Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis Quantitative Finance

The impact of extreme wind on US mortgage defaults in the face of climate change

Author:
T.J. Veringa Weiler

Company:
Triple A Risk Finance

Supervisor:
Prof. dr. C. Zhou

Company supervisor:
Ms. Sanne Visser

Second assessor:
Dr. P. Wan

October 25, 2023

Abstract

This research delves into the relationship between extreme wind events and residential mortgage defaults in the southeast United States. Employing a theoretical framework introduced by Einmahl et al. (2022), we perform statistical tests across a range of weather stations in key cities. Results indicate varying total frequencies of extreme wind across cities, with two out of 22 cities exhibiting heteroscedastic frequency of extreme wind. Logistic regression analysis, incorporating four essential predictors and an extreme wind step function, reveals a statistically significant 14% increase in the probability of mortgage default following an extreme wind event. Despite its statistical significance, the added explanatory power of extreme wind within the regression model is limited. However, an XGBoost algorithm demonstrates improved predictive performance when extreme wind is included, highlighting the potential benefits of incorporating this factor into mortgage default prediction models.

Contents

List of abbreviations

Abbreviation	Description
ASOS	Automated surface observing system
AUC	Area under curve
EVT	Extreme value theory
FICO	Fair, Isaac, and Company (score)
GEV	Generalised extreme value (distribution)
IID	Independent and identically distributed
LR	Likelihood ratio (test)
LTV	Loan-to-value (rate)
MDR	Mortgage default rate
PD	Probability of default
pdf	Probability distribution function
POT	Peaks-over-threshold
ROC	Receiver operating characteristic
TC	Tropical cyclone

1 Introduction

Due to the human-caused globally changing climate, weather and climate extremes appear increasingly frequently and are becoming more severe (IPCC, 2023). The United States (US) are not immune to the impact of climate change on such weather and climate extremes, including those related to wind like tropical cyclones (TCs) and tornadoes. For example, Chand et al. (2022) find that while tropical cyclone frequency is decreasing in the rest of the world, it is increasing in the North Atlantic region. Popular examples of TCs such as hurricane Katrina or Harvey exemplify the destructive capabilities of TCs. The repercussions of TCs often include extreme flooding, destroying roads, houses and sometimes entire cities. However, despite experiencing persistent record-breaking losses from TC induced flooding, the US remains ill-prepared for TC flood events, with the deficiency in residential flood insurance being a prime example of this (Tonn & Czajkowski, 2022).

At first glance these extreme wind events should not be a large problem, seeing as in the US homeowner insurance generally covers all forms of wind damage. However, this insurance does not protect citizens from secondary effects of extreme wind events. These secondary effects, such as floods and storm surges, can thus cause home owners huge financial losses, potentially leading to mortgage defaults. This also holds for other forms of extreme wind such as tornadoes, leading to many citizens losing their homes.

This thesis investigates the relationship between extreme wind and mortgage defaults in the US. It focuses on the southeast region of the US which has continuously experienced significant impacts from extreme wind events. Utilising an extreme value theory (EVT) approach, we use the theoretical framework introduced by Einmahl et al. (2022) to statistically test for historical trends in extreme wind speed in a spatial setting. Having established (potential) historical trends in extreme wind speed, we investigate the effect extreme wind has on the probability of default (PD) in the same region.

Our research contributes to the existing literature in several ways. Firstly, considering extreme wind in the US, we are the first paper to investigate trends in extreme wind in general instead of focusing on specific events of extreme wind. Secondly, we apply a novel methodology to evaluating data on extreme wind as well as incorporating it as a causal factor

for mortgage defaults. We build upon this further by studying the effect of extreme wind on mortgage defaults in the southeast region of the US, a region for which studies on extreme wind have been limited to specific hurricanes.

Our analysis reveals several key insights. Notably, the frequency of extreme wind events varies among the 22 cities in the southeast US, with two of them exhibiting heteroscedasticity of extreme wind frequency between 2000 and 2020. Employing a logistic regression, extreme wind emerges as a statistically significant predictor, suggesting that a home that has experienced an event of extreme wind has an average 14% higher PD. This indicates its capacity to explain variations in mortgage defaults alongside several benchmark predictors. However, the practical value of incorporating extreme wind as a predictor of mortgage defaults within the benchmark model is limited. This finding is consistent across most individual cities. In contrast, we find that including extreme wind as a predictor in an XGBoost model enhances predictive accuracy, implying intricate interactions between the predictors and their influence on mortgage defaults. This highlights the importance of considering the secondary consequences of extreme wind events, even when homeowners are protected against direct damages.

The rest of the paper is structured as follows. Earlier performed research that is relevant for this paper is discussed in Section 2. We examine the data used in this research in Section 3. In Section 4, an outline is given of the methodology used in this research. We provide and discuss the results in Section 5. We end with concluding remarks in Section 6.

2 Literature review

This section outlines pre-existing literature on trends in extreme values, trends of wind speed in the US and the effect of extreme wind speed on mortgage defaults in the US.

2.1 Trends in extreme values

There have been several research papers studying trends in extreme values in the past, often in the context of weather. Recently, Einmahl, de Haan, and Zhou (2016) proposed a flexible theoretical framework to study so-called heteroscedastic extremes, or time-varying extremes,

through their scedasis function. The advantages of their framework are first of all that they do not deal with any limit situation but instead do an extreme value analysis based on the domain of attraction. They also do not impose any (local) parametric model on the scedasis function and provide asymptotic properties of the estimators.

Ferreira et al. (2017) expand upon the heteroscedastic extremes framework, extending it to the domain of space and time by introducing a space-time scedasis function. This scedasis function accommodates temporal trends in a spatial setting. To address the time-varying behaviour of extremes, Ferreira et al. (2017) propose a homogenisation procedure that yields stationary pseudo observations. These pseudo observations capture the evolving nature of extremes and facilitate their incorporation into the estimation of (joint) exceedance probabilities using the non-parametric scedasis function. This comprehensive approach enables a thorough examination of the time and space dynamics of extreme events, enhancing the accuracy and reliability of their probability estimation.

Einmahl et al. (2022) further extend this work this by introducing a test statistic for constant frequency of extremes in a spatial setting. They also introduce a test statistic for constant total frequency of extremes across space. They apply their model to rainfall in Northern Germany. This paper is of particular interest to this research, as its two statistical tests aid in gaining insight into the behaviour of extremes in a set of observations. We therefore closely follow Einmahl et al. (2022), performing both proposed tests to gain a better understanding of the behaviour of our sample of wind data.

2.2 Extreme wind speed events

There have been numerous studies investigating trends in specific types of extreme wind events in both the entire world and specific to the US. For example, Chand et al. (2022) investigate trends of TC frequency for all relevant parts of the world. They find that, with the exception of the North Atlantic region, TC frequency has been declining between 1900 and 2012. This research confirms the study by Lima et al. (2021), in which they find a significant linear upward trend of TC and post-TC frequency in the North Atlantic region over the period 1978-2019. As hurricanes can cause severe floods, Sajjad et al. (2020) specifically study spatial heterogeneities in hurricane flood risk along the US Atlantic and Gulf coasts.

They find that the areas that run the highest risk of hurricane floods are mostly found in the Gulf region, particularly along the west coast of Florida.

On the other hand, Tippett et al. (2016) study temporal trends in the frequency of so-called tornado outbreaks, sequences of tornadoes that occur in close succession. Using extreme value distributions with time-linear parameters, they find that the frequency of U.S. outbreaks with many tornadoes is increasing and that it is increasing faster for more extreme outbreaks. Gensini and Brooks (2018) study temporal trends in tornado frequency in the US and investigate spatial variability across the country. They find that, while the temporal trend is negative for the middle and southern Great Plains, there is a robust positive temporal trend in tornado frequency for the midwestern and southeastern US.

These studies show that the two main types of extreme wind that the US falls prey to, TCs and tornadoes, appear most frequently and increasingly so in the southeastern region of the US. We therefore focus our research on the southeast US.

Despite extensive research on trends in specific extreme wind events (especially in the US), studies concerning extreme wind in general are scarce. Young et al. (2012) investigate global trends in extreme marine wind speeds by determining the pdf that represents the data and seeking changes in the pdf's parameter values over time. They find proof of a positive trend in 100 year return period values of wind speed. However, these trends have very high uncertainty, which leaves room for debate on the magnitude of the trend. Breivik et al. (2014) test different models in their estimation of historical wind extreme trends. Among other things they find that the trend in annual maximum surface wind speed is not statistically significant for many regions of the world, including the northern hemisphere in which the US is situated. However, no definitive conclusion can yet be made for the case of the US alone. Our research thus expands upon existing literature, as we investigate trends of extreme wind specifically in the southeastern US.

2.3 Effect of extreme wind on mortgage defaults in the US

Studying the impact of extreme wind events on mortgage defaults in the US is compelling given the fact that most homeowners insurances cover wind damage (including extreme

events such as tornadoes and TCs) but do not cover their indirect consequences. Tornadoes and TCs are often accompanied by extreme precipitation and landfalling TCs can cause large storm surges, both of which increase the probability of flooding. Hurricane Harvey was a prime example of the enormous amounts of damages that TCs can cause and the detrimental impact it can have on mortgage credit risk (Kousky et al., 2020).

Research on the effect of extreme wind speed on mortgage defaults in the US has thus far been limited to studies concerning TCs. Kousky et al. (2020) find that, in areas where flood insurance was not required, the likelihood of a loan being 180 or more days delinquent or in default during the two years following Hurricane Harvey rose with increasing property damage. Rossi (2020) evaluate the influence of hurricane frequency and intensity on the risk of mortgage defaults. Their findings reveal that mortgage loans associated with the occurrence of category 3, 4, or 5 hurricanes during the loan period exhibit a 13% to 18% higher likelihood of becoming delinquent for 180 days or 90 days or more, compared to loans unaffected by such severe hurricanes.

These studies solidify the hypothesis that, even though damage to property by extreme wind is insured in the US, hurricanes still have a high impact on mortgage defaults. It does however remain unclear whether such conclusions can be drawn about *all* cases of extreme wind. Our research attempts to fill that gap.

3 Data

For this research, two different types of data are collected. To study extreme wind speeds, we need wind speed data for different locations. To study the impact of extreme wind speeds on the probability of mortgage default, we need a comprehensive set of mortgage data. In the US there is public access to both. As mentioned in Section 2.2, we focus on the southeast US. Specifically, we study the states depicted in Figure 1.

To investigate the effect of extreme wind on the PD over time as well as space, we collect wind and mortgage data from a grid of 22 of the most highly populated cities in the southeast US. These are shown in Figure 1 by red dots. A full list of the cities in question can be found in Table 1. Both wind and mortgage data is collected for the year 2000 through 2020, inclusive.

Figure 1: Map of the US showing the states that this research is focused on. The red dots show the cities for which wind and mortgage data is collected.

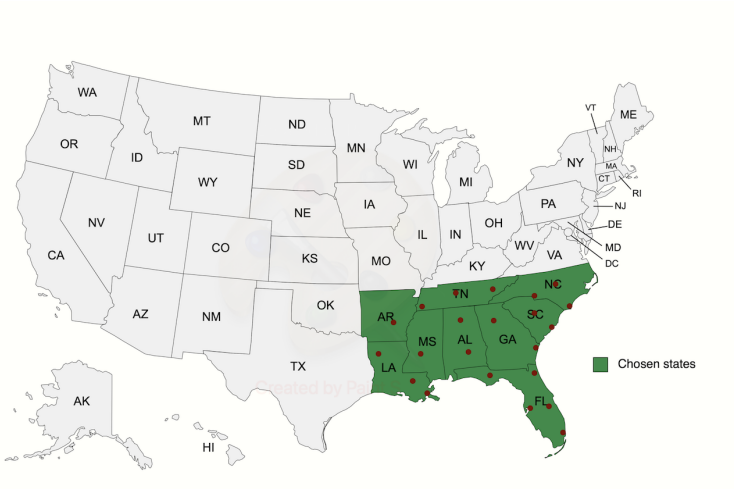


Table 1: List of cities that are part of this research, per state.

State	**List of cities**
Alabama	Huntsville, Montgomery
Arkansas	Little Rock
Florida	Jacksonville, Miami, Orlando, Tallahassee, Tampa
Georgia	Atlanta, Savannah
Louisiana	Baton Rouge, New Orleans, Shreveport
Mississippi	Jackson
North Carolina	Charlotte, Raleigh, Wilmington
South Carolina	Charleston, Columbia
Tennessee	Knoxville, Memphis, Nashville

3.1 Wind speed data

3.1.1 Data preprocessing

The wind speed data used in this research is provided by the automated surface observing systems (ASOS) program. It is a joint project between the US National Weather Service, the US Federal Aviation Administration and the US Department of Defence (NCEI, 2022). The ASOS have been the US's primary surface weather observing network since the completion of its grid 2004 and has over 900 weather stations across the US. Each station measures a wide range of weather variables, updated every minute. An hourly report of this weather data is made available through so-called meteorological aerodrome reports, which are used predominantly by meteorologists and aircraft pilots.

In the context of this research, we pay particular attention to weather data pertaining to wind speed. Two wind speed related variables are measured by the ASOS: the regular wind speed and the wind gust speed. The wind speed is reported as the two-minute average measured wind speed. The wind gust speed is reported as the maximum 5-second sustained wind speed and is only reported if this maximum exceeds the regular wind speed by a minimum of 10 knots. For both these variables, the hourly report shows the value at that particular point in time. In the hourly report a third variable is also reported: the peak wind speed. It is defined as the highest wind speed observed or recorded since the last scheduled hourly observation.

The official wind speed scale is defined as the maximum one-minute sustained wind speed. As such, the peak wind speed would be the only variable needed. However, the peak wind speed is only reported if the maximum wind speed since the last hourly report exceeds 25 knots. This means that the value is often not reported. Therefore, the wind speed $v_{wind,t}$ used in this research is calculated by

$$v_{wind,t} = \max_{v}(v_t, v_{peak,t}), \tag{1}$$

where v_t is the regular wind speed and $v_{peak,t}$ is the peak wind speed at time t. We collect data and calculate $v_{wind,t}$ for each city in our grid, resulting in a time series of wind speed data for each of the 22 cities in our chosen grid.

The ASOS can measure wind speeds of up to 125 knots. However, in the hourly report, we find some values exceeding this threshold of 125 knots. These values are mostly preceded and succeeded by values between 0 and 10 knots. We suspect these extreme values to be due to an ASOS failing, so we set all values above 125 knots to be the average of the preceding and succeeding value. In doing so, we account for the case in which the extreme value is actually *not* due to a failing ASOS, since actual extreme wind speeds are usually preceded and succeeded by other relatively high wind speeds.

Lastly, we group our wind data into daily maxima. This cuts the size of the dataset by approximately an order of 24, which greatly increases manageability of the wind data. We expect that converting the hourly data to daily data should not decrease the accuracy of the dataset too much, seeing as two different extreme wind events (such as tornadoes or TCs) rarely occur within the same day. This means that with daily maxima we can still study the behaviour of extreme wind reliably.

3.1.2 Data description

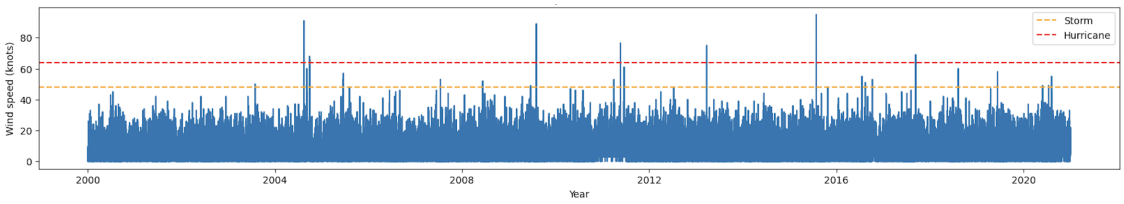
Descriptive statistics of daily maxima of wind for all cities can be found in Table 2. Firstly, we find that the wind speed is positively skewed for all cities. Since wind speed is bounded below by zero but can reach high values, we expect a positive skew. Secondly, there is excess kurtosis for all cities. This suggests that wind speed data has a heavy right tail.

An example of a time series of wind speed for Orlando, Florida can be found in Figure 2. Two things stand out from the plotted wind speed. Firstly, we can distinguish the extreme values, demonstrating the extreme wind conditions Orlando has faced. An example is the three closely succeeding peaks in 2004, when three hurricanes passed Orlando within a few weeks. Secondly, we notice the sawtooth pattern in the time series. This is not unique for Orlando. Most cities' wind speed time series show seasonality. As wind speed has previously been shown to exhibit seasonal patterns (e.g. Chen, Li, and Pryor (2013); Torres Silva dos Santos and Santos e Silva (2013)), we group data from all cities and make boxplots of wind speeds for each month. The result of this is shown in Figure 3.

Table 2: Descriptive Statistics of wind speed for all cities.

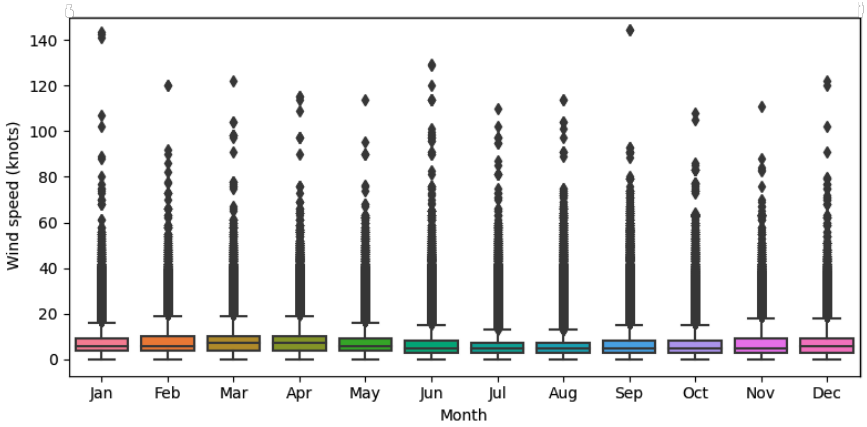
State	City	Mean	Median	Std.	75%	Max	Skewness	Kurtosis
Alabama	Huntsville	15.57	14	8.50	20	122	1.66	8.77
	Montgomery	14.31	11	8.16	19	114	1.50	5.34
Arkansas	Little Rock	16.19	14	8.36	21	122	1.61	7.98
Florida	Jacksonville	17.27	15	9.09	22	144	2.94	20.39
	Miami	18.41	18	7.43	22	95	1.54	6.82
	Orlando	18.01	17	7.92	22	95	1.29	4.80
	Tallahassee	15.46	14	8.37	20	145	2.42	18.23
	Tampa	15.54	15	7.29	19	130	2.16	17.77
Georgia	Atlanta	17.37	16	8.29	22	104	1.02	2.50
	Savannah	15.00	13	7.25	19	62	1.13	1.59
Louisiana	Baton Rouge	15.61	14	8.44	20	120	1.76	9.66
	New Orleans	19.06	17	9.31	26	77	1.20	2.24
	Shreveport	14.75	12	8.31	19	115	1.63	6.31
Mississippi	Jackson	15.16	14	8.37	20	79	1.30	3.11
North Carolina	Charlotte	14.68	14	7.67	19	99	1.32	4.72
	Raleigh	15.21	15	7.74	20	75	0.97	1.40
	Wilmington	17.56	17	8.36	22	95	1.10	3.01
South Carolina	Charleston	17.01	16	7.69	21	75	1.15	2.07
	Columbia	15.20	13	8.30	20	72	1.22	2.02
Tennessee	Knoxville	15.14	12	9.18	20	104	1.29	2.27
	Memphis	18.02	17	8.62	23	114	1.17	4.54
	Nashville	16.29	16	8.69	21	111	1.45	6.52

Figure 2: Historical wind speed for Orlando, Florida.



Note: the yellow dashed line shows a wind speed of 48 knots, signifying wind storm speed. The red dashed line shows a wind speed of 60 knots, signifying hurricane wind speed.

Figure 3: Monthly boxplots for all cities' data grouped together.



The boxplot indicates that, after excluding the outliers, wind speed exhibits seasonality (monthly boxplots excluding outliers can be found in Figure 9 in Appendix A). The summer and autumn season have lower wind speed averages than the winter and spring season. On the other hand, we do not observe such seasonality considering the outliers in the box-plots. This can be attributed to the hurricane season that is around June to November and the tornado season, which is both in early spring as well as autumn. Since we plan to study extreme wind, for which there is no clear sign of seasonality, we do not split the data to accommodate seasonality.

3.1.3 Removing serial dependence of wind speed data

Wind speed is generally found to be serially correlated (Liu & Hu, 2019). This is supported by our data, in which we find that, for all cities, daily maxima are serially correlated up to

one day. To be able to assume independent observations, we remove the serial dependence of the wind speed data. We follow an approach suggested by Einmahl et al. (2022).

In this approach, we start by grouping the observations from all 22 weather stations. For each day, we extract the daily maximum wind speed across all weather stations, the 'station-wise maximum'. We then arrange the station-wise maxima in descending order. We remove the serial dependence by first identifying the pair of calendar days with the highest and second highest station-wise maxima. If the second maximum's recording date is within two days of the first maximum, we remove this observation from the entire dataset. That is, across all 22 weather stations, the wind speed corresponding to this date is removed. Otherwise, both days are retained. This procedure is then applied to the remaining ordered station-wise maxima, removing any day that falls within two days of any previously retained day. After this, we are left with wind speed data that we assume to be independent.

3.2 Mortgage data

The mortgage data comes from the database offered by the Federal Home Loan Mortgage Corporation, or Freddie Mac. Freddie Mac is, besides Fannie Mae, the largest government-sponsored enterprise operating in the mortgage market. It purchases mortgages from lenders, packages them into mortgage-backed securities and sells them to investors to provide liquidity to the housing market. They also provide public access to much of their mortgage data.

We collect a data subsample of over one million unique mortgages, with mortgage originations distributed uniformly over the years 2000 until 2020. The mortgage data offers data concerning both the origination of the mortgage as well as its monthly performance. We remove mortgages for which the postal code does not correspond to one of the cities that are considered in our research or for which there are missing values for the different relevant features. This leaves us with a sample of 51,169 unique mortgages.

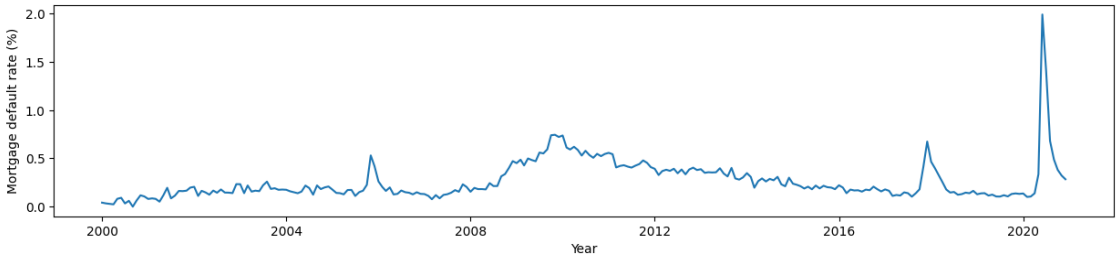
3.2.1 Mortgage default rate

In the US, there is no general consensus or rule on how many days of delinquency in payments define a mortgage default. We therefore follow the definition of the European Banking Authority and consider a mortgage to be in default if the borrower is delinquent in payment

for at least three months. This definition is also widely used in research on mortgage defaults in the US (e.g. Noordewier, Harrison, and Ramagopal (2001); Ambrose and Capone (2000)).

To gain insight into the mortgage defaults over time, we plot the time series of the mortgage default rate (MDR) of our sample of mortgages, shown in Figure 4.

Figure 4: Monthly MDR for the entire sample of mortgages



Four peaks stand out. The large peak between 2004 and 2008 is around 90-180 days after hurricane Katrina, which caused many homes to be destroyed. The gradual peak between 2008 and 2012 is a consequence of the 2008 financial crisis, which had a very large negative impact on the housing market. The peak between 2016 and 2020 follows the 2017 Atlantic hurricane season, one of the most devastating hurricane seasons in US history, including many well known TCs such as hurricane Irma, Maria and Harvey. The last peak after 2020 is a consequence of the COVID-19 pandemic, where many people had difficulty paying their mortgages to keep their property due to difficult financial situations (Benfer et al., 2020).

To summarise, Figure 4 shows two economy-related peaks in the MDR and two peaks related to extreme wind conditions and the consequences thereof. It is thus a first peek into the potentially significant effect of extreme wind on mortgage defaults.

4 Methodology

In this section, we discuss the methodology used in this research. The methodology consists of two parts. Firstly, we study the behaviour of wind extremes in our sample. We introduce the concept of (testing for) heteroscedastic extremes. Then, we extend towards heteroscedastic extremes in a spatial setting and a statistical test for constant total frequency of extremes over space, as well as a test for homoscedastic extremes in a spatial setting. Both of these tests are performed on our wind data.

Secondly, we lay out the regression setting that we will use to predict mortgage defaults. We establish an appropriate regression model and consider key benchmark predictor variables. Then, we address the implementation of the frequency of wind extremes into our benchmark predictor space. Lastly, we discuss methods to analyse our model and investigate whether wind speed extremes have a significant effect on mortgage defaults.

4.1 Estimation and test of extreme wind frequency

4.1.1 Introduction to heteroscedastic extremes

For the purpose of studying extreme observations, we use extreme value theory (EVT). Introduced by Fisher and Tippett (1928), EVT is often used for observations for which the distribution function is heavy-tailed, as their tails are not modelled well by commonly used distribution functions. For this, consider a sample of independent and identically distributed (IID) observations $(X_1, \ldots, X_n)$ with common distribution function $F(x) = \mathbb{P}(X \leq x)$. A distribution function $F(x)$ is said to be heavy-tailed if it satisfies the relation

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \tag{2}$$

where α is the tail index. The tail index determines the shape of the tail distribution.

To allow for time-varying extremes, we deviate from the notion of IID observations and allow the observations to be drawn from different distributions over time. In introducing time-varying (or heteroscedastic) observations, we closely follow Einmahl et al. (2016).

Consider a set of independent observations $X_1, \ldots, X_n$ following various continuous distributions $F_1(x), \ldots, F_n(x)$ with common right end point x^*. Define the positive scedasis function c on $[0, 1]$ such that

$$\lim_{x\uparrow x^*} \frac{1 - F_i(x)}{1 - F_0(x)} = c\left(\frac{i}{n}\right) \tag{3}$$

for all $1 \leq i \leq n$, where $F_0(x)$ is a continuous distribution function with the same right end point as $F_i(x)$. Notice the similarities with limit relation (2). Relation (3) only compares the tails of the distributions and thus makes no assumption on the centres of the distributions. By imposing the condition

$$\int_0^1 c(s)\mathrm{d}s = 1, \tag{4}$$

we can interpret c as the relative frequency of extremes over the interval. For example, if $c(s) = 1 \; \forall \; 0 \leq s \leq 1$, then the relative frequency is constant (and equal to 1) over the interval and we can thus speak of time-constant (or homoscedastic) extremes.

We also assume F_0 is in the maximum domain of attraction of the generalised extreme value distribution (GEV), or $F_o \in \text{MDA}(G_\gamma)$. This means that for all $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{5}$$

for a positive scale function $a(t)$ and the extreme value index γ. Here,

$$U := \left(\frac{1}{1 - F_0}\right)^{\leftarrow},$$

where '$\leftarrow$' denotes the left continuous inverse function. We also define

$$U_i := \left(\frac{1}{1 - F_i}\right)^{\leftarrow},$$

so that combining relation (3) and relation (5), one can show that

$$\lim_{t\to\infty} \frac{U_i(tx) - U_i(t)}{a(t)\{c(i/n)\}^\gamma} = \frac{x^\gamma - 1}{\gamma}.$$

This means that all distributions F_i belong to the domain of attraction of the same GEV (with the same extreme value index γ). Consequently, one can model the relative frequency of extreme values over time with the scedasis function c.

4.1.2 Introduction to testing for homoscedastic extremes

To test for homoscedastic extremes, or $H_0 : c(s) = 1 \; \forall \; 0 \leq s \leq 1$, Einmahl et al. (2016) use the integral of the scedasis function c, fittingly called the integrated scedasis C:

$$C(s) := \int_0^s c(u)\mathrm{d}u \text{ for } s \in [0, 1].$$

To obtain a finite sample estimator of C, Einmahl et al. (2016) transform the set of independent observations $X_1, \ldots, X_n$ into order statistics $X_{(1)} \leq \ldots \leq X_{(n)}$. With a suitable intermediate sequence $k = k(n)$, for which

$$\lim_{n\to\infty} k = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{k}{n} = 0, \tag{6}$$

define the estimator

$$\hat{C}(s) := \frac{1}{k} \sum_{i=1}^{[ns]} \mathbb{1}_{\{X_i > X_{(n-k)}\}}. \tag{7}$$

Since $c(s)$ can be seen as the relative frequency of extremes at point s, $C(s)$ should be proportional to the number of extremes in the first $[ns]$ observations. Testing $H_0 : c(s) = 1 \; \forall \; 0 \leq s \leq 1$ is then equivalent to testing

$$C(s) := \int_0^s c(u)\mathrm{d}u = \int_0^s 1\mathrm{d}u = s \text{ for } s \in [0, 1].$$

This leads to the testing problem

$$\begin{cases} H_0 : C(s) = s, \\ H_1 : C(s) \neq s. \end{cases}$$

Einmahl et al. (2016) use a Kolmogorov-Smirnov-type test statistic

$$T_1 := \sup_{0 \leq s \leq 1} |\hat{C}(s) - C_o(s)|,$$

where $C_0(s) = s$. Assuming the same conditions as in Theorem 1 in Einmahl et al. (2016) are met, then as $n \to \infty$,

$$\sqrt{k} T_1 \xrightarrow{\mathrm{d}} \sup_{0 \leq s \leq 1} |B(s)|, \tag{8}$$

where $B(s)$ is a standard Brownian bridge. Rejecting H_0 in favour of $H_1 : C(s) \neq s$ with confidence level α, entails $\sqrt{k}T_1$ being larger than the $(1-\alpha)$-th quantile of $\sup_{0 \leq s \leq 1} |B(s)|$. Rejection of the null hypothesis means rejection of homoscedastic extremes, meaning that the frequency of extreme wind speed is not constant over time.

4.1.3 Expansion to heteroscedastic extremes in a spatial setting

The aforementioned estimation of the scedasis and corresponding test for heteroscedastic extremes works well for local sets of observations. However, it does not take into account potential spatial dependencies, which in this research manifest as dependencies between weather stations. In this study, we follow Einmahl et al. (2022), who extend upon Einmahl et al. (2016) by taking into account (potentially) spatially dependent observations.

Consider the set of independent random vectors $X_{i,1}, \ldots, X_{i,m}$, where $i \in [1, 2, \ldots, n]$. In our research, n is the number of observations in each weather station and m is the number

of weather stations. Denote $F_{i,j}(x)$ as the distribution function of $X_{i,j}$. We extend relation (3) to include space such that

$$\lim_{x \uparrow x^*} \frac{1 - F_{i,j}(x)}{1 - F_0(x)} = c\left(\frac{i}{n}, j\right) \in (0, \infty) \tag{9}$$

holds for observations $i \in [1, 2, \ldots, n]$ in weather stations $j \in [1, 2, \ldots, m]$. Here, $F_0(x)$ is some baseline continuous distribution function in the domain of attraction of an extreme value distribution and $c(\cdot, j)$ is the space-time scedasis. For the space-time scedasis we also introduce its integrated scedasis, defined as

$$C_j(s) := \frac{1}{m} \int_0^s c(u, j) \mathrm{d}u, \tag{10}$$

for $0 \leq t \leq 1$ and $j = 1, 2, \ldots, m$. We then impose the condition

$$\sum_{j=1}^{m} C_j(1) = 1,$$

to ensure that the function c is uniquely defined. Now, $c(i/n, j)$ can be interpreted as the relative frequency of extremes at observation i and location j. Additionally, $C_j(s)$ should be proportional to the number of extremes in the first $[ns]$ observations of location j. In order to compare the different C_j's across weather stations, we use the same extreme value threshold for all weather stations. This *common* threshold is established by taking a high empirical quantile of the total number of observations $N := n \cdot m$. We define a new order statistic threshold $X_{(N-k)}$, which is the $(N-k)$-th order statistic of the set of all observations $\{X_{i,j}\}_{i=1,j=1}^{n,m}$, for an intermediate sequence $k = k(n)$ with the same conditions as in condition (6). However, the value of k is now likely different from the one discussed in Section 4.1.2. We can then define the new estimator

$$\hat{C}_j(s) := \frac{1}{k} \sum_{i=1}^{ns} \mathbb{1}_{\{X_{i,j} > X_{(N-k)}\}}. \tag{11}$$

We assume that $F_0(x) \in MDA(G_\gamma(x))$ for $\gamma \in \mathbb{R}$. Together with definition (11), this means that γ is now the *common* extreme value index and is thus assumed constant over time and space. Mathematically, $F_{i,j} \in MDA(G_\gamma(x))$ for all $i \in [1, 2, \ldots, n]$ and $j \in [1, 2, \ldots, m]$.

Now that we have obtained the individual distributions, we model the dependence in space. For this, we require some prerequisites. First, let $F_i(x_1, x_2, \ldots, x_m)$ be the distribution

function of $X_{i,1}, X_{i,2}, \ldots, X_{i,m}$. We assume that the distribution function $F_{i,j}$ of observation $X_{i,j}$ is continuous and set $U_{i,j}(s) := F_{i,j}^{\leftarrow}(1 - 1/s)$. Here, '$\leftarrow$' denotes the generalised inverse function. We also assume that

$$\tilde{F}(x_1, x_2, \ldots, x_m) := F_i\left(U_{i,1}(x_1), U_{i,2}(x_2), \ldots, U_{i,m}(x_m)\right)$$

does not depend on i and is in the domain of attraction of a multivariate extreme value distribution. This assumption implies that the multivariate tail dependence structure does not depend on i, which in the case of a time series of observations means it is independent of time. We can now write the tail copula R_{j_1,j_2} of the components j_1 and j_2 ($j_1 \neq j_2$) as

$$R_{j_1,j_2}(x, y) = \lim_{s \downarrow 0} \frac{1}{t} \mathbb{P}(1 - F_{i,j_1}(X_{i,j_1}) \leq tx, 1 - F_{i,j_2}(X_{i,j_2}) \leq ty),$$

for $(x, y) \in [0, \infty]^2 \backslash \{(\infty, \infty)\}$.

4.1.4 Testing for homoscedastic extremes in a spatial setting

In our new multivariate setting we can test whether the total frequency of extreme events is constant over various locations. Using the multivariate integrated scedasis from definition (10), we notice that constant total frequency of extreme events would mean that, over a time interval, every weather station would have the same number of total extremes. If we let the total amount of extremes be 1, then every station should have a total number of extremes of $1/m$. Mathematically, our testing problem becomes

$$\begin{cases} H_0 : C_j(1) = \dfrac{1}{m} & \text{for all } j = 1, \ldots, m, \\ H_1 : C_j(1) \neq \dfrac{1}{m} & \text{for some } j = 1, \ldots, m. \end{cases}$$

We can now define the test-statistic. Let $\mathbb{1}_m$ be the m-unit vector and I_m be the m-dimensional identity matrix. Let $M := I_m - \dfrac{1}{m}\mathbb{1}_m\mathbb{1}_m^{'}$. We also define the symmetric covariance matrix $\Sigma = \Sigma(s_1, s_2, t_1, t_2)$ with entries

$$\sigma_{j_1,j_2}(s_1, s_2, t_1, t_2) = \frac{1}{m} \int_0^{t_1 \wedge t_2} R_{j_1,j_2}(s_1 c(u, j_1), s_2 c(u, j_2)) \mathrm{d}u,$$

for $1 \leq j_1, j_2 \leq m$. In this study, we use the finite sample estimator $\hat{\Sigma}$ with entries

$$\hat{\sigma}_{j_1,j_2} = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}_{\{X_{i,j_1} > X_{(N-k)}, X_{i,j_2} > X_{(N-k)}\}}.$$

Following the proof in Einmahl et al. (2022), under H_0, $D = \sqrt{k}(\hat{C}_1(1) - \frac{1}{m}, \ldots, \hat{C}_m(1) - \frac{1}{m})'$ is asymptotically m-variate normal with zero vector mean and covariance matrix $M\Sigma_1 M'$, where $\Sigma_1 = \Sigma(1,1,1,1)$. If we assume that Σ_1 is invertible, then $\text{rank}(M\Sigma M') = m - 1$. This means we can limit our attention to the first $m-1$ elements. Let A_{m-1} denote a matrix consisting of the first $m-1$ rows and columns of some square matrix A. We define the test statistic as

$$T_2 := D'_{m-1}((M\hat{\Sigma}_1 M')_{m-1})^{-1} D_{m-1}. \tag{12}$$

As shown by Einmahl et al. (2022), under H_0,

$$T_n \xrightarrow{d} \chi^2_{m-1}, \text{ as } n \to \infty.$$

Besides testing for constant total frequency of extremes across space, we test for homoscedastic extremes in the multivariate setting. For all weather stations $j \in [1, 2, \ldots, m]$, the test problem becomes

$$\begin{cases} H_0 : C_j(s) = sC_j(1), \\ H_1 : C_j(s) \neq sC_j(1). \end{cases}$$

The notable difference from the test problem in Section 4.1.2 is the addition of $C_j(1)$, which stems from the fact that we group observations of all weather stations together and use the common threshold $X_{(N-k)}$. To test the null hypothesis, we use a Kolmogorov-Smirnov-type test statistic, based on the process $\sqrt{k}(\hat{C}_j(s) - t\hat{C}_j))/\sqrt{\hat{C}_j(1)}, s \in [0,1]$; under H_0,

$$\left\{ \sqrt{k\hat{C}_j(1)} \left(\frac{\hat{C}_j(s)}{\hat{C}_j(1)} - s \right) \right\}_{s \in [0,1]} \xrightarrow{d} \{B(s)\}_{s \in [0,1]}.$$

If we define the test statistic as

$$T_3 := \sup_{0 \leq s \leq 1} \left| \hat{C}_j(s) - s\hat{C}_j(1) \right| \frac{1}{\sqrt{\hat{C}_j(1)}}, \tag{13}$$

then, under $H_0 : C_j(s) = sC_j(1)$, as $n \to \infty$,

$$\sqrt{k} T_3 \xrightarrow{d} \sup_{0 \leq s \leq 1} |B(s)|, \tag{14}$$

where $B(s)$ is a standard Brownian bridge. We can notice similarities between this test and the univariate test in Section 4.1.2. Performing the two aforementioned tests for all weather stations aids us in understanding the behaviour of wind extremes in our dataset.

Choosing k for heteroscedastic extremes in a spatial setting

To estimate C and test for homoscedastic extremes, we need an input k (e.g. (10), (14)). For this, we follow Einmahl et al. (2022) and use the peaks-over-threshold (POT) method. Consider again a set of IID observations $(X_1, \ldots, X_n)$ with common distribution function $F(x)$. Pickands (1975) shows that the conditional excess distribution over a threshold u follows approximately the generalised Pareto distribution (GPD), given by

$$G_{\gamma,\sigma}(x) = \begin{cases} 1 - (1 + \gamma x/\sigma)^{-1/\gamma}, & \gamma \neq 0, \\ 1 - \exp(-x/\sigma), & \gamma = 0, \end{cases} \tag{15}$$

where γ is the same extreme value parameter and $\sigma > 0$ is the scale parameter. In this equation $x \geq 0$ when $\gamma \geq 0$ and $0 \leq x \leq 1 - \sigma/\gamma$ when $\gamma < 0$.

In our data sample, we let the estimator of the threshold u be $\hat{u} := X_{(N-k)}$, where $N = n \cdot m$ is the set of grouped observations. $k = k(N)$ is a suitable intermediate sequence as defined in (6). Choosing k too small leaves little observations left above the threshold, resulting in an extreme value distribution with high variance. Choosing k too large results in a low threshold, which can cause biased extreme value parameters. To choose k, we estimate the extreme value and scale parameter γ and σ from equation (15). As our sample size is larger than 500, we can use maximum likelihood to obtain reliable estimates (Hosking & Wallis, 1987). The log-likelihood corresponding to the GPD for a set of observations is given by

$$\ell(\gamma, \sigma; x) = -\log \sigma - (1 + \frac{1}{\gamma}) \log(1 + \gamma \frac{x}{\sigma}), \quad 0 < x < \frac{\sigma}{-\gamma_-}.$$

Using this log-likelihood function, the pseudo log-likelihood can be written as

$$L_{N,k}(\gamma, \sigma) = \sum_{i=1}^{k} \ell(\gamma, \sigma, X_{N-i+1} - X_{N-k}).$$

In order to maximise the pseudo log-likelihood, we solve the score equations

$$\begin{cases} \dfrac{\partial}{\partial \gamma} L_{N,k}(\gamma, \sigma) = 0, \\ \dfrac{\partial}{\partial \sigma} L_{N,k}(\gamma, \sigma) = 0. \end{cases} \tag{16}$$

Under conditions (i)-(iv) stated in Einmahl et al. (2022) (which we assume met) and for $\gamma_{true} > -1/2$, one can prove that there exists a unique set of estimators $(\hat{\gamma}, \hat{\sigma})$ that maximise

the score equations (16). Since the extreme value index is bounded below by zero, we use the BFGS algorithm to obtain the estimates $\hat{\gamma}$ and $\hat{\sigma}$ numerically for each value of k. We then plot $\hat{\gamma}$ against k to obtain stability plots. We choose the highest value of k for which the value of $\hat{\gamma}$ remains stable.

4.2 Regression

4.2.1 Choice of model

A commonly used method for studying the drivers of mortgage default is a logistic regression or its inverse function, the logit model (e.g. Altman and Saunders (1997); Episcopos et al. (1998); Qi et al. (2021)).

At any time, a mortgage i is in any of two states, denoted

$$y_i = \begin{cases} 0 & \text{Not in default} \\ 1 & \text{In default.} \end{cases}$$

Since y_i is a binary variable, it follows a Bernoulli distribution

$$y_i \sim \mathrm{B}(p_i),$$

where $p_i := \mathbb{P}(y_i = 1|\mathbf{x}_i)$ is the probability of default (PD) given the set of regressors $\mathbf{x}_i$. In the logistic regression model the probability distribution of y_i is given by

$$p_i := \mathbb{P}(y_i = 1|\mathbf{x_i}) = \frac{\exp(\beta'\mathbf{x}_i)}{1 + \exp(\beta'\mathbf{x}_i)},$$

where β is the parameter vector belonging to the vector of predictor variables $\mathbf{x}_i$. For sake of readability, we write this in the form of the logit model, meaning that for the log odds ratio we have that

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta'\mathbf{x}_i.$$

This is an important equation, as it shows that the log odds ratio is a linear function of predictor space $\mathbf{x}_i$. In our study, the predictor space $\mathbf{x}_i$ consists of four control variables, to which the model wind data is added (see Sections 4.2.2 and 4.2.3, respectively).

We estimate the regression coefficients β by maximum likelihood. It can be shown that the log-likelihood function of y_i is given by

$$\ell(\beta) = \frac{1}{n} \sum_{i=1}^{n} y_i \beta' \mathbf{x}_i - \ln\left(1 + \exp(\beta' \mathbf{x}_i)\right).$$

We find the coefficient space β by maximising the log-likelihood function using the Newton-Raphson procedure.

Class imbalance

A well-known issue for logistic regression is the situation where $\mathbb{P}(y_i = 0)$ and $\mathbb{P}(y_i = 1)$ are not close to 50%. If there are many more observations where $y_i = 1$ (or $y_i = 0$ for that matter), there is no (approximate) balance between the two states of y_i. This is called class imbalance in the sample. This has negative effects on the ML estimation, as it affects its predictive capability (Oommen et al., 2011). Given that the mortgage default rate in our sample ranges from 0% to 2% (see Figure 4), the probability of default in our sample is relatively low. We are thus dealing with a sample with class imbalance, wherein one class (default) is substantially underrepresented in comparison to the other class (non-default).

A valuable method of solving class imbalance is the synthetic minority oversampling technique (SMOTE), as introduced by Chawla et al. (2002), which has since then become the de facto standard in the framework of learning from imbalanced data (Fernández et al., 2018). The SMOTE preprocessing algorithm generates synthetic default observations utilising a nearest neighbour algorithm, which are incorporated into the original data to estimate the coefficients of the logit models. This technique enhances the predictive power of the models by supplementing the dataset with artificial observations. An in-depth explanation on the SMOTE algorithm can be found in Appendix B. To diminish the negative effects of class imbalance in our data sample, we perform the SMOTE preprocessing algorithm before the maximum likelihood estimation of our coefficient space β.

4.2.2 Predictor space of the benchmark model

To be able to draw definitive conclusions on the significance of the frequency of wind extremes, we need a predictive benchmark model. A model's performance is in large part affected by the quality of its predictors. We thus seek to incorporate a set of key predictors

that have been widely recognised as influential factors in mortgage default risk assessment.

Jones and Sirmans (2015) offer a comprehensive examination of the determinants of residential mortgage default, drawing on an extensive review of nearly 100 empirical and theoretical studies in the field. After careful examination of the relevant studies found in Jones and Sirmans (2015) and the data that is available to us, we find four variables to be crucial predictors within our benchmark model.

The Fair, Isaac, and Company (FICO) score of the borrower acts as an indicator of the borrower's credit background and is used to assess the borrower's creditworthiness. A higher FICO score is associated with higher creditworthiness and thus lower default risk. The initial loan-to-value (LTV) rate represents the proportion of the property's value that the borrower holds as equity at the time loan origination. It has been shown to have a significant and positive relation with default risk. The mortgage age is also included as a predictor. Mortgage age is found to have a non-linear relation with default risk. It has been shown that default risk shows an ascending pattern during the early stages of the mortgage agreement, maintains a relatively stable level during the intermediate years, and experiences an additional rise towards the conclusion of the mortgage term. Lastly, we include the unemployment rate. Studies show that an increased unemployment rate is associated with a higher default risk. Intuitively, individuals who have recently experienced job loss may encounter challenges in meeting their mortgage payments. Adding unemployment as a benchmark predictor has the added benefit of acting as a control variable for extreme wind. There are two main ways that extreme wind can cause mortgage defaults, either by destroying the property of the owner, or by destroying their workplace, which can end the owner's state of employment. Adding the unemployment rate as a benchmark predictor can thus in part control for the second cause of mortgage default by extreme wind.

The FICO score, initial LTV rate, and loan age are provided with the mortgage data. We obtain the monthly unemployment rate from the website of the US Bureau of Labor Statistics. Not all predictors remain unaltered for the regression. Our target variable is 3-month delinquency, which by definition is lagged by 3 months. In our regression setting we want to be able to predict the '3-month ahead default' with concurrent predictors. For

that reason we lag the time-varying unemployment rate predictor by 3 months to match the target variable. The FICO score, initial LTV rate and loan type remain constant over time and thus also remain unaltered. By introducing lag to our time-varying predictors, we transform the regression from a contemporaneous to a predictive one.

4.2.3 Incorporation of wind extreme frequency

To incorporate extreme wind into our predictor space, we use a simplified version of the scedasis function. However, a discrepancy arises due to the difference in frequency between the wind data (daily granularity) and the mortgage data (monthly granularity). To align the frequencies, we transform the wind extremes frequency from daily to monthly granularity. We achieve this by integrating the scedasis function over a month. We can do this by using the integrated scedasis function and its estimator $\hat{C}_j(t)$ (as defined in equation (10)). Using the fact that our dataset consists of 20 years, or 240 months of data, we arrive at the monthly frequency of extreme wind at location j in month $t \in [0, 239]$ by

$$\widehat{\text{Freq}}_j(t) := \hat{C}_j\left(\frac{t+1}{240}\right) - \hat{C}_j\left(\frac{t}{240}\right) = \frac{1}{k}\sum_{i \in O_t} \mathbb{1}_{\{X_{i,j} > X_{(N-k)}\}},$$

where O_t corresponds to the set of observations in month t. With this notation, the monthly frequency $\widehat{\text{Freq}}_j(t)$ is given as a fraction of the total frequency, which is 1. To aid interpretability of the coefficient value of $\widehat{\text{Freq}}_j(t)$, we multiply $\widehat{\text{Freq}}_j(t)$ by k. This results in the step function $\hat{W}_j(t)$, which represents the total frequency of extreme wind in month t:

$$\hat{W}_j(t) := k \cdot \widehat{\text{Freq}}_j(t) = \sum_{i \in O_t} \mathbb{1}_{\{X_{i,j} > X_{(N-k)}\}}. \tag{17}$$

We will refer to $\hat{W}_j(t)$ as the 'wind step function'. By matching location j to mortgage i, we can incorporate the wind step function into the regression analysis as an additional predictor in the predictor space $\mathbf{x}_i$. Before integrating it into the predictor space, we lag the wind step function by 3 months, to assess its predictive power. Intuitively, an event of extreme wind today would mean an increased probability to become delinquent in payments. However, the increased probability of being 3 months delinquent only arises 3 months later. We will refer to the model with the benchmark predictors and monthly integrated scedasis as the 'full model'.

4.2.4 Model analysis and performance assessment

The first step of each model's analysis is performing a logistic regression on the entire dataset and analysing the statistical significance of each predictor by means of a t-test. Specifically, we are interested at the statistical significance of the monthly integrated scedasis aside from the benchmark predictors. We perform a likelihood ratio (LR) test to test for a statistically significant difference between the fit of the benchmark model and the full model.

To assess the predictive power of both logit models, we split the data into an 80/20 train/test split and compare predicted (non-)defaults in the test set to their actual status. We consider several evaluation metrics. Firstly, we employ a confusion matrix to evaluate the performance of both models. Then, we compute various accuracy metrics (accuracy, precision, recall and F_1 score) to assess the effectiveness and reliability of the model. As a last measure, we plot the receiver operating characteristic (ROC) curve to calculate the area under the curve (AUC) score to gauge the model's ability to distinguish between defaults and non-defaults. We can compare the different accuracy metrics of the benchmark and full model to investigate potential differences in the predictive power of both models.

5 Empirical Results

In this section, we present and discuss the results obtained from the research. Firstly, we discuss the tests that were performed concerning potential heteroscedasticity in wind extremes. Secondly, we briefly discuss the results of the benchmark model regression. Thirdly, we incorporate extreme wind as a predictor and investigate the potential improvement this provides over the benchmark model. We perform robustness checks to ensure the reliability and consistency of our findings. Lastly, we synthesise all results and discuss their implications.

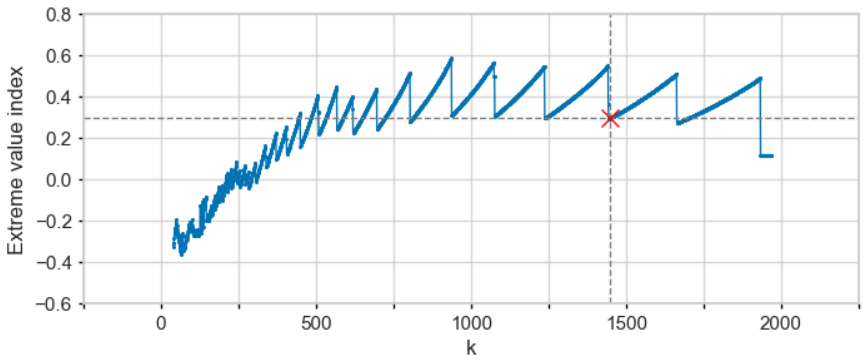
5.1 Tests on wind speed extremes

Our initial focus revolves around scrutinising extreme wind speeds across the southeast United States. We perform two tests: whether the total frequency of extremes is constant across all cities and whether the frequency of extremes of each city is homoscedastic.

Choosing k

To perform our tests we must first choose an appropriate value of k. For this, we group all $N = m \cdot n$ observations. We then solve the score equations (16) for a range of values of k. From this we construct a stability plot, in which the estimated extreme value index $\hat{\gamma}$ is plotted against different values of k. The stability plot can be found in Figure 5.

Figure 5: Stability plot of the grouped dataset, showing $\hat{\gamma}$ for different values of k.



Note: the chosen value of k and corresponding $\hat{\gamma}$ is marked with a red cross.

Inspecting the stability plot, we observe the instability of $\hat{\gamma}$ at lower values of k. As k increases, $\hat{\gamma}$ gradually increases to a more stable value. However, we also observe a sawtooth pattern emerge. This can be attributed to the fact that our wind speed data has unit accuracy, leading to many values being observed multiple times. Each sawtooth corresponds to values of k for which the value of $X_{(N-k)}$ remains the same. As k increases further, we see that the value of $\hat{\gamma}$ gradually decreases. For example, the rightmost sawtooth is generally lower than the one to the left of it. We want to choose k as large as possible, but low enough so that there is no bias in the corresponding value of $\hat{\gamma}$. We therefore choose $k = 1450$, marked by a red cross in Figure 5. This translates to a threshold $X_{(N-1450)} = 38$ knots.

Test for constant total frequency over all cities

With the chosen $k = 1450$ and corresponding threshold $X_{(N-1450)} = 38$ knots, we estimate $C_j(1)$ using the estimator defined in equation (10). The estimation results can be found in Table 3.

We observe disparities in $\hat{C}_j(1)$ depending on the city. New Orleans, a coastal city

Table 3: Value of $\hat{C}_j(1)$ for each city.

State	**City**	$\hat{C}(1)$	**State**	**City**	$\hat{C}(1)$
Alabama	Huntsville	0.047	Louisiana	New Orleans	0.121
	Montgomery	0.031		Shreveport	0.046
Arkansas	Little Rock	0.042	Mississippi	Jackson	0.040
Florida	Jacksonville	0.063	North Carolina	Charlotte	0.022
	Miami	0.043		Raleigh	0.026
	Orlando	0.037		Wilmington	0.056
	Tallahassee	0.040	South Carolina	Charleston	0.041
	Tampa	0.020		Columbia	0.046
Georgia	Atlanta	0.050	Tennessee	Knoxville	0.063
	Savannah	0.019		Memphis	0.057
Louisiana	Baton Rouge	0.039		Nashville	0.051

known for having suffered many extreme wind events, has by far the highest total frequency of extreme wind. Notably, while some inland cities such as Charlotte and Raleigh exhibit relatively low frequency of extreme wind, this pattern is not consistent amongst the sampled cities. The three cities in the state of Tennessee, for example, still have relatively high total extreme wind frequency, demonstrating that high frequency of extreme wind is not exclusive to coastal cities.

To test for constant total frequency of extreme wind across all cities, we calculate the test statistic (12) for $k = 1450$. We find a p-value of 0.000, leading us to the conclusion that the total frequency of extreme wind is not constant over all cities. In fact, we obtain a p-value of 0.000 even when deviating from our chosen value of k, which solidifies our conclusion. These results do not surprise us as the different cities are hundreds of kilometres apart from each other and the climates in which they are situated vary substantially.

Test for homoscedasticity in each city

Next, we examine the behaviour of wind extremes in each city in a spatial setting. We calculate the test statistic (13) and calculate the p-value of homoscedastic extremes for each

city. The results can be found in Table 4. From the results we conclude that with a confidence level of 95%, 13 cities' wind extremes show heteroscedasticity. This means that for more than half of the cities in this research, the frequency of extremes has not remained constant over the sample time period between 2000 and 2020. However, since we test for a large number of cities simultaneously, we may encounter a multiple test problem. To account for this, we perform a Bonferroni correction. Specifically, we divide the significance level by the number of tests, such that $\alpha_{new} = 5\%/22 = 0.23\%$. With the Bonferroni correction in mind, only Jackson and Nashville show heteroscedastic extremes, only two of the 22 considered cities.

Table 4: Value of test statistic T_3 and p-value of homoscedasticity for each city.

State	**City**	T_3	p-**value**	**State**	**City**	T_3	p-**value**
Alabama	Huntsville	1.783	0.003	Louisiana	New Orleans	0.828	0.469
	Montgomery	0.864	0.417		Shreveport	1.717	0.005
Arkansas	Little Rock	1.547	0.015	Mississippi	Jackson	2.221	0.000
Florida	Jacksonville	0.944	0.312	North Carolina	Charlotte	1.428	0.030
	Miami	1.015	0.237		Raleigh	0.829	0.472
	Orlando	1.424	0.032		Wilmington	1.090	0.172
	Tallahassee	1.720	0.005	South Carolina	Charleston	1.503	0.019
	Tampa	1.004	0.247		Columbia	1.985	0.001
Georgia	Atlanta	1.067	0.189	Tennessee	Knoxville	1.364	0.044
	Savannah	1.594	0.011		Memphis	1.635	0.009
Louisiana	Baton Rouge	1.148	0.133		Nashville	2.318	0.000

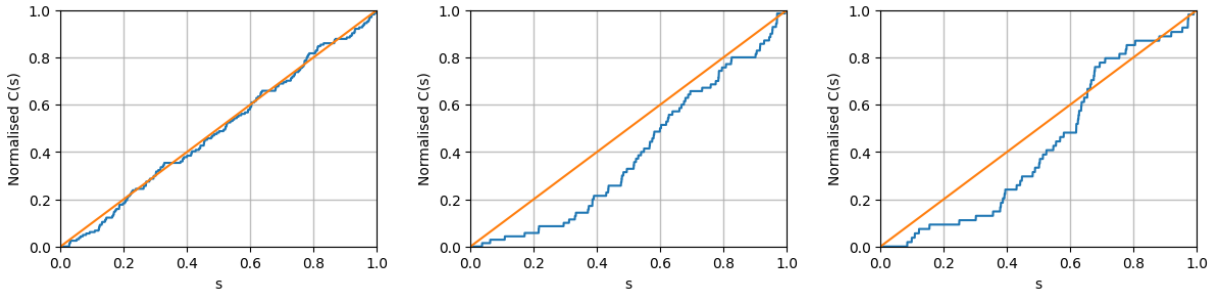
Note: p-values that are below 5% are highlighted in grey.

However, we are not only interested in the result of the test for homoscedasticity in each city. We also study the shape of the integrated scedasis function over time. The null hypothesis states that $C_j(s) = sC_j(1)$, meaning that the integrated scedasis increases linearly over time. Taking New Orleans as an example, Table 4 shows that we fail to reject H_0. Thus, the integrated scedasis should increase fairly linearly. The normalised integrated scedasis of New Orleans is plotted in Figure 6a. Consistent with the statistical test, we observe that the integrated scedasis of New Orleans closely follows the integrated scedasis under H_0.

The shape of the integrated scedasis is less straightforward for Jackson and Nashville for which there is heteroscedasticity of wind extremes. For example, the integrated scedasis of Nashville is shown in Figure 6c. Here, the integrated scedasis deviates from the straight line under H_0. It seems that for Nashville, the frequency of extremes is low during the first years of our sample period and increases with time. For Jackson, we see no clear pattern.

The inability to reject the null hypothesis for 20 of the 22 cities suggests that, for most cities in the southeast US, the frequency of extreme wind is constant over time. This means that the statement by the IPCC (2023) concerning the increasing frequency of weather extremes does not hold for extreme wind in most cities in the southeast US.

Figure 6: Estimated integrated scedasis of New Orleans (left), Nashville (middle) and Jackson (right) against the integrated scedasis under H_0 (orange).



(a) New Orleans **(b)** Nashville **(c)** Jackson

Note: the integrated scedasis is normalised such that $C_j(1) = 1$.

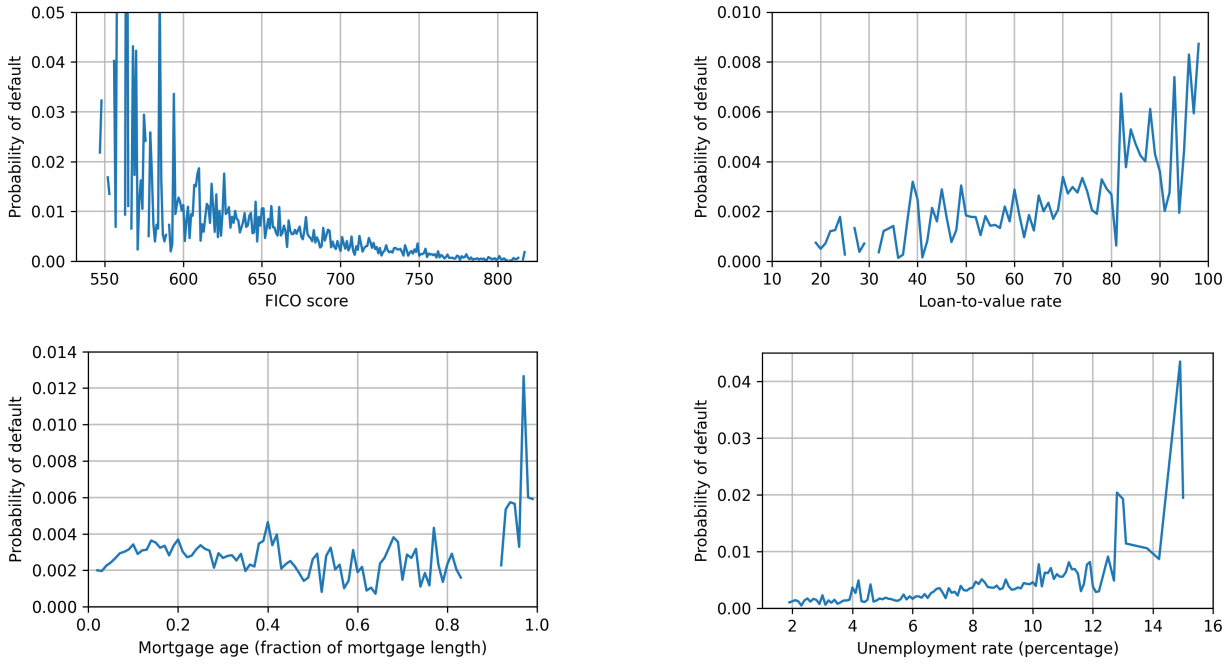
5.2 Benchmark model analysis

Having studied the dynamics of our data on extreme wind, we continue to study its potential effect on mortgage defaults. Before incorporating extreme wind as a predictor, we first analyse our benchmark predictors and benchmark model.

Benchmark predictor analysis

We examine the relation between our benchmark predictors (FICO score, LTV rate, mortgage age and unemployment rate) and the PD. The plots can be found in Figure 7.

Figure 7: Plots of PD against the different predictors of the benchmark model.



Note: mortgage age is taken as a fraction of the total mortgage length.

Evaluating the plots left-to-right and top-to-bottom, we first observe how the PD varies across different FICO scores. As expected, the PD declines as the FICO score increases, suggesting that mortgages with higher credit scores have a lower probability of defaulting. We also see that the PD varies greatly for low FICO scores between 550 and 600. This can be attributed mostly to the fact that there is a relatively small amount of mortgage loans issued to mortgage takers with a low FICO score. The increased variability of the PD is thus due to a low sample size and a resulting imprecise measurement. In the plot of PD against LTV rate, we observe that PD increases with LTV rate, as expected. We see a higher variability for high LTV rates, because there are not many high LTV rate mortgages given out. The relation between PD and mortgage age is less straightforward. The PD increases over the first quarter of the mortgage age, then drops slightly until around three quarters of the mortgage age, at which point it increases again. This is fairly consistent with the results of Hendershott and Schultz (1993), who find that the PD increases over the first quarter of mortgage age, then remains fairly constant until the last quarter, at which point it increases again. It is also in large part consistent with the studies outlined in Quercia

and Stegman (1992), who similarly find an increase in PD over the first quarter, a decrease over the next half but a decline in default risk in the final years (instead of an increase in the final years). There are multiple reasons for the PD to increase in the later years of a mortgage. An example of this is if the property has so-called negative equity. This occurs when the value of the property falls below the outstanding mortgage balance. In this case, the homeowner may strategically default the mortgage and let the lender capture ownership of the property (Foote, Gerardi, & Willen, 2008). Lastly, we observe that the PD increases with unemployment rate, again, in line with expectations.

Based on the observations from Figure 7, we split the mortgage age predictor into three parts: the first quarter of mortgage age, the middle part and the last quarter. This is to capture the non-linear relationship between mortgage age and probability of default. We include FICO score, LTV rate and unemployment rate as continuous predictors.

Benchmark model analysis

We perform a logistic regression of the probability of default onto the entire dataset with the benchmark predictors FICO score, LTV rate, the early, middle and final phase of mortgage age and the unemployment rate. The results can be found in Table 5. All coefficients are statistically significant, with all $p-$values smaller than 0.01. Most coefficients have the sign we expect. However, the coefficient of the variable Mortgage age (middle) is positive, contrary to our conclusions based on Figure 7.

The pseudo R^2 of the benchmark model is low, with a value of 26.6%. This metric suggests that our model accounts for approximately 27% of variation in the PD. It is important to note that mortgage default prediction is a complex task influenced by many different factors. The stochastic nature of different influences can make precise predictions challenging. As such, the low pseudo R^2 value aligns with our expectations.

Next, we examine the predictive performance of our benchmark model. Using an 80/20 train/test split, we predict (non-)defaults in our test sample. The relevant statistics can be found in Table 6 (the confusion matrix of the predictions can be found in Figure 10a in Appendix A). The accuracy and AUC exhibit substantial values, suggesting a good predictive model. However, as the test set is imbalanced, evaluating only the accuracy and AUC can

Table 5: Regression statistics for the benchmark logistic regression.

Variable	Coef.	SE	Variable	Value
Intercept	5.647	0.021	Pseudo R^2	0.266
FICO score	-0.016	0.000	Log-Likelihood	$-1.481 \cdot 10^6$
LTV rate	0.026	0.000	AIC	2962783
Mortgage age (early)	0.021	0.000		
Mortgage age (middle)	0.008	0.000		
Mortgage age (final)	0.004	0.000		
Unemployment rate	0.293	0.001		

Note: p-values of all coefficients are < 0.01.

Note: the reported pseudo R^2 *is the McFadden's* R^2.

give a skewed image of the model performance. Intuitively, predicting only non-defaults for a dataset with 2% defaults would give an accuracy of 98% and a small false negative rate, resulting in a high AUC. We therefore also inspect the precision, recall and F_1 score. We see that the recall is 57%, meaning that of all true defaults, most are accurately predicted to be defaults. The precision is however very low: our model predicts many non-defaulted mortgages to be in default. We attribute this to a number of reasons. Firstly, while performing SMOTE on the training set improves predictive performance on the test set, it does not completely mitigate the negative effects of an unbalanced dataset. Additionally, having few predictors of default can result in lower performance. We also assume a linear relation between the log-odds of default and our predictors, since we are using a logistic regression. This relation may however be non-linear. All these factors can negatively impact the predictive performance of our benchmark model.

Table 6: Accuracy statistics for the default predictions of the benchmark model.

	Accuracy	Precision	Recall	F_1 score	AUC
Value	86.11%	9.61%	57.48%	16.47%	83.69%

5.3 Full model analysis

Having established knowledge on our benchmark model, we examine the potential effect of extreme wind on the PD. For this, we incorporate the wind step function (17) into our mortgage dataset. We employ a standard reference point for extreme wind — hurricane wind speed, which is conventionally defined as 60 knots. This choice allows us to comprehensively examine the influence of extreme wind on mortgage defaults using a well-accepted threshold. We therefore set the threshold value $X_{(N-k)}$ equal to 60 knots.

Table 7: Regression statistics for the logistic regression including wind step function.

Variable	**Coef.**	**SE**	**Variable**	**Value**
Intercept	5.653	0.021	Pseudo R^2	0.266
FICO score	-0.016	0.000	Log-Likelihood	$-1.481 \cdot 10^6$
LTV rate	0.026	0.000	AIC	2962413
Mortgage age (early)	0.021	0.000		
Mortgage age (middle)	0.008	0.000		
Mortgage age (final)	0.004	0.000		
Unemployment rate	0.293	0.001		
Wind step function	0.131	0.008		

Note: p-values of all coefficients are < 0.01.

Note: the reported pseudo R^2 *is the McFadden's* R^2.

We first perform a logistic regression of the PD onto our full predictor space, using the entire dataset. The results can be found in Table 7. Comparing Table 7 to Table 5, it becomes evident that the inclusion of the wind step function does not alter the coefficients of any benchmark predictor. This observation suggests that the presence of the wind step function does not influence the estimated impact of the other predictors. Further, we can observe that the wind step function is statistically significant. This leads us to conclude that the wind step function, and thus extreme wind, contributes to explaining the PD of our sample of mortgages. The pseudo R^2 is the same in both regressions, suggesting that the wind step function does not add significant explanatory power to the benchmark model. We

suspect that this originates from the fact that both defaults as well as non-zero values of the wind step function are very rare. The union of a default and a non-zero value of the wind step function is present in only 0.072% of observations. However, the LR test comparing the benchmark and full model rejects the null hypothesis of equal fit. The slightly lower AIC of the full model confirms our conclusion that the full model is different from (and slightly superior to) the benchmark model.

Due to the linear relation between the predictor space and the log-odds of default, we can interpret the value of 0.131 of the wind step function coefficient. Specifically, a one-unit increment in the wind step function corresponds to on average a 14.00% increase in the PD, ceteris paribus. In the context of our dataset, this suggests that homes experiencing extreme wind events have on average a 14% higher probability of being in default, holding all other factors constant.

The presence of a statistically significant wind step function as an explanatory variable within the benchmark model holds substantial implications. As discussed in Section 2.3, most homeowners' insurance policies cover wind damage. Intuitively, if homeowners' insurance covers wind damage, then extreme wind should not have an impact on mortgage defaults, resulting in a statistically insignificant extreme wind variable. The fact that the wind step function is significant suggests that the explanatory capacity of extreme wind originates from its secondary effects, which is primarily due to flooding (Tonn & Czajkowski, 2022). A notable implication emerges from this observation. Flood insurance requirements are restricted to regions designated as high flood risk areas in government-issued flood maps, overseen by the Federal Emergency Management Agency. Consequently, only properties located outside of these high-risk zones would be susceptible to the secondary impacts of extreme wind. The statistical significance of extreme wind as a predictor prompts us to question whether these flood maps sufficiently account for the evolving weather patterns. This conclusion aligns with the findings of Tonn and Czajkowski (2022), offering a valuable perspective on the need for revised flood risk assessments.

Next, we predict (non-)defaults with the same 80/20 train/test split, this time including the wind step function as predictor. The results can be found in Table 8. The confusion

matrix corresponding to the predictions can be found in Figure 10b in Appendix A.

Table 8: Accuracy statistics for default predictions of the benchmark and full model.

	Accuracy	**Precision**	**Recall**	F_1 **score**	**AUC**
Benchmark	86.11%	9.61%	57.48%	16.47%	83.69%
Full	86.12%	9.58%	57.20%	16.41%	83.67%

Comparing the accuracy statistics of both models, we see almost no difference. We conclude from this that while extreme wind *does* contribute to explaining the PD, it *does not* contribute to predicting mortgage defaults. Again, we suspect this to be due to the fact that only a very small fraction of observations have both non-zero wind step function and are in default.

This conclusion is drawn from our analysis of the entire dataset. However, it is important to note that alternative insights may emerge when we delve into different subsamples of the data. By considering subsamples based on season, year and geographical location, we aim to gain a more comprehensive understanding of the impact of extreme wind on mortgage defaults. Aside from different subsamples, we also study the potential presence of non-linear relations between predictors and the target by using an XGBoost algorithm.

Season

Firstly, we investigate the effect of seasonality. We know from the literature that most extreme wind events occur during tornado season (between March and May) and hurricane season (between June and November). We may therefore see extreme wind have a higher coefficient value and more predictive power during these periods, since tornadoes and hurricanes have shown to have particularly destructive consequences for housing.

We split our dataset into the 'extreme wind season' (March until November), including tornado and hurricane season, and the winter season (December until February). Then, we perform our logistic regression individually on both samples of data. We find that the wind step function is statistically significant in both seasons, with a p-value of 0.000. The coefficient of the wind step function takes a value of 0.119 in extreme wind season and 0.406

in winter. Contrary to our belief, extreme wind events in winter have a larger impact on the PD than extreme wind events in the tornado and hurricane season. This may be due to a delayed effect of extreme wind on mortgages defaulting. The coefficient value of 0.406 corresponds to an increase in PD by 50% for a house experiencing an event of extreme wind, compared to a house that has not experienced such an event. This seems high. Additional research based on a larger dataset could prove useful in (dis)proving this conclusion.

Secondly, we split both datasets into an 80/20 train/test split and investigate the relative performance metrics of the benchmark and full model. In both seasons we see no improvement by including the wind step function.

Year

As discussed in Section 3.2.1, two visible peaks in the MDR time series are due to extreme weather events. We investigate whether extreme wind may improve the predictive power of the benchmark model in years with many extreme wind events. Specifically, we investigate whether the wind step function improves performance of the benchmark model during and up to one year after hurricane Katrina, as well as during and up to one year after the hurricane season in 2017. However, we find no substantial increase in predictive power by including the wind step function in any of these two periods of time.

Geographical location

Some states are much windier than others, and have experienced more events of extreme wind than others. We investigate for each state separately if the wind step function is statistically significant and if extreme wind has predictive power. The results can be found in Table 9.

We find that the wind step function is statistically significant and positive in six of the nine states, suggesting that extreme wind has no significant impact in Arkansas and Mississippi. These two states have not experienced many extreme wind events. We also notice the difference in coefficient value between states. The highest coefficient values of wind are found in Louisiana and North Carolina, two coastal states that have been subjected to multiple hurricanes. A coefficient value of 0.468 suggests an increase in PD by roughly 60%, following an event of extreme wind. While Louisiana is known for having suffered

major losses from examples such as hurricane Katrina, this value may be inaccurately high due to the small sample size of only one state. We attribute the negative coefficient value in Georgia to this same limitation.

The predictive performance of the benchmark model does not improve substantially by adding extreme wind in any state. Including only the six states for which the wind step function is statistically significant in the predictive regression also does not improve upon the predictive performance of the benchmark model.

Table 9: Regression and accuracy statistics for different states in the southeast US.

State	**Coef.**	**Accuracy**	**Precision**	**Recall**	F_1 **score**	**AUC**
Alabama	0.212***	85.57 (-0.05)	5.58 (-0.18)	55.36 (-1.79)	10.13 (-0.33)	84.01 (-0.32)
Arkansas	-	85.60 (-0.16)	6.39 (-0.07)	60.67 (0.00)	11.56 (-0.11)	85.05 (0.32)
Florida	0.019*	84.46 (-0.14)	16.22 (-0.02)	58.16 (0.62)	25.36 (0.03)	83.04 (-0.02)
Georgia	-0.063*	87.04 (-0.02)	8.68 (-0.04)	62.40 (-0.20)	15.24 (-0.06)	86.62 (-0.01)
Louisiana	0.468***	87.16 (1.89)	6.29 (-1.06)	31.53 (-13.05)	10.48 (-2.13)	74.56 (-2.74)
Mississippi	-	84.48 (0.00)	10.46 (0.00)	55.17 (0.00)	17.58 (0.00)	85.70 (-0.01)
North Carolina	0.430***	87.30 (-0.03)	5.98 (-0.03)	62.82 (-0.15)	10.92 (-0.05)	86.03 (-0.11)
South Carolina	0.173***	86.04 (-0.01)	7.60 (-0.08)	61.06 (-0.76)	13.52 (-0.15)	83.70 (-0.25)
Tennessee	0.203***	86.39 (0.11)	5.53 (-0.07)	56.67 (-1.31)	10.08 (-0.14)	83.56 (0.01)

*Note: significance levels are indicated as follows: * signifies* $p < 0.05$*, ** signifies* $p < 0.01$ *and *** signifies* $p < 0.001$*.*

Note: values between brackets show the accuracy metric with respect to the benchmark.

We can combine the results from testing for homoscedasticity of wind in each state with the results from the logistic regression run on each city. Nashville has shown signs of increasing frequency of extreme wind over the years 2000 until 2020. Simultaneously, we find a positive and statistically significant wind step function for Nashville. Over time, its frequency of extreme wind has been increasing and extreme wind (in form of the wind step function) is statistically significant in explaining PD. Thus, our findings suggest that as the frequency of extreme wind conditions increases, so does the total amount of mortgages defaulting in Nashville. However, the threshold for extreme wind in the statistical test and logistic regression are not consistent; we use a threshold of 38 knots in the statistical tests

but a threshold of 60 knots in our logistic regression. To make definitive conclusions on specific cities such as Nashville, more in-depth analysis is required.

Non-linear relationships

Parametric regressions are limited by assuming certain relations amongst predictors, and between the predictors and the target variable. However, many non-parametric approaches have been discovered that can capture complex relationships and non-linearities in the data. Due to its shown performance in predicting mortgage defaults (Coşer et al., 2019), we employ an extreme gradient boosting (XGBoost) algorithm. We choose the hyperparameters by a 5-fold cross-validation over a grid of hyperparameter values. The chosen grid and hyperparameter values of the XGBoost algorithm can be found in Table 14 in Appendix A.

We train and test the benchmark and full model (including wind step function) using the same 80/20 train/test split. The relevant accuracy statistics can be found in Table 10. The confusion matrices for both models can be found in Figure 11 in Appendix A.

From Table 10, we observe that the benchmark model performs better with the XGBoost algorithm than a standard logistic regression on all metrics except recall. We also find that the XGBoost algorithm including wind step function performs better than the benchmark XGBoost algorithm on all metrics. To get a better understanding of the impact of the wind step function on the XGBoost algorithm, we depict feature importance plots, shown in Figure 8.

Based on the feature weight importance plot, we conclude that the wind step function feature is seldom used in predicting defaults, since the feature is rarely non-zero. Based on the feature gain importance plot, we conclude that the wind step function has a significant effect on predicting mortgage defaults, when the feature is used. These conclusions provide more insight into the predictive power of extreme wind: while extreme wind does not occur often, it can be a valuable feature for predicting mortgage defaults using an XGBoost algorithm.

This result is compelling, as it suggests that extreme wind *can* add to the predictive accuracy of a predictive model. The capacity to foresee mortgage defaults with a considerable lead time holds substantial advantages for mortgage lenders. The most important benefit is the ability to accurately determine the expected loss of their mortgage portfolio. Lenders can

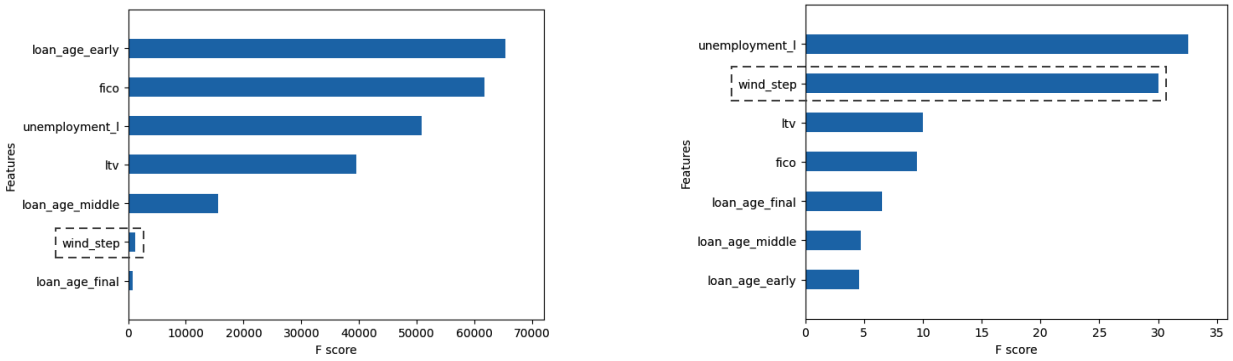
Table 10: Accuracy statistics for the default predictions of the benchmark logistic model and XGBoost model.

	Accuracy	**Precision**	**Recall**	F_1 **score**	**AUC**
Benchmark logistic	86.11%	9.61%	57.48%	16.47%	83.69%
Benchmark XGBoost	98.17%	68.14%	42.43%	52.29%	96.46%
Full XGBoost	98.25%	70.68%	44.30%	54.46%	96.83%

Note: the XGBoost algorithm including wind step function is denoted "Full XGBoost".

Note: best scores are highlighted in grey.

Figure 8: Feature importance plots of the XGBoost algorithm based on weight (left) and gain (right).



(a) Feature weight importance

(b) Feature gain importance

Note: the wind step function is highlighted and denoted by 'wind_step'.

also use this information to proactively reach out to support potentially defaulting mortgage borrowers, offering support and risk mitigation measures. This can reduce the likelihood of default among high-risk mortgages. A methodological drawback is that many non-parametric models, such as the XGBoost model, are considered 'black box models'. Many companies champion explainability of predictions made by their models, meaning that they often do not consider Machine Learning models. A good example of this is the XGBoost model we use, for which we do not know if a higher frequency of extreme wind in fact has a higher probability of default as an effect. Exploring this is beyond the scope of this paper and thus left for further research.

5.4 Robustness checks

In this section, we subject our research outcomes to a plethora of robustness checks to validate the reliability of our findings and enhance our understanding of the relationship between extreme wind events and mortgage defaults. Specifically, we perform threshold analysis, investigate temporal lags and perform K-fold cross-validation, each offering distinct perspectives on our research.

5.4.1 Threshold analysis

Firstly, we vary the definition of extreme wind by changing the threshold for what constitutes an extreme event. We then re-estimate the logistic regression model with different definitions of extreme wind and observe how the coefficient and significance level of the wind step function changes with respect to the initial threshold of 60 knots. We also investigate how the impact of extreme wind changes the predictive power of the benchmark model through the wind step function. The relevant statistics can be found in Table 11.

Table 11: Regression and accuracy statistics for different thresholds of extreme wind.

Threshold	**Coef.**	**Accuracy**	**Precision**	**Recall**	**F_1 score**	**AUC**
50 knots	0.009	85.98%	9.40%	56.51%	16.11%	83.15%
55 knots	0.154	85.97%	9.43%	56.84%	16.17%	83.31%
60 knots	0.131	86.12%	9.58%	57.20%	16.41%	83.67%
65 knots	0.321	86.07%	9.58%	57.21%	16.41%	83.48%
70 knots	0.371	86.11%	9.60%	57.20%	16.44%	83.52%

Note: p-values of all coefficients are < 0.01.

Note: best scores are highlighted in grey.

We find that the p-values of all coefficients of the wind step function are smaller than 0.01. On average, the coefficient value increases with the threshold. Intuitively, events of higher wind speeds should on average have more damaging effects than lower wind speeds. This solidifies the hypothesis that only very high wind speeds have a tangible negative effect on mortgage defaults. Our results are thus robust to an increase and (to a lesser

extend) a decrease of the threshold. The performance metrics of the predictive model support this conclusion. Accuracy statistics are generally lower for lower thresholds, though the differences are small.

5.4.2 Temporal lag analysis

We also analyse whether there is a temporal lag between extreme wind and its impact on mortgage defaults. We examine different lag periods to identify if the effect is immediate or delayed. We include a lag of 4, 5 and 6 months and compare it to the lag of 3 months that has been used throughout this research.

We perform a logistic regression with the full dataset, using each lag of wind step function separately. The results can be found in Table 12. We find that all lags are statistically significant, with all p-values smaller than 0.01. In fact, the coefficient values at the 4 and 5 month lags are higher than at the original 3 month lag, suggesting there is a sizeable lagged effect of extreme wind on mortgage defaults. This can be due to mortgage owners being able to pay their mortgage for a short period of time, before having to default the mortgage. The accuracy statistics suggest an opposing conclusion, since the accuracy statistics of the 4 month lag are smallest. However, the differences in accuracy statistics are low.

Table 12: Regression and accuracy statistics for lags of the wind step function.

Lag	**Coef.**	**Accuracy**	**Precision**	**Recall**	**F_1 score**	**AUC**
3 months	0.131	86.12%	9.58%	57.20%	16.41%	83.67%
4 months	0.236	86.04%	9.29%	56.07%	15.94%	82.69%
5 months	0.178	86.12%	9.55%	57.54%	16.37%	83.69%
6 months	0.117	86.04%	9.56%	57.80%	16.40%	83.90%

Note: p-values of all coefficients are < 0.01.

Note: best scores are highlighted in grey.

5.4.3 K-fold cross-validation

We implement 5-fold cross-validation to assess the model performance with different subsets of our data. This can help ensure that the wind step function variable retains significance

when the model is applied to various random partitions of our dataset. We find that extreme wind is statistically significant in each random subset of data. This is in support of our initial result in Section 5.3, where we also found extreme wind to be statistically significant. The mean coefficient value of the wind step function is 0.180 ± 0.008, which deviates significantly from our initial coefficient value of 0.131. This discrepancy shows the sensitivity of our coefficient estimates to fluctuations in data composition and size, underscoring how the magnitude of the coefficient values obtained from the different subsamples of data may be inaccurate. Using a larger dataset in future research could thus provide more accurate results.

We also study the different accuracy metrics. The results can be found in Table 13. We find that the accuracy metrics of both the benchmark and full model do not differ significantly from the original values. Lastly, we find that the mean of each accuracy metric of the benchmark model falls within one standard deviation of the corresponding mean of each accuracy metric of the full model. This supports our conclusion that extreme wind does not significantly improve the predictive performance of the benchmark model in predicting (non-)defaults.

Table 13: Mean accuracy statistics with std. for the benchmark and full model.

Metric	**Benchmark (%)**	**Full (%)**
Accuracy	86.10 ($\pm$0.03)	86.10 ($\pm$0.03)
Precision	9.65 ($\pm$0.06)	9.65 ($\pm$0.05)
Recall	57.96 ($\pm$0.36)	58.01 ($\pm$0.36)
F_1 score	16.54 ($\pm$0.10)	16.55 ($\pm$0.09)
ROC	83.83 ($\pm$0.12)	83.85 ($\pm$0.11)

6 Conclusion

In this study, we investigate trends in extreme wind occurrences and their potential impact on mortgage defaults within the southeast United States. Employing the theoretical framework introduced by Einmahl et al. (2022), we conduct statistical tests across selected weather stations in major cities of the region. These tests assess both the total frequency of extreme wind across cities and the presence of homoscedastic (constant-frequency) extreme wind patterns in a spatial context. Subsequently, we explore the influence of extreme wind on residential mortgage defaults through logistic regression. This analysis incorporates four key benchmark predictors and examines the explanatory capacity of extreme wind, represented by the wind step function. We further evaluate predictive performance by comparing the benchmark model against a full model that includes both benchmark predictors and a step function of extreme wind when applied to a test dataset. Additionally, our investigation encompasses variations in subsets of data and explores non-standard relationships using an XGBoost algorithm. We validate the reliability of our finding with robustness checks.

The test for total frequency of extreme wind across all cities reveals that some cities have significantly higher frequency of extreme wind than other cities. Additionally, two of the 22 considered cities show heteroscedasticity of extreme wind over the years 2000 until 2020. This suggests that for most cities the frequency of extreme wind is constant over time.

In a logistic regression setting, we find that our four benchmark variables (the FICO score, LTV rate, mortgage age and unemployment rate) are key predictors of the PD of residential mortgages. Including extreme wind into the logistic regression yields a statistically significant wind step function, signifying an average 14% increase in PD for a home that has experienced an event of extreme wind, ceteris paribus. The statistical significance of extreme wind holds substantial implications. Since wind damage is covered in most homeowners' insurance, the explanatory power of extreme wind must come from its secondary effects on housing such as flooding. This suggests that the US is ill-prepared for the secondary effects of extreme wind. While this shows that extreme wind has explanatory power of mortgage defaults, its added impact to explaining mortgage defaults in a regression setting is low.

While adding a step function within the benchmark model does not improve upon its

predictive power, doing so in an XGBoost algorithm *does* improve the predictive power of the benchmark model. This shows the potential benefit of using extreme wind in a model predicting mortgage defaults. The practical benefit remains highly dependent on mortgage lenders' willingness to use so-called 'black box models' such as the XGBoost algorithm.

There are several limitations in our research that could form the basis of new research. Firstly, the ASOS is not a very reliable source of meteorological data and can only measure wind speeds up to 125 knots. More severe events of extreme wind have been shown to have a higher impact on housing damage (Rossi, 2020). Using more reliable wind speed measuring systems that can also accurately measure wind speeds higher than 125 knots, allows for including severity of extreme wind as a weighing factor. This could provide more valuable insights into the effect of extreme wind on mortgage defaults.

Secondly, our research encompasses a very large region of the US within which we only consider a small number of cities. Studying a broader range of cities could provide more in-depth knowledge on the effect of extreme wind on mortgage defaults. Increasing the granularity of the considered region also improves understanding of the behaviour of extreme wind events. While hurricanes typically span large areas, tornadoes often have very high but very local impact on housing.

For our mortgage data, a more extensive mortgage dataset could improve the model. Other mortgage characteristics, like the probability of negative equity and borrower income have also been shown to be reliable predictors of mortgage defaults (Jones & Sirmans, 2015). Including additional predictors can fortify the benchmark model performance, leading to a potential improvement in its pseudo R^2. Additionally, an extreme wind variable that attains statistical significance within a better benchmark model reinforces the reliability of the result.

A last extension to this research is to study the areas where extreme wind has significant explanatory power. By mapping out the areas where extreme wind has particularly high explanatory power and comparing this to the local flood maps, an indication can be made on which local areas have suffered increasing risk of flooding by extreme wind. Based on this, flood maps could be improved to meet today's risks of flooding.

References

Altman, E. I., & Saunders, A. (1997). Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, *21*(11-12), 1721–1742.

Ambrose, B. W., & Capone, C. A. (2000). The Hazard Rates of First and Second Defaults. *The Journal of Real Estate Finance and Economics*, *20*, 275–293.

Benfer, E., Robinson, D. B., Butler, S., Edmonds, L., Gilman, S., McKay, K. L., ... Steinkamp, N. (2020). COVID-19 Eviction Crisis: An Estimated 30-40 Million People in America Are at Risk.

Breivik, , Aarnes, O. J., Abdalla, S., Bidlot, J. R., & Janssen, P. A. (2014). Wind and wave extremes over the world oceans from very large ensembles. *Geophysical Research Letters*, *41*(14), 5122–5131.

Chand, S. S., Walsh, K. J., Camargo, S. J., Kossin, J. P., Tory, K. J., Wehner, M. F., ... Murakami, H. (2022). Declining tropical cyclone frequency under global warming. *Nature Climate Change*, *12*(7), 655–661.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Chen, L., Li, D., & Pryor, S. C. (2013). Wind speed trends over China: quantifying the magnitude and assessing causality. *International Journal of Climatology*, *33*(11), 2579–2590.

Coşer, A., Maer-Matei, M., & Albu, C. (2019). Predictive models for loan default risk assessment. *Economic Computation & Economic Cybernetics Studies & Research*, *53*(2).

Einmahl, J. H., de Haan, L., & Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *78*(1), 31–51.

Einmahl, J. H., Ferreira, A., De Haan, L., Neves, C., & Zhou, C. (2022). Spatial dependence and space-time trend in extreme events. *Annals of Statistics*, *50*(1), 30–52.

Episcopos, A., Pericli, A., & Hu, J. (1998). Commercial Mortgage Default: A Comparison

of Logit with Radial Basis Function Networks. *Journal of Real Estate Finance and Economics*, *17*(2), 163–178.

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905.

Ferreira, A., Friederichs, P., de Haan, L., Neves, C., & Schlather, M. (2017). Estimating space-time trend and dependence of heavy rainfall. *arXiv preprint arXiv:1707.04434*.

Fisher, R. A., & Tippett, L. H. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190.

Foote, C. L., Gerardi, K., & Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics*, *64*(2), 234–245.

Gensini, V. A., & Brooks, H. E. (2018). Spatial trends in United States tornado frequency. *NPJ Climate and Atmospheric Science*, *1*(1), 38.

Hendershott, P. H., & Schultz, W. R. (1993, 12). Equity and Nonequity Determinants of FHA Single-Family Mortgage Foreclosures in the 1980s. *Real Estate Economics*, *21*(4), 405–430.

Hosking, J. R., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, *29*(3), 339–349.

IPCC. (2023). Climate Change 2023: Synthesis Report. In *Contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change* (pp. 35–115). Geneva, Switzerland: IPCC.

Jones, T., & Sirmans, G. S. (2015). The underlying determinants of residential mortgage default. *Journal of Real Estate Literature*, *23*(2), 169–205.

Kousky, C., Palim, M., & Pan, Y. (2020). Flood Damage and Mortgage Credit Risk: A Case Study of Hurricane Harvey. *Journal of Housing Research*, *29*(sup1), 86–120.

Lima, M. M., Hurduc, A., Ramos, A. M., & Trigo, R. M. (2021). The Increasing Frequency of Tropical Cyclones in the Northeastern Atlantic Sector. *Frontiers in Earth Science*, *9*, 74–115.

Liu, L., & Hu, F. (2019). Long-term Correlations and Extreme Wind Speed Estimations.

Advances in Atmospheric Sciences, *36*(10), 1121–1128.

NCEI. (2022). *What's an Automated Surface Observing System (ASOS)? — News — National Centers for Environmental Information (NCEI).* Retrieved from `https://www.ncei.noaa.gov/news/whats-automated-surface-observing-system-asos`

Noordewier, T. G., Harrison, D. M., & Ramagopal, K. (2001). Semivariance of property value estimates as a determinant of default risk. *Real Estate Economics*, *29*(1), 127–159.

Oommen, T., Baise, L. G., Vogel, R. M., Baise, G., Vogel, R. M., & Oommen, T. (2011). Sampling Bias and Class Imbalance in Maximum-likelihood Logistic Regression. *Math Geosci*, *43*, 99–120.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, *3*(1), 119–131.

Qi, M., Scheule, H., & Zhang, Y. (2021). Positive Payment Shocks, Liquidity and Refinance Constraints and Default Risk of Home Equity Lines of Credit at End of Draw. *The Journal of Real Estate Finance and Economics*, *62*, 423–454.

Quercia, R., & Stegman, M. (1992). Residential mortgage default: a review of the literature. *Journal of Housing Research*, *3*(2), 341–379.

Rossi, C. V. (2020). Assessing the impact of hurricane frequency and intensity on mortgage default risk. *Journal of Risk Management in Financial Institutions*, *14*(4), 426–442.

Sajjad, M., Lin, N., & Chan, J. C. (2020). Spatial heterogeneities of current and future hurricane flood risk along the U.S. Atlantic and Gulf coasts. *Science of the Total Environment*, *713*, 136704.

Tippett, M. K., Lepore, C., & Cohen, J. E. (2016). More tornadoes in the most extreme U.S. tornado outbreaks. *Science*, *354*(6318), 1419–1423.

Tonn, G., & Czajkowski, J. (2022). US tropical cyclone flood risk: Storm surge versus freshwater. *Risk Analysis*, *42*(12), 2748–2764.

Torres Silva dos Santos, A., & Santos e Silva, C. M. (2013). Seasonality, Interannual Variability, and Linear Tendency of Wind Speeds in the Northeast Brazil from 1986 to 2011. *The Scientific World Journal*, *2013*, 1–10.

Young, I. R., Vinoth, J., Zieger, S., & Babanin, A. V. (2012). Investigation of trends in extreme value wave height and wind speed. *Journal of Geophysical Research: Oceans*,

117(C11).

A Tables and figures

Table 14: Hyperparameter space for XGBoost tuning.

Hyperparameter	**Search space**	**Chosen value**
Learning rate	0.05, 0.1, 0.2, 0.5	0.1
Maximum depth	10, 15, 20, 25, 30	25
No. of estimators	25, 50, 100, 150	50
Subsample ratio	0.7, 0.8, 0.9	0.7
Column sample	0.7, 0.8, 0.9	0.9
Gamma	0, 0.1, 0.2	0.1
Scale pos. weight	1, 2, 3	1

Figure 9: Monthly boxplots for all cities' data grouped together, excluding outliers.

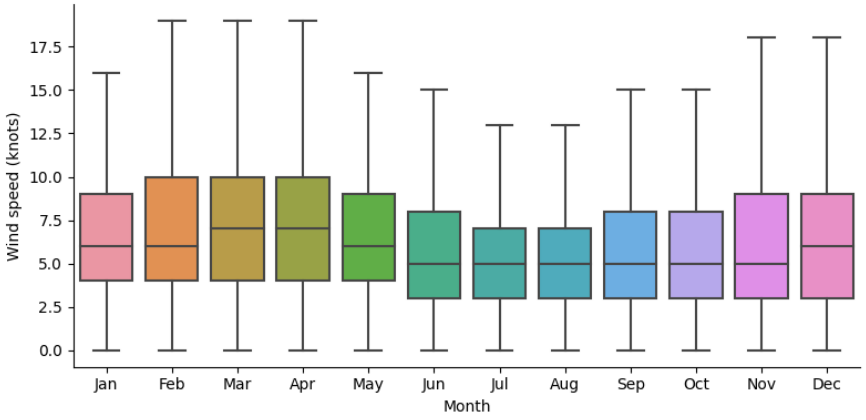
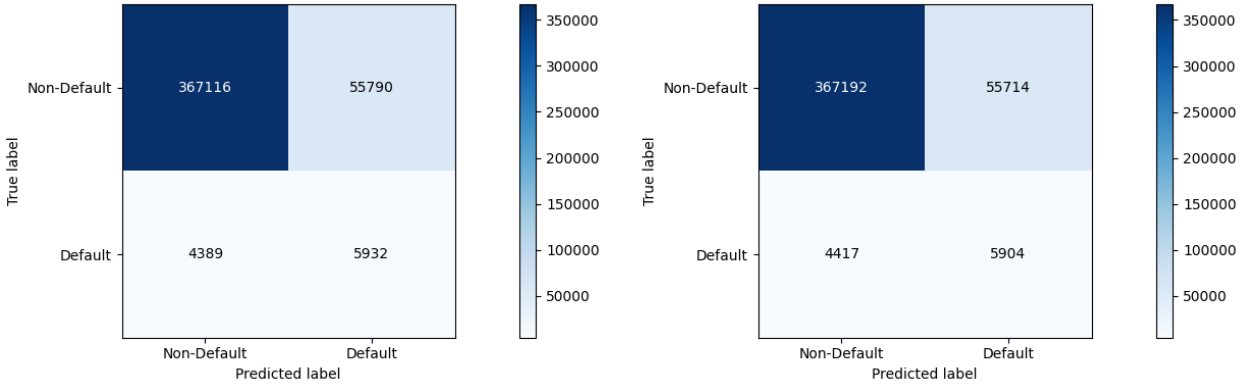
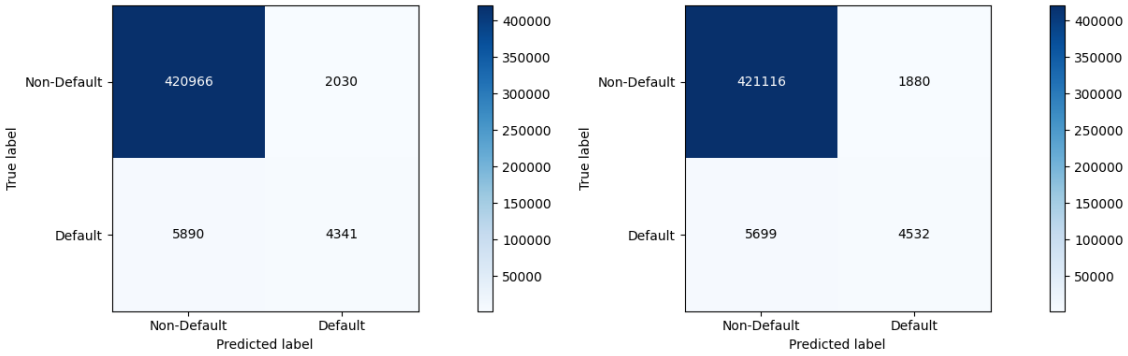


Figure 10: Confusion matrix for the default predictions of the logistic benchmark model (left) and full model (right)



(a) Benchmark model **(b)** Full model

Figure 11: Confusion matrix for the default predictions of the XGBoost benchmark model (left) and full model (right)



(a) Benchmark model **(b)** Complete model

B SMOTE algorithm

In the realm of binary classification, particularly when dealing with imbalanced datasets, the issue of class imbalance can pose substantial challenges. Class imbalance arises when one of the two classes within the dataset is notably underrepresented compared to the other. This imbalance can severely affect the predictive performance of models, often leading to biased outcomes and a diminished ability to accurately predict the minority class. In our research, the minority class is the default class.

Class imbalance can have profound consequences for predictive modelling. When one class significantly outnumbers the other, models tend to exhibit a bias towards the majority class. Consequently, they may struggle to generalise effectively to the minority class, leading to sub-optimal results.

To mitigate the adverse effects of class imbalance, we employ the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is a data augmentation method designed to counteract class imbalance by generating synthetic instances for the minority class. Unlike traditional oversampling methods that merely duplicate existing minority class samples, SMOTE introduces entirely new and artificial observations. These synthetic instances enrich the dataset with valuable information, enabling models to make more informed predictions.

For each minority class observation, SMOTE identifies the k nearest neighbours. Subsequently, for each feature in the dataset, the algorithm selects one of the k neighbours and generates a synthetic observation. This process entails interpolating feature values between the original observation and the chosen neighbour. The interpolation procedure is applied to every feature. This results in the creation of new synthetic instances that closely resemble the minority class. These synthetic instances inject fresh insights into the dataset, thereby enhancing the model's ability to differentiate between classes.

Our parameter choices of k = 3 and a sampling strategy of 0.5 have been determined through a grid search process. This optimisation procedure ensures that SMOTE aligns seamlessly with the objectives of our research and the intricacies of our benchmark model.

In this research, we have selected the number of nearest neighbours and the sampling

strategy through a grid search over both variables. We have opted for a parameter value of $k = 3$, signifying that for each minority class instance, the algorithm selects the three closest neighbours in the feature space. The sampling strategy parameter dictates the ratio of synthetic minority class instances to actual ones. By setting this parameter to 0.5, the SMOTE algorithm creates 5 synthetic observations for every 10 real observations in the default class. This does not completely balance out the dataset, but does improve model performance. For instance, using the SMOTE algorithm increases the F_1 score of the benchmark model from 9.66% to 16.47%.