Insurance Portfolio Analysis in the State of Florida

Daniel Ricardo Sarmiento[1,1,*]

[a]*Bogota D.C Colombia*

Abstract

Insurers have and manage portfolios that are simply a set of policies, in this article we seek to analyze a portfolio that contains policies designed and acquired by the construction sector in the essential objective is to make proposals that may benefit the insurance sector according to the information contained in the Dataset.

Keywords: keyword1, keyword2

1. Introduction

Strategic decisions do not occur at the contract level. They occur in the boardroom, where managers review available data and possibly launch new strategies. From a portfolio perspective, insurers want to plan their capacity, establish management policies, and balance the mix of products being distributed to increase revenue while controlling volatility.

Conceptually, one can think of an insurance company as nothing more than a collection, or portfolio, of insurance contracts. It has been learned about the modeling of insurance portfolios as the sum of individual contracts, taking into account hypotheses of independence between the contracts. Given their importance, this chapter focuses squarely on portfolio distributions.

- Insurance portfolios represent the obligations of insurers and, therefore, are particularly interested in the probabilities of large risks.
- Insurance portfolios represent the company's obligations and therefore insurers maintain an equivalent amount of assets to meet those obligations. Risk Measures summarize the distribution of the insurance portfolio and are used to quantify the amount of assets that an insurer needs to have to meet its obligations.

  With the available dataset we seek to answer the questions posed above in addition to being able to present a spatial analysis of the available information in search of possible patterns that allow us to make decisions related to the improvement of the portfolio.
- **Theoretical framework**

  The traditional approach to modeling the distribution of aggregate losses begins by separately fitting a frequency distribution to the number of losses and a severity distribution to the size of the losses. The estimated aggregate loss distribution combines the distribution for the frequency and the distribution for the severity of convolution losses.

  Discrete distributions, often referred to as count distributions or frequency distributions, to describe the number of events, such as the number of driver accidents or the number of insured claims. Lifetimes, asset values, losses and claims sizes are often modeled as continuous random variables and

*Corresponding author

Email address: dsarmientosar@unbosque.edu.co (Daniel Ricardo Sarmiento)

[1]Database Design and Analysis

as such are modeled using continuous distributions, often referred to as loss or severity distributions. A mixed distribution is a weighted combination of simpler distributions that is used to model an investigated phenomenon in a heterogeneous population, such as modeling more than one type of liability insurance claim (small but frequent claims and large but relatively rare claims). . This explores the use of continuous and mixed distributions to model the random size of losses. Key attributes are presented that characterize continuous models and that are also a means to create new distributions from existing ones. The effect of coverage modifications is also explored.

- *Moment generating function.*

  The moment generating function, denoted by $M_x(t)$ uniquely characterizes the distribution of x.While it is possible for two different distributions to have the same moments and still be different, this is not the case with the moment generating function. That is, if two random variables have the same moment generating function, then they have the same distribution. The moment generating function is given by

$$M_x(t) = E(e^{tx}) = \int_0^\infty e^{tx} f_x(x) dx$$

for all t for which the expected value exists. The moment generating function is a real function for which the

kth derivative at zero is equal to the kth ordinary moment of X. In symbols, this is:

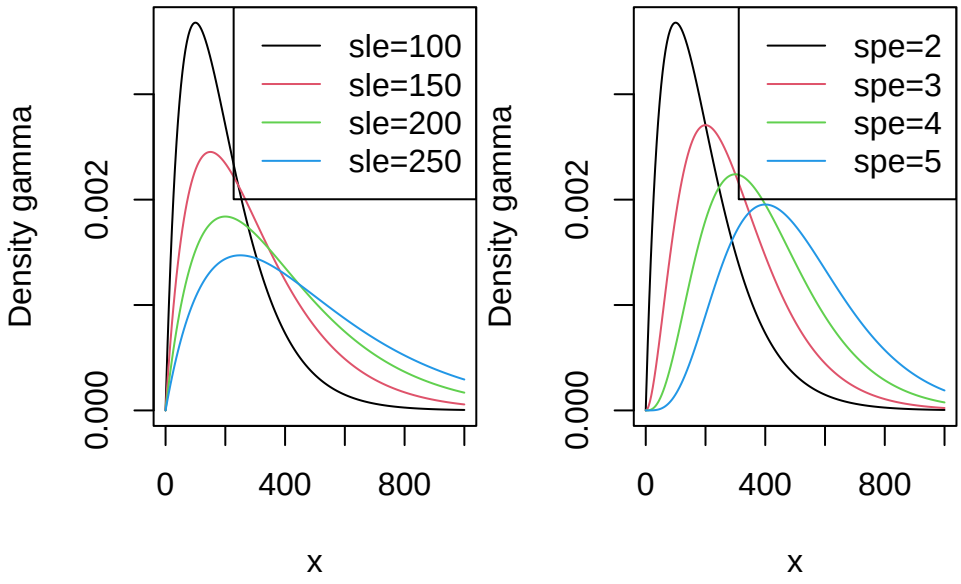
$$\frac{d^k}{dt^k} M_x(t)|_{t=0} = E(X^k)$$

- *Continuous Distributions to Model the Severity of Losses*

  The definition and application of four fundamental distributions for severity will be presented:

  1. **Gamma**
     The traditional approach to loss modeling is to fit separate models for frequency and severity. Let X be a continuous variable and have a gamma distribution with form parameter a and scale parameter O if its probability density function is given by

$$f_x(x) = \frac{(x/\theta)^\alpha}{x\gamma(\alpha)} e^{-x/\theta}$$

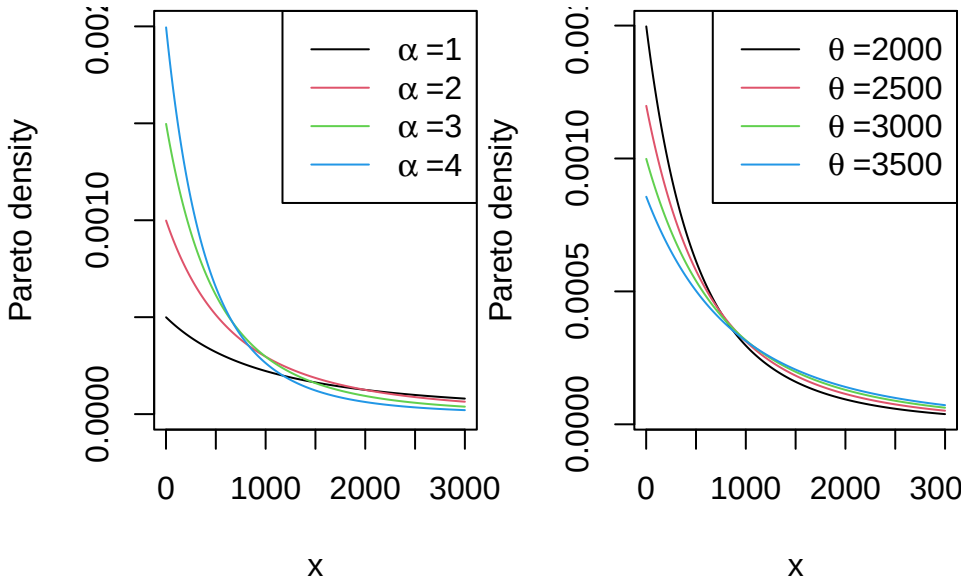


2. **Pareto**

The Pareto distribution, named after the Italian economist Vilfredo Pareto (1843-1923), has many economic and financial applications.

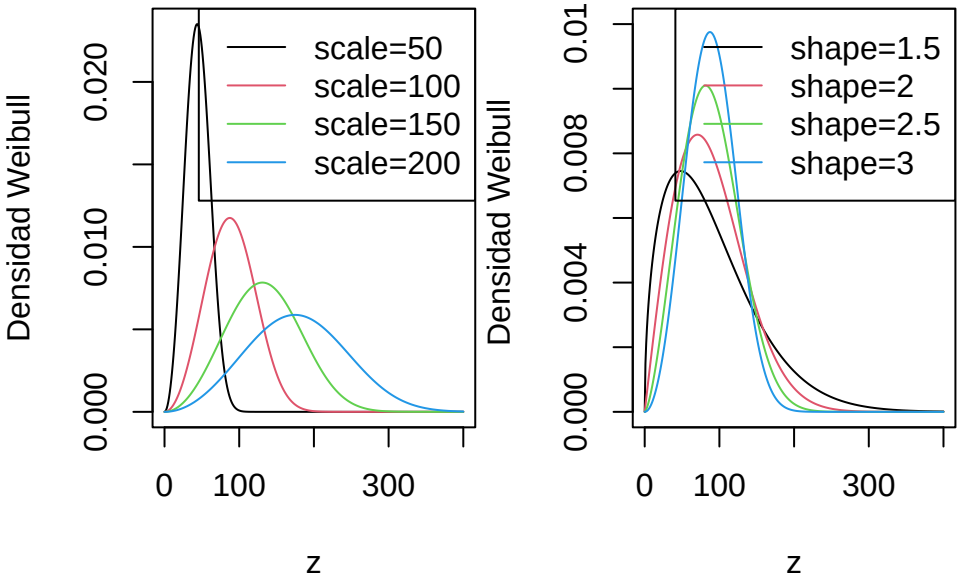
The continuous variable X is said to have a Pareto distribution with shape parameter α and scale parameter θ if its pdf is given by

$$f_x(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$$



3. **Weibull**

The Weibull distribution, named after the Swedish physicist Waloddi Weibull (1887-1979), is widely used in reliability, life time analysis, weather forecasting, and general insurance claims for the frequency of claims and the gamma distribution for the severity. An alternative approach to loss modeling that has recently gained popularity is to create a single model for pure premium (average cost of claims).



- **Analysis**

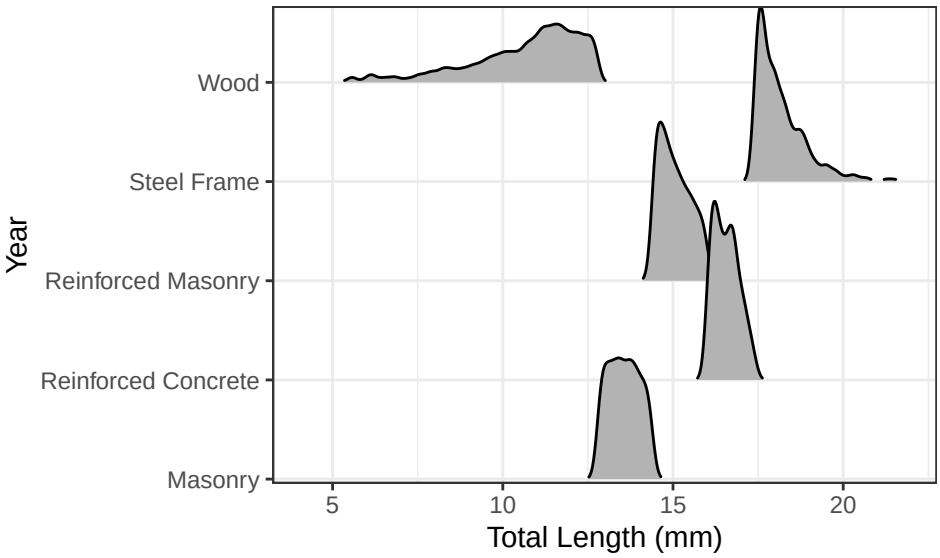
The data was taken for the years 2020 - 2021 in the state of Florida, these describe the characteristics of an insurance portfolio where the following variables are indexed:

Variable	Description
PolicyID:	Identification of the Insured
Line:	Type of Construction (Residential - Commercial)
Contruction:	Main Material of Construction ()
County:	
Ubicacion:	Type of coordinates (Longitude - Latitude)
TIV.2020:	Total Insurance Value 2020
TIV.2020:	Total Insurance Value 2021
Growth Rate:	

1. *Descriptive dataset analysis*

Construction	Mean	Sd	Sk	Total
Masonry	860868.28	390929.63	0.6548163	7969057670
Reinforced Concrete	16651177.83	6458192.69	0.9718453	21629880000
Reinforced Masonry	3824856.84	1863189.91	1.0211299	16160020160
Steel Frame	116984117.65	199863281.36	6.6785818	31819680000
Wood	93715.07	93531.86	1.0973706	2022464932

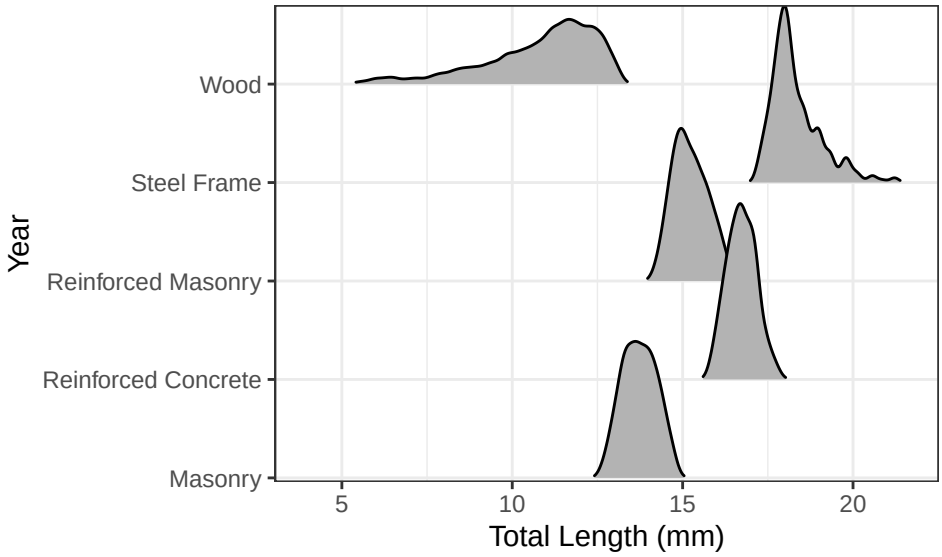
We can see the basic descriptions related to the central moments and the asymmetry of our portfolio depending on the type of main material that the construction is made of.



For the year 2020 we can show how the distributions according to the type of construction material are quite different from each other in all senses such as asymmetry and centrality, since it would not be

correct to assume that the distributions are indented, for the above by means of the generating function At times we would not obtain a real representation of the distribution of the portfolio.

Construction	Mean	Sd	Sk	Total
Masonry	1041986	533665.9	0.9714503	9645665598
Reinforced Concrete	20212429	9062551.9	1.2134574	26255944858
Reinforced Masonry	4621373	2510370.3	1.2343862	19525300845
Steel Frame	133492500	190403263.8	5.0966033	36309960000
Wood	113493	118893.5	1.3606895	2449292801



For the year 2021 we have the same situation and therefore the same problem, in order to solve this problem it is decided to carry out MonteCarlo processes to approximate the proportion of expected claims, we will model the sum of the random variables through convulsion and finally we will estimate the expected losses for the two years and the expected losses in what would be a bad year.

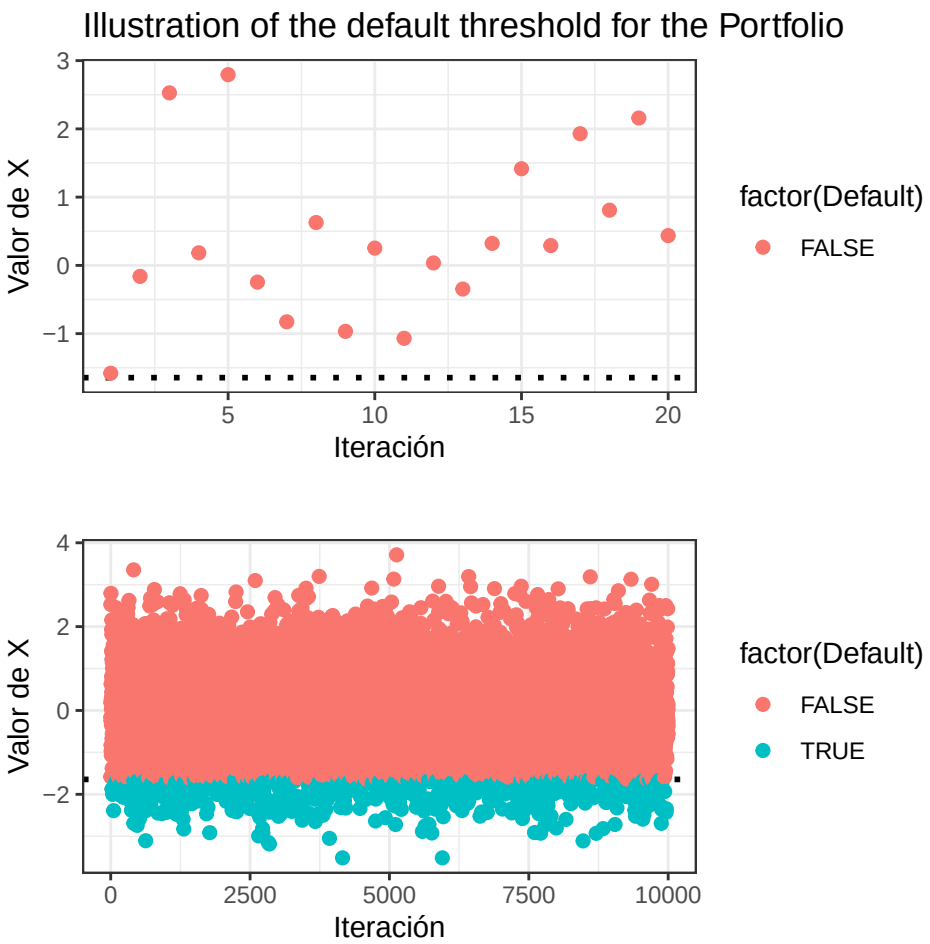
- **Vasicek single factor model**

  The critical concept is that, from a banker's perspective and at a fundamental level, a company exists to pay off its debt. When the company's assets fall below debt levels, it can no longer pay its debt and will therefore default. We know that, at a macro level, there is one thing that affects all companies, large and small: the state of the economy. To reflect that fact, we can let Z be a random number drawn from a standard normal distribution (with mean 0 and standard deviation 1) that represents the state of the economy. If we get a high number, then we are in a good economic state, if we get a low number, then we are in a bad economic state. To reflect the idiosyncrasies of an individual firm, we can let Zvar be a random number (also selected from a standard normal distribution with mean 0 and standard deviation 1) specific to firm i.

  When we run enough simulations, the loss scenarios we collect will eventually converge on the expected loss for the portfolio and we can then look at the full loss distribution and associated probabilities for the entire range of losses.

  To illustrate how this would work for a single insured, let's set the PD for our insurer to 5% and run the simulation using the latent factor Xi to determine how many times in M iterations we get a default. We will note that when M is small, the claim rate is probably not 5%, but as M gets higher, we will start to approach 5% (ie, in the long run, our claim rate will converge to the expected rate).

```r
## Monte Carlo
M <- 10000
## Defining variables
rho <- 0.09    ## correlation factor of portfolio...assuming at 0.09 for trial
X <- numeric(M)
threshold <- numeric(M)
iteration <- numeric(M)
set.seed(777)
Z <- rnorm(M, mean=0, sd=1)   ## generating common risk factor
Zvar <- rnorm(M, mean=0, sd=1)
  for (m in 1:M) {
    iteration[m] <- m
    X[m] <- sqrt(rho)*Z[m] + sqrt(1-rho)*Zvar[m]
    threshold[m] <- qnorm(0.05, mean=0, sd=1)  ## PD set at 5%
    }
sim <- as.data.frame(cbind(iteration,X,threshold))
library(dplyr)
sim <- mutate(sim, Default = (X < threshold))
```



```
drate2 <- sum(sim$Default)/(dim(sim)[1])
drate2
```

```
## [1] 0.0504
```

The default rate after running many iterations is now 0.0504. And all the default events are indicated in light blue. As we compute the simulation for all the borrowers in the portfolio we can expect the same result as we've seen here for one borrower.

- **Data Preparation**

  The first step is to read in our prepared portfolio that it has a PD, LGD and EAD for each policyholder and verify the final product. For illustrative purposes. The development of PDs and LGDs is not covered here, but PDs are taken from the historical risk grade transition matrix and historical loss LGDs and assigned to each insured in the portfolio according to their risk grade and/or industry.

```
N <- 1000
ID <- seq(from=1, to=N, by = 1)
PD <- rep_len(c(0.00001, 0.01, 0.08, 0.0002), length.out = N)
LGD <- rep(.5, N)
EAD <- rep(1000000, N)
Portfolio <- data.frame(ID, PD, LGD, EAD)
```

N	is the number of loans in the portfolio
rho	Is the portfolio correlatio
M	Is the number of iterations in the simulation
x	Will be the loss (in dollars) for each iteration
Rate	Will be the default rate per iteration

```
N <- dim(Portfolio)[1]  ## gives us the number of loans in the dataset
rho <- 0.09    ## sets the portfolio correlation to be used in the simulation
M <- 20000   ## number of iterations
x <-numeric(M)   ## initializes loss vector
rate <- numeric(M)  ## initializes rate vector
```

```
set.seed(777)
for (m in 1:M) {
    Loss <- 0
    DefaultCount <- 0
    DefaultRate <- 0

    Z <- rnorm(1, mean=0, sd=1)   ## generating common risk factor
    Zvar <- rnorm(N, mean=0, sd=1)  ## generating N idiosyncratic risk factors

    for (i in 1:N) {
      X <- sqrt(rho)*Z + sqrt(1-rho)*Zvar[i]   ## evaluating X for each loan i
      threshold <- qnorm(Portfolio$PD[i], mean=0, sd=1)
      if (X < threshold) {
        Loss <- Loss + Portfolio$LGD[i]*Portfolio$EAD[i]
```

```
        DefaultCount <- DefaultCount + 1      ## counting +1 for a defaulted loan
        }
      DefaultRate <- DefaultCount/N
      }
    x[m] <- Loss       ## capturing total portfolio loss per iteration
    rate[m] <- DefaultRate     ## capturing total default rate per iteration
    }
```

```
## [1] 11276250
```

```
## [1] 11222250
```

The two values are very close as we have run enough simulations for our mean to converge on the most likely outcome. Additionally, we now have a whole range of losses and their associated likelihood to consider.