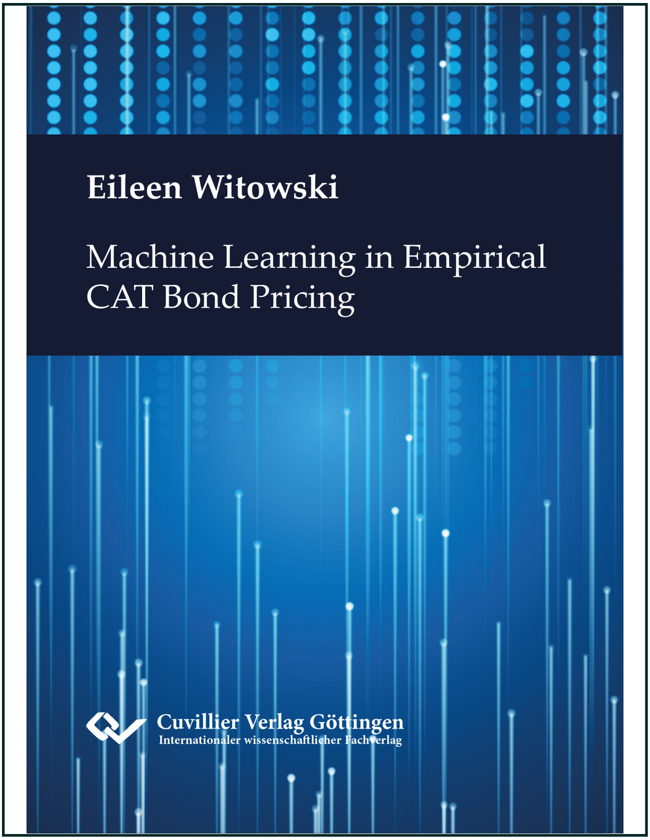
Eileen Witowski (Autor)

Machine Learning in Empirical CAT Bond Pricing



https://cuvillier.de/de/shop/publications/9045

1 Introduction

1.1 Motivation and Aims of the Thesis

In recent decades, the frequency and the amount of losses from natural catastrophes have increased continuously. Since 1970, the total losses caused by natural disasters have increased. The highest loss amounts were driven by Hurricane Katrina in 2005, the Tohoku earthquake in 2011, and the 2017 hurricane season with many strong hurricanes. In 2021, natural catastrophes caused an estimated USD 111 billion in insured losses worldwide, the fourth-highest amount since 1970 (Bevere, 2022). According to the annual Allianz Risk Barometer, which is published by the Allianz Global Corporate & Speciality (2023), natural catastrophes are among the top 10 risks for 2023 globally, making them one of the most feared risks since the first report in 2012.[1] Reinforcing this, climate change, which is often cited in the literature as a trigger for natural catastrophes, is listed among the top 10 risks as well (Allianz Global Corporate & Speciality, 2023). Additionally, the Global Risk Report 2023 ranks the risk of natural disasters and extreme weather events as one of the three most relevant global sources of risk in terms of severity (World Economic Forum, 2023). Because natural catastrophes are very difficult to predict, and their consequences are devastating, their prevention is challenging from an economic perspective.

(Re)Insurance companies play an essential role in mitigating the consequences of natural catastrophes as they provide insurance against natural catastrophe risks. However, offering

[1] The Allianz Risk Barometer Reports can be derived from `https://www.agcs.allianz.com/news-and-insights/reports/allianz-risk-barometer.html`.

coverage against losses from natural disasters faces insurers with a challenge. In the event of natural disasters, a large number of major losses occur (usually) on a regional basis, which makes it difficult for insurance companies to cover these risks. To overcome this challenge, insurers transfer (parts of) their risk to reinsurance companies. Reinsurance companies generally have a greater regional diversification than insurance companies, making it easier to cover natural catastrophe losses. However, the reinsurance market is limited. The increasing occurrences and losses due to catastrophe events result in increased risk for (re)insurance companies and rising prices of traditional natural catastrophe insurance and can sometimes lead to capacity bottlenecks in the (re)insurance market. (Re)Insurance companies are thus faced with the challenge of dealing with this development. One option is transferring the increasing risks from the (re)insurance market to the capital market and securitizing them through catastrophe (CAT) bonds.

Against the background of the limited capacity of the reinsurance market, CAT bonds were first issued in the 1990s. From 1997 onwards, the CAT bond market became established, leading to increased scientific interest. As is the case for other asset classes, the pricing of CAT bonds is one of the main research topics. Two strands of research have developed in the literature. The first deals with the determinants of CAT bond premiums[2], and the second with their forecast. The studies can be further subdivided depending on the market segment under consideration. The initial issue of CAT bonds takes place on the primary market, while they continue to be traded on the secondary market. Both markets are examined in the literature. This has already developed a good understanding of the most important price determinants of CAT bonds as well in the

[2] The risk premium of a CAT bond is the portion of the coupon that is paid over and above the risk-free interest rate.

primary market (Lane, 2000; Wang, 2000, 2004; Galeotti et al., 2013; Braun, 2016; Trottier et al., 2018) as in the secondary market (Gürtler et al., 2016; Herrmann and Hibbeln, 2023, 2021; Götze and Gürtler, 2020). A literature review shows that although a large number of exogenous variables are considered, past risk premiums have not yet been considered as price determinants for CAT bonds. This seems surprising, as studies for other asset classes have demonstrated the relevance of past prices for the current price level. For example, some stock market studies (e.g., Lo and MacKinlay (1988); Frennberg and Hansson (1993); Khoojine and Han (2020)) show that stock prices have an autoregressive structure.

In addition to models explaining the factors influencing CAT bond premiums, forecasting models have already been used on the primary market (Galeotti et al., 2013; Braun, 2016; Trottier et al., 2018) to forecast the risk premium as accurately as possible, thus providing practitioners with a basis for making decisions when buying and selling CAT bonds. Out-of-sample forecasting models on the secondary market have not yet been examined. Regardless of their objective, most studies rely on traditional linear regression models. However, the literature has shown that advanced machine learning approaches have produced promising results in the context of asset price forecasting (e.g., Gu et al. (2020); Bianchi et al. (2021a)). With the exception of Makariou et al. (2021), such methods have not yet been tested in the CAT bond pricing literature. Makariou et al. (2021) implemented a linear regression and a random forest model to predict CAT bond primary market premiums. However, they did not include macroeconomic variables in their models, although they have shown relevance in previous CAT bond pricing studies (e.g., Braun and Kousky (2021); Gürtler et al. (2016)). Further, they did not consider the time structure of their data set, using random subsampling to split the data into training and test data. Especially the

CAT bond primary market is relatively small compared to some other asset classes such as stocks. This poses the risk of overfitting, which may result in poor out-of-sample forecasts. Against this background, it is interesting whether advanced machine learning methods, which are mainly well-known for their use with large data sets, can also be applied to the CAT bond market and achieve good forecasting results. Compared to primary market premiums, CAT bond secondary market premiums have significantly lower conditional variance (given the available price information). Additionally, the panel structure of the CAT bond secondary market data set makes more information available than in the primary market. These two facts indicate that the CAT bond secondary market is a low-uncertainty environment. It still needs to be determined whether advanced machine learning methods provide significantly more accurate forecasts than linear regression models in a market with low uncertainty.

Compared to traditional linear regression, advanced machine learning methods have a disadvantage: their black-box nature. A lack of transparency in price forecasting methods can limit their applicability, especially for institutional investors, as they are often required by regulation to use interpretable and explainable models (see European Insurance and Occupational Pensions Authority (2021); Basel Committee on Banking Supervision (2013)).

In summary, the present thesis addresses the following research questions:

- Can advanced machine learning approaches be applied in a comparatively small market environment and improve the prediction accuracy for CAT bond premiums in the primary market?
- Which model and which set of explanatory variables provide accurate forecasts for

CAT bond secondary market premiums?

- How can regulatory requirements for transparent models be met when using advanced machine learning methods and how can the model results and determinants of CAT bond secondary market premiums be interpreted?

- Do CAT bond secondary market premiums follow an autoregressive process, and if so, is an autoregressive model sufficient to predict CAT bond premiums?

The above research questions add to the literature, improve the understanding of the CAT bond market, and represent important recommended actions for practitioners. This thesis answers the above research questions based on empirical analyses in terms of three separate projects. The first study (Chapter 3) compares different advanced machine learning approaches to a traditional linear regression model regarding forecasting accuracy in the primary CAT bond market. The second study (Chapter 4) builds the first out-of-sample forecasting model for CAT bond secondary market premiums, comparing different methods, and shows how to identify important pricing variables in the context of random forest forecasting, which provides the most precise prediction results. The third empirical analysis (Chapter 5) shows that CAT bonds follow an autoregressive pattern since CAT bond secondary market premiums have particularly low conditional volatilities for given historical premiums. The autoregressive model can be used for precise out-of-sample prediction that does not statistically leak behind a random forest approach. If, however, the aim is to understand the CAT bond determinants, one should use a broader random forest model in combination with variable importance measures.

1.2 Course of Investigation

To answer the research questions stated in Section 1.1, the thesis is structured as follows.

In Chapter 2, the basics needed for the thesis are explained. First, the functionality of CAT bonds is discussed in Subsection 2.1.1. This is followed by a description of their historical development in Subsection 2.1.2. Section 2.2 then explains the models used in the following three studies.

Chapter 3 answers the first research question of whether advanced machine learning methods improve the prediction accuracy for CAT bond primary market premiums. Therefore, Section 3.1 motivates the analysis in more detail. Section 3.2 presents the data used in the empirical analysis conducted in Section 3.3. The results of this first empirical study are then summarized in Section 3.4.

To answer the second research question, Chapter 4 compares different out-of-sample forecasting models with regard to their accuracy in predicting CAT bond secondary market premiums. In order to meet the regulatory requirements for transparency, methods for interpreting the model results and CAT bond determinants are presented, thereby answering the third research question. Section 4.1 gives an overview of the relevant literature and motivates the study. Subsequently, the data is reviewed in Section 4.2. Section 4.3 presents the empirical analysis and its results, which Section 4.4 outlines in the following.

Based on the findings from Chapter 4 that price changes (in the form of momentum variables) are important factors influencing CAT bond premiums and that the secondary market has very low variance, Chapter 5 examines the fourth research question of whether CAT bond premiums follow an autoregressive structure that can be exploited for accurate forecasting. The study is justified in Section 5.1. Section 5.2 outlines the procedure of the study. The empirical in-sample and out-of-sample analysis results are then presented and evaluated in 5.3. Finally, Section 5.4 summarizes the study.

Chapter 6 concludes this thesis.

2 Scientific Background[1]

The intention of this chapter is to develop an understanding of CAT bonds and their pricing. Therefore, Section 2.1 explains the basics of CAT bonds. Section 2.2 then introduces various machine learning methods[2] that will be used for price forecasting in the further course of the thesis.

2.1 CAT Bonds

2.1.1 Definition and Functionality

Severe natural catastrophes have led to growing losses in recent years. Figure 2.1 shows the economic losses caused by natural disasters since 1980. While in the 1980s, the inflation-adjusted (2020 prices) economic costs of natural catastrophes (e.g., earthquakes, hurricanes, windstorms) were on average USD 35 billion per year, in the 1990s, the annual average was about USD 120 billion. This rose to USD 154 billion in the 2000s and further to USD 201 billion in the 2010s.

[1] The following sections are based on the studies "Improving CAT Bond Pricing Models via Machine Learning," published in the Journal of Asset Management in 2020, and "Forecasting Accuracy of Machine Learning and Linear Regression – Evidence from the Secondary CAT Bond Market," published in the Journal of Business Economics in 2023. Both articles were funded by the Deutscher Verein für Versicherungswissenschaft e.V., for which we are thankful.

[2] While linear regression methods and the types of methods introduced throughout this chapter belong to the same model family, we will subsequently use the term "advanced machine learning method" to describe methods that extend or generalize linear regression methods.

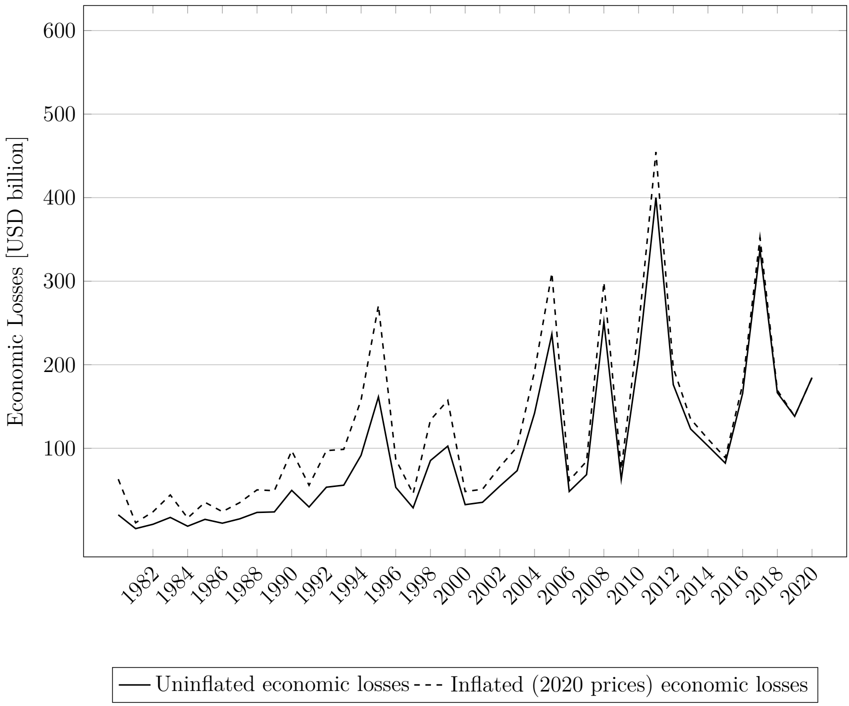


Figure 2.1: **Economic losses from natural catastrophes.**
This figure is based on Bevere (2021).

As a result, insured losses worldwide have also grown continuously, leading to increased capacity shortages in the (re)insurance market and policyholders paying rising premiums (Gürtler et al., 2016). To counter this capacity problem, CAT bonds have been issued since the 1990s (Gürtler and Rehan, 2009). CAT bonds are bonds that (re)insure against natural catastrophe risks by transferring the catastrophe risk from the insurance market to the capital market. Thus, they belong to the class of insurance-linked securities. CAT bonds are private placement transactions that are not publicly traded. They are structured by two parties: the sponsor, which is typically a (re)insurer, and the investor, such as a bank.