re-estimate Equation 21 using sub-samples based on firm size. The results are presented in Table 19.

In Table 19, Model (1) is the same as Model (13) in Table 7 and is included to aid comparison to the other models in the table. In Model (2), the sample is comprised of the largest 90% of firm-year observations. The coefficient on *DR* is positive and significant at the 0.10 level. Relative to Model (1), the coefficient is about 50% larger and the standard error is about 75% smaller, suggesting that the large standard error on *DR* in Model (13) is indeed driven by the inclusion of small insurers facing little variation in DR over time. In Models (3) - (6) in Table 19, we reduce the data further. In each model, we remove the bottom 10% of the observations (by *SIZE*) from the previous model. In Model (3), which uses only the top 80% of firm-years, the coefficient on *DR* is significant at the 0.05 level, and in Models (4) - (6), the coefficient is positive and significant at the 0.01 level. In Models (1) - (5), the coefficient on *DR* increases in magnitude, while the standard error on the coefficient decreases. While moving from Model (1) to Model (6), which includes only the upper 50% of all firm-year observations, the coefficient tends to increase in addition to increasing significance level, suggesting that larger firms derive greater performance benefit (as measured by ROA) from increasing diversification.[40]

When using DR as the measure of diversification, our findings are inconsistent with those that find support for the diversification discount. When measuring diversification as previously proposed in the literature, the diversification discount persists.

10 Conclusion

Using a rich and detailed dataset to overcome the data limitations present in many prior studies, we construct a measure of diversification, the Diversification Ratio (DR), in the U.S Property-Liability insurance industry that accounts for the interplay between geographic and line of business activities and includes both the risk and correlation present in an insurer's underwriting portfolio. Through a comparison of prior diversification measures used in the literature for the 10 largest insurers in 2014, we verify that decreased concentration of business activities is associated with increased diversification, in general. Using two of these firms, we also illustrate that this general relationship need not hold in all cases, thus highlighting one potential drawback of using concentration measures as proxies for diversification.

To demonstrate the value added by DR in addressing questions specifically related to diversification, we apply the measure to the basic Liebenberg and Sommer (2008) model. In their study of the effects of diversification on firm performance in the U.S. P-L insurance industry, Liebenberg and Sommer find evidence that insurers operating in a single line of business

---

[40] The coefficient on *DR* in Model (6) is approximately equal to that in Model (5), though they are similar in magnitude. Continuing the analyses through all remaining size deciles (not shown) results in consistently positive and significant *DR* coefficients, suggesting that the firm-year observations from the smallest insurers are driving the large standard errors and relatively smaller coefficient magnitudes. Similar outcomes are observed using quartiles (rather than deciles).

outperform insurers operating in multiple lines. We find that the application of DR results in a simplification of the model, as DR encompasses several dimensions of insurer activity. Using DR, rather than prior measures, results in an empirical outcome that differs from any previous study we have seen in insurance. Specifically, we find that greater diversification in an insurer's underwriting portfolio is not associated with a decrease in firm performance, a relationship identified in the literature as the diversification discount. Furthermore, when limiting our sample to larger insurers, we find evidence of a diversification premium. We also find that greater diversification is associated with lower variability in firm performance (as measured by the standard deviation in ROA) and with higher risk-adjusted ROA.

While an improved measure of underwriting risk and diversification can shed new light on outstanding questions, the measure proposed in this work allows us to address questions that simply cannot be answered using any prior measure of diversification that we have seen. One example that could be explored is the effect on diversification of insurer expansion into new states or lines of business. By comparing the diversification outcomes that would result from each of a selection of potential expansions for an insurer, we can determine which specific expansion decision would result in the greatest diversification. Using proportional measures, all expansions are identical. A second example involves the relationship between reinsurance usage and diversification as complementary mechanisms for risk management. As was the case in the prior example, addressing such a question requires a measure that can differentiate between specific insurer underwriting activities.

In addition to being an explicit point of study, diversification is a staple control variable in economic models investigating organizational form, capital structure and utilization, regulation, agency, insurance demand, and many other topics. The value of an improved measure of firm diversification in these settings is extensive. Beyond this, the intermediate measures of risk within and correlation among state-lines in the insurance industry provide additional fertile ground for future study. While we understand that catastrophic loss events, legal and regulatory environments, consumer demographics, and industry market structure all influence the relationships among state-line loss ratio outcomes, the measurement of those relationships (as correlations) is a necessary first step in assessing the influence of these factors.

The larger the number of unique state-line pairs (relative to industry wide observations) a firm-year portfolio contains, the less accurate is the measurement. When specific state-line pairs are rarely observed, the data used to determine the correlations is reduced. The manifestation of this data limitation is of greater concern when the weights placed on those state-lines within an insurer's underwriting portfolio is large. The smaller are the weights on those state-lines within the portfolio, the smaller is the effect on the measure.[41] This is less of a concern when measuring DR for existing portfolios, as the data limitation itself arises only when these combinations rarely occur, but should be addressed when measuring diversification in counterfactual portfolios that are constructed in ways that do not reflect the observed behavior of insurers.

When a small number of large (small) insurers is present in a state-line dominated by small (large) insurers, the risk determined in that state-line may not accurately reflect the risk faced by the minority set of insurers. This occurs because insurers writing a large number of policies within a state-line benefit more from the law of large numbers than do insurers writing fewer policies. Therefore, when the risk in the state-line is determined using observations of mostly "large n" insurers, the resulting risk measure may understate the risk faced by insurers writing relatively fewer policies. However, due to the role of correlation in the measure of portfolio risk, the effect

[41] As observed in Equation 2, each correlation appears in the measure of portfolio risk multiplied by the product of the weights in the two state-lines.

that this would have on the resulting measurement of an insurer's diversification is unclear, because it depends on the other state-lines present in the specific insurer-year portfolio.

In the literature, it is common for studies to use business concentration measures, such as the count of business segments (lines or states of operation) and product and geographic Herfindahls, as proxies for diversification. While such measures are often necessary due to data limitations, these measures do not account for the extent of the similarities across business segments, nor do they capture the relationships that exist among those segments. We improve upon these measures by including risk and correlation, moving from measures of concentration to a direct measure of diversification. Despite the limitations of the measure outlined above, DR represents a significant advancement in the measurement of business diversification for insurers, aligning the measurement of business diversification with the portfolio theory based concept of diversification, increasing confidence in the results of economic models employing diversification as a control, and opening new opportunities for study within insurance and beyond.

Bibliography

Amit, Raphael, and Joshua Livnat. “Diversification Strategies, Business Cycles and Economic Performance.” *Strategic Management Journal* 9, no. 2 (1988): 99–110.

Berger, Allen N., J.David Cummins, Mary A. Weiss, and Hongmin Zi. “Conglomeration versus Strategic Focus: Evidence from the Insurance Industry.” *Journal of Financial Intermediation* 9, no. 4 (2000): 323–62. https://doi.org/10.1006/jfin.2000.0295.

Berger, Philip G., and Eli Ofek. “Diversification’s Effect on Firm Value.” *Journal of Financial Economics* 37 (1995): 39–65.

Berry-Stölzle, Thomas R., Robert E. Hoyt, and Sabine Wende. “Capital Market Development, Competition, Property Rights, and the Value of Insurer Product-Line Diversification: A Cross-Country Analysis.” *Journal of Risk and Insurance* 80, no. 2 (2013): 423–459.

Berry-Stölzle, Thomas R., Andre P. Liebenberg, Joseph S. Ruhland, and David W. Sommer. “Determinants of Corporate Diversification: Evidence From the Property-Liability Insurance Industry.” *Journal of Risk and Insurance* 79, no. 2 (2012): 381–413. https://doi.org/10.1111/j.1539-6975.2011.01423.x.

Bishara, Anthony J., and James B. Hittner. “Reducing Bias and Error in the Correlation Coefficient Due to Nonnormality.” *Educational and Psychological Measurement* 75, no. 5 (2015): 785–804. https://doi.org/10.1177/0013164414557639.

Chatterjee, Sayan, and Birger Wernerfelt. “The Link between Resources and Type of Diversification: Theory and Evidence.” *Strategic Management Journal* 12, no. 1 (1991): 33–48.

Che, Xin, and Andre P. Liebenberg. “Effects of Business Diversification on Asset Risk-Taking: Evidence from the U.S. Property-Liability Insurance Industry.” *Journal of Banking & Finance* 77 (2017): 122–36. https://doi.org/10.1016/j.jbankfin.2017.01.004.

Chen, Xuanjuan, Helen Doerpinghaus, Bing-Xuan Lin, and Tong Yu. “Catastrophic Losses and Insurer Profitability: Evidence from 9/11.” *Journal of Risk and Insurance* 75, no. 1 (2008): 39–62.

Choi, Byeongyong Paul, and Mary A. Weiss. “An Empirical Investigation of Market Structure, Efficiency, and Performance in Property-Liability Insurance.” *Journal of Risk and Insurance* 72, no. 4 (2005): 635–673.

Choueifaty, Yves, and Yves Coignard. “Toward Maximum Diversification.” *The Journal of Portfolio Management* 35, no. 1 (2008): 40–51. https://doi.org/10.3905/JPM.2008.35.1.40.

Elango, B., Yu-Luen Ma, and Nat Pope. “An Investigation into the Diversification–Performance Relationship in the US Property–Liability Insurance Industry.” *Journal of Risk and Insurance* 75, no. 3 (2008): 567–591.

Goetz, Martin R., Luc Laeven, and Ross Levine. “Identifying the Valuation Effects and Agency Costs of Corporate Diversification: Evidence from the Geographic Diversification of U.S. Banks.” Review of Financial Studies 26, no. 7 (2013): 1787–1823. https://doi.org/10.1093/rfs/hht021.

Hoyt, Robert E., and James S. Trieschmann. “Risk/Return Relationships for Life-Health, Property-Liability, and Diversified Insurers.” *The Journal of Risk and Insurance* 58, no. 2 (1991): 322. https://doi.org/10.2307/253240.

King, Alan L. “The Market Performance of Diversified and Non-Diversified Organizations within the P-L Insurance Industry.” *The Journal of Risk and Insurance* 42, no. 3 (1975): 471. https://doi.org/10.2307/251702.

Laeven, Luc, and Ross Levine. “Is There a Diversification Discount in Financial Conglomerates?” *Journal of Financial Economics* 85, no. 2 (2007): 331–67. https://doi.org/10.1016/j.jfineco.2005.06.001.

Lamm-Tennant, Joan, and Laura T. Starks. “Stock versus Mutual Ownership Structures: The Risk Implications.” *Journal of Business* (1993): 29–46.

Lamont, Owen A., and Christopher Polk. “Does Diversification Destroy Value? Evidence from the Industry Shocks.” *Journal of Financial Economics* 63, no. 1 (2002): 51–77.

Lang, Larry H. P., and René M. Stulz. “Tobin’s q, Corporate Diversification, and Firm Performance.” *Journal of Political Economy* 102, no. 6 (1994): 1248–1280.

Liebenberg, Andre P., and David W. Sommer. “Effects of Corporate Diversification: Evidence from the Property–Liability Insurance Industry.” *Journal of Risk and Insurance*, 75, no. 4 (2008): 893–919.

Markides, Constantinos, and Peter Williamson. “Related Diversification, Core Competences and Corporate Performance.” *Resources, Firms and Strategies: A Reader in the Resource-Based Perspective*, (1997).

Markowitz, Harry. “Portfolio Selection.” *The Journal of Finance*, 7, no. 1 (1952): 77. https://doi.org/10.2307/2975974.

Martin, John D., and Akin Sayrak. “Corporate Diversification and Shareholder Value: A Survey of Recent Literature.” *Journal of Corporate Finance*, 9, no. 1 (2003): 37–57.

McCullough, Kathleen A., and Robert E. Hoyt. “Implications of Corporate Diversification and Focus Strategies.” In *Risk Theory Seminar*, (2002): 1–24.

Morris, Brandon CL, Stephen G. Fier, and Andre P. Liebenberg. “The Effect of Diversification Relatedness on Firm Performance.” *Journal of Insurance Issues* 40, no. 2 (2017): 125–158.

Servaes, Henri. “The Value of Diversification during the Conglomerate Merger Wave.” *The Journal of Finance* 51, no. 4 (1996): 1201–1225.

Villalonga, Belen. “Diversification Discount or Premium? New Evidence from the Business Information Tracking Series.” *The Journal of Finance* 59, no. 2 (2004): 479–506.

Wernerfelt, Birger, and Cynthia A. Montgomery. “Tobin’s q and the Importance of Focus in Firm Performance.” *The American Economic Review*, (1988): 246–250.

Figures

Figure 1: Histogram of Loss Ratio Outcomes

This histograms presented in this figure are constructed from all line of business level loss ratios (horizontal axes) observed within 1991-2014 that are less than or equal to 1. Loss ratios larger than 1 are observed in the data and are included in subsequent figures. The purpose of this figure is to provide a snapshot within the range of "usual outcomes" and illustrate that the line-level loss ratio distributions are heterogeneous. The top panel is the distribution of loss ratios for personal auto, the bottom panel is for homeowners'. The y-axes are the percentages of observations falling in each of the 50 bins. Although these distributions are clearly different from one another, prior measures of diversification do not account for these differences. A boxplot summarizes each histogram (directly above each). The circle marks the mean loss ratio in the distribution.

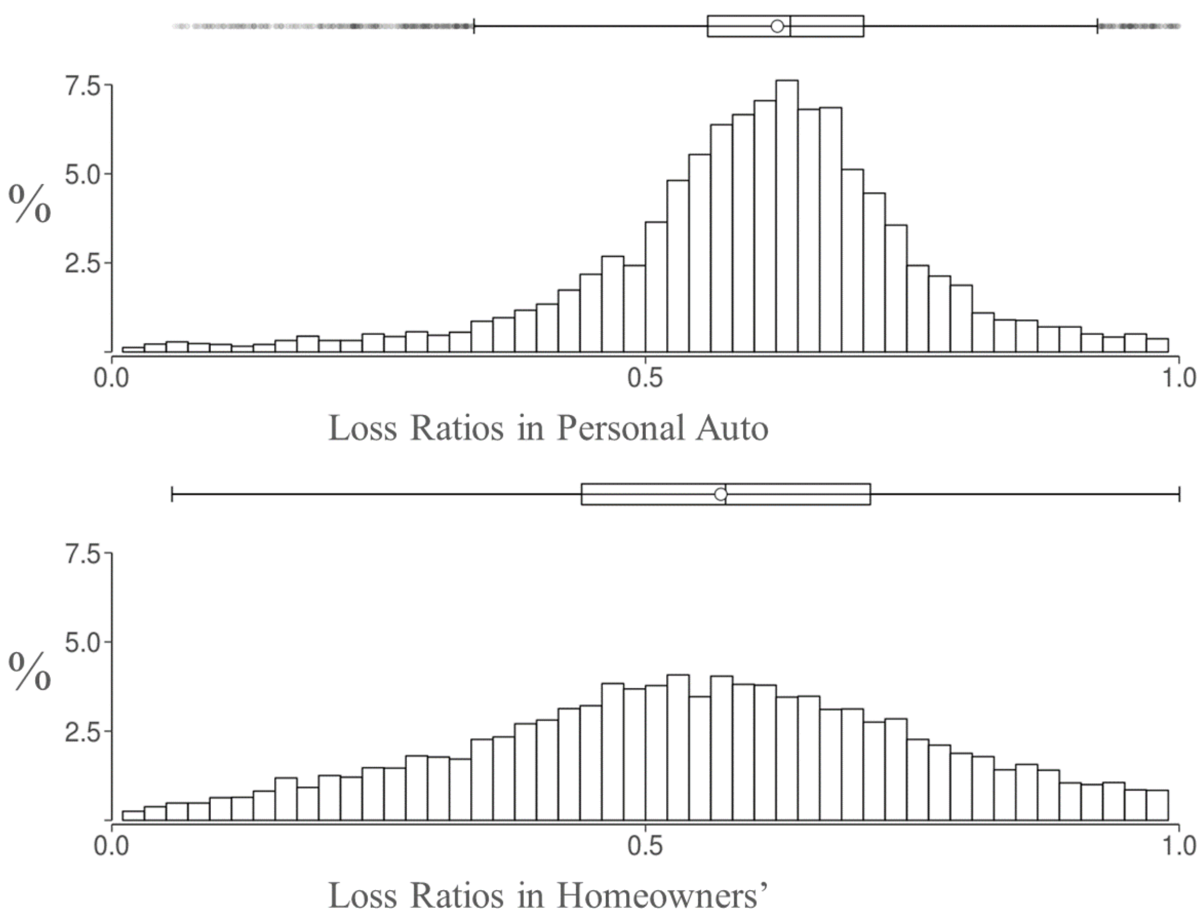


Figure 2: Distribution of Loss Ratios Across Lines of Business

This figure is a boxplot of the distribution of loss ratios within each line of business. While loss ratios greater than 2 are observed in our data, this figure is bounded to illustrate the differences in loss ratios across lines of business that fall in the common range of outcomes. The 2 lines of business summarized in Figure 1 are shaded in this figure for comparison. The lines are labeled in the figure as: *air* is Aircraft, *allied* is Fire and Allied lines, *bm* is Boiler and Machinery, *bt* is Burglary and Theft, *cauto* is Commercial Auto, *cmp* is Commercial Multiple Peril, *credit* is Credit, *earth* is Earth Movement, *fidelity* is Fidelity, *fin* is Financial Guaranty, *fo* is Farmowners, *ho* is Homeowners, *inland* is Inland Marine, *mpl* is Medical Professional Liability, *ocean* is Ocean Marine, *oliab* is Other Liability, *pauto* is Personal Auto, *product* is Products Liability, *surety* is Surety, *war* is Warranty, and *wc* is Workers' Compensation.

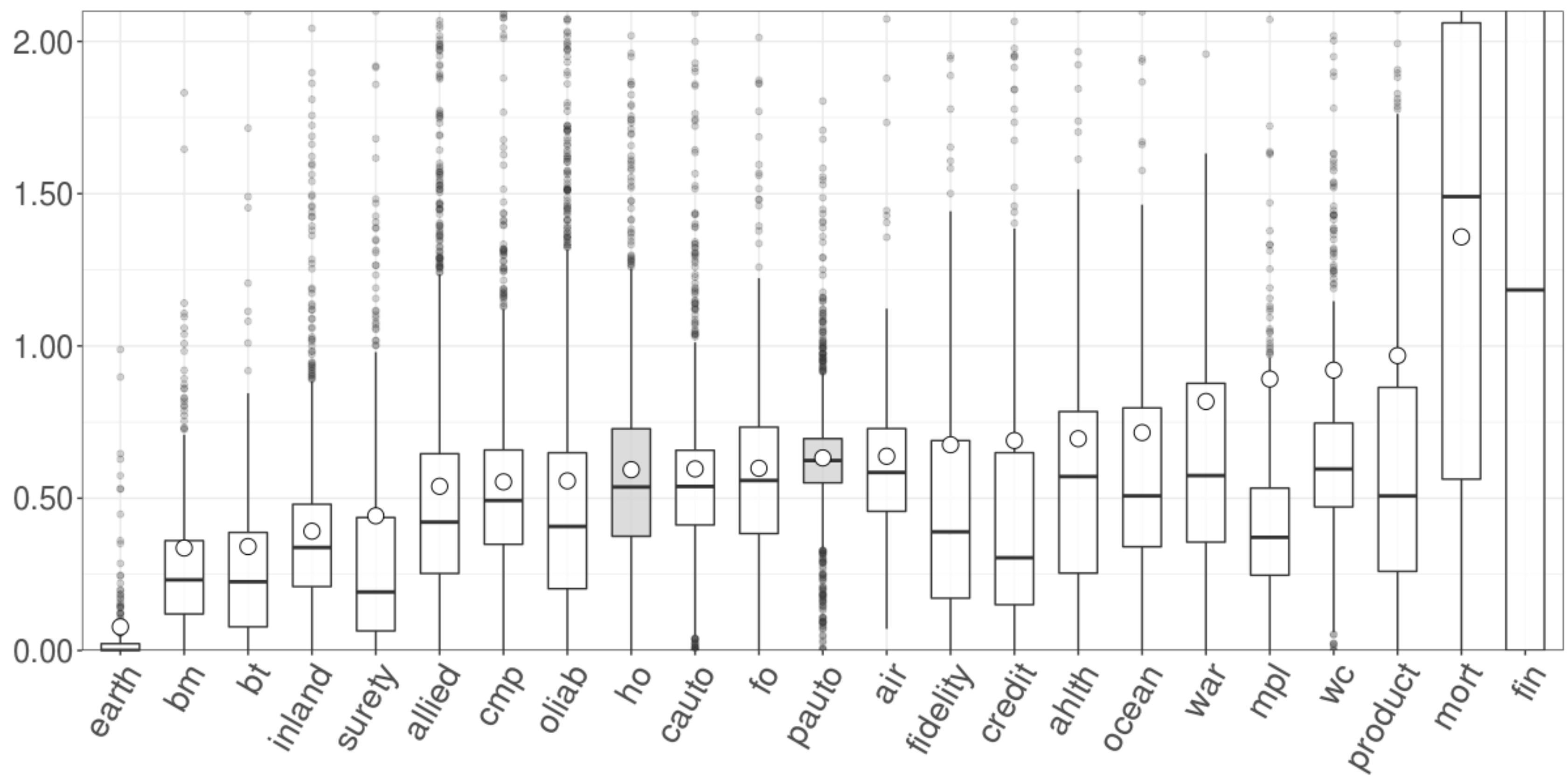
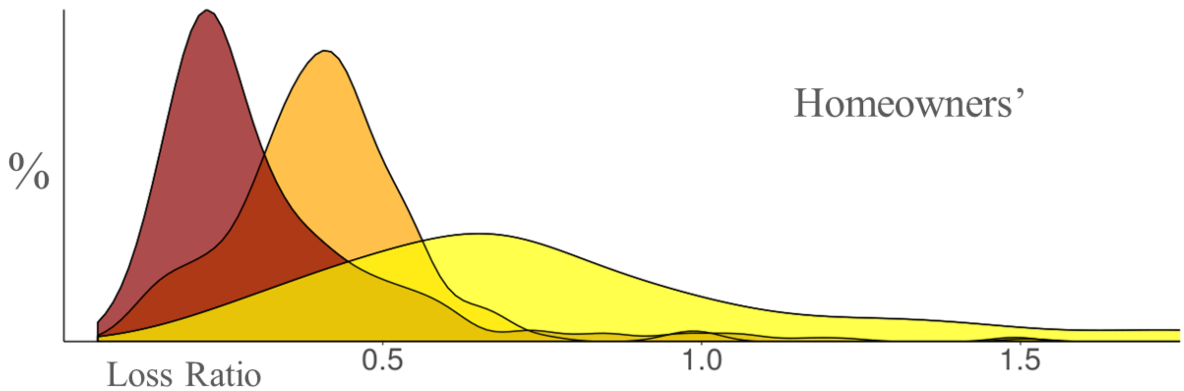


Figure 3: Geographic Variation of Loss Ratios

This figure illustrates the geographic variation in loss ratios observed across 3 states: California (orange), Florida (red), and Georgia (yellow) in 2009-2013 using density plots. The horizontal axes are loss ratio outcomes. The top panel is constructed from the loss ratio outcomes for the homeowners' line of business in each of the three states, while the bottom panel is for commercial multiple peril.



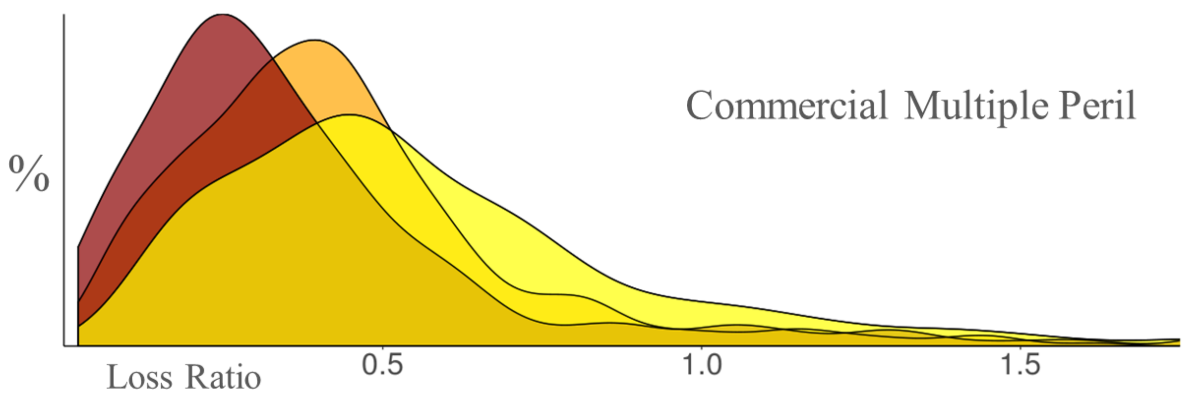


Figure 4: Differences Across States Within Homeowners'

This figure shows the extent of the differences in loss ratio distributions within homeowners' for Florida and Georgia. The top panel contains the empirical distribution of loss ratios in Florida in homeowners' (bars) and the best fit distribution (red line). The bottom panel shows the same best-fitting distribution, a Burr distribution (red line), and the empirical distribution for loss ratios in Georgia in homeowners' (bars). We use a chi-squared goodness-of-fit test to determine that the observed loss ratios in Georgia homeowners' do not come from the same distribution as Florida homeowners'. This process is repeated for all states within homeowners' against this same best-fit distribution, and the p-values and χ^2 test statistic from each test are presented in Table 1.

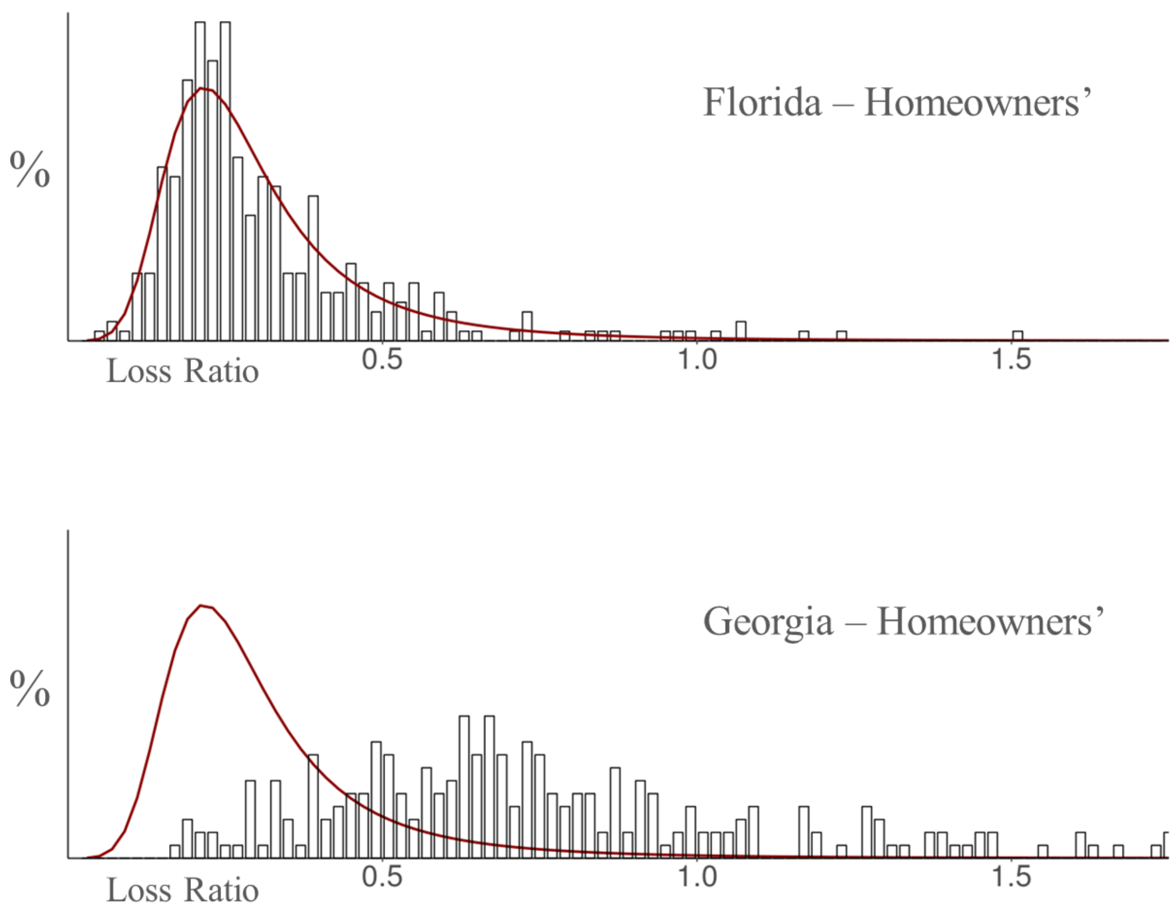


Figure 5: State-Line Comparison of Loss Ratio Outcomes (2009-2013)

This figure provides histograms and density plots for loss ratio outcomes in 2 state-lines: Texas-Allied Lines (top) and Texas-Homeowners' (middle). The horizontal axes represent loss ratios. The bottom panel is an overlaid density plot for the 2 state-lines (Texas-Allied Lines in red, and Texas-Homeowners' in blue). Though the distributions look similar, correlation is about the relationship existing between draws from these distributions, rather than the similarities existing between the distributions, themselves.

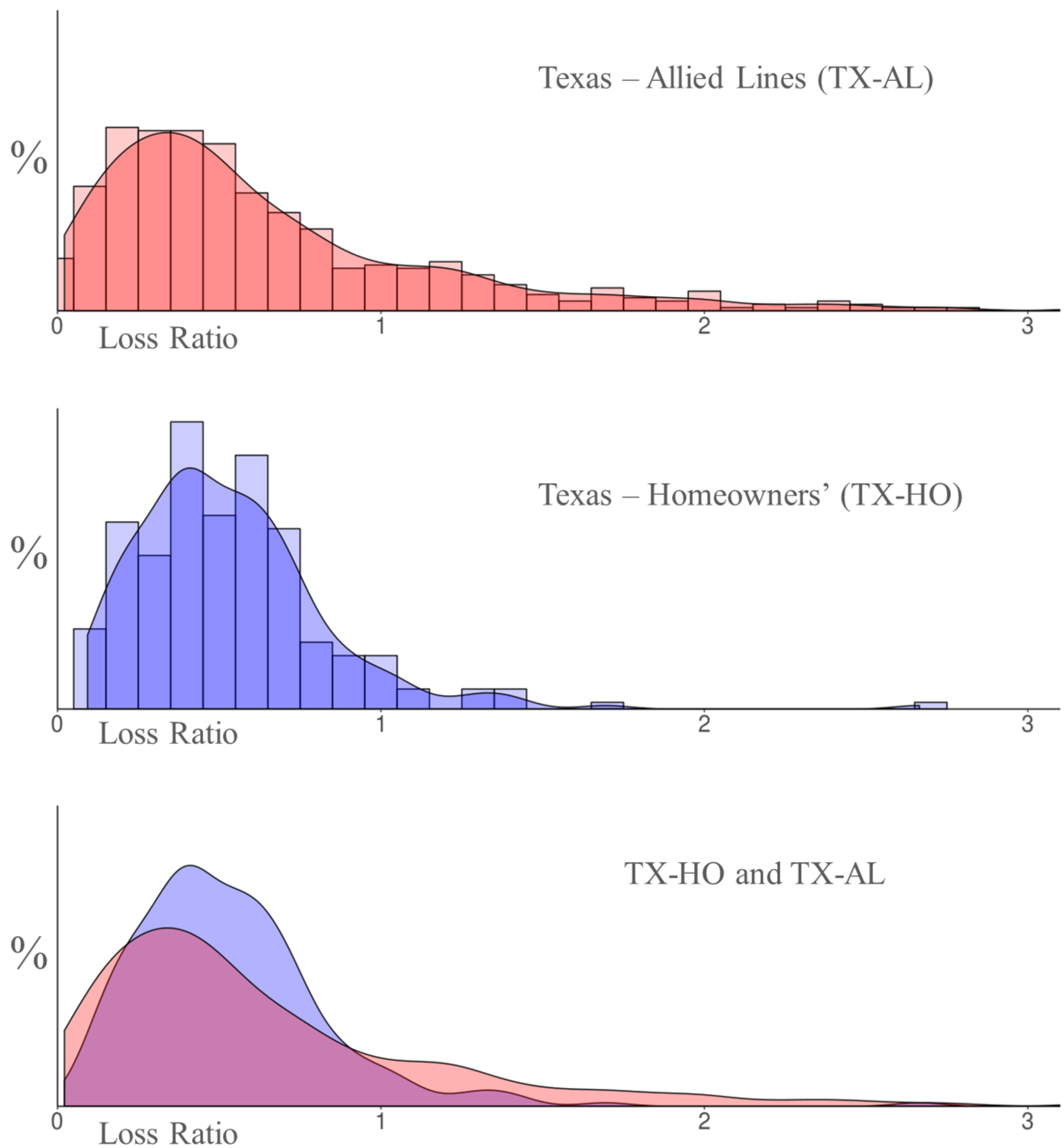


Figure 6: Organization of the Data Structure

As described in Section 6, the full data structure utilized in the matrix approach to the determination of the Diversification Ratio and its intermediaries, correlation and risk, contains 1173 rows and 1173 columns, each corresponding to one state-line. The order of the rows (columns) is arbitrary in the calculations, provided that the columns (rows) take the same order. Consequently, each diagonal element is the intersection of the row and column assigned to a particular state-line. There are two types of elements in the data structure: diagonal elements which hold loss ratios observed for individual insurers and the year the loss ratio was observed (labeled as "1"), and off-diagonal elements which hold loss ratio pairs (Equation 16). In this simplified example, the diagonal elements shows correspond to Texas-Allied Lines (TX-AL) and Texas-Homeowners (TX-HO). The off-diagonal element labeled "2" will contain loss ratio pairs wherein one loss ratio is found in TX-AL and another in TX-HO, and both will come from the same insurer in the same year.

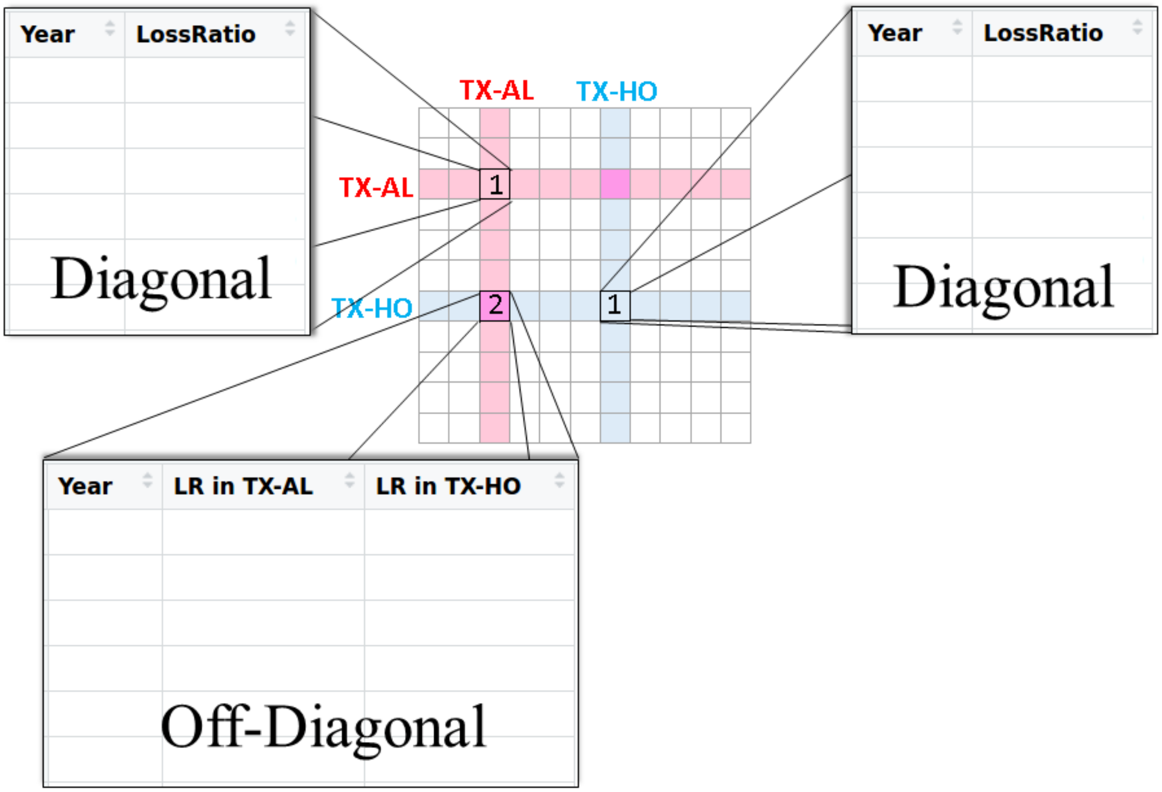


Figure 7: Processing the Data – First Insurer

Following the creation of the data structure, each insurer-year portfolio is processed. This figure illustrates the procedure for processing a single insurer. For illustrative purposes, the first insurer processed is Colonial Insurance Group, which is the insurer used in the example described in Section 5. Colonial operates in 2 state lines in 2014. The 2 state lines in which Colonial operates, Texas-Allied Lines (TX-AL) and Texas-Homeowners' (TX-HO), can be seen alongside Colonial's resultant state-line loss ratios for the two state-lines in the bottom right of the figure. The TX-AL loss ratio is recorded on the diagonal (where the TX-AL row meets the TX-AL column). The same is done for TX-HO. The loss ratio pair (Equation 16) resulting from Colonial's operation in precisely 2 state-lines is recorded on the off diagonal where the TX-AL column intersects the TX-HO row. Due to the symmetry of the structure, a second off-diagnal element exists wherethe TX-HO column intersects the TX-AL row (not labeled).

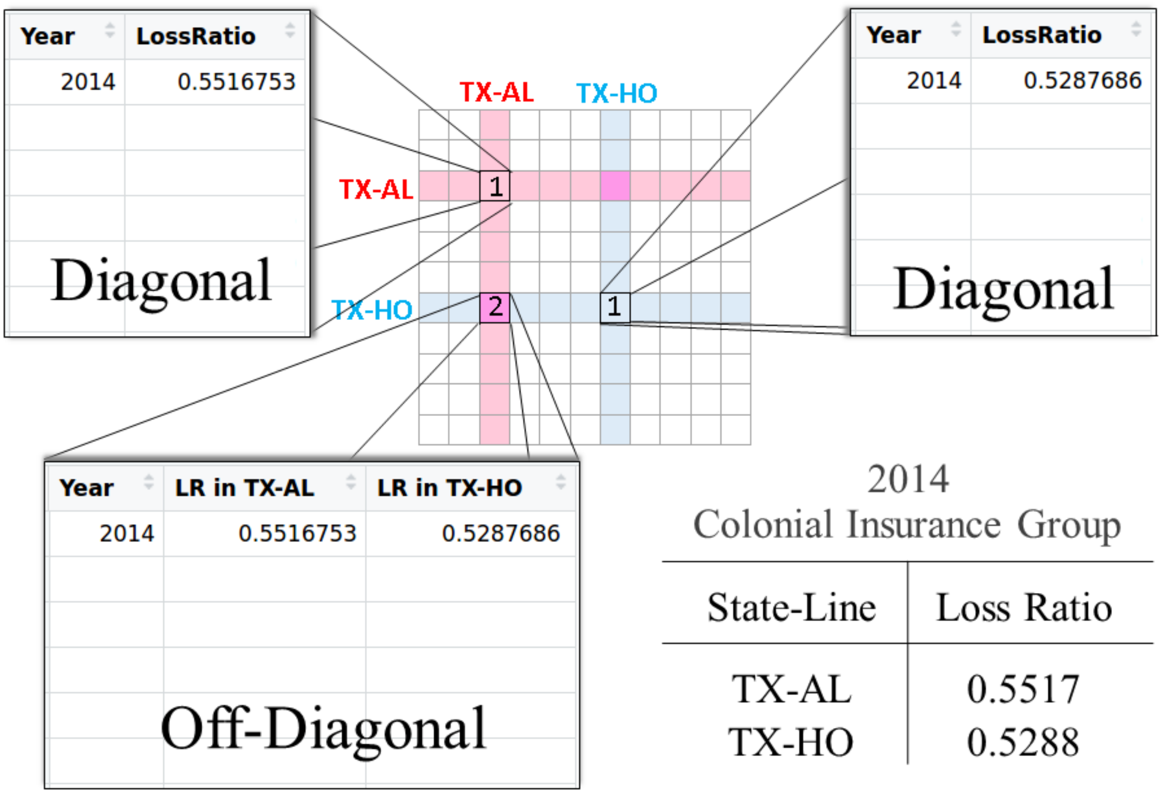


Figure 8: Processing the Data – Second Insurer

In this figure, we illustrate the processing of a second insurer. Here, we use the state-line underwriting portfolio of Wellington (the insurer) in 2011, which is provided in the bottom right of the figure. Wellington, like Colonial (the example firm used in Section 5 and described in Figure 7), operates in both Texas-Allied Lines (TX-AL) and Texas-Homeowners' (TX-HO). These two state-lines are processed as they were for Colonial. The loss ratio resulting within each state-line is individually recorded in the proper diagonal element of the data structure, adding to the loss ratios recorded when Colonial was processed, and the loss ratio pair resulting from these two state lines (TX-HO / TX-AL) is added in the off-diagonal element. Wellington, however, operates in a third state-line, Texas-Other Liability (TX-OL). This loss ratio is recorded in the data structure where the TX-OL row meets the TX-OL column (not shown). Additionally, 2 additional state-line / state-line pairs are observed with the addition of the third state-line: TX-HO / TX-OL and TX-AL / TX-OL. The resulting loss ratio pairs (Equation 16) are recorded in the correct off-diagonal elements (not shown). Hence, the processing of a 3 state-line portfolio yields 3 state-line loss ratios and "3 choose 2" loss ratio pairs (Equation 17).

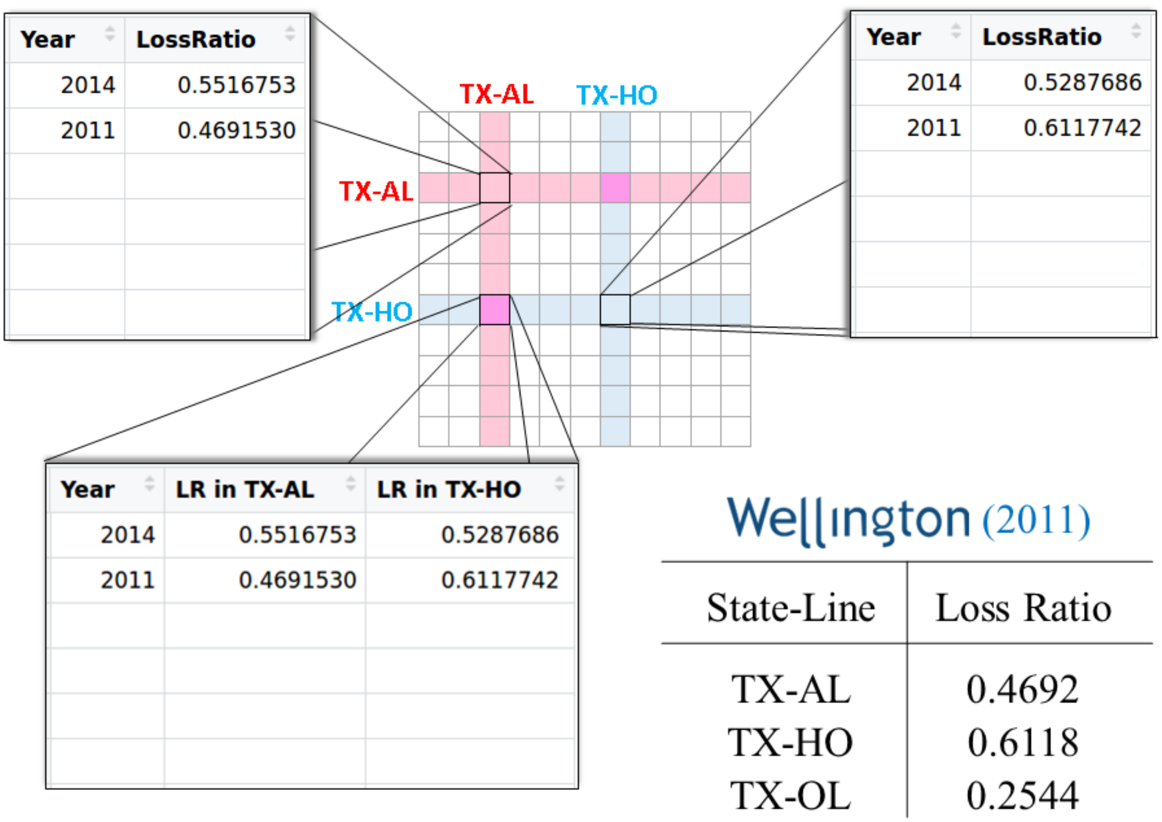


Figure 9: State-Line Level Loss Ratios in Homeowners' by State (2009-2013)

This figure provides the state-line (SL) level loss ratios observed in homeowners' insurance (HO) grouped by state. Loss ratios are displayed on the vertical axis, with states on the horizontal axis. Each data point corresponds to one observed SL level loss ratio in HO. The mean of each SL distribution is marked with a circle. The states are organized from left to right in order of increasing average loss ratio. These loss ratios are used to determine the SL risk in HO within each state. For example, the standard deviation of the observations that are summarized by the leftmost box (Hawaii) is the state-line risk for Hawaii-Homeowners' in 2014.

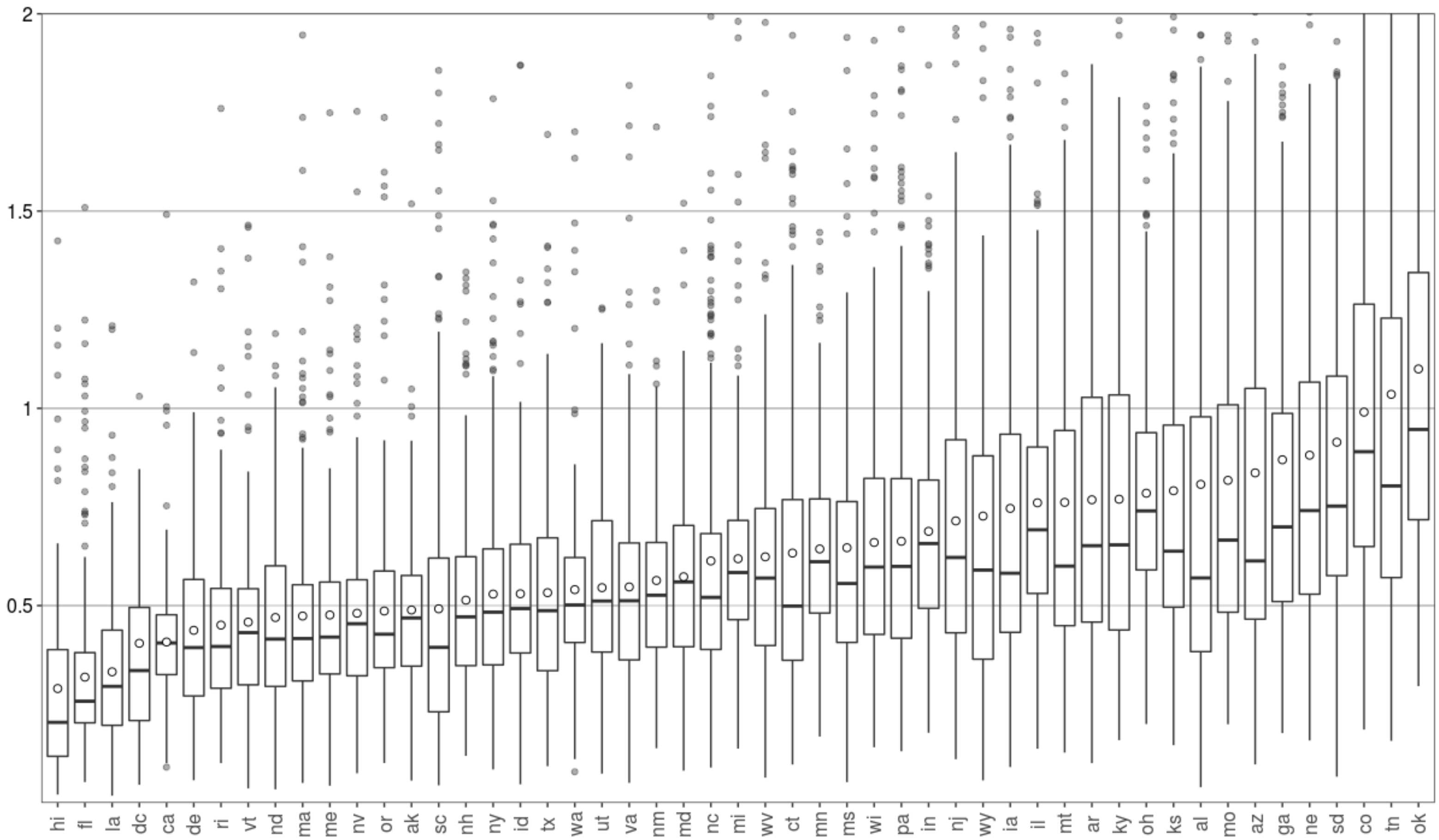


Figure 10: State-Line Risk Measures by Line of Business (2014)

This figure illustrates the variation in state-line (SL) level risk measures. Within each line, the SL level risk measures (standard deviation of insurer-year-state-line loss ratios observed within 2009-2013) from each of the 51 states are collected and are summarized by boxes. The lines are organized from left to right in increasing mean risk. The variation in SL risk *within* each line of business is indicative of the importance of considering in *which* states insurers operate within the line. For example, there is relatively little variation in the SL risk measures within personal auto, indicating that the variation in loss ratios within the line across states is relatively low. This does not necessarily imply that particular state participation profiles matter less for personal auto; such a determination requires observation of both SL risk (pictured) as well as correlation. In this figure: *air* is Aircraft, *allied* is Fire and Allied lines, *bm* is Boiler and Machinery, *bt* is Burglary and Theft, *cauto* is Commercial Auto, *cmp* is Commercial Multiple Peril, *earth* is Earthquake, *fo* is Farmowners, *ho* is Homeowners, *inland* is Inland Marine, *mpl* is Medical Professional Liability, *ocean* is Ocean Marine, *oliab* is Other Liability, *pauto* is Personal Auto, *product* is Products Liability, *wc* is Workers' Compensation.

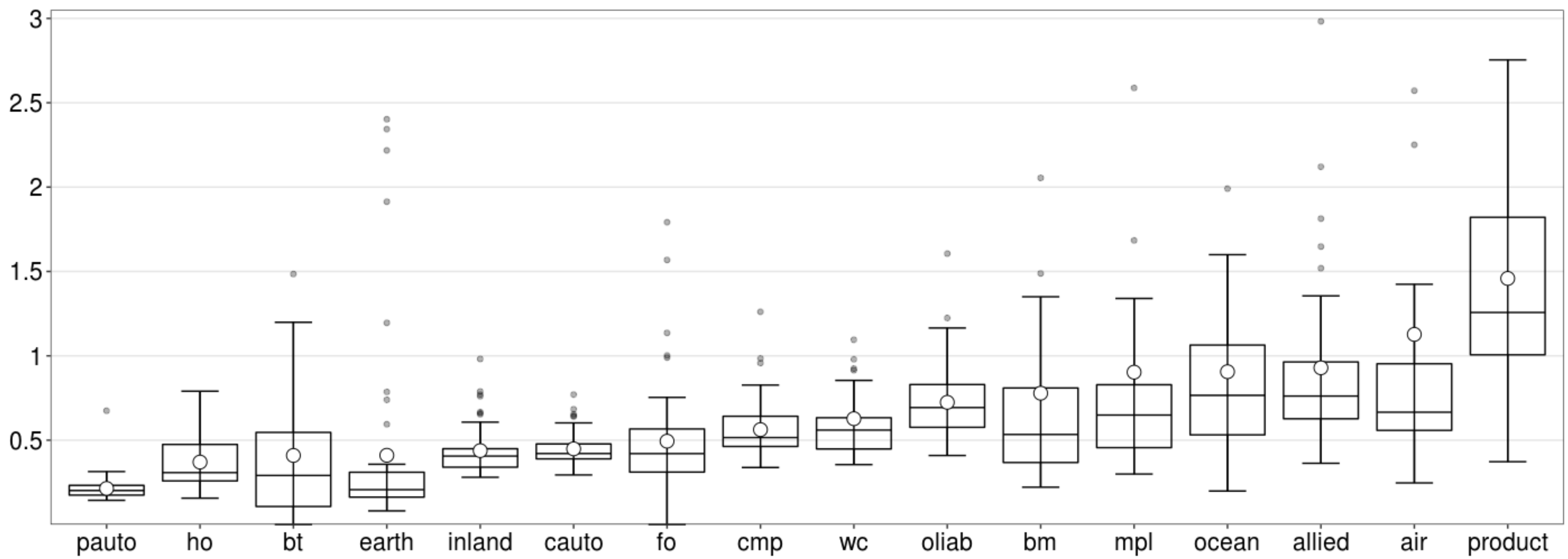


Figure 11: Organizing the Correlation Matrix for Descriptive Purposes

The full state-line / state-line (SLSL) correlation matrix is 1173×1173. There is one row and one column for each state and line combination (51 states and 23 lines of business). All calculations can be done regardless of the ordering, so long as the matrix is symmetric. For the purposes of describing the correlations in the full matrix, we create a subset at the intersection of (all states in) personal auto and (all states in) homeowners'. This section is shaded in the figure. The result is not a correlation matrix, but a collection of SLSL correlations, wherein one SL is a state in personal auto and the other is in a state in homeowners'. The states may be the same or different. Due to symmetry, this subset exists twice in the matrix. The choice of homeowners' and personal auto is somewhat arbitrary, and we explore other line/line combinations later in subsequent analyses.

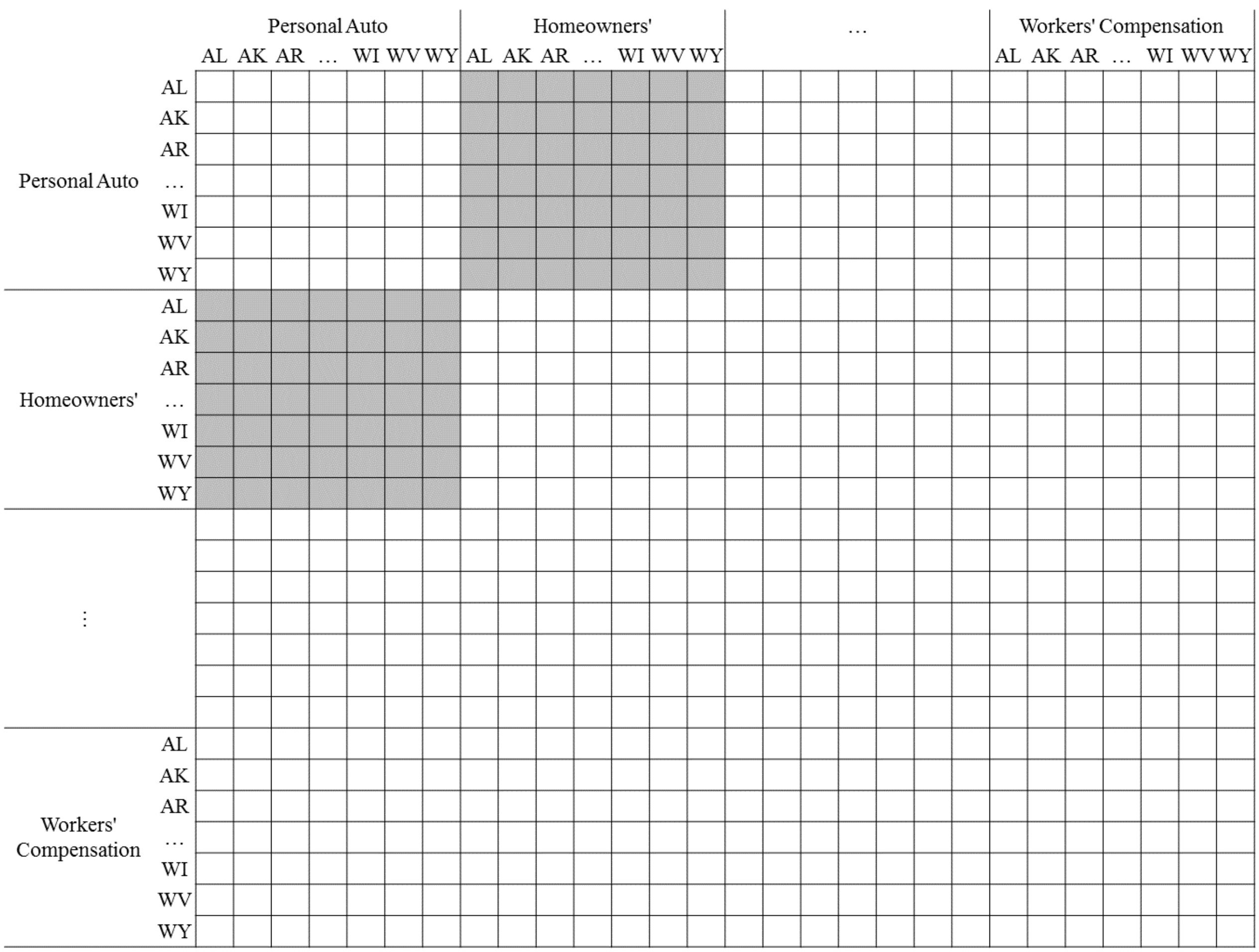


Figure 12: Visualization of State-Line / State-Line Correlations

This figure contains 3 panels, each illustrating a different aspect of state-line / state-line (SLSL) correlations observed wherein one line is personal auto and the other is homeowners'. Panel 1 (Raw SLSL Correlations) is a boxplot constructed of the SLSL correlations in 2014. Summary statistics for this chart are found in Table 7, in the panel of the same name. Panel 2 (Credibility) illustrates the number of SLSL pairs (Equation 16) that were used in determining each SLSL correlation in 2014. One horizontal bar is found at each resulting correlation, and the horizontal length of the bar is based on the number of pairs. The number of pairs is then used to weight the observations in Panel 1, resulting in Panel 3 (Weighted SLSL Correlations). The vertical axes of all panels is correlation.

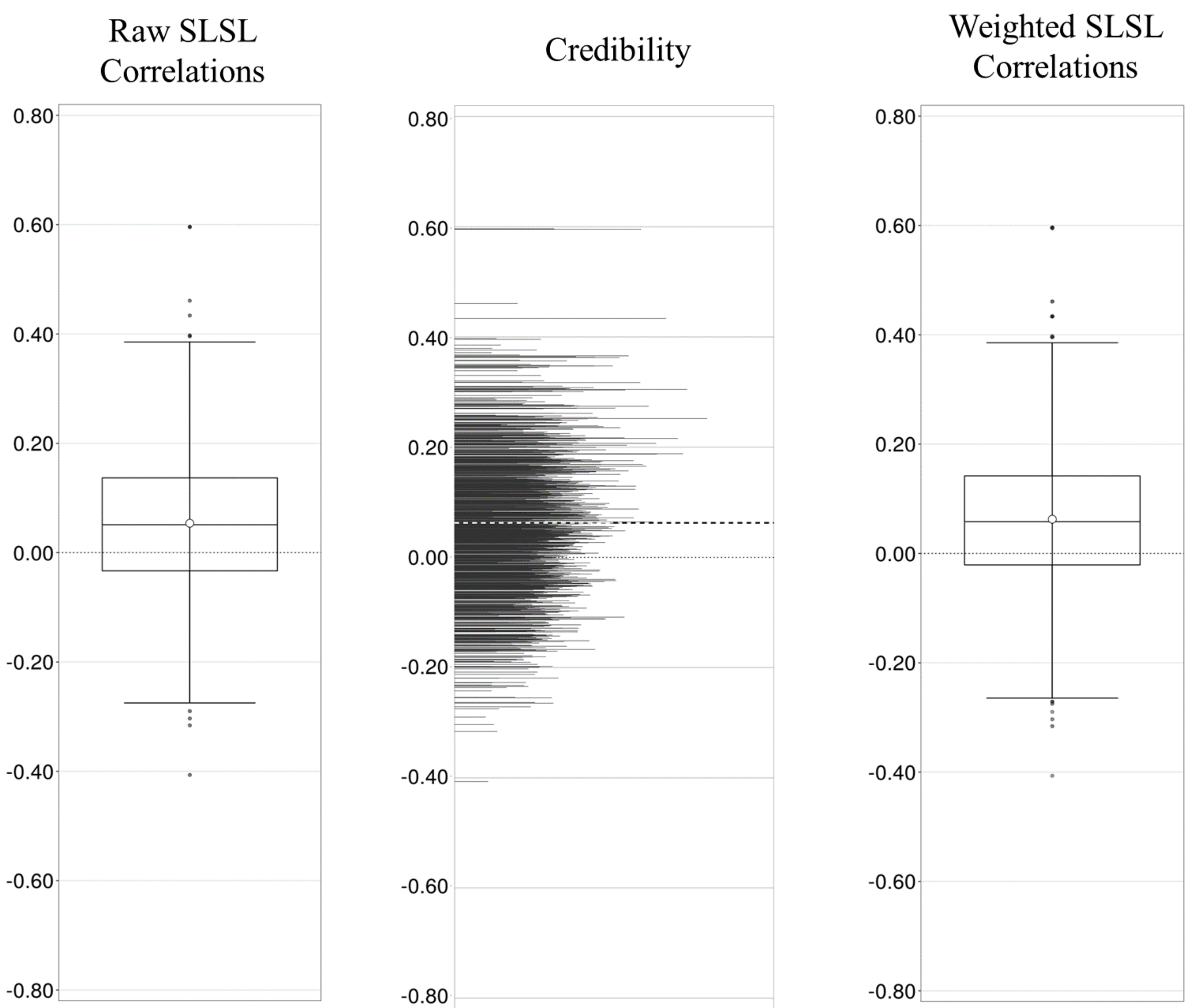


Figure 13: State-Line / State-Line Correlation for HO and PA Over Time

This figure contains a series of weighted state-line / state-line (SLSL) correlation boxplots (see Figure 12 for meaning) wherein one line is homeowners' and the other is personal auto. Each point represents a single SLSL correlation, determined from all insurer-year-state-line loss ratios from the 5 years prior to the year displayed. All even numbered years from 1996-2014 are provided here. Summary statistics for every 4th year are provided in Table 8. Variation in correlations over time is expected, as the loss ratios used are influenced by the occurrence of loss events as well as demographics and legal and regulatory environments within the state, all of which are subject to change over time.

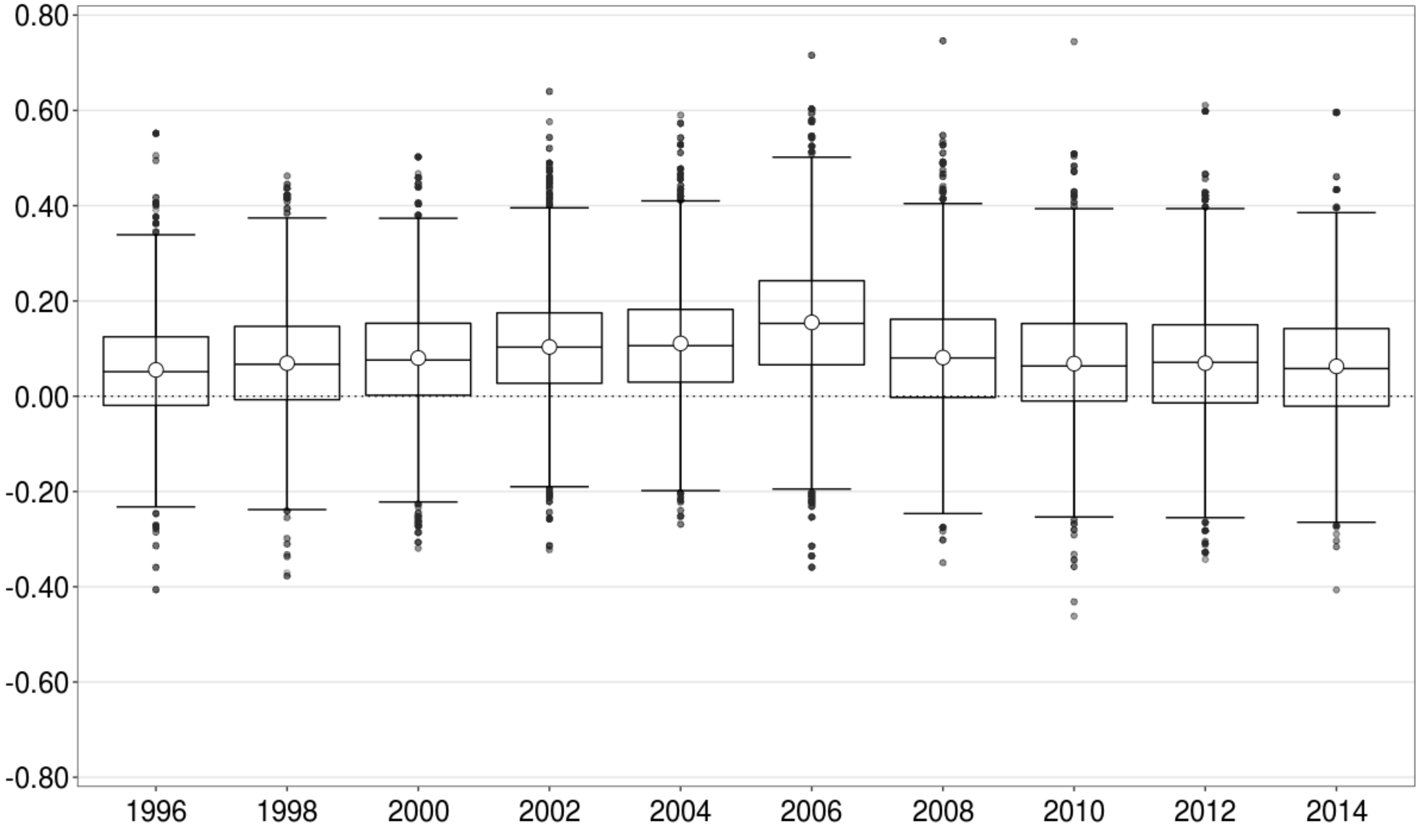


Figure 14: Tracking State-Line / State-Line Correlation Pairs for HO and PA Over Time

This figure is constructed in the same manner as is Figure 13. However, rather than summarize the observed state-line / state-line (SLSL) correlations with boxes, we track 4 particular SLSL combinations over time. Each correlation in the figure exists between one state in personal auto (PA) and another (or the same) state in homeowners'. Arizona-PA and North Carolina-HO (in red) has the most stable correlation over the entire time period (1996-2014), while Oregon-PA and South Dakota-HO (blue) have the most variable. Hawaii-PA and North Dakota-HO (green) have the lowest average correlation, and Iowa-PA and Iowa-HO (orange) have the largest average correlation. As described in Section 5 and further shown in Table 9, many of the largest positive SLSL correlations in PA×HO occur when both lines are seated in the same state.

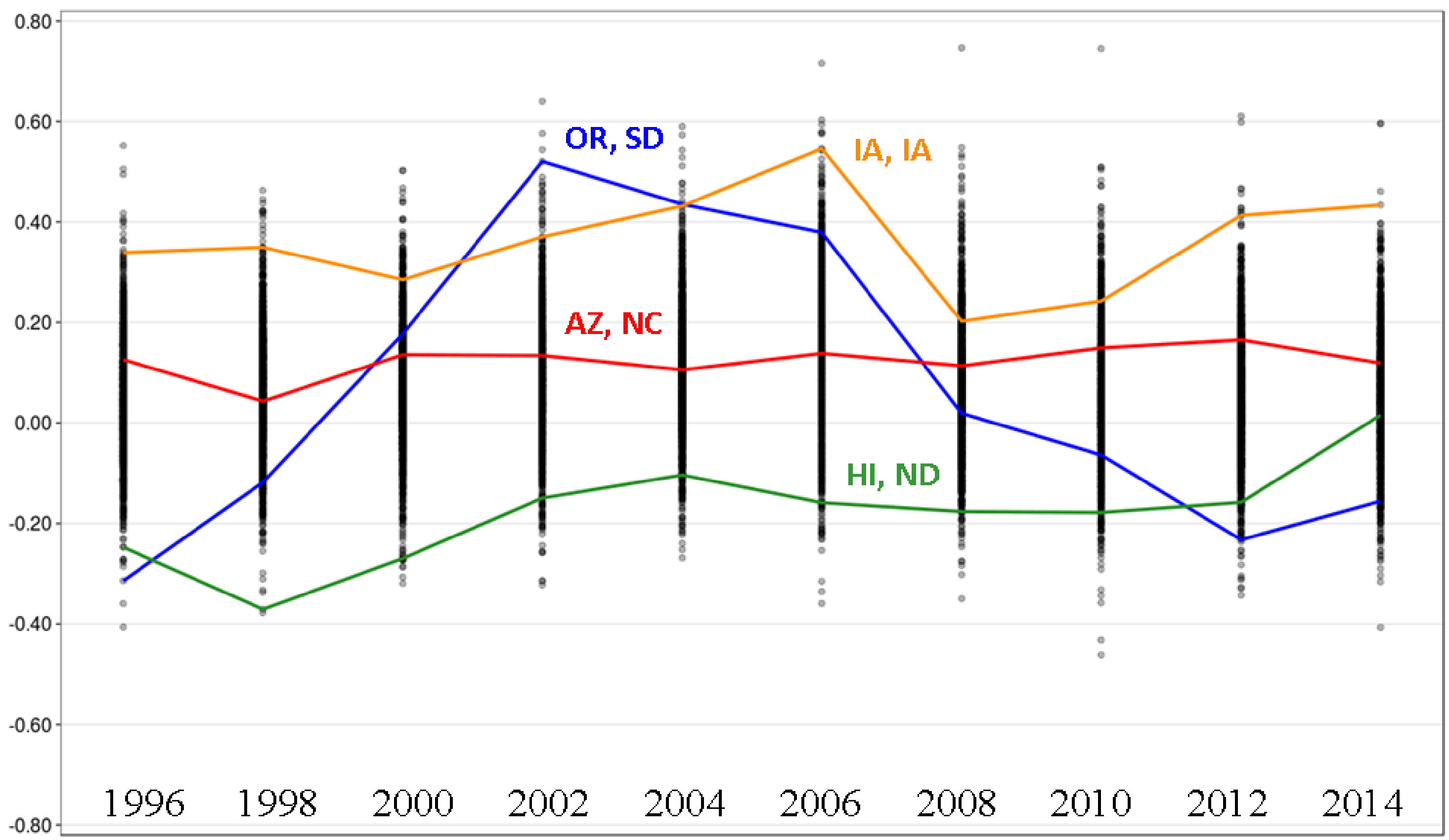


Figure 15: State-Line / State-Line Correlations Between Personal Auto and Other Lines in 2014

This figure demonstrates the variation in state-line / state-line (SLSL) correlations when one line of business is personal auto. The relatively large spread number of tail outcomes in the leftmost distribution (personal auto (PA) × workers' compensation (WC)) indicates that the relationship between the loss ratio outcomes between states across these two lines depends more heavily on which states in the lines are being considered. The rightmost distribution, with a notably higher average correlation, implies that the average between-state correlation value within PA is larger than the interstate correlations between PA and a different line of business. To clarify, this rightmost distribution contains only 2550 data points, as same-state comparisons within PA (having a correlation of 1) are omitted.

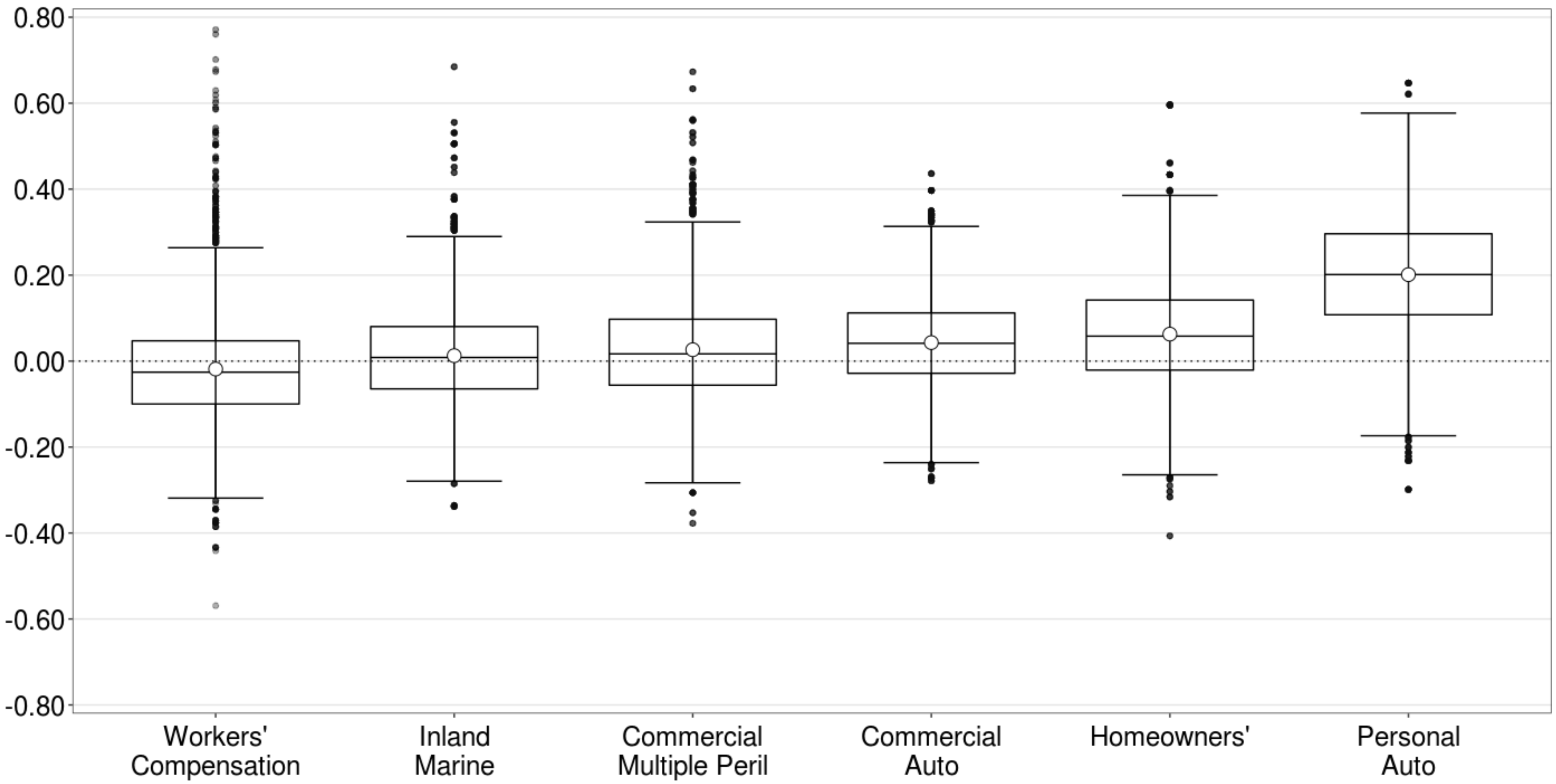


Figure 16: Tracking State-Line / State-Line Correlation Pairs for WC and PA Over Time

This figure is identical to Figure 14, but is constructed from state-line / state-line (SLSL) correlations between a state in personal auto (PA) and a state in workers' compensation (WC). We track 4 particular SLSL combinations over time. Each correlation in the figure exists between one state in PA and another (or the same) state in WC. Iowa-PA and New York-WC (in red) has the most stable correlation over the entire time period (1996-2014), while Oklahoma-PA and Wyoming-WC (blue) have the most variable. Alaska-PA and West Virginia-WC (green) have the lowest average correlation, and Illinois-PA and North Dakota-WC (orange) have the largest average correlation. As described in Section 5 and further shown in Table 10, many of the largest positive SLSL correlations in PA×WC occur when the WC state is Wyoming.

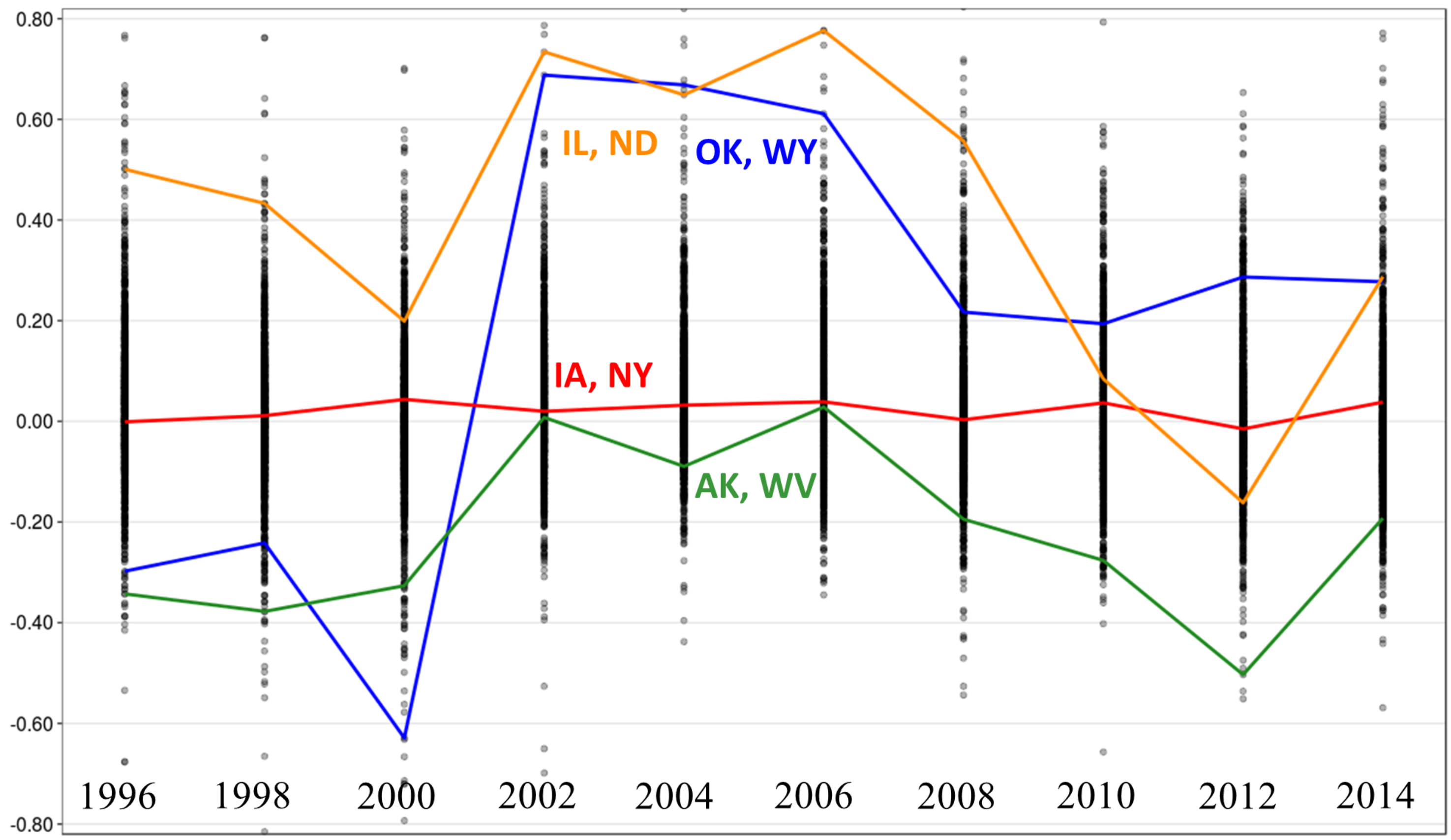


Figure 17: Number of Observations in Workers' Compensation Across States (2009-2013)

This figure shows the number of insurer year observations in Workers' Compensation (WC) across states from 2009-2013, the time period used to determine state-line risk in 2014. North Dakota (nd), Wyoming (wy), Washington (wa), and Ohio (oh) are the only states in the U.S. WC market with monopolistic state funds. These states, therefore, have relatively fewer observations of insurer loss ratio outcomes than do states with more conventional WC insurance markets. As the number of insurer-year observations within a state-line increases, so does the importance of the state-line risk when determining portfolio risk and diversification for insurers. That is, a low number of participating insurers in a state-line or pair of state-lines results in less data available to determine risk and correlation, but this also means that the resulting risk and correlation are used less often in determining insurer underwriting risk and diversification.

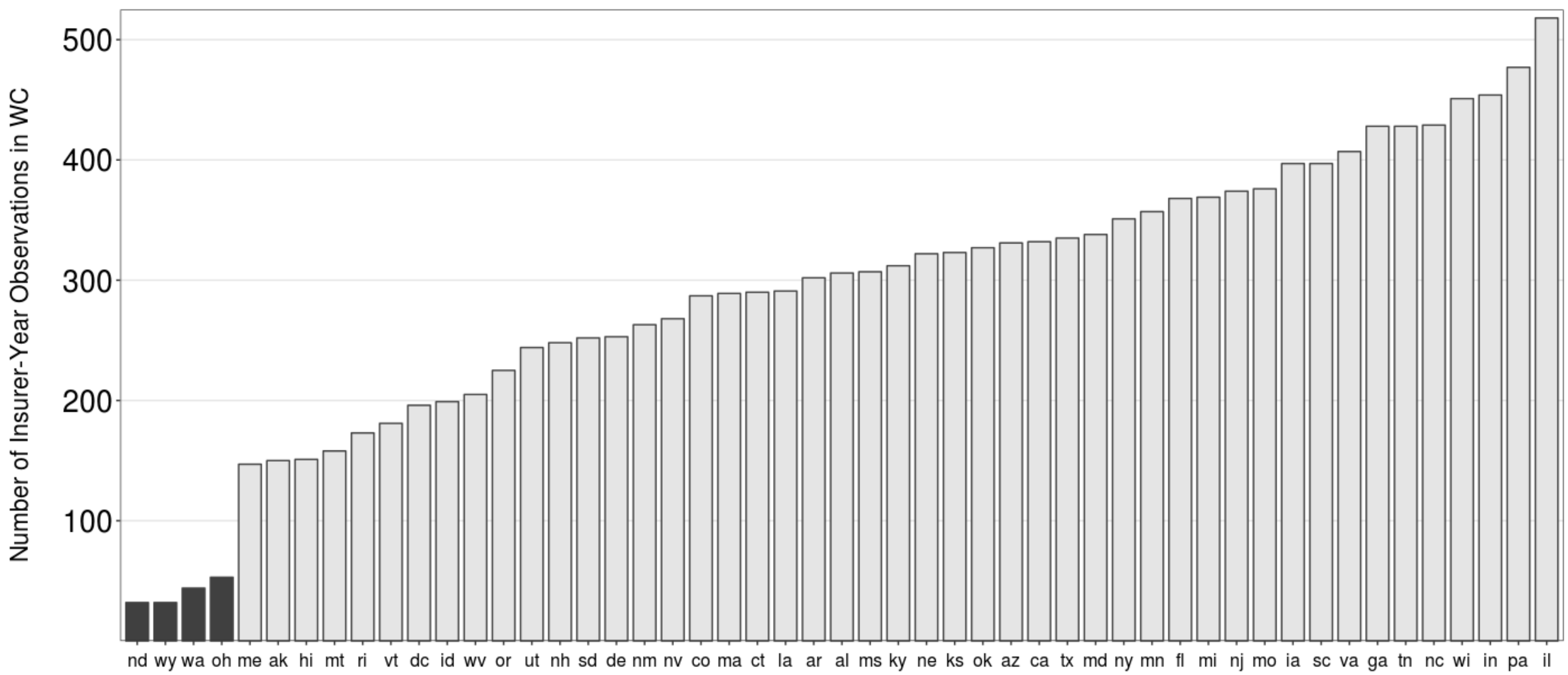


Figure 18: Diversification vs. Concentration for the 10 Largest Insurers in 2014

This figure is a scatterplot of diversification (Diversification Ratio, DR) and concentration (line of business Herfindahl) for the 10 largest U.S. property-liability insurers (by direct premiums written) in 2014. Note the range of the vertical axis to facilitate display. A general relationship between diversification and concentration is observable on the chart: diversification increases as concentration decreases. This is not surprising, as concentration is a factor of diversification. The insurers are represented by their logos. From left to right: Travelers Group , American International Group, Liberty Mutual Group, Hartford Group, Nationwide Group, Farmers Insurance Group, State Farm Group, Allstate Insurance Group, Berkshire Hathaway Group, and Progressive Group. Of particular interest to this work is the relationship between Liberty Mutual and Hartford. These two insurers are similarly concentrated but differentially diversified. Details for this chart are presented in Table 12.

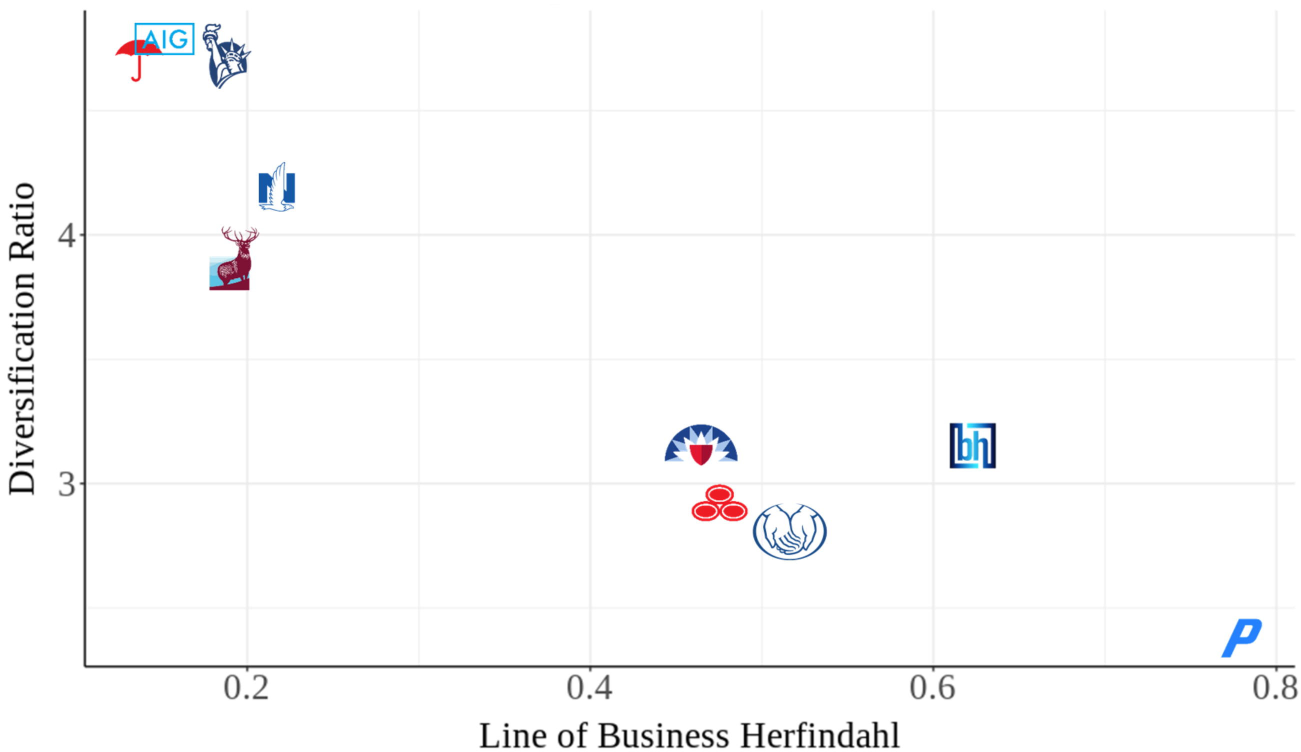


Figure 19: Concentration Plots for Travelers, State Farm, and Progressive in 2014

This figure contains concentration plots for 3 of the largest insurers in 2014: Travelers (left), State Farm (center), and Progressive (right). These concentration plots allow the visualization of concentration at 3 different levels. Each concetnration plot is divided into 3 sections, separated by dotted lines. These 3 sections correspond to personal lines, commercial lines, and financial lines. The pink bars form a proportional bar chart of premiums coming from each of these line groups. Here, Travelers does more than twice as much business in commercial lines as it does in personal lines. State Farm and Progressive are much more focused on personal lines of business. The vertical axis is divided into 23 segments, one for each line of business. The blue bars illustrate the proportion of total direct premiums written within each line of business. Travelers has a number of small blue bars, while State Farm has 2 dominant bars for homeowners' and personal auto. Progressive's business is dominated by personal auto, though there is a noticeable bar in commercial auto. Finally, within each line of business, premiums can come from any one of 51 states. The proportion of premiums *within each line* across states is captured by a similar bar chart in black. Therefore, we can easily observe insurers' distribution of business across lines and across states within each line, as well as across the personal, commercial, and financial groupings. The horizontal axis is the proportion of business.

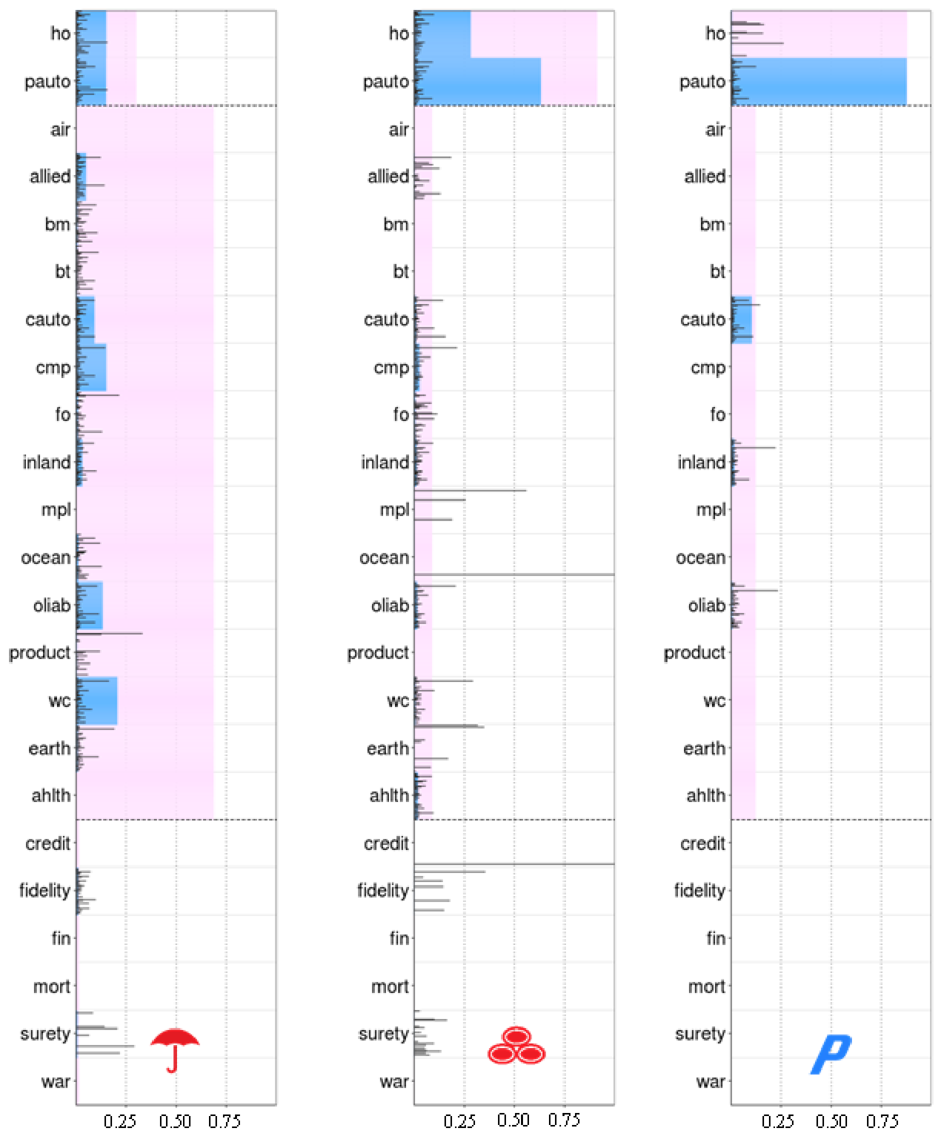
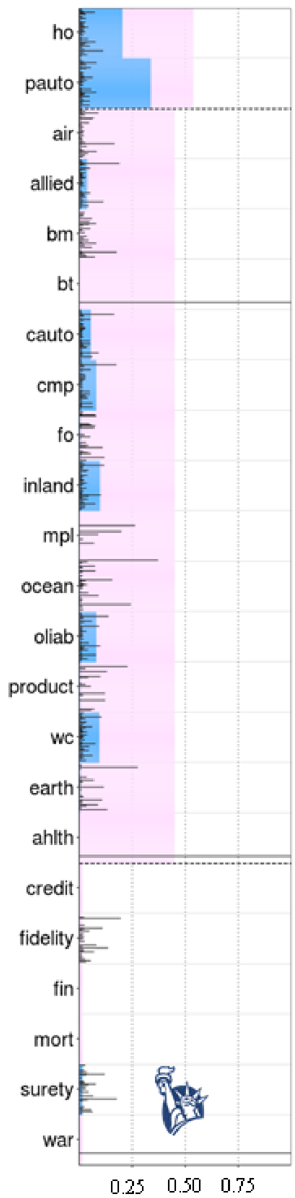


Figure 20: Concentration Plots for Liberty Mutual and Hartford in 2014

An explanation of concentration plots in general is provided in Section 6 and in Figure 19. This figure contains concentration plots for two insurers (Liberty Mutual, left, and Hartford, right) in 2014. These two insurers are similarly concentrated but differentially diversified. The similarity in concentration is evident from these plots. Both have roughly the same number of blue bars (8 clearly visible), and they exist in the same lines, though the weighting in each line is different. We can see that Liberty Mutual is operates dominantly in personal lines (though the allocation between personal and commercial is similar) and Hartford operates dominantly in commercial lines. With few exceptions, both insurers appear to have little state concentration within lines. The exact proportions of premiums within these 8 lines as well as the calculation of the Herfindahl measures determined from these proportions for both insurers are provided in Table 13.



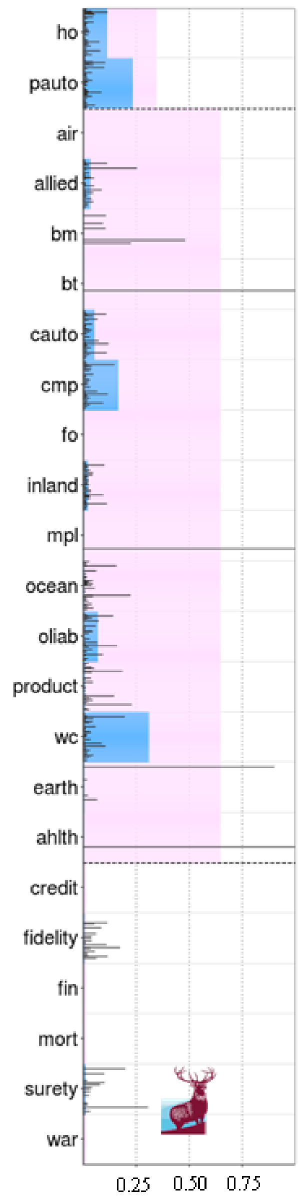


Figure 21: Geographic Concentration in Workers' Compensation for Hartford in 2014

This figure expands the proportional allocation of direct premiums written for Hartford *within* the workers' compensation line of business that is observed in Figure 22. In the portfolio plot, this chart appears vertically (rotated 90 degrees from this layout) inside the blue bar at workers' compensation (WC). Hartford writes positive premiums in WC in each of 49 states (no business in Wyoming or North Dakota), but Hartford's WC business is relatively larger in California (CA), New York (NY), and New Jersey (NJ).

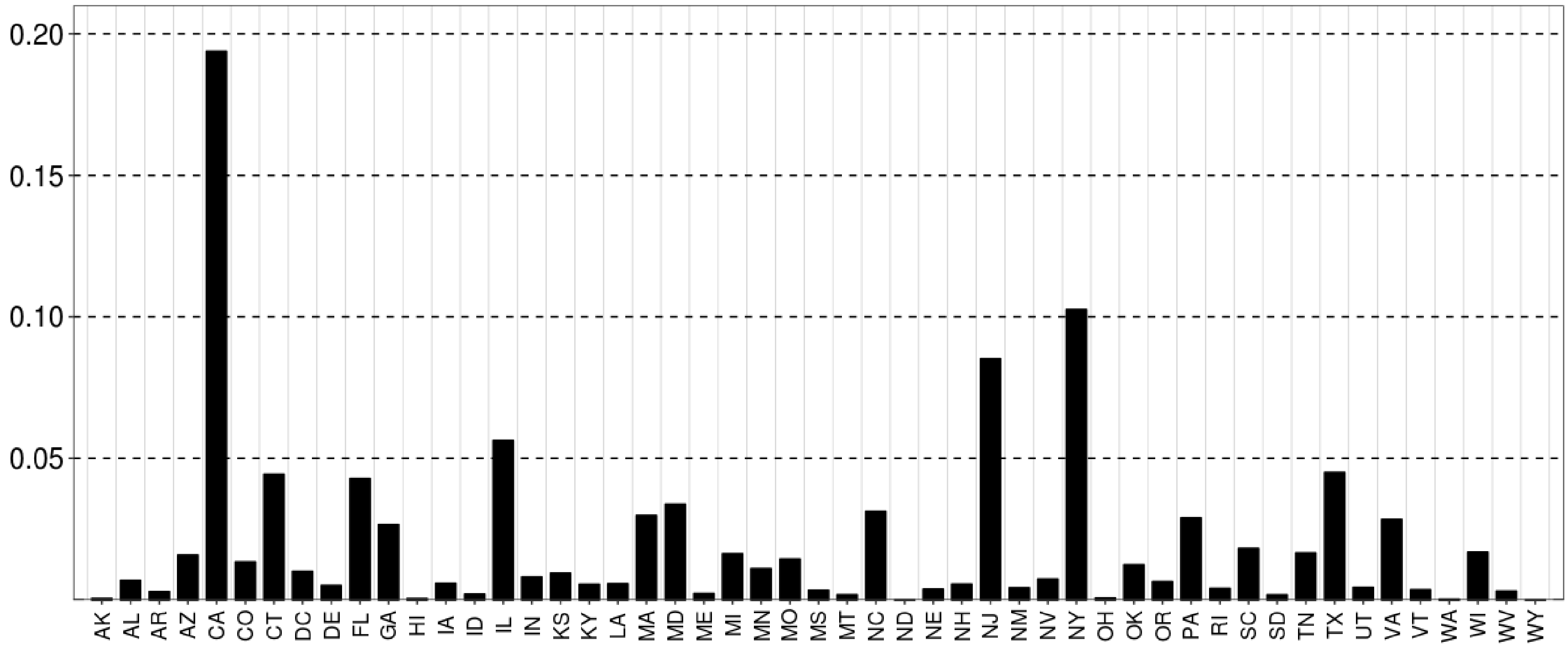
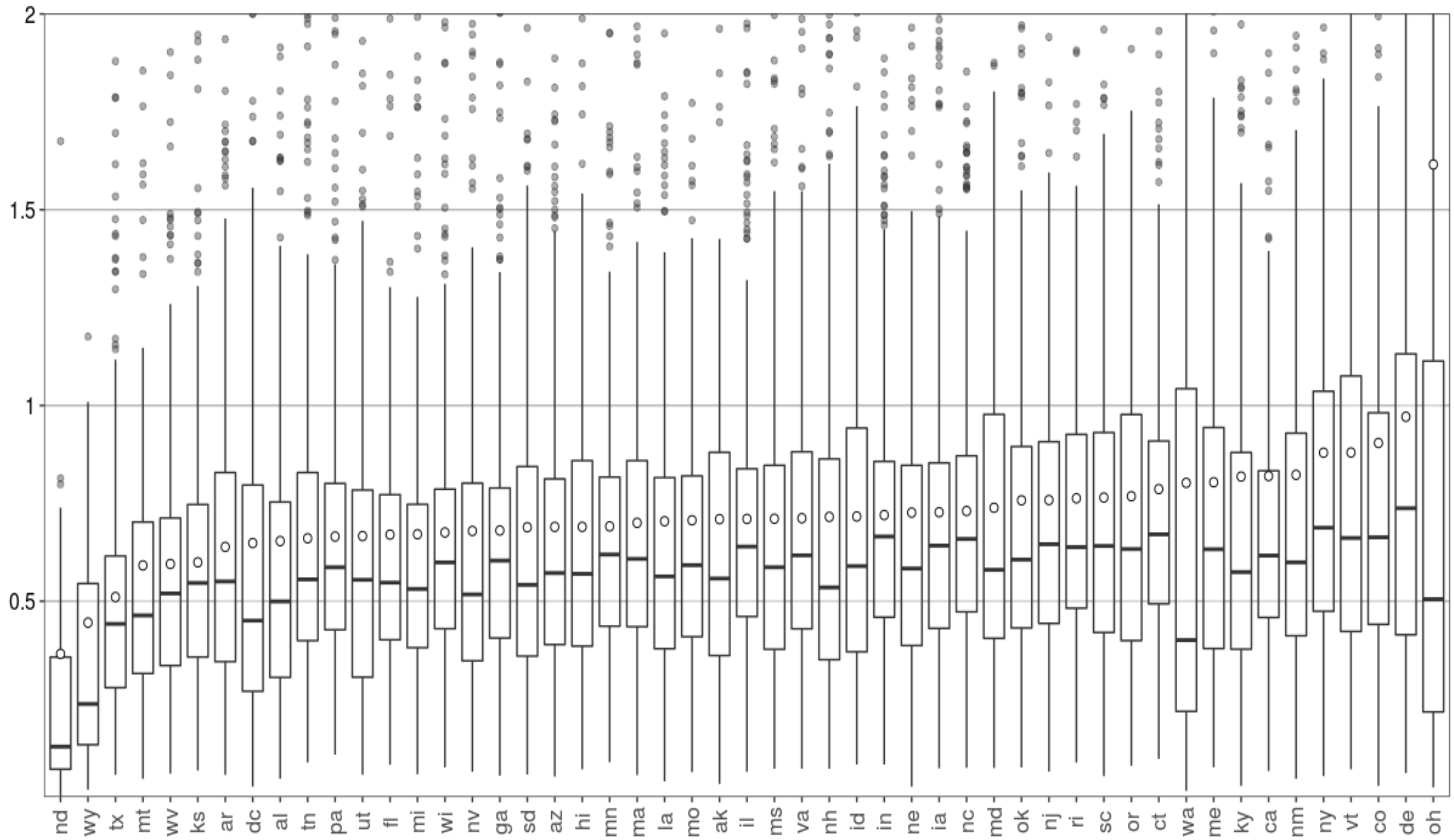


Figure 22: State-Line Level Loss Ratios in Workers' Compensation by State (2009-2013)

This figure is identical in construction to Figure 9, but it provides the state-line (SL) level loss ratios observed in workers' compensation (WC) grouped by state. Loss ratios are displayed on the vertical axis, with states on the horizontal axis. Each data point corresponds to one observed SL level loss ratio in WC. The mean of each SL distribution is marked with a circle. The states are organized from left to right in order of increasing average loss ratio. These loss ratios are used to determine the SL risk in WC within each state. For example, the standard deviation of the observations that are summarized by the rightmost box (Ohio) is the state-line risk for Ohio-Workers' Compensation in 2014.



Tables

Table 1: Chi-Squared Goodness-of-Fit Test Results for States in Homeowners'

After determining the best fit distribution for loss ratios observed in the state of Florida *within* the homeowners' line of business, we perform a chi-squared goodness-of-fit test using the fitted distribution and the empirical distribution of loss ratios for all 51 states *within* homeowners'. This table contains the χ^2 statistic and the p-value for each test. The results of the tests show that the best fitting distribution for Florida does not statistically fit the results for 44 out of the remaining 50 states, though it is a good fit for the observed outcomes in Florida. That is, variation exists across states within a line. The top panel contains the results of the tests for the states in which the p-value > 0.1, that is, the states for which the distribution of loss ratio outcomes within homeowners' is not significantly different from Florida.

State	p-value	χ^2	State	p-value	χ^2
Florida	1.0000	14.4611	Delaware	0.8221	85.0184
Washington, DC	0.9965	64.3631	Rhode Island	0.5060	97.1242
California	0.9902	68.3287	Nevada	0.1376	113.3510
Massachusetts	0.8271	84.7734			

State	p-value	χ^2	State	p-value	χ^2
New York	0.0371	124.3923	West Virginia	0.0000	364.6822
Maine	0.0173	129.8302	Indiana	0.0000	385.2839
Oregon	0.0105	133.1831	Connecticut	0.0000	388.0019
New Hampshire	0.0100	133.4840	Iowa	0.0000	423.2731
North Dakota	0.0031	140.6542	New Jersey	0.0000	431.9223
Vermont	0.0026	141.7753	Illinois	0.0000	444.1471
Texas	0.0008	148.3591	Kansas	0.0000	508.3256
Virginia	0.0002	156.4149	Arizona	0.0000	558.2975
South Carolina	0.0001	157.6482	Georgia	0.0000	583.0781
Washington	0.0000	176.4256	Wyoming	0.0000	596.1010
Louisiana	0.0000	200.5727	Alabama	0.0000	602.0840
Maryland	0.0000	206.8173	Ohio	0.0000	603.2783
Utah	0.0000	208.4174	Arkansas	0.0000	609.9303
New Mexico	0.0000	231.4742	Missouri	0.0000	610.0542
Idaho	0.0000	261.0350	Kentucky	0.0000	617.2474
Michigan	0.0000	284.9310	Nebraska	0.0000	629.1429
Wisconsin	0.0000	294.0054	Montana	0.0000	658.8133
Pennsylvania	0.0000	296.4607	South Dakota	0.0000	863.7477
North Carolina	0.0000	304.2714	Tennessee	0.0000	1081.1188
Alaska	0.0000	309.0131	Colorado	0.0000	1119.0024
Minnesota	0.0000	318.5257	Oklahoma	0.0000	1594.9028
Mississippi	0.0000	356.2541	Hawaii	0.0000	2352.5795

Table 2: Premium and Loss Data for Colonial Insurance Group (2014)

In 2014, Colonial Insurance Group wrote business in one state, Texas, and in two lines of business in that state, allied lines and homeowners'. This table provides the direct premiums written (Premiums) and the direct losses incurred (Losses), which includes losses and loss related expenses, in the two state-lines as reported in the insurer's annual financial statements. The rightmost column contains the state-line level loss ratios (Losses/Premiums) for Colonial.

	Colonial Insurance Group (2014)		
	Premiums	Losses	Loss Ratio
Texas - Allied Lines	$3,840,047	$2,118,459	0.552
Texas - Homeowners'	$2,655,468	$1,404,128	0.529
Total	$6,495,515	$3,522,587	0.542

Table 3: Sample Loss Ratio Observations (Texas – Allied Lines)

This table contains a sample of the state-line level loss ratios observed in Texas in allied lines from 2009 to 2013. The loss ratios are determined as direct losses incurred (including losses and loss related expenses) over direct premiums written. Each observation is for a single insurer (group or unaffiliated single) operating in the state-line within the time period. The full set from which this sample is drawn contains 483 state-line-insurer-year observations. This dataset is used to determine the standard deviation (risk) in loss ratios in allied lines in Texas in 2014, the year directly following the 5-year time period over which the loss ratios are observed.

Year	Line of Business	State	Loss Ratio
2009	Allied Lines	Texas	1.1355
2010	Allied Lines	Texas	1.6872
2011	Allied Lines	Texas	1.2034
2012	Allied Lines	Texas	0.3857
2013	Allied Lines	Texas	1.1880

Table 4: Sample Loss Ratio Observations (Texas – Homeowners')

This table contains a sample of the state-line level loss ratios observed in Texas in homeowners' from 2009 to 2013. The loss ratios are determined as direct losses incurred (including losses and loss related expenses) over direct premiums written. Each observation is for a single insurer (group or unaffiliated single) operating in the state-line within the time period. The full set from which this sample is drawn contains 237 state-line-insurer-year observations. This dataset is used to determine the standard deviation (risk) in loss ratios in homeowners' in Texas in 2014, the year directly following the 5-year time period over which the loss ratios are observed.

Year	Line of Business	State	Loss Ratio
2009	Homeowners'	Texas	0.4837
2010	Homeowners'	Texas	0.6407
2011	Homeowners'	Texas	0.3969
2012	Homeowners'	Texas	0.4543
2013	Homeowners'	Texas	0.5078

Table 5: Sample of Loss Ratio Pair Observations for (TX-AL, TX-HO)

This table contains a sample of the state-line/state-line loss ratio pairs observed in Texas in allied lines (TX-AL) and in Texas in homeowners' (TX-HO) from 2009 to 2013. The loss ratios are determined as direct losses incurred (including losses and loss related expenses) over direct premiums written in each state-line. Each observed pair is for a single insurer (group or unaffiliated single) operating in both state-lines in the same year within the time period. The full set from which this sample is drawn contains 171 insurer-year-state-line observed pairs. This dataset is used to determine the correlation in loss ratios between TX-AL and TX-HO in 2014, the year directly following the 5-year time period over which the loss ratios are observed.

Year	Loss Ratio in TX-AL	Loss Ratio in TX-HO
2009	1.1355	0.0929
2010	0.6647	0.1025
2011	0.3648	0.1040
2012	0.7464	0.2062
2013	0.2937	0.2390

Table 6: Summary Statistics for State-Line Level Risk

This table shows the summary statistics for state-line level risk measures within each line of business, wherein risk is measured as the standard deviation in the reported loss ratios (direct losses incurred over direct premiums written) across all insurers operating in any state within the line of business in any of the 5-years prior to the year reported in the table. For example, in 1996, the reported mean risk measure in homeowners' is 0.5254. To arrive at this value, we begin by collecting each insurer's reported losses and premiums in homeowners' within a particular state from 1991-1995, inclusive. We determine the loss ratio within the state-line for each insurer. We then determine the standard deviation in these loss ratios, the state-line risk. Finally, we take the mean of the state-line risk measures across all states within homeowners' to arrive at the value reported in the table.

	1996			2014		
Line of business	mean	median	stdev	mean	median	stdev
Accident and Health	0.5266	0.4732	0.2597	0.5508	0.5172	0.1995
Aircraft	1.2590	0.9650	0.9408	1.1270	0.6659	1.6798
Fire and Allied lines	0.9624	0.7367	1.1462	0.9289	0.7616	0.5637
Boiler and Machinery	1.3212	0.8416	1.8409	0.7784	0.5336	0.8786
Burglary and Theft	0.1756	0.1582	0.1659	0.4094	0.2909	0.5377
Commercial Auto	0.4225	0.4027	0.0853	0.4485	0.4206	0.0960
Commercial Multiple Peril	0.5787	0.4519	0.5488	0.5625	0.5155	0.1679
Credit	0.4262	0.3485	0.3103	0.9082	0.6929	1.0615
Earthquake	0.3851	0.0334	1.0935	0.4103	0.2067	0.5634
Fidelity	1.0424	0.7319	0.8166	1.3549	1.0877	0.9215
Financial Guaranty	2.8721	0.0015	7.5494	11.0623	~ 0.0000	34.3486
Farmowners	0.3824	0.3674	0.2083	0.4940	0.4203	0.3268
Homeowners	0.5254	0.3127	0.9243	0.3702	0.3076	0.1475
Inland Marine	0.5901	0.3498	1.3444	0.4381	0.4056	0.1489
Mortgage Guaranty	0.3300	0.2507	0.1738	0.9692	0.5618	1.1328
Medical Professional Liability	1.0451	0.7192	1.1061	0.9030	0.6489	0.9136
Ocean Marine	0.7231	0.6041	0.5521	0.9058	0.7657	0.7841
Other Liability	0.8884	0.8452	0.2754	0.7243	0.6934	0.2142
Personal Auto	0.2229	0.2119	0.0532	0.2138	0.2016	0.0751
Products Liability	1.8021	1.2753	1.5871	1.4586	1.2568	0.7316
Surety	0.7405	0.6883	0.3453	0.8661	0.8298	0.4056
Warranty	Not reported prior to 2008			0.7462	0.5951	0.5565
Workers' Compensation	0.6485	0.4919	0.4108	0.6273	0.5598	0.3831

Table 7: State-Line / State-Line (SLSL) Summary Statistics for HO × PA (2014)

This table contains the summary statistics for SLSL correlations in 2014 wherein one state is in homeowners' and the other in personal auto. The summary measures in the four panels of this table correspond to the 3 panels of Figure 12. ***Raw SLSL Correlations***: These values summarize the raw SLSL correlations observed. Each correlation value is determined using industry wide observed loss ratios occurring between 2009 and 2013, inclusive, as described in Section 6. ***Credibility***: These values summarize the number of observed loss ratio pairs (Equation 16) that determine each of the SLSL correlations. ***Weighted SLSL Correlations***: The weighted mean, median, and standard deviation of SLSL correlations are achieved through the credibility weighting of observed correlations. Each observation is weighted according to the number of observed loss ratio pairs that were used in its determination. In this way, we account for potential spurious correlation outcomes due to limited data.

Mean SLSL correlation	0.0538	Raw SLSL Correlations
Median SLSL correlation	0.0511	
Standard deviation of SLSL correlations	0.1264	
Total number of SLSL pairs	147351	Credibility
Mean number of SLSL pairs per correlation	111.12	
Weighted mean SLSL correlation	0.0627	Weighted SLSL Correlations
Weighted median SLSL correlation	0.0581	
Weighted standard deviation of SLSL correlations	0.1237	
Largest positive SLSL correlation	0.5963	
Number of SLSL pairs for largest positive correlation	119	
Largest negative SLSL correlation	-0.4067	
Number of SLSL pairs for largest negative correlation	40	
Max number of SLSL pairs for any SLSL correlation	302	
SLSL correlation at max number of SLSL pairs	0.2519	

Table 8: State-Line / State-Line (SLSL) Summary Statistics for HO × PA Over Time

This table contains the summary statistics for SLSL correlations over time wherein one state is in homeowners' and the other in personal auto (from Figure 13). In each year, the SLSL correlation values are determined using insurer level observed loss ratios (taken across the industry) occurring in the 5 years prior to the correlation measure. While Figure 13 is based on the weighted correlations only, the table provides the full set of summary statistics for all 3 panels of Table 7 and Figure 12. Organization is as described in Table 7.

	1998	2002	2006	2010	2014
Mean SLSL correlation	0.0639	0.0956	0.1438	0.0596	0.0538
Median SLSL correlation	0.0617	0.0963	0.1438	0.0554	0.0511
Standard deviation of SLSL correlations	0.1223	0.1256	0.1451	0.1345	0.1264
Total number of SLSL pairs	169392	169630	156835	142405	147351
Mean number of SLSL pairs per correlation	127.75	127.93	118.28	107.39	111.12
Weighted mean SLSL correlation	0.0693	0.1035	0.1549	0.0684	0.0627
Weighted median SLSL correlation	0.0670	0.1031	0.1529	0.0634	0.0581
Weighted standard deviation of SLSL correlations	0.1144	0.1183	0.1407	0.1289	0.1237
Largest positive SLSL correlation	0.4626	0.6398	0.7157	0.7445	0.5963
Number of SLSL pairs for largest positive correlation	46	67	71	46	119
Largest negative SLSL correlation	-0.3775	-0.3224	-0.3592	-0.4617	-0.4067
Number of SLSL pairs for largest negative correlation	72	49	112	40	40
Max number of SLSL pairs for any SLSL correlation	339	349	322	297	302
SLSL correlation at max number of SLSL pairs	0.2206	0.0901	0.2286	0.2429	0.2519

Table 9: Largest Positive SLSL Correlations for HO × PA in 2014

This table contains the 20 largest positive state-line / state-line (SLSL) correlation values observed in 2014, wherein one state is in homeowners' and the other is in personal auto. There are 2601 total correlations from which these 20 were taken. Observations for which the homeowners' state and the personal auto state are the same are shaded. Of the 20, 12 correspond to such within-state correlation between homeowners' and personal auto loss ratio outcomes.

Correlation	Personal Auto State	Homeowners' State
0.3613	Iowa	Iowa
0.3511	North Dakota	North Dakota
0.3443	Colorado	Washington, DC
0.3362	South Dakota	South Dakota
0.3178	Massachusetts	South Dakota
0.3143	Nebraska	Nebraska
0.3118	Missouri	Missouri
0.3012	Tennessee	Tennessee
0.2758	Colorado	Colorado
0.2727	Connecticut	Washington, DC
0.2723	Georgia	Maryland
0.2681	Texas	Texas
0.2638	Nevada	West Virginia
0.2600	Ohio	Ohio
0.2588	Arkansas	Ohio
0.2563	Washington, DC	Oregon
0.2474	Mississippi	Mississippi
0.2470	Oklahoma	Oklahoma
0.2463	Louisiana	Louisiana
0.2448	Alaska	Connecticut

Table 10: Largest Positive SLSL Correlations for WC × PA in 2014

This table contains the 20 largest positive state-line / state-line (SLSL) correlation values observed in 2014, wherein one state is in workers' compensation and the other is in personal auto. There are 2601 total correlations from which these 20 were taken. Observations for which the workers' compensation state is Wyoming are shaded. Of the 20, 9 correspond to SLSL correlations wherein the workers' compensation state is Wyoming.

Correlation	Personal Auto State	Workers' Compensation State
0.3613	Illinois	North Dakota
0.3511	Vermont	Wyoming
0.3443	Washington, DC	Nevada
0.3362	South Carolina	Wyoming
0.3178	Wyoming	Wyoming
0.3143	Michigan	Wyoming
0.3118	New Hampshire	Washington
0.3012	Illinois	Wyoming
0.2758	New Jersey	Wyoming
0.2727	Missouri	North Dakota
0.2723	Alaska	California
0.2681	Tennessee	Wyoming
0.2638	Louisiana	North Dakota
0.2600	California	North Dakota
0.2588	Idaho	Ohio
0.2563	Nevada	Pennsylvania
0.2474	Texas	Washington
0.2470	Texas	Wyoming
0.2463	Pennsylvania	Wyoming
0.2448	Washington, DC	Louisiana

Table 11: Diversification Ratio (DR) Summary Statistics

In this table, we provide the summary statistics for DR at 6-year intervals throughout our time period. The values are reported at the line level and are determined using all insurer-year observations in which the insurer participated in the line of business.

	1996			2002			2008			2014		
Line of business	mean	median	stdev	mean	median	stdev	mean	median	stdev	mean	median	stdev
Accident and Health	2.2055	1.8313	1.0685	2.3029	2.2462	0.9358	2.3591	2.2757	0.9702	2.9281	2.9812	1.2391
Aircraft	3.5080	3.7521	0.6949	3.2141	3.3308	0.7174	3.2781	3.3251	0.5111	3.5437	3.6736	0.9411
Boiler and Machinery	3.3733	3.7100	0.7961	3.0080	3.0680	0.6966	2.7509	2.7323	0.7764	3.0779	3.2707	1.1683
Burglary and Theft	2.8331	3.0725	1.0539	2.8824	3.0328	0.8834	3.0949	3.3296	0.7926	3.3681	3.7140	1.0901
Commercial Auto	2.2200	2.0876	0.8996	2.2288	2.1307	0.7685	2.2005	2.0891	0.7814	2.4673	2.3375	1.0173
Commercial Multiple Peril	2.2457	2.0097	0.8380	2.2295	2.0610	0.7572	2.1694	1.9616	0.7591	2.4466	2.1744	0.9910
Credit	2.8284	2.7600	1.0697	2.6202	2.5379	0.8860	2.5837	2.4415	0.9216	2.8848	2.9359	1.1439
Earthquake	2.9555	3.1511	0.9489	2.9068	3.0094	0.7521	2.8569	3.0504	0.7745	3.3144	3.1624	0.9440
Farmowners	2.1644	1.8685	0.8332	2.0637	1.8268	0.7261	2.0420	1.7870	0.7251	2.1005	1.7252	0.8980
Fidelity	3.4539	3.7095	0.6103	3.1541	3.2479	0.6612	3.1437	3.3269	0.7260	3.6057	3.7238	0.8168
Financial Guaranty	2.2619	1.7518	1.0785	1.2066	1.0448	0.2932	1.7659	1.2144	1.0524	1.2086	1.1406	0.2021
Fire and Allied lines	2.0491	1.7892	0.8592	2.0424	1.8730	0.7608	1.9747	1.7851	0.7648	2.1462	1.7509	0.9805
Homeowners	2.1009	1.8172	0.8408	2.0290	1.8264	0.7505	1.9235	1.7305	0.7400	2.0684	1.6794	0.9323
Inland Marine	2.3130	2.1853	0.8661	2.3049	2.2448	0.7366	2.3152	2.2384	0.7392	2.5650	2.4456	0.9836
Medical Professional Liability	1.9559	1.4060	1.0900	1.9852	1.4665	1.0157	1.8220	1.3095	0.9744	2.1821	1.7340	1.2444
Mortgage Guaranty	2.3094	1.8342	1.1460	2.0667	1.3035	1.1724	2.3207	1.3160	1.3788	2.5491	1.5634	1.3623
Ocean Marine	2.8346	2.8819	0.9717	2.7524	2.9336	0.8615	2.8837	3.0484	0.7970	3.3695	3.2687	0.9719
Other Liability	2.1503	1.9794	0.8619	2.1224	2.0179	0.7691	2.0849	1.9405	0.7625	2.2798	2.0543	0.9788
Personal Auto	2.0805	1.8285	0.8947	2.0213	1.8882	0.8023	1.9624	1.7682	0.8072	2.2290	1.9713	1.0143
Products Liability	2.9722	3.0725	0.8485	2.7840	2.8771	0.7365	2.8562	2.9653	0.6757	3.3957	3.4442	0.8393
Surety	2.5883	2.7114	1.0753	2.4117	2.4890	0.9730	2.3080	2.2757	0.9267	2.8567	2.9674	1.1640
Warranty	Not reported prior to 2008									2.7582	2.7511	1.2608
Workers' Compensation	2.2132	2.0302	0.9921	2.1409	2.0387	0.8733	2.0285	1.8710	0.8884	2.3122	1.9828	1.1423

Table 12: Concentration and Diversification Measures for the 10 Largest Insurers in 2014

This table provides the total amount of direct premiums written, the Diversification Ratio, and the line of business level Herfindahl concentration measure for each of the 10 largest (as determined by direct premiums written) insurers in the U.S. Property-Liability insurance market in 2014. The logo for each insurer (leftmost column) are those used in Figure 18. This table is organized in order of decreasing concentration (left to right in Figure 18).

Insurer	NAIC Group Code	Total DPW (in $B)	Diversification Ratio	Herfindahl (Line-level)
Travelers Group	3548	21.2317	4.7285	0.1374
American Int'l Group	12	16.3207	4.7832	0.1513
Liberty Mutual Group	111	28.3314	4.7133	0.1876
Hartford Group	91	10.4898	3.9098	0.1900
Nationwide Group	140	18.8289	4.1957	0.2179
Farmers Insurance Group	69	12.8273	3.1399	0.4645
State Farm Group	176	54.5919	2.9186	0.4759
Allstate Insurance Group	8	28.4236	2.8067	0.5138
Berkshire Hathaway Group	31	26.2097	3.1374	0.6232
Progressive Group	155	18.8816	2.3836	0.7787

Table 13: Line of Business Proportions for Liberty Mutual and Hartford (2014)

This table contains the proportional breakdown of premium sources across the 8 largest lines for Liberty Mutual and Hartford in 2014. The proportions here account for 97.2 and 97.9 percent of the business written by Liberty Mutual and Hartford, respectively. The line of business (LOB) Herfindahl for each insurer is determined as the sum of the squared proportions of business written in each line. Liberty Mutual and Hartford have similar Herfindahl concentration measures, but the insurers are differentially diversified, with Liberty Mutual being more diversified than Hartford.

	Liberty Mutual		Hartford	
Line of business	Proportion	Squared Proportion	Proportion	Squared Proportion
Personal Auto	0.336	0.113	0.232	0.054
Homeowners'	0.202	0.041	0.111	0.012
Inland Marine	0.096	0.009	0.020	0.000
Commercial Auto	0.053	0.003	0.051	0.003
Commercial Multiple Peril	0.078	0.006	0.163	0.026
Workers' Compensation	0.094	0.009	0.307	0.094
Other Liability	0.078	0.006	0.065	0.004
Allied Lines	0.035	0.001	0.031	0.001
LOB Herfindahl		0.1875		0.1900

Table 14: The Largest SLSL Contributors to the Detriment of Hartford's Diversification in 2014

This table contains the 10 largest state-line / state-line (SLSL) contributors (out of almost 230,000 components) to the reduction of Hartford's diversification in 2014. As the denominator in the Diversification Ratio (the underwriting portfolio risk) for increases, the insurer's DR decreases. The SLSL components of the denominator are isolated, and the largest identified as those that contribute most to the reduction in diversification. Each SLSL component is the product of the weights in each SL, the risks in each SL, and the correlation between the two SLSLs. Diversification suffers most when the product of these is largest. Consequently, diversification is reduced to a greater extent when business is concentrated in a particular SL or when a large proportion is allocated to each SL in a pair, when the risk in a SL (or each in a pair) is disproportionately large relative to the rest of the insurer's underwriting portfolio, or the correlation between a pair of SLs is large relative to the other SL pairings found in an insurer's underwriting portfolio. *State 1 (2)* and *LOB 1 (2)* describe a state-line in the pairing. *Weight in SL_1 (2)* and *Risk in SL_1 (2)*, describe the weight (as a percentage) and risk (standard deviation in insurer level loss ratios taken across the industry from 2009-2013) for state-line 1 (2), respectively. The correlation between the SL loss ratio outcomes is provided as *SLSL Correlation*. The *DR Denominator Component* is determined using the percentages, rather than the proportions, for ease of interpretation. In the table workers' compensation is abbreviated as "WC" and homeowners' is abbreviated as "HO". The states are provided as the common state abbreviation.

DR Denominator Component	State 1	LOB 1	State 2	LOB 2	Weight in SL_1	Weight in SL_2	SLSL Correlation	Risk in SL_1	Risk in SL_2
6.6746	CA	WC	NJ	WC	5.9302	2.6052	0.4309	0.9246	0.5422
4.8618	CA	WC	NY	WC	5.9302	3.1402	0.1925	0.9246	0.7334
2.4871	CA	WC	FL	WC	5.9302	1.3130	0.3639	0.9246	0.4747
2.3929	NJ	WC	NY	WC	2.6052	3.1402	0.3678	0.5422	0.7334
1.8910	CA	WC	CT	WC	5.9302	1.3575	0.2269	0.9246	0.5598
1.6032	CA	WC	AZ	HO	5.9302	0.4062	0.6019	0.9246	0.5980
1.4138	CA	WC	GA	WC	5.9302	0.8159	0.3494	0.9246	0.4523
1.0443	CA	WC	IL	WC	5.9302	1.7238	0.1304	0.9246	0.4237
0.8642	CA	WC	MD	WC	5.9302	1.0326	0.1494	0.9246	0.5110
0.6309	CT	WC	NY	WC	1.3575	3.1402	0.1802	0.5598	0.7334
0.5945	CA	WC	MA	WC	5.9302	0.9121	0.1338	0.9246	0.4443

Table 15: Liebenberg and Sommer (2008) Variable Description

This table contains the variable names and meanings for the diversification model used in Liebenberg and Sommer (Equation 19). Also included is the mapping of those variables to the variables used in our model including our measures of diversification and underwriting risk (Equation 20).

Variable	Definition	Included in our model as:
MULTLINE	Binary indicator for whether the insurer operates in a multiple lines of business	DR
SIZE	Natual logarithm of total admitted assets	SIZE
CAPASSET	Policyholder surplus / total admitted assets	CAPASSET
SDROA	Standard deviation of ROA over the prior 5 years	SDROA
GEODIV	1 - Geographic HHI	DR, GEODIV
WCONC	Weighted sum of market share per line multiplied by line specific Herfindahl	*Not included*
PCTLH	Percentage of premiums from life-health insurance	*Not included*
MUTUAL	= 1 if firm is a mutual, 0 otherwise	Firm fixed-effect
PUBLIC	= 1 if firm is publicly traded, 0 otherwise	Firm fixed-effect
GROUP	= 1 if firm is a group, 0 otherwise	Firm fixed-effect
YEAR	Year fixed-effect	Year fixed-effect
LINE	Indicators for line of business participation for 23 lines	DR
STATE	Indicators for participation in state (any line of business)	DR

Table 16: Summary Statistics for Model Components

This table contains the name, meaning, and summary statistics for variables included in the final regression model in Section 8. The total number of observations is 14,667 insurer-years, with 1,593 unique insurers, each averaging 7.22 year observations in the data.

Variable	Variable definition	Min	Max	Mean	Median	Standard Deviation
ROA	Return on assets	-0.6646	0.6668	0.0187	0.0236	0.0597
SDROA	5-year standard deviation in ROA	0.0001	0.6084	0.0350	0.0248	0.0352
RAROA	ROA / UPR	-23.7350	5.6700	0.0668	0.0651	0.3058
DR	Diversification Ratio	1.0000	5.1786	1.7914	1.4744	0.7691
SPREAD	1 - HHI	0.0000	0.8981	0.3550	0.3935	0.3055
GEODIV	1 - Geographic HHI	0.0000	0.9650	0.3358	0.1266	0.3730
UPR	Underwriting portfolio risk	0.0022	49.4937	0.4394	0.2891	0.7206
SIZE	ln(total assets)	12.5680	26.3279	18.2028	17.9886	2.3193
CAPASSET	policyholder surplus / total assets	-1.7407	1.0000	0.4569	0.4277	0.1857

Table 17: Regression Results Using HHI

The table presents regression estimates using HHI measures of diversification in 3 panels. In each panel we provide the estimates of 3 models, each using the same measure (or set of measures) of diversification. The first, second, and third models in each panel correspond to the dependent variables *ROA* (return on assets), *SDROA* (5-year standard deviation of ROA), and *RAROA* (risk-adjusted ROA, which is ROA divided by the underwriting risk, *UPR*), respectively. Panel 1 uses *SPREAD* (1-HHI) as the measure of diversification, while Panel 2 uses *GEODIV* (1-geographic HHI), and Panel 3 uses *SPREAD* and *GEODIV*. Other independent variables include: *UPR* (underwriting portfolio risk), *SIZE* (log of total assets), and *CAPASSET* (policyholder surplus / total assets). Each model includes firm and year fixed-effects. The constant and the fixed effects are omitted from the table to conserve space. Robust standard errors, clustered at the firm level, are reported below the coefficients in parentheses. ***, **, and * indicate two-tailed statistical significance at the 0.01, 0.05, and 0.10 levels.

	Panel 1			Panel 2			Panel 3		
Model :	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. variable:	ROA	SDROA	RAROA	ROA	SDROA	RAROA	ROA	SDROA	RAROA
SPREAD	-0.0163***	-0.0028	-0.0351				-0.0152***	-0.0031	-0.0409
	(0.000)	(0.202)	(0.162)				(0.000)	(0.163)	(0.105)
GEODIV				-0.0097***	0.0018	0.0381*	-0.0081**	0.0021	0.0425*
				(0.010)	(0.365)	(0.084)	(0.032)	(0.287)	(0.056)
UPR	0.0099***	0.0089***		0.0099***	0.0090***		0.0098***	0.0089***	
	(0.000)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)	
SIZE	0.0232***	-0.0136***	0.0839***	0.0231***	-0.0139***	0.0793***	0.0237***	-0.0138***	0.0810***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAPASSET	0.1330***	-0.0335***	0.4930***	0.1320***	-0.0334***	0.4950***	0.1320***	-0.0334***	0.4950***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	14667	14667	14667	14667	14667	14667	14667	14667	14667
R^2	0.137	0.083	0.058	0.137	0.083	0.058	0.137	0.083	0.058

Table 17 (continued): Regression Results Using DR

The table continues the regression estimates, now using DR as the diversification measure. In each panel we provide the estimates of 3 models, each using the same measure (or set of measures) of diversification. The first, second, and third models in each panel correspond to the dependent variables *ROA* (return on assets), *SDROA* (5-year standard deviation of ROA), and *RAROA* (risk-adjusted ROA, which is ROA divided by the underwriting risk, *UPR*), respectively. Panel 4 uses *DR* (Diversification Ratio), and Panel 5 includes *DR*, *SPREAD* (1-HHI), *GEODIV* (1-geographic HHI). Other independent variables include: *UPR* (underwriting portfolio risk), *SIZE* (log of total assets), and *CAPASSET* (policyholder surplus / total assets). Each model includes firm and year fixed-effects. The constant and the fixed effects are omitted from the table to conserve space. Robust standard errors, clustered at the firm level, are reported below the coefficients in parentheses. ***, **, and * indicate two-tailed statistical significance at the 0.01, 0.05, and 0.10 levels.

	Panel 4			Panel 5		
Model :	(10)	(11)	(12)	(13)	(14)	(15)
Dependent variable:	ROA	SDROA	RAROA	ROA	SDROA	RAROA
DR	-0.00215	-0.00277***	0.0496***	0.00217	-0.00384***	0.0685***
	(0.176)	(0.001)	(0.000)	(0.252)	(0.000)	(0.000)
SPREAD				-0.0172***	0.000416	-0.105***
				(0.000)	(0.862)	(0.000)
GEODIV				-0.0101**	0.00557***	-0.0206
				(0.015)	(0.010)	(0.399)
UPR	0.00985***	0.00869***		0.00992***	0.00866***	
	(0.000)	(0.000)		(0.000)	(0.000)	
SIZE	0.0227***	-0.0134***	0.0748***	0.0237***	-0.0136***	0.0787***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAPASSET	0.133***	-0.0336***	0.495***	0.132***	-0.0333***	0.495***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
N	14667	14667	14667	14667	14667	14667
R^2	0.136	0.084	0.06	0.137	0.084	0.061

Table 18: Summary Statistics for DR and DR Variation by Insurer Size Decile

This table contains the summary statistics for DR and the 5-year standard deviation in DR at the insurer level. The deciles, shown in the left column, are determined based on insurer size (as log of total assets). Within each decile, the number of insurer-year observations is reported in Column 2. Column 3 contains the number of single state-line insurer-year portfolios (those portfolios with a DR = 1) in each decile. The mean DR determined using all insurer-year portfolios in the decile is shown in Column 4. For each insurer-year observation, we determine the standard deviation in DR over the 5 years prior to the observation for the insurer. Within each decile of insurer-year observations, we report the mean of the standard deviations in Column 5.

Decile (by *SIZE*)	# of Obs	# of Single SL Insurers	Mean DR	Mean SD of DR
1	1,107	498	1.2131	0.0512
2	1,106	427	1.2622	0.0619
3	1,106	337	1.3193	0.0713
4	1,107	278	1.4257	0.0822
5	1,106	154	1.5306	0.1084
6	1,106	121	1.6421	0.1103
7	1,107	104	1.8597	0.1286
8	1,106	81	1.9119	0.1344
9	1,106	44	2.1790	0.1551
10	1,106	49	2.8734	0.2056

Table 19: Regression Results Using DR with a *SIZE*-limited Sample

The table presents the regression estimates using the full set of diversification measures as in Model (13) of Table 17 under increasingly limited samples based on insurer size (*SIZE*). Model (1) includes all firm-years and is identical to Model (13) of Table 17. Model (2) uses only the top 90% (by *SIZE*), Model (3) the top 80%, Model (4) the top 70%, Model (5) the top 60%, and Model (6) the top 50%. Each model uses *ROA* (return on assets) as the dependent variable. Each model includes firm and year fixed-effects. The constant and the fixed effects are omitted from the table to conserve space. Robust standard errors, clustered at the firm level, are reported below the coefficients in parentheses. ***, **, and * indicate two-tailed statistical significance at the 0.01, 0.05, and 0.10 levels.

Model :	(1)	(2)	(3)	(4)	(5)	(6)
Percentile of firm-years by SIZE:	All	Top 90%	Top 80%	Top 70%	Top 60%	Top 50%
Dependent variable:	ROA	ROA	ROA	ROA	ROA	ROA
DR	0.00217	0.00328*	0.00421**	0.00574***	0.00625***	0.00623***
	(0.252)	(0.061)	(0.013)	(0.000)	(0.000)	(0.000)
SPREAD	-0.0172***	-0.0183***	-0.0179***	-0.0159***	-0.0133***	-0.0125***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.004)	(0.010)
GEODIV	-0.0101**	-0.00902**	-0.0120***	-0.0169***	-0.0220***	-0.0125***
	(0.015)	(0.020)	(0.002)	(0.000)	(0.000)	(0.004)
UPR	0.00992***	0.0136***	0.0127***	0.0147***	0.0169***	0.0178***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SIZE	0.0237***	0.0207***	0.0179***	0.0163***	0.0114***	0.00821***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAPASSET	0.132***	0.133***	0.144***	0.151***	0.147***	0.156***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
N	14667	13200	11733	10266	8800	7333
R^2	0.137	0.157	0.172	0.199	0.215	0.229