Adaptive Runtime Fairness Monitoring for Credit Scoring During Economic Fluctuations

Eigard Alstad[1]
Supervisor(s): Anna Lukina[1]
[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Final project course: CSE3000 Research Project
Thesis committee: Anna Lukina, Christoph Lofi

Abstract

The need for fair automated decision making is increasing as algorithms continue to have a growing impact on humans. Runtime fairness monitors are algorithms that detect fairness violations of fairness constraints as an algorithm is being run on real-world data, through computing statistics over observed data instances. These monitors however often assume that the input data comes from a static distribution.

Our research aims to build a runtime fairness monitor for credit scoring algorithms that tracks distribution shifts in the underlying economic state to detect fairness violations more quickly. To accomplish this, we have designed a synthetic credit scoring dataset that simulates economic fluctuations. Our main contribution is the development of an adaptive fairness monitoring algorithm that dynamically adjusts to economic fluctuations, through a sliding window that disregards older, less representative samples when fluctuations in the underlying distribution our detected.

We have tested our adaptive monitor against a baseline monitor that simply computes fairness metrics over all observed data instances. Our preliminary results show that our adaptive approach improves the speed of detecting fairness violations compared to the more traditional monitoring method. In conclusion, there is potential for incorporating economic state detection into credit score fairness monitoring. Future work should validate these findings with real-world data and explore additional fairness metrics.

1 Introduction

Fairness in automated decision-making has become a crucial area of research as the influence of algorithms on humans continues to grow. Automated systems are now commonly implemented in various domains such as finance, healthcare and energy [1]. These systems make decisions that significantly affect individuals. They rely on historical data which can propagate biases and discrimination if not monitored and managed to be fair. There are various fairness definitions and metrics, for the purpose of this research we focus on demographic parity, which states that ... .

Addressing biases in credit scoring is particularly important. Credit scores are a measure of how reliable individuals are in paying back loans on time, and are important in determining individual's access to financial services. Unfair credit scoring practices can amplify socio-economic disparities and have, in the past, disproportionately affected minority groups [17]. This issue is further complicated by economic fluctuations, which are periods of economic growth, or booms, and economic decline, or recessions. Economic fluctuations can alter the distribution of data that the credit scoring algorithms are applied to. The impact of a period of extreme economic fluctuation, the 2008 financial crisis, on credit prediction is explored in [12]. There is a two-way relationship between inequality in credit scoring processes and the frequency and volatility of economic fluctuations, where an increase in one can lead to an increase in the other [14] [18]. This relationship motivates incorporating detection of economic fluctuations into strategies for ensuring credit scoring practices are fair.

Different approaches to ensuring domain-agnostic algorithmic fairness have been proposed. Tools such as FairSquare and VeriFair utilize techniques such as static analysis and random sampling to provide probabilistic fairness guarantees about programs, but they do not consider the actual data being applied in real-time, which is crucial in domains with significant human impacts such as credit scoring [2] [5]. Runtime fairness monitors are a more hands-on approach that check for fairness specification violations as data is processed in real-time [3]. These runtime monitors assume a static data distribution, and [3] introduces the general problem of handling a data input stream with non-static distributions. For the scope of our research we narrow this general problem down to the credit scoring domain and non-static nature of an economy.

Our research tackles the questions of whether we can track the underlying economic state through a credit scoring dataset, and whether this information can be incorporated to monitor fairness of credit scoring at run-time.

The main algorithmic challenge lies in detecting and adapting to changes in the data distribution without prior knowledge of when these changes occur. Our main contribution is an adaptive fairness monitoring algorithm that utilizes a sliding window mechanism to detect shifts in the economic state through income distribution and adjust fairness checks accordingly. This approach ensures that the algorithm remains responsive to non-static data distributions, which is critical for maintaining fairness in dynamic environments.

We evaluate the new monitoring algorithm on a credit scoring model trained on a synthetic dataset representing a stable economy. We then compare our algorithm to a baseline fairness monitor when applied on a larger synthetic dataset representing economic fluctuations. Our primary results indicate that the adaptive algorithm detects fairness violations faster than the baseline approach, demonstrating its effectiveness in dynamic economic scenarios.

The structure of the paper is as follows: Section 2 provides an overview of related work in credit scoring, algorithmic fairness and runtime monitoring. Sections 3 and 4 describes the problem context in more detail. Section 5 details our methodology for generating our synthetic datasets and our main contribution of the adaptive fairness monitoring algorithm. Section 6 presents the experimental results, comparing the adaptive algorithm to the baseline. Section 6 discusses the limitations of our approach and potential directions for future research. Section 8 highlights our contributions towards responsible research and section 9 provides a conclusion of the paper and avenues for further work.

2 Related Work

There is extensive research on identifying and mitigating discrimination of algorithmic credit scoring models. A recent overview is provided by [10], which addresses research settings, solutions and open challenges. The impact of economic fluctuations, particularly the 2008 financial crisis, on credit prediction is explored in [12], providing valuable insight into the connection between algorithmic decision-making and economic conditions. This work provides insight into how macroeconomic changes can influence algorithmic credit decisions, highlighting the need for adaptive fairness monitoring. Further contributions to ensuring that automated credit scoring systems promote fairness and equity include [11], [13] and [20]. Furthermore, the effect of long term fairness monitoring itself on the underlying distribution of credit scoring is explored in [15] and [7].

Algorithmic fairness research, independent of any particular domain, is broadening as the influence of automated decision-making on human lives grows. Tools such as FairSquare and VeriFair have contributed significantly to this field by utilizing static analysis and random sampling techniques to provide probabilistic fairness guarantees. These tools have contributed by using advanced computational methods for automated fairness verification, and achieving scalability to large models.

An alternative technique for verifying the fairness of an algorithm is run-time fairness monitors. A framework proposed by [3] allows developers to specify fairness expectations alongside their algorithms, and reports any violations of these constraints. These monitors, similarly to FairSquare and VeriFair, simplify the process of checking fairness and prevent developers from needing to write their own code for this purpose. The advantage over tools such as FairSquare is that runtime monitors track algorithms in real-world conditions, which is useful when the real-world data distribution is different from the training dataset or anticipated data input distributions.

3 Hypothetical Lending Scenario

To illustrate the ideas and techniques presented in this paper, a fictional bank known as FairBank will be used as a running example. FairBank is experimenting with different automated credit scoring algorithms for approving or denying loan applications. Table 1 describes the data collected by FairBank on their customers, that they intend to process for deciding credit scores and loan approvals/rejections for their customers.

Attribute	**Type**	**Description**
Income	Numerical	The applicant's annual income.
Employment Status	Boolean	Indicator of applicant's current employment status.
Minority Status	Boolean	Indicator representing whether the applicant belongs to an arbitrary minority group.
Loan Approval	Boolean	Indicator representing the decision to approve or deny the loan application.

Table 1: Dataset Attributes

Formally, Fairbank are in possession of a dataset D of loan applicants characterized by the attributes in Figure 1. Fairbank wants to monitor that their credit scoring algorithms do not discriminate against minority groups, and hence want to monitor demographic parity. Let

- X_i represents the set of features for applicant i.

- S_i represents the sensitive attribute for applicant i (e.g., minority status).
- Y_i is the loan approval decision for applicant i (Y/N).

Demographic parity requires that the probability of a loan approval should be similar across different demographic groups, formally defined as

$$P(Y = \text{Yes} \mid S = \text{Minority}) \approx P(Y = \text{Yes} \mid S = \text{Non-Minority})$$

For our purposes it is not important what the minority group is, we consider it an arbitrary classification for the sake of analysing discrimination. It is against the law in the EU for financial institutions such as Fairbank to be discriminatory in practices such as credit scoring [9]. For this reason it is crucial to detect fairness violations both reliably and quickly.

The research question's with regard to this context are:

1. Can fluctuations in economic state be detected through a credit scoring dataset?
2. How can incorporating economic fluctuations into demographic parity monitoring algorithms for credit scoring models affect the speed of fairness violation detection?

As will be explored further in the coming sections, Fairbank are monitoring three credit scoring models, each with different objectives. These objectives are maximizing profit, minimizing risk, and maximizing profit while upholding demographic parity. The models are trained on a credit scoring dataset corresponding to a stable economy, which may not always represent the distribution of data the models are applied to.

4 Preliminaries

Before describing the methodology and results of our research, this section will take the opportunity to define in more detail some central concepts.

Fairness Metrics

Fairness in automated decision-making can be quantified in different ways, with different metrics representing different mathematical formulations and context. Common metrics include Demographic Parity, Equal Opportunity, and Equal Odds. Demographic parity is chosen for this study, it requires that the approval rate for each class of the sensitive attribute is equal. This metric is relevant in credit scoring to ensure that all demographics receive the same opportunities and access to financial products.

More detail on economic fluctuations

Economic recessions and booms can impact consumer behavior and financial stability, and in turn affect the quality and reliability of credit scoring data [8]. During fluctuations, traditional models may be prone to makinng potentially unfair and biased credit decisions. Our research considers the incorportation of economic indicators such as employment rate to better understand and adjust to these fluctuations. By integrating these indicators, the proposed fairness monitoring algorithm can dynamically recalibrate fairness assessments in response to economic downturns or booms.

Sliding Window Protocols

In the context of fairness monitoring, sliding window protocols are used to focus on the most recent data, acknowledging that older data may not reflect the current underlying distribution and current economic conditions. This approach involves continuously updating the decisions used to calculate fairness metrics by introducing a window that moves over time in order to include only the newer data entries. This helps mitigate the effect of outdated information on current monitoring.

5 Methodology

This section describes the setup for our experiments. First we will motivate and explain the synthetic data generation process, including design choices, features selected, and simulation of economic conditions. Following that, we will detail our main contribution, the adaptive fairness monitoring algorithm.

5.1 Synthetic Credit Scoring Dataset

To evaluate the effectiveness of the proposed adaptive fairness monitoring algorithm, we generated a synthetic credit scoring dataset. This synthetic dataset allows us to control various factors and simulate different economic conditions, providing a robust and controlled testing ground for our experiments.

We chose to create synthetic data due to privacy concerns associated with real-world credit scoring data, which can contain sensitive information such as income and employment history. Moreover, obtaining real-world data that reflects various economic conditions over time can be difficult, as it requires extensive data collection over long durations. By generating a synthetic dataset, we ensure that our research does not compromise individual privacy but still allows us to test the adaptive fairness monitoring algorithm under diverse economic scenarios.

The dataset is designed to reflect and mimic real-world credit scoring scenarios, including important features that influence loan approval decisions. The dataset includes the following features for each applicant:

- **Income**: The applicant's annual income.
- **Employment Status**: Boolean indicator of applicant's current employment status.
- **Minority Status**: Boolean indicator representing whether the applicant belongs to a minority group.
- **Loan Approval**: Boolean indicator representing the decision to approve or deny the loan application.

The selection of the features is based on practical relevance and previous literature. Income and employment status are crucial indicators of financial stability and ability to repay loans and are used as features in similar studies [11] [15]. The inclusion of minority status allows us to assess demographic parity and detect potential bias against minority groups. Finally, the loan approval decision is the outcome variable that the fairness of the credit scoring algorithm is evaluated against. While real-life systems will also consider features such as education and employment sector, we limit ourselves to four features in order to avoid over-complicating the experiment and monitoring model.

Economic fluctuations are simulated using time-series data to model periods of economic recession, boom, and stability. This simulation affects features in the following ways:

- **Recession**: During a recession, we expect:
  - Lower average income across the entire population [6].
  - Higher unemployment rates, disproportionately affecting different groups [16].
- **Boom**: During a boom, we expect:
  - Higher average income [6].
  - Lower unemployment rates [16].
- **Stability**: In stable periods, economic indicators remain relatively constant, and provide a baseline for comparison.

Next, we delve into the numerical details of the data generation process for each economic state. We strive to generate realistic values and statistics in order to improve the reliability of our research. We utilise a log-normal distribution for generating the income data as it reflects the positive skewness that is usually observed in real-world income data, reflecting that a larger proportion of the population earns lower incomes while a smaller proportion earn high incomes. To represent the income distribution for the standard economic state, we take the median and standard deviation for income in the United States in 2022 in order to compute the two parameters for the log-normal distribution [19]. To simulate recessions, the median incomes are adjusted downwards. In reality, these adjustments disproportionately affect lower-income and minority groups, which we reflect in the adjustments to our medians [4].

We utilise a Bernoulli distribution for generating the binary variables of employment status, minority status and loan approval as it allows us represent states with two-outcomes and incorporate real-world probabilities. The unemployment rate is typically a lagging indicator for economic fluctuations, meaning we have incorporated a delay before the employment rates fall during a recession and before they rise during a boom [6].

The final dataset we generate represents a stable economy followed by a recession and finally a boom, in order to explore diverse transitions. The different economic states are interpolated to simulate linear drifts. Our dataset consists of 1900 data points and represents a time-series of instances for our credit scoring model to act upon. The data was generated in Python, Pandas and NumPy were used for creating and handling data structures, while SciPy was utilised for statistical distributions and random sampling.

5.2 Adaptive Fairness Monitoring Algorithm

The main contribution of this paper is the adaptive fairness monitoring algorithm described in Algorithm 1. The algorithm builds upon the monitoring algorithm proposed by [3], which is the baseline we will compare to in the next section. The new contributions to this algorithm are the use of a sliding window and monitoring the income distribution to detect significant changes in the economic state. As mentioned earlier, the (un)employment rate is a lagging indicator of economic fluctuations, we therefore chose to only use the income feature to detect fluctuations. The data parameter consists of instances of the credit scoring dataset, including the decision to grant/deny a loan.

The motivation behind including a sliding window in the algorithm is that the credit scoring instances will not be arising from a static distribution. A sliding window allows us to disregard older instances that are less representative of the current distribution. The algorithm iterates through the decisions made by the model, maintaining a sliding window of recent decisions. We utilise a counter to ensure that we compare input distributions periodically instead of after every new instance. We use the Kolmogorov-Smirnov test on line 16 to compare the two sample distributions. If we find that the distributions are significantly different then we reset the sliding window to only the newest instance. At line 27 we begin to check that the demographic parity specification is satisfied dependent on there being enough data in the current window.

Algorithm 1 Adaptive Fairness Monitoring with Sliding Window for Credit Scoring Model

```
1: procedure MONITOR_FAIRNESS(data, max_window_size, min_window_size, fairness_ratio, check_interval)
2:     sliding_window ← empty list
3:     previous_income_dist ← None
4:     counter ← 0
5:     for i ← 0 to length(data) − 1 do
6:         Add data[i] to sliding_window
7:         if length(sliding_window) > max_window_size then
8:             Remove oldest sample from sliding_window
9:         end if
10:        counter ← counter + 1
11:        if counter ≥ check_interval then
12:            counter ← 0
13:            if length(sliding_window) ≥ min_window_size then
14:                current_income_dist ← income from sliding_window
15:                if previous_income_dist ≠ None then
16:                    (ks_stat, p_value) ← KS test on current_income_dist, previous_income_dist
17:                    if p_value < 0.05 then
18:                        print "Income distribution change detected at index", i
19:                        Reset sliding_window to contain only data[i]
20:                        previous_income_dist ← current_income_dist
21:                        continue
22:                    end if
23:                end if
24:                previous_income_dist ← current_income_dist
25:            end if
26:        end if
27:        if length(sliding_window) ≥ min_window_size then
28:            Calculate approval rates for majority and minority in sliding_window
29:            if minority approval rate / majority approval rate < fairness_ratio then
30:                print "Fairness violation at index", i
31:            else
32:                print "No fairness violation at index", i
33:            end if
34:        end if
35:    end for
36: end procedure
```

6 Experimental Results

In this section, we present the experimental results gathered from testing our adaptive fairness monitoring algorithm. We compare it to the baseline fairness monitoring algorithm, which in this case is the same as Algorithm 1 but without the sliding window or income distribution comparison. We compare the two algorithms in terms of when the violations of demographic parity are reported. We utilise the synthetic dataset and adaptive algorithm described in the previous section, with a fairness ratio of 0.8, meaning that the probability of approval for a minority must be at-least 80% that of the majority class.

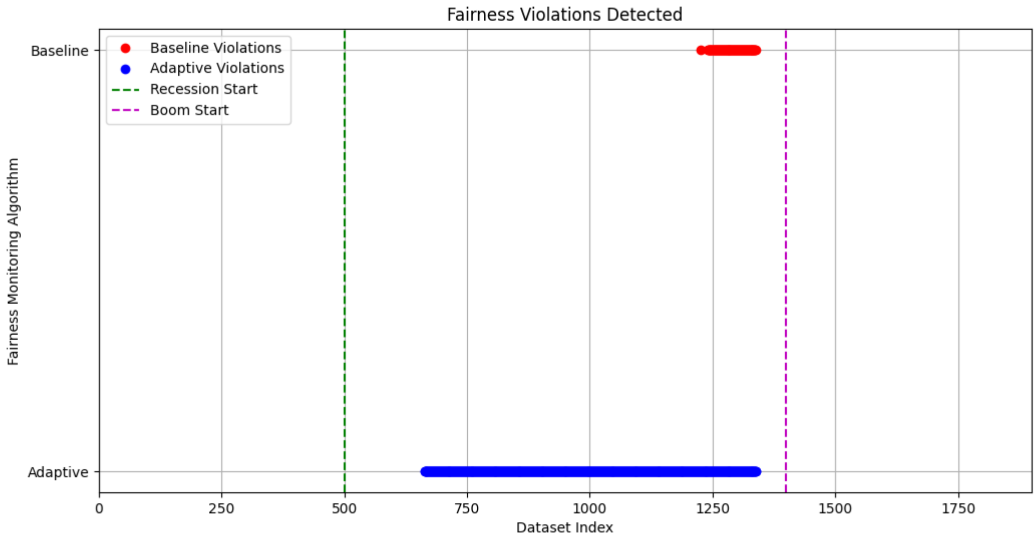

Figure 1: Fairness Violations throughout the Synthetic Dataset for the Adaptive and Baseline Fairness Monitoring Algorithms

The results are displayed in Figure 1. For the case of our synthetic dataset, it seems as though the adaptive algorithm detected the unfairness of the credit scoring model earlier by approximately 500 data instances. The vertical dashed lines in Figure 1 indicate the start of the recession and boom periods, with the first dashed line representing the start of the recession. The data was generated in a way where the recession and boom affected the minority group disproportionately in comparison with the majority group, it is therefore logical that the credit scoring model began to violate the demographic parity constraint during the simulated recession.

Referring back to the research questions, we have shown that for our synthetic dataset, detecting the economic state through the credit scoring dataset is indeed possible and did lead to detecting fairness violations faster. There is, however, a number of limitations to our experiments that must be acknowledged.

The synthetic nature of the dataset, while allowing us to control various factors and simulate different conditions, does likely not capture the real-life complexities of real-world economic data. Real-world datasets likely contain more complex relationships and are both messier and more noisy, which would affect the performance of the adaptive algorithm. The reliability of the study would have been enhanced by applying the monitoring algorithms on multiple datasets rather than only one.

The parameters used to generate the synthetic dataset, such as income distributions and employment probabilities, were based on simplified assumptions. Although we tried to inform these choices by real-life data, they may not accurately represent the true distribution of real life economic data. This simplification could lead to an overestimation of the algorithm's effectiveness in our synthetic environment compared to a complex real-world scenario.

Our experimental setup has also assumed a clear distinction between three distinct economic states (stable, recession, boom). In reality, economic transitions are likely to be more complex. This assumption likely limits the generalizability of our findings, as the adaptive algorithm might not perform as well if the economic changes are more nuanced and less detectable by our statistical tests.

7 Discussion

The results indicate that our algorithm could help detect fairness violations more quickly. This early detection has significant implications in real-world scenarios. For financial institutions, quickly identifying unfairness in credit scoring models can prevent discriminatory practices, improve compliance with laws and regulations, and increase relationships with stakeholders. Our algorithm has shown potential to help provide more equitable outcomes for bank loan applicants, particularly those from minority groups.

The scope of our research was limited in several ways. Firstly, we focused on a single fairness metric, while the field of fairness in automated decision-making is vast, with various other metrics that could also be considered and evaluated against. Additionally, our results do not consider the computational complexity of the monitoring algorithms. The sliding window mechanism and periodic distribution comparisons will result in higher overall computational costs. We consider this out-of-scope for our research, and it was not part of our research questions, yet it is an important factor to consider before applying any of these techniques in real-world systems, where resources may be limited.

The robustness of our results could have been improved through experimenting with generating multiple synthetic datasets and testing against different parameter values for maximum window size and the other adaptive algorithm parameters. This would provide a more comprehensive understanding of how these parameters affect the performance of the algorithm and a more confident conclusion on the algorithms ability to quickly detect fairness violations.

Our results fit into the broader context of existing research through highlighting that it is possible to incorporate economic state detection into fairness monitoring for credit scoring models. We have also shown an indication towards this enhancing the ability of the fairness monitor to detect fairness violations more quickly, but were not able to show this reliably within the scope of our project. This contribution both advances theoretical understanding and has practical implications for improving equity and trust in automated systems.

8 Responsible Research

8.1 Ethical Considerations

Algorithmic bias refers to systematic error in processing data that can lead to unfair outcomes. This is especially prevalent when historical data, containing discriminatory patterns, is used for training models. This research addresses the problem of bias in automated credit scoring by proposing a novel adaptive fairness monitoring algorithm. Our approach dynamically adjusts to economic conditions, ensuring that our model does not perpetuate existing inequalities but rather works towards mitigating them. This proactive approach to bias mitigation is crucial for maintaining fairness in automated financial decisions that impact people's lives.

Transparency in algorithms is also critical for trust and accountability. Our research contributes to this by detailing the techniques and criteria used by the proposed fairness monitoring algorithm. By publishing our workings, we allow for external validation and scrutiny. This transparency can also help stakeholders in understanding evaluating the decisions made by our fairness monitoring algorithm.

The embrace of open science principles allows other researchers to replicate our findings, or even extend our work. The accompanying codebase is documented, ensuring both transparency and reproducability. Our code also contains seeds to ensure results are reproducable.

8.2 Data Privacy

To avoid potential misuse of sensitive personal data, this study uses simulated datasets that mirror real-world scenarios without the individual privacy risks of real data. These datasets are designed to reflect realistic economic conditions and demographic distributions, in order to ensure the findings are both relevant and ethical.

9 Conclusions and Future Work

This research aimed to investigate whether economic state detection could be incorporated into fairness monitoring for credit scoring models. Our primary findings have indicated that this is indeed feasible and in fact enhances the speed at which fairness monitors can detect violations. While our findings are limited by the use of synthetic data, they do provide a proof of concept.

The key contribution of this work is the development of an adaptive fairness monitoring algorithm for credit score models. Our algorithm uses a sliding window mechanism together with periodic comparisons of the underlying economic state. The algorithm detects economic fluctuations and uses this knowledge to adjust a sliding window to focus on newer, more representative samples.

Despite the promising results, there are several open issues and areas for improvement. Future research should explore additional fairness metrics in order to provide a more comprehensive assessment of fairness in credit scoring models. Incorporating a wider range of metrics can help indicate the robustness of the fairness monitor.

Moreover, the computational complexity of our monitoring algorithms needs to be addressed. It has been developed without regard to computational costs. Future work should focus on optimizing the use of the sliding window mechanism and distribution comparisons in order to ensure the techniques can be applied in real-world systems.

Furthermore, our research primarily focused on credit scoring models. Extending the techniques to other domains, such as hiring processes or healthcare, could provide insights and show whether they are more broadly applicable. The techniques are not limited to economic fluctuations, but could be applied anywhere where the underlying data distribution is not static.

In summary, while this study has made contributions to fairness monitoring in credit scoring, addressing the limitations and expanding the scope of the research can enhance the robustness and applicability of these findings to more practical scenarios. Future research should continue to explore and improve these techniques, ensuring that fairness in automated decision-making remains a priority in all domains.

References

[1] Bilal Al-Khateeb. Intersection of artificial intelligence, machine learning, and big data in transforming healthcare management, financial services, and sustainable energy solutions. *Emerging Trends in Machine Intelligence and Big Data*, 15(11):12–19, 2023.

[2] Aws Albarghouthi, Loris D'antoni, Samuel Drews, and Aditya V. Nori. Fairsquare: probabilistic verification of program fairness. *Proceedings of the ACM on Programming Languages*, 1:1 – 30, 2017.

[3] Aws Albarghouthi and Samuel Vinitsky. Fairness-aware programming. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 211â219, New York, NY, USA, 2019. Association for Computing Machinery.

[4] R. Scott Baird. Economic growth slows in 2023: Implications and policy responses. `https://www.everycrsreport.com/reports/R44705.html`, 2023. Accessed: 2024-06-23.

[5] Osbert Bastani, Xin Zhang, and Armando Solar-Lezama. Probabilistic verification of fairness properties via concentration, 2019.

[6] Corporate Finance Institute. Recession. `https://corporatefinanceinstitute.com/resources/economics/recession/`, 2024. Accessed: 2024-06-15.

[7] Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, D. Sculley, and Yoni Halpern. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 525â534, New York, NY, USA, 2020. Association for Computing Machinery.

[8] Cristian deRitis. Economically conditioned credit scores: An example from auto loans. Analysis report, Moody's Analytics, 2024. Available at Moody's Analytics [Accessed on 16.05.2024].

[9] European Union Agency for Fundamental Rights. Article 21 - non-discrimination. `https://fra.europa.eu/en/eu-charter/article/21-non-discrimination`, 2024. Accessed: 2024-06-15.

[10] Ana Cristina Bicharra Garcia, Marcio Gomes Pinto Garcia, and Roberto Rigobon. Algorithmic discrimination in the credit domain: what do we know about it? *AI & SOCIETY*, 2023.

[11] Christophe Hurlin, Christophe PÃ©rignon, and SÃ©bastien Saurin. The fairness of credit scoring models, 2024.

[12] Amir E. Khandani, Adlar J. Kim, and Andrew W. Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking Finance*, 34(11):2767–2787, 2010.

[13] Eric L. Lee, Jing-Kai Lou, Wei-Ming Chen, Yen-Chi Chen, Shou-De Lin, Yen-Sheng Chiang, and Kuan-Ta Chen. Fairness-aware loan recommendation for microfinance services. In *Proceedings of the 2014 International Conference on Social Computing*, SocialCom '14, page 1â4, New York, NY, USA, 2014. Association for Computing Machinery.

[14] Michelle Seng Ah Lee and Luciano Floridi. Algorithmic fairness in mortgage lending: from absolute conditions to relational trade-offs. *Minds and Machines*, 31(1):165–191, 2021.

[15] Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. *CoRR*, abs/1803.04383, 2018.

[16] Niall O'Higgins. Who is hit hardest during a financial crisis? the vulnerability of young men and women to unemployment in an economic downturn. https://www.iza.org/publications/dp/4359/who-is-hit-hardest-during-a-financial-crisis-the-vulnerability-of-young-men-and-women-to-unempl 2009. Accessed: 2024-06-23.

[17] Lisa Rice and Deidre Swesnik. Discriminatory effects of credit scoring on communities of color. *Suffolk UL Rev.*, 46:935, 2013.

[18] Joseph E. Stiglitz. Macroeconomic fluctuations, inequality, and human development. In *Macroeconomics and Human Development*, page 28. Routledge, 1st edition, 2013.

[19] United States Census Bureau. Income in the united states: 2023. https://www.census.gov/library/publications/2023/demo/p60-279.html, 2023. Accessed: 2024-06-15.

[20] Mark Weber, Mikhail Yurochkin, Sherif Botros, and Vanio Markov. Black loans matter: Distributionally robust fairness for fighting subgroup discrimination, 2020.