PSTAT 296B
Winter 2023

How Much Surplus Should A Health Insurer Hold To Avoid Insolvency?

Students
Kai Wa Ho
Joanna Cai
Xinyin Huang
Yifan Wu

Advisors
Ian Duncan PhD, FSA, FIA, FCIA, FCA, CSPA, MAAA
Nickloas Thiessen, M.S. of Actuarial Science

Table of Contents

1. Executive Summary

The insurance regulators always require insurance companies to be well capitalized to support its operations and reduce the insolvency risk. The requirement of minimum excess capital has been a major tool implemented since the early days to protect the insurance companies from becoming insolvent. However, choosing the “right” level of capital that is equally acceptable to all insurance companies is not an easy problem. As a result, the topic of this research project is how much surplus a health insurer should hold to avoid insolvency using the risk-based capital requirements proposed by National Association of Insurance Commissioners (NAIC).

By constructing a dynamic model that connects all the assumptions and variables involved, we are able to simulate the financial conditions of health insurers by projecting their premium revenues, membership enrollments, claims amount, expenses, investment incomes, etc. and finally the surplus into the future and examine if any of them are susceptible to high insolvency risk indicating by an RBC ratio below 200%. Our study shows that health insurers are extremely unlikely to become insolvent if maintaining an RBC ratio above 1250%, while the risk of insolvency can go up to 30% if the RBC ratio is very close to the bottom line where intervention of insurance regulators starts to become necessary.

However, the report also highlights some limitations of the model. For example, the model may not account for all potential risks that could lead to insolvency, such as unexpected changes in the regulatory environment or shifts in market conditions. Additionally, the model assumes that insurers will have access to capital markets to raise additional funds if needed, which may not always be the case. Insurance companies must be aware of their limitations and use it in conjunction with other risk management strategies to ensure their long-term solvency.

2. Introduction

2.1 Sponsor

Santa Barbara Actuaries (SBA), is an actuarial and healthcare consulting firm specializing in cutting-edge predictive modeling and analytics. The company's services include financial outcomes evaluation, shared savings, financial value prop assessment, sales support, engagement optimization, expert witness, and general actuarial support. We hope that through our research, we can help the sponsor improve the consulting services on choosing a level of capitalization for the insurers.

2.2 Motivation

Limiting the insolvency risk of insurance companies has always been one of the top objectives of insurance regulations. The requirement of minimum capital has been a major tool implemented since the early days to protect the companies from becoming insolvent. However, choosing the "right" level of capital that is equally acceptable to all insurance companies is not an easy problem to solve because the financial risks are affected by a variety of factors including but not limited to the size of the companies, the type of business written, success of management, and degree of risk aversion or desire for profits. The solution proposed by the National Association of Insurance commissioners (NAIC) was to develop a hybrid risk-based capital (RBC) requirement. Instead of setting a fixed minimum RBC, the NAIC requires companies to hold minimum percentages of different assets and liabilities as capital, with these percentages based on their historical variability. The RBC requirementIt allows companies to operate freely at any given level above the minimum threshold and poses regulatory actions on companies that fall beneath the threshold. While (under the RBC regime) insurers rarely become insolvent, there have been recent claims that some companies are holding excessive surplus. Therefore, our research was motivated by the following prompt "How Much Surplus Should a Health Insurer Hold?" based on the RBC guidelines from the NAIC.

RBC requirements differ for different lines of business. Our focus on this project is health insurers.

2.3 Risk Based Capital & Regulation

Risk Based Capital (RBC) is the minimum amount of excess capital or surplus that an insurance company should hold based on its risk profile to avoid insolvency.

The NAIC health RBC formula classifies the risks faced by a health insurer into five major categories:

- H-0: Insurance Affiliates & Miscellaneous Other
  - This is the risk from declining value of insurance subsidiaries as well as risk from off-balance sheet and other miscellaneous accounts. Miscellaneous assets include derivative instruments held for purposes other than trading that have a positive fair value, computer software, and bullion.

- H-1: Asset Risk
  - This is the risk of assets' default of principal and interest or fluctuation in market value. The cash flow connected with assets will differ from expectations or assumptions for reasons other than a change in investment rate of return.

- H-2: Underwriting Risk
  - This is the risk of underestimating liabilities from business already written or inadequately pricing business to be written in the coming year. 80% of the required surplus under the RBC formula arises from Underwriting Risk.

- H-3: Credit Risk
  - This is the risk of recovering receivable amounts from creditors. Credit risk is the possibility of a loss resulting from a borrower's failure to repay a loan or meet contractual obligations. Traditionally, it refers to the risk that a lender may not receive the owed principal and interest, which results in an interruption of cash flows and increased costs for collection[1].

- H-4: Business Risk
  - This is the risk of general business. Business risk is the exposure a company or organization has to factor(s) that will lower its profits or lead it to fail. Anything that threatens a company's ability to achieve its financial goals is considered a business risk[2].

The formula for calculating the RBC is:

$$RBC = H_0 + \sqrt{{H_1}^2 + {H_2}^2 + {H_3}^2 + {H_4}^2}$$

[1] Kenton, W. (2022, September 13). *What is business risk? definition, factors, and examples*. Investopedia. Retrieved December 4, 2022, from https://www.investopedia.com/terms/b/businessrisk.asp#:~:text=Business%20risk%20is%20the%20exposure,is%20considered%20a%20business%20risk.

[2] Team, T. I. (2022, September 20). *Credit risk: Definition, role of ratings, and examples*. Investopedia. Retrieved December 4, 2022, from https://www.investopedia.com/terms/c/creditrisk.asp.

The aggregation of H1 to H4 using the “square root rule” can be interpreted as an assumption that there is zero correlation among those risks. It can also be interpreted as those risks under the square root are calibrated to yield a more appropriate result for “typical” insurers.

Under the RBC system, regulators have the legal authority to take preventive and corrective measures. These measures vary depending on the capital deficiency indicated by the RBC ratio, which is defined as the total adjusted capital (TAC) divided by the authorized control level (ACL). As we can see from Table 1, if the ratio is at or above 200%, no regulatory intervention is needed. Below that ratio, there are four levels of regulatory intervention ranging from submission of action plans to a regulatory takeover of the management of the company. If the ratio is below 70%, a regulator is obligated to take over management of the company. These preventive and corrective measures are designed to provide for early regulatory intervention to correct problems before insolvencies become inevitable, thereby minimizing the adverse impact of insolvencies.

Table 1: Risk-based Capital Authoritative Action Outcomes

Table 1: Risk-based Capital Authoritative Action Outcomes

RBC Ratio (= TAC/ACL)	Outcome
>200%; No Action Level	No action is required.
150% to 200%; Company Action Level	The health care insurer is required to submit a business plan to improve financial strength.
100% to 150%; Regulatory Action Level	The health care insurer is required to submit a business plan to improve financial strength. Also, the regulator is authorized to perform a review of practices.
70% to 100%; Authorized Control Level	The regulator is authorized to take actionable steps to improve the financial strength of the health care insurer.
<70%; Mandatory Control Level	The regulator is required to take actionable steps to control the health care insurer.

One issue with the current RBC regulation is that it is static: that is, it assesses the current capital adequacy of a company. It does not project the company’s business into the future to determine whether the company’s capital base is adequate for its current and expected future business. To perform this assessment requires a dynamic model, which we consider in the next section.

3. Data Summary

3.1 NAIC Data

The data obtained from the National Association of Insurance Commissioners has data on Health/Life Insurance Companies' Annual Statements. The dataset contains the relevant data of 101 insurance companies from the year 2005-2019.

NAIC important fields include:

1. Company Name
2. State(s) of Business
3. Statutory Minimum Capital and Surplus Requirement
4. Total Liabilities, Admitted Assets, Capital/Surplus, (Premium) Revenues
5. Total Administrative Expenses, Medical and Hospital Expenses
6. Underwriting Gain(Loss), Net Investment Gain(Loss), Total Other Income, Net Income
7. Total Adjusted Capital, Authorized Control Level, Risk-Based Capital
8. Total Member Months
9. Loss Ratio
10. RBC Ratio

3.2 Data Cleaning Process

This dataset is extremely important to the development of our dynamic model as most of the state variables of the model come from this dataset. Therefore, one of the initial main tasks this quarter is to perform a thorough data screening. We examined the data anomalies including the data ones that are missing, mis-recorded, and abnormal but justifiable.

Some data are missing or recorded as 0 because the company might be in the start-up stage (or may have ceased writing business). For instance, Ambetter of Peach State, Inc. doesn't have a positive total revenue (all zeroes) until 2015. And if we also look at its total admitted assets or total members months, the numbers only start to accumulate after 2014. Thus, we believe that Ambetter of Peach State didn't underwrite any policies until 2015, and we will exclude its data prior to 2015 when modeling.

Some data are mis-recorded and can be fixed by recalculating using the correct formula. For instance, from 2006 to 2007, the total admitted assets of Blue Cross Blue Shield of Michigan are down by one digit (the others are almost ten times higher). After calculating the total admitted assets by the sum of total liabilities and total capital and surplus, we found that the new numbers actually make a lot more sense, and thus we believe that the original data is indeed misrecorded.

On the other hand, Cigna Health and Life Insurance Company has relatively volatile total revenues, even a negative revenue in 2008. When we looked at the total number of months, the entire row was “not applicable”, and we could not justify the negative value. Therefore, we neutralized the negative number to the average of 2007 and 2009.

Last but not least, we carefully removed some of the companies because the data anomalies were more than a few and they were hardly justifiable. One example is the John Alden Life Insurance Company, which has extremely small or negative revenues and expenses between 2016 and 2019, while its total member months are not accessible and its total admitted assets, liabilities, and total capital and surplus look totally fine. Therefore, we decided to remove John Alden Life when constructing our model.

The method we used to identify those data anomalies above is to use the trend or change between consecutive years. The supplementary criterias include expense ratio (within 5-10% of total premium revenues), total administrative expenses over total premium revenues (around 10-15%, ideally below 20%), stable loss ratio and RBC ratio etc.

3.3 Other Data

- **Trend and Concentration Index**

The concentration index tells us how evenly market share is distributed across insurers in the market.

- **MarketScan Sample Claims Data**

MarketScan Sample Claims Data is a family of health research databases primarily including commercial private medical insurance paid claims data. We will use this database to project stochastically the level of annual claims in our dynamic model.

- **MarketScan 2017_100k_sample**

This data is a random 100k MarketScan 2017 sample with 12-month continuous eligibility.

- **Towers Watson annual report annual trend report**

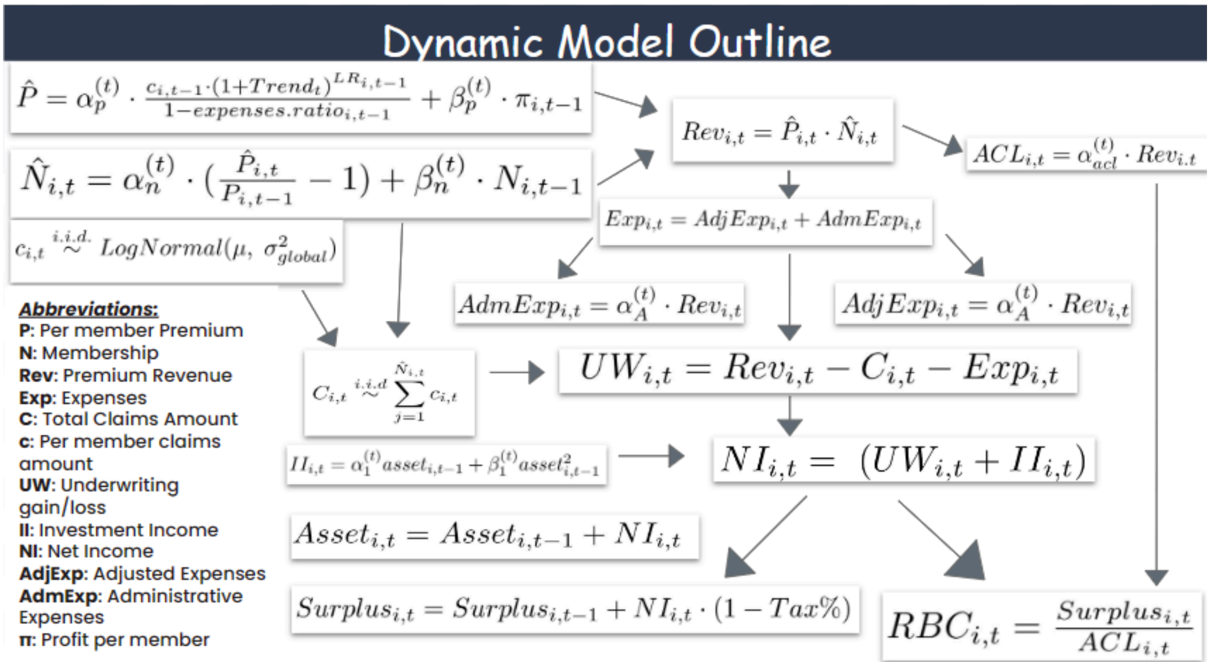
Healthcare Cost Trend is defined as the relative increase in cost per member per month of a health plan. For the dynamic model we will need to project costs annually and we will require a trend model. ItThe Towers Watson survey was completed by 687 U.S employers with at least 100 employees between June and July 2018. It reflects respondents’ 2018 health plan decisions and strategies as well as expected changes for 2019 and 2020. For consistency with prior years’ surveys, results provided in this report are primarily based on 554 employers, each with at least 1,000 employees. Collectively they employ 11.4 million employees and operate in all major industries.

4. Model Development

4.1. Dynamic model outline

The objective of our research is to develop a dynamic model to determine the risk-based capital for health insurers. The reason that the model is dynamic is because the system of the model changes with time. For instance, the ultimate claims amount varies each year, causing the pool of capital or surplus to change, which in turn affects the premiums charged during the next policy year, and leads to an increase or decrease in membership. Relationships between variables emerge from the model as the result of the underlying process, which is expressed as a set of equations specifying how the surplus changes over time as a function of the current and past values of the state variables such as premiums, memberships, claims, etc.. The complete system of model is shown in the following figure:

Figure 1: Dynamic Model Outline



4.2 Analysis on model components

The following section will list all the components in the dynamic model. For explanation purposes, we plugged the weighted averages into the coefficient parameters.

4.2.1. Premium

$$\hat{P}_{i,t} = 1.0044 \cdot \frac{c_{i,t-1} \cdot (1+Trend_t)^{LR_{i,t-1}}}{1-expenses.ratio_{i,t-1}} + 0.71523 \cdot \pi_{i,t-1} \quad \textbf{(1)}$$

$\hat{P}$ is the premium per member per month. c is the average claim. Trend is the inflationary factor of the claims in the following year. LR is the loss ratio which is defined by the total claim amount over total premium revenues. π is the profit margin, which can be defined as the profits gained from each member.

The weighted average of α is 1.0044 and β is 0.75123 respectively. The fraction in this formula reflects the expected premiums after the companies trend their average claims to the current level. So if α equals to 1, it means my trended results are most reasonable and accurate. If α is greater or less than 1, then it means my trended results are under or over estimated. The coefficient of π means for every dollar the insurer profits from each member, the average premiums will increase by approximately $0.75.

Table 2: values of the (premium) coefficients from year 2006-2019

	Year	adj_claimpm	pmpm
1	2019	0.99643	0.80081
2	2018	1.01000	1.79229
3	2017	0.99770	0.88430
4	2016	0.99840	0.32550
5	2015	0.97035	-0.18180
6	2014	0.99279	0.75459
7	2013	1.10053	-1.48166
8	2012	0.97020	1.33934
9	2011	0.97717	0.83584
10	2010	0.98931	0.77919
11	2009	1.03060	0.58797
12	2008	1.01331	1.33780
13	2007	0.97541	2.10647
14	2006	1.04200	0.19222

4.2.2 Membership

$$\hat{N}_{i,t} = -3.1 \cdot 10^6 \cdot \left(\frac{\hat{P}_{i,t}}{P_{i,t-1}} - 1\right) + 1.0314 \cdot N_{i,t-1} \quad \textbf{(2)}$$

$\hat{N}$ is the total member months (also known as the membership enrollment) and $\hat{P}$ (with or without hat) is the premium per member per month. $\frac{\hat{P}_t}{P_{t-1}} - 1$ represents the change in the "observed premium" last year compared to the current estimated premium from equation (1). The first coefficient means for every 1% increase in premium, the insurer will lose around 30,000 members. The second coefficient means what's the proportion of members from last year who stay in the policies for the following year. If the value is larger than 1, it means new members are joining from the group of "insurables" which are people eligible for insurance but do not have insurance due to reasons like poverty, age, and unemployment. The membership model could be refined because first, it is impossible for every company to lose 30,000 members if they increase their premium by 1%. But the idea is for large-sized companies with a large membership base, the insureds are less sensitive to the increase in premium compared to smaller companies (i.e. a company with 10 million members vs. a company with 100,000 members, then 1% increase in premium will change -1% and -30% of the membership respectively). The results indicate that smaller companies are not encouraged to raise their premiums as they are already at a competitive disadvantage compared to larger companies with a larger base of membership (or loyal insureds). Second, we actually had the "insurables" data but it has a very poor linear correlation with the membership. We would like to include this piece of information, yet our multiple attempts of incorporating it into our model all ended in failure.

Table 3: values of the (membership) coefficients from year 2006-2019

	Year	Change_in_ pmpm	old_member
1	2019	-4256000	1.0400
2	2018	20300	1.0310
3	2017	-459900	1.0070
4	2016	-3079000	0.9714
5	2015	-6090000	1.0560
6	2014	100000	1.0520
7	2013	-941500	1.0240
8	2012	-1392000	1.0310
9	2011	-1181000	0.9824
10	2010	-3371000	1.0290
11	2009	-9124000	1.1220
12	2008	-7371000	1.0480
13	2007	-5771000	1.0270
14	2006	-1237000	1.0500

4.2.3 Premium Revenue

$$Rev_{i,t} = \hat{P}_{i,t} \cdot \hat{N}_{i,t} \qquad \textbf{(3)}$$

Total premium revenues = premium per member per month * total member months

4.2.4 Expenses

$$Exp_{i,t} = AdjExp_{i,t} + AdmExp_{i,t} \quad \textbf{(4)}$$

Total expenses equal the sum of claims adjustment expenses and administrative expenses. Both claim adjustment expenses and administrative expenses can be expressed as percentages of the total premium revenues.

$$AdmExp_{i,t} = 0.083 \cdot Rev_{i,t} \quad \textbf{(5)}$$

$$AdjExp_{i,t} = 0.037 \cdot Rev_{i,t} \quad \textbf{(6)}$$

Table 4: values of the (expenses) coefficients from year 2005-2019

	year	cd_coe1		year	ad_coe1
1	2019	0.0366171	1	2019	0.083270
2	2018	0.0354348	2	2018	0.097223
3	2017	0.0348189	3	2017	0.078029
4	2016	0.0318183	4	2016	0.091385
5	2015	0.0338090	5	2015	0.095160
6	2014	0.0377430	6	2014	0.098411
7	2013	0.0337750	7	2013	0.079995
8	2012	0.0339700	8	2012	0.078916
9	2011	0.0338050	9	2011	0.084714
10	2010	0.0407670	10	2010	0.080745
11	2009	0.0426300	11	2009	0.083311
12	2008	0.0415500	12	2008	0.072755
13	2007	0.0419150	13	2007	0.070065
14	2006	0.0393140	14	2006	0.071622
15	2005	0.0359720	15	2005	0.073295

4.2.5 ACL Model

$$ACL_{i,t} = 0.038 \cdot Rev_{i,t} \quad \textbf{(7)}$$

Authorized Control Level is calculated as a percentage of total premium revenues which will be used in calculating the final RBC ratios.

Table 5: values of the (expenses) coefficients from year 2005-2019

	year	acl_coe1
1	2019	0.0383836
2	2018	0.0369819
3	2017	0.0365917
4	2016	0.0388190
5	2015	0.0391340
6	2014	0.0381900
7	2013	0.0386420
8	2012	0.0381310
9	2011	0.0355490
10	2010	0.0362770
11	2009	0.0382190
12	2008	0.0376740
13	2007	0.0382800
14	2006	0.0400880
15	2005	0.0398960

4.2.6 Claims

$$C_{i,t} \overset{i.i.d}{\sim} \sum_{j=1}^{\hat{N}_{i,t}} c_{i,t} \quad \textbf{(8)}$$

Most of our coefficients are deterministic based on linear regressions. However, we choose to model the non-zero claims stochastically because it is one of the most uncertain quantities in our model that can leverage the RBC ratios. We assume that in each company, individual members will generate non-zero claims which approximately follow a probability distribution. If the company has N members, then we will generate N samples from the distribution and then calculate the sum and mean to obtain the total claims amount and average claims respectively. Details of this model will be discussed later.

4.2.7 Underwriting Gains/Losses

$$UW_{i,t} = Rev_{i,t} - C_{i,t} - Exp_{i,t} \quad \textbf{(9)}$$

Underwriting gains or losses are premium revenues minus the total claim amount and all other expenses.

4.2.8 Investment Income

$$II_{i,t} = 0.034 \cdot asset_{i,t-1} + 4.89 \cdot 10^{-13} \cdot asset^2_{i,t-1} \quad \textbf{(10)}$$

For investment income, we proposed a linear model with linear and quadratic terms of assets. The above equation (with the weighted averages of all coefficients in different years) can be interpreted as the more assets a company holds, the more investment income it gains. Nonetheless, we noticed some negative values of the quadratic term in several years (shown in Table(4)) which indicate that for large enough assets, a further increase in assets will lead to less growth in investment income. It is known as the marginal diminishing returns in economics. Our investment income model has several limitations. First, we observed in our data that negative investment incomes accompany both small and large levels of total assets. But our proposal fails to generate negative returns. Second, we chose this model with two terms because it has a better fit overall than the ones with linear term only and quadratic term only. Adding terms with higher degrees will not yield much better fit and it will lose the ability of interpretation. If we had more information about what the investment incomes of each company consisted of, then we are able to take more variables into account.

Table 6: values of the (investment income) coefficients from year 2005-2019

	year	inv_coe1	inv_coe2
1	2019	0.02820	9.192e-13
2	2018	0.02403	-4.556e-14
3	2017	0.02218	2.846e-12
4	2016	0.01874	-8.966e-14
5	2015	0.02061	1.261e-12
6	2014	0.02979	4.993e-13
7	2013	0.03415	2.113e-12
8	2012	0.02690	2.078e-12
9	2011	0.03403	-6.582e-13
10	2010	0.03868	1.791e-13
11	2009	0.08649	-3.770e-12
12	2008	0.01196	-1.746e-13
13	2007	0.04273	2.944e-12
14	2006	0.05730	-3.356e-12
15	2005	0.03773	2.590e-12

4.2.9 Net Income

$$NI_{i,t} = \left(UW_{i,t} + II_{i,t}\right) \quad \textbf{(11)}$$

Net Income is the sum of underwriting gains or losses and the investment incomes.

4.2.10 Asset

$$Asset_{i,t} = Asset_{i,t-1} + NI_{i,t} \cdot (1 - Tax\%) \quad \textbf{(12)}$$

Here we assume that the next year's assets will be the previous year's assets plus the net income minus taxes which is usually a significant factor in change of assets. But this is still a rough approximation because we don't have enough information to study the actual changes in liabilities and equities.

4.2.11 Surplus

$$Surplus_{i,t} = Surplus_{i,t-1} + NI_{i,t} \cdot (1 - Tax\%) \quad \textbf{(13)}$$

Similar to assets, the next year's surplus will be the previous year's surplus plus the net income minus taxes.

4.2.12 RBC

$$RBC_{i,t} = \frac{Surplus_{i,t}}{ACL_{i,t}} \quad \textbf{(14)}$$

Using the authorized control level calculated in equation (7), we can derive the RBC ratios by surplus over ACL.

4.3 Modeling claims

Considering the fact that claims are heavy-tailed, we choose four possible distributions: log-normal, weibull, pareto, and inverse gamma. We fit the data into these distributions in R and use a summary of the fit with visual diagnostics to determine which one is the best. (Plots are on the next page)

After comparing the empirical and theoretical densities, QQ-plots, empirical and theoretical cumulative distributions functions and P-P plots, we believe that the log-normal distribution might be the best fit. (It also has the second lowest AIC.) Nonetheless, we would like to double check the normality of our log-transformed claims. As it turns out, the log-transformed claims do not have a good normal approximation for small quantiles of claims. It has a left tail clearly deviated from the normal line indicating a possible left skew. It implies that the log-normal density underestimates the actual density of small-size claims as we can see from the zoomed-in histogram (the blue line represents the actual density and the red line represents the log-normal density). But for most part of the log-transformed claims, the QQ-plot shows a very good fit.

Figure 2: Diagnostic plots for Log-normal fit

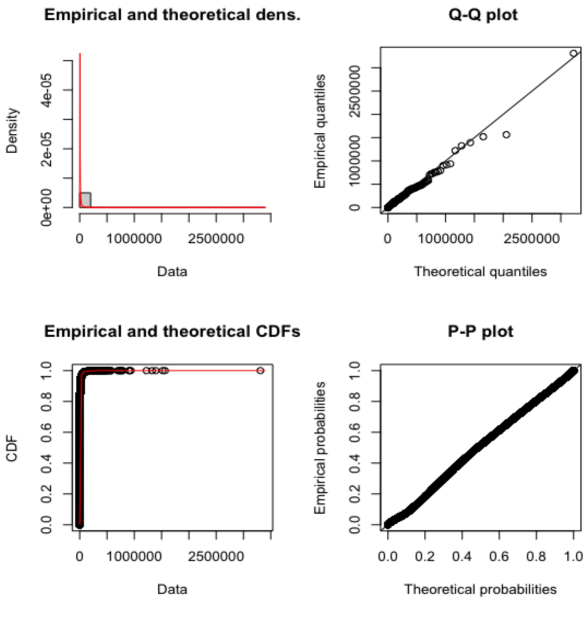


Figure 3: Diagnostic plots for Weibull fit

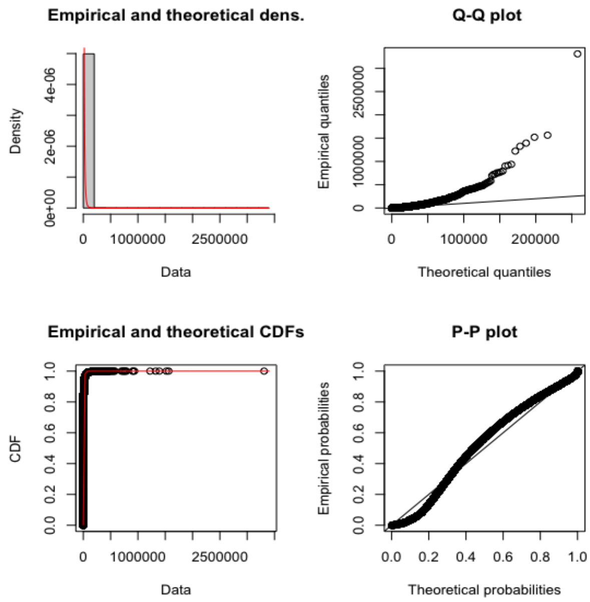


Figure 4: Diagnostic plots for Pareto fit

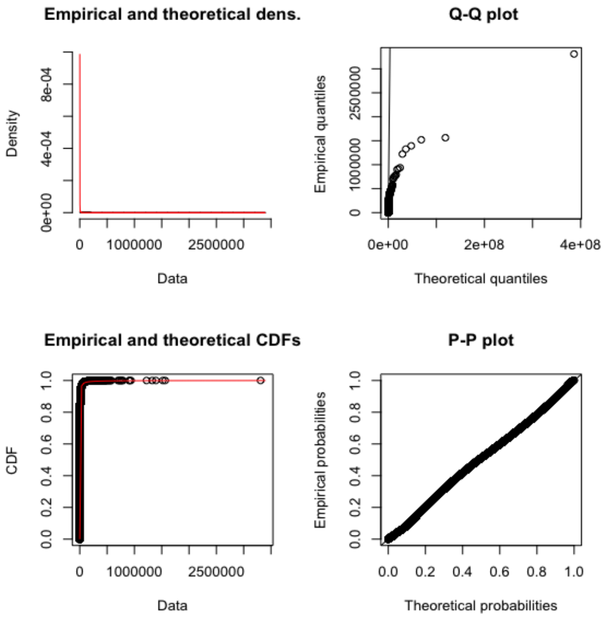


Figure 5 :Diagnostic plots for Inverse-gamma fit

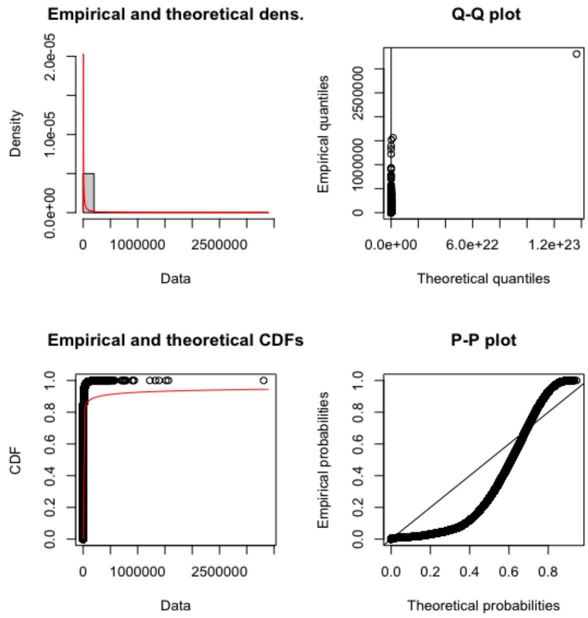
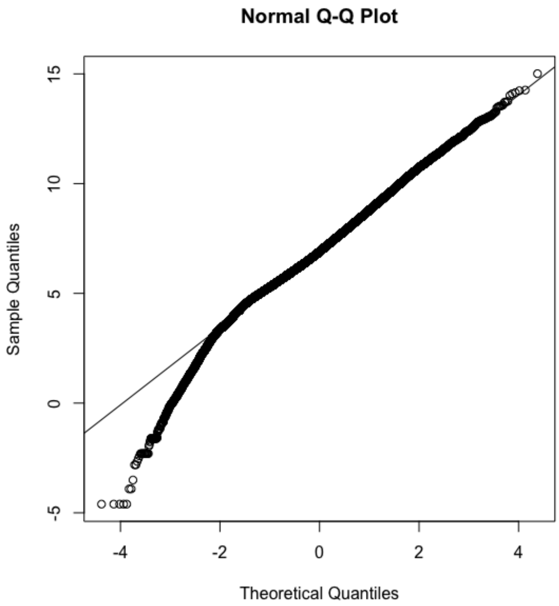


***Figure 6 is the QQ plot for log transformed claims. Figure 7 is the histogram showing the group of log-transformed claims that explains the deviation of the left tail from the normal line.**

Figure 6: QQ plot for log-normal fit **Figure 7: Histogram of non-log-normal samples**



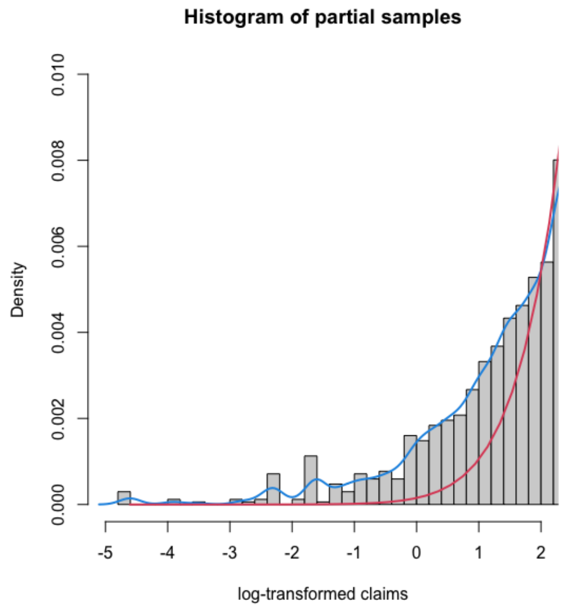


Figure 8: Histogram of log-transformed claims

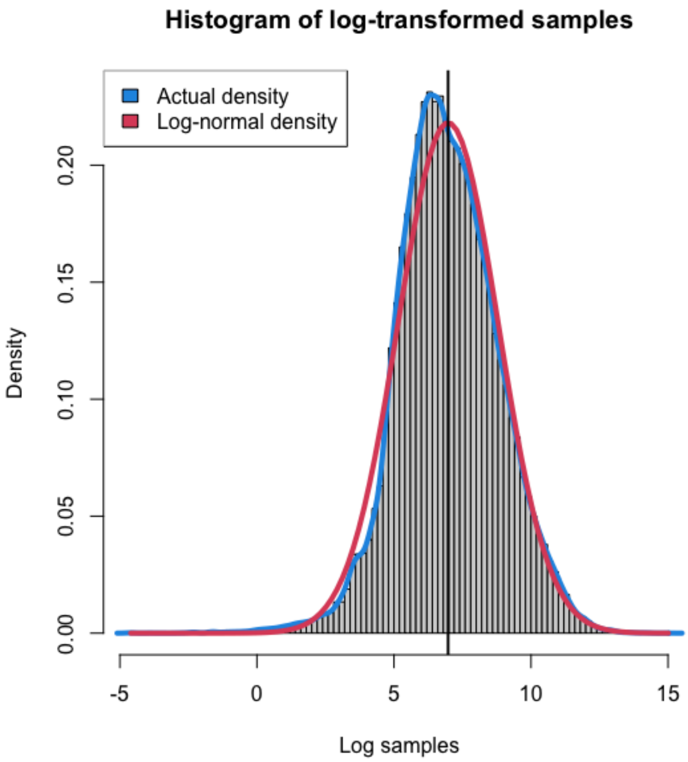


***Figure 8 above shows the actual density (in blue) and the log-normal density (in red) of the claims**

Other than the diagnostic results, we decided to use the log-normal distribution because it gives us a promising scaling method. Our data analysis shows that the mean of the 100,000 sample claims in the market is approximately $4,862. But the mean in our NAIC dataset only ranges from $19 to $883. The simplest way of scaling is to use the ratio of the average claims of each company to the average claims in the market:

$\widehat{X}_{i(scaled)} = \bar{X}_{samples} * \frac{\bar{X}_i}{\bar{X}_{market}}$ where $\widehat{X}_{i(scaled)}$ is our projected average claim of the i^{th} company, $\bar{X}_{samples}$ is the mean of claims generated from the log-normal distribution, $\overline{X}_i$ is the average claim of the i^{th} company, and $\bar{X}_{market}$ is the average claim in the market.

However, this scale method is a little naive because the only variability in this scaling method is $\bar{X}_{samples}$ which comes from the market data. After doing some research, we came up with an alternative approach which is fixing the dispersion around the mean, also known as the coefficient of variation, of claims for all individual companies.

Here are the methodologies. Suppose we have the market claims following a log-normal distribution with mean μ and standard deviation σ (both parameters can be estimated using maximum likelihood estimators) and the claim variable of a company is denoted by X. We can describe the dispersion of claims with the ratio of standard deviation of the claims to the mean of the claims defined as COV (coefficient of variation). According to the algebra, we discovered that the COV only depends on σ, so a constant COV can be calculated by $\sqrt{e^{\sigma^2} - 1}$ with a fixed σ. Interestingly, the mean and variance of the original claims (without log transformation) depend on both μ and σ. Therefore, decrease in μ will shift the distribution to the left, but will also decrease both the actual mean and variance of the claims. These results enabled us to create miniatures of the market that can represent each individual company with their own claim distributions (i.e. $X_i \sim LogNormal(\mu_i, \sigma^2)$, $X_i s$ are not identical but independent). A simple example to explain this result is if a company with an

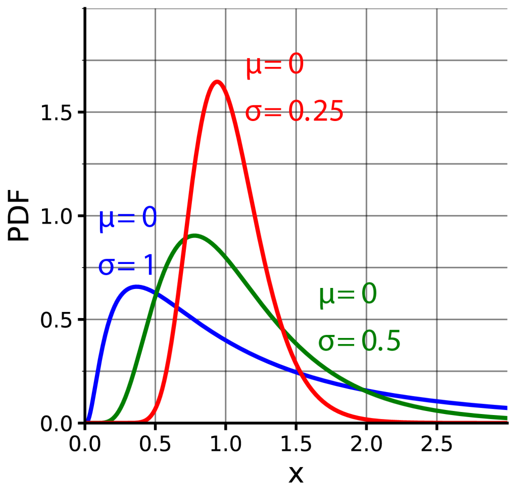
$$\mathrm{E}[X] = e^{\mu + \frac{1}{2}\sigma^2},$$
$$\mathrm{E}[X^2] = e^{2\mu + 2\sigma^2},$$
$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = (\mathrm{E}[X])^2(e^{\sigma^2} - 1) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1),$$
$$\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]} = \mathrm{E}[X]\sqrt{e^{\sigma^2} - 1} = e^{\mu + \frac{1}{2}\sigma^2}\sqrt{e^{\sigma^2} - 1},$$

The arithmetic coefficient of variation $\mathrm{CV}[X]$ is the ratio $\frac{\mathrm{SD}[X]}{\mathrm{E}[X]}$. For a log-normal distribution it is equal to[2]

$$\mathrm{CV}[X] = \sqrt{e^{\sigma^2} - 1}.$$

average claim of $100 has claims ranging from $0 to $200, then a company with an average claim of $1000 might have claims ranging from $0 to $2000. Clearly, the latter company has both higher mean and variance, but the dispersions around the mean for those two companies are the same. We believe under the log-normal distribution setting, this scaling method is plausible in the sense that larger insurance companies are more likely to underwrite higher risks and receive filings of extremely large claims (more significant tail) that raise the claims average way above those small or mid sized companies. Nonetheless, we have no idea how the claims in each company are actually distributed so this method requires further examination if we have more data.

Figure 9: Probability Density Function of Log-Normal Distribution



[3]

5. Simulation

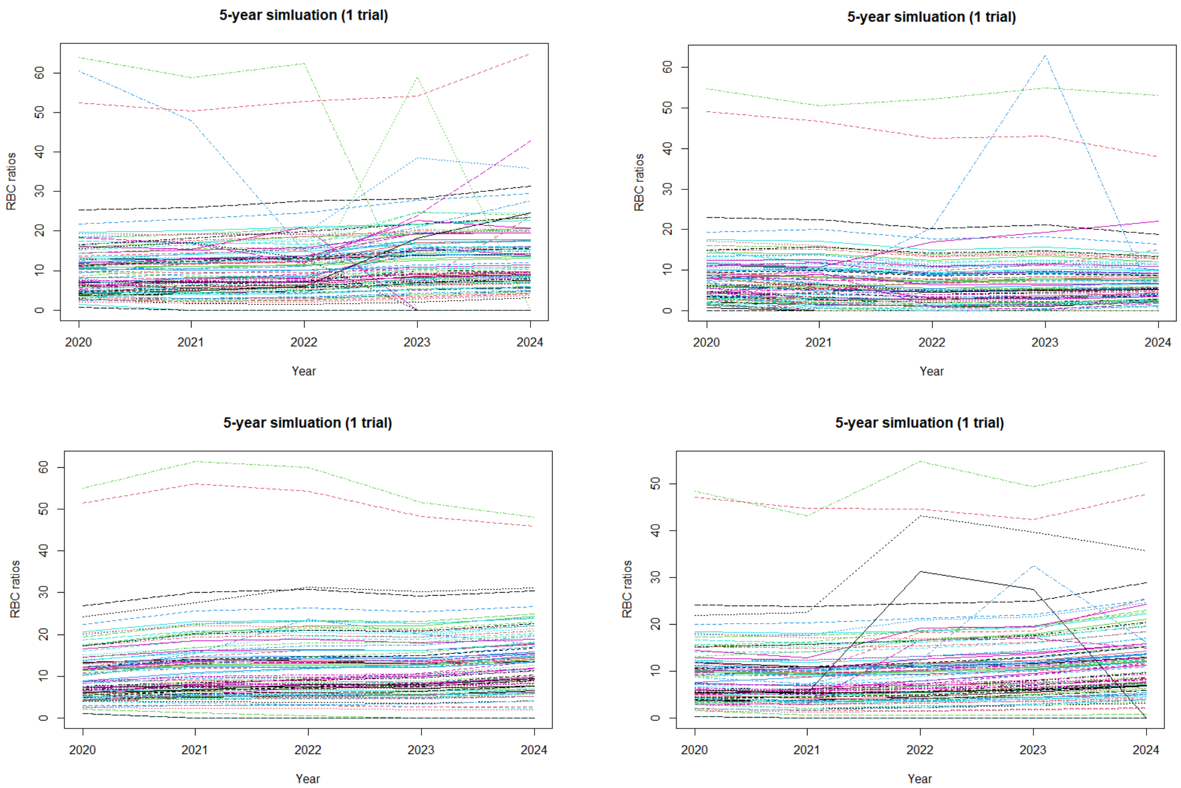
[3] https://en.wikipedia.org/wiki/Log-normal_distribution

For simulation, we chose to bootstrap the sets of coefficients (from the tables produced above) instead of using weighted averages. There are two main reasons:

1. The weighted averages are constants from a complete pooling result, and thus it can only simulate the "expected" (not necessarily the most possible) scenarios for health insurers. It doesn't align with our purpose of simulations, which is to measure the uncertainty of how state variables evolve over time for each company and examine their risks of insolvency.
2. "History repeats itself." For each model we created, all sets of coefficients are unique from year 2005 to 2019 (even if some coefficients from different years are very close). Bootstrapping the coefficients enables us to repeat one of the settings including the pricing strategies of insurance companies indicated by change in premiums, the economic environment indicated by investment incomes, etc. in the future.

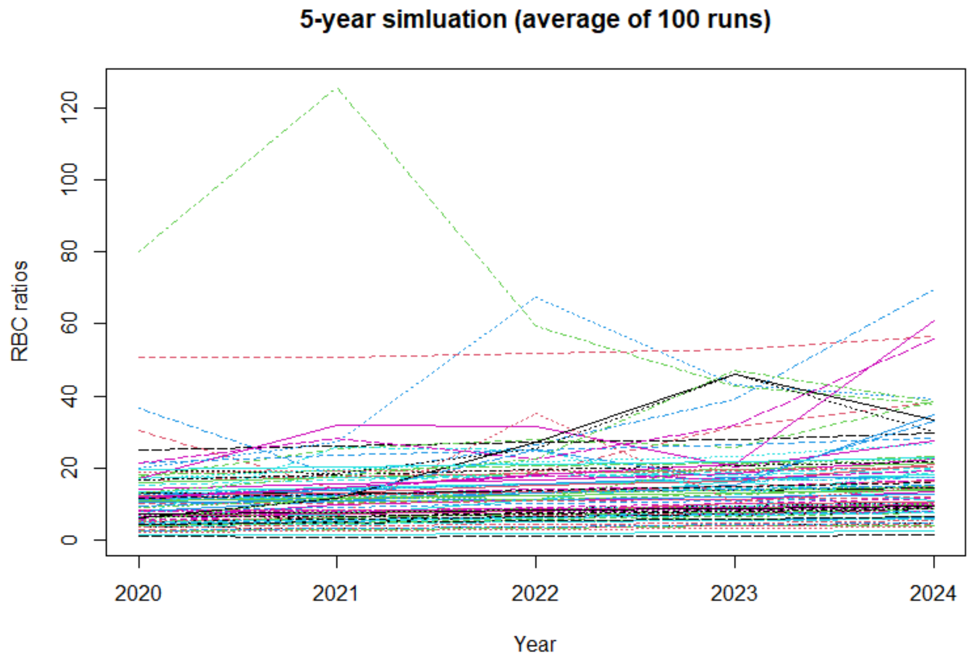
The following are some trace plots of RBC ratios under one simulation.

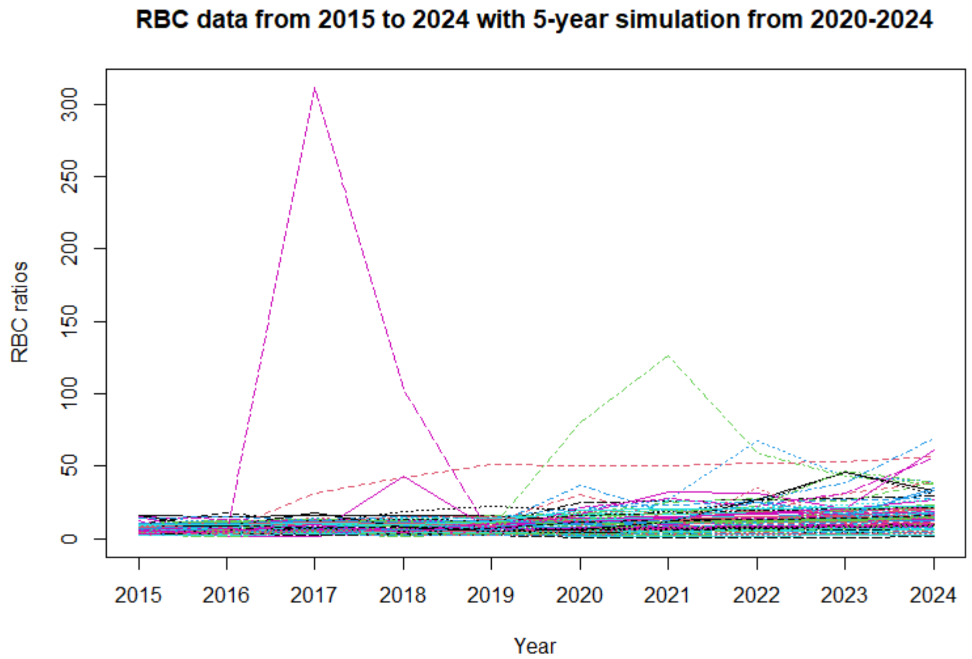
Figure 10-13: Trace plots of RBC simulations from 2020-2024



During the 5-year simulation period, the RBC ratios don't have a large fluctuation for most of the companies. Those "sharp" turns in the plots might indicate that the bootstrapped coefficients don't describe the corresponding companies well in those specific years. Therefore, we decide to run 100 times and observe how their averages behave.

Figure 14 & 15: 100-trial averages of RBC ratios from 2020 to 2024 (with the observed data from 2015 to 2019 and simulated data from 2020-2024





Running 100 trials stabilizes the RBC ratios except for the one in green which might be an outlier. Out of 101 companies, only two will have an RBC ratio less than 200% on average after 5 years. The results are within our expectation because generally companies with proper management should be capitalized well enough to withstand insolvency.

6. Conclusion

Using the 100 runs of 5-year simulation above, we count the times that the companies become insolvent and compute a table of probabilities based on their RBC levels from the observed data. Our simulations show that within the next 5 years, companies with an RBC ratio above 1250% are very unlikely to become insolvent. Companies with an RBC ratio between 800% and 1250% have about 10% probability of being insolvent. Companies with an RBC ratio below 800% have an increasingly high probability of insolvency. Finally, companies with an RBC ratio just above 200% have the maximum 28% probability of being insolvent.

Table 7: Insolvency probabilities based on RBC levels

RBC Ratios	Insolvency probability
RBC >=2000%	*1%*
1500% <= RBC < 2000%	*1%*
1250% <= RBC < 1500%	*3%*
1000% <= RBC < 1250%	*6%*
800% <= RBC < 1000%	*4%*
600% <= RBC < 800%	*12%*
400% <= RBC < 600%	*20%*
200% <= RBC < 400%	*28%*

The table provides some insights to how much surplus a health insurer should hold based on their current risk-based capital. However, our calculation of the insolvency probabilities does not take into account the fact that companies with similar RBC ratios can have very different data such as premiums and membership due to the size of the businesses. Therefore, the computation of insolvency probabilities can be improved if we classify the companies by adding extra criteria such as the combination of RBC levels with revenues or membership.

7. Future Research

The dynamic model we use is a simplified model from the dynamic linear model which incorporates multivariate time series analysis. Since we only have 15 years of data, it is not feasible for us to identify any underlying seasonal trend among the coefficients. If we have more years of data, then we may model the coefficients as a function of time. We can also use forecasting to adjust any assumptions or formula if necessary.

Another extension to our dynamic model is studying the conditional distributions. Because the structure of our model contains the relationship between all state variables such as premiums, memberships, expenses, surplus, etc., one can write out, for instance, the conditional value of premiums in the current year based on the average claims and profit margins from last year. In this case, we can use a non-parametric distribution-free model instead of a linear model to describe the relationships. This method is not generally preferred over the parametric model because of the computational complexity. Thus, an alternative is the Bayesian hierarchical modeling with (Gaussian) linear regressions which assumes that there exists conditional probability distributions between the variables. It's similar to the idea of a dynamic linear model, except that the coefficients do not follow functions of time anymore. Instead, it is conditioned on the values of other parameters.

On the Bayesian side, we can also use Monte Carlo Markov Chain if we continue to assume that the coefficients can be summarized by probability distributions. This might be the hardest method to realize because for MCMC, we need to choose the prior distributions for premiums, memberships, claims, and so on, but we don't have a reasonable "belief" about what their prior distributions are. We can put the simplest uniform prior on the positive real line for all variables, but the computational costs will be well beyond our imagination. However, it is possible to obtain a more reasonable prior from the results of research papers or experienced "guesses" from insurance specialists. Then, we can compute the posterior distributions of the (linear) coefficients through our data, and carry out the simulations with the posterior predictive distributions of our variables.

8. Bibliography

1. NAIC DATA, Santa Barabra Actuaries(SBA)
2. Trend and Concentration Index, Santa Barabra Actuaries(SBA)
3. MarketScan Sample Claims Data, Santa Barabra Actuaries(SBA)
4. MarketScan 2017_100k_sample , Santa Barabra Actuaries(SBA)
5. Towers Watson annual report annual trend report , Santa Barabra Actuaries(SBA)
6. Kenton, W. (2022, September 13). *What is business risk? definition, factors, and examples*. Investopedia. Retrieved December 4, 2022, from https://www.investopedia.com/terms/b/businessrisk.asp#:~:text=Business%20risk%20is%20the%20exposure,is%20considered%20a%20business%20risk.
7. Team, T. I. (2022, September 20). *Credit risk: Definition, role of ratings, and examples*. Investopedia. Retrieved December 4, 2022, from https://www.investopedia.com/terms/c/creditrisk.asp.
8. "Log-Normal Distribution." Wikipedia, Wikimedia Foundation, 6 Feb. 2023, https://en.wikipedia.org/wiki/Log-normal_distribution.

9. Appendix

9.1 Data Dictionary

Variable	***Type***	***Important***	***Summary of Variables***
Company	Text	Yes	Indication of the data of different companies
State	Text	No	The state that companies located
NAIC#	Numeric	No	File number
Statement type	Numeric	No	Type of the companies
Public/Privacy	Numeric	No	Differentiate the owner of the companies
Line	Numeric	No	The line of the item
Line Item	Categorical	Yes	Variables
Date	Date	Yes	Range from 2005 - 2019
Total Admitted Assets	Numeric	Yes	Data given by NAIC
Total Liabilities	Numeric	Yes	Data given by NAIC
Statutory minimum capital and surplus requirement	Numeric	Yes	Minimum surplus required to be hold by companies according to different state
Total capital and surplus	Numeric	Yes	Data given by NAIC
Total Revenue	Numeric	Yes	Data given by NAIC
Total Medical and hospital expenses	Numeric	Yes	Data given by NAIC
Claims adjustment expenses	Numeric	Yes	Data given by NAIC
TOtal administrative expenses	Numeric	Yes	Data given by NAIC
Net underwriting gain(loss)	Numeric	Yes	Data given by NAIC
Net underwriting gain	Numeric	Yes	Data given by NAIC
Net investment gain	Numeric	Yes	Data given by NAIC

Total Other income	Numeric	Yes	Data given by NAIC
Net income or loss	Numeric	Yes	Data given by NAIC
Net cash from operations	Numeric	Yes	Data given by NAIC
Total adjusted capital	Numeric	Yes	Data given by NAIC
Authorized control level risk-based capital	Numeric	Yes	Data given by NAIC
Total members at the end of period	Numeric	Yes	Data given by NAIC
Total members months	Numeric	Yes	Data given by NAIC
Loss ratio	Numeric	Yes	$\frac{Total\ Medical\ and\ Hospital\ expense}{Total\ Revenues}$
RBC ratio	Numeric	Yes	$\frac{Total\ Adjusted\ Capital}{Authorized\ Control\ Level\ Risk{-}Based\ Ca}$
TAC PMPM	Numeric	Yes	$\frac{Total\ Adjusted\ Capital}{Total\ Member\ Months}$
TAC % Premomim	Numeric	Yes	$\frac{Total\ Adjusted\ Capital}{Total\ Revenue}$
Net income % of premium	Numeric	Yes	$\frac{Net\ Income\ or\ Loss}{Total\ Revenue}$
Net Underwriting gain or loss % or premium	Numeric	Yes	$\frac{Net\ Underwriting\ Gain\ or\ Loss}{Total\ Revenue}$
Total Amin % of Premium	Numeric	Yes	$\frac{Claims\ Adjustment\ Expenses}{Total\ Revenue} + \frac{Total\ Administrative\ Expenses}{Total\ Revenue}$
Total Admin PMPM	Numeric	Yes	$\frac{Claims\ Adjustment\ Expenses}{Total\ members\ months} + \frac{Total\ Administrative\ Expenses}{Total\ members\ months}$
Surplus Months of Claims	Numeric	Yes	$Total\ Adjusted\ Capital \div \frac{Total\ Medical\ and\ Hospital\ Expenses}{Net\ Income\ or\ Loss}$

* RBC ratio and Surplus Months of Claims are our focus of this project

End of Quarter Report