The Cohort Shapley value to measure fairness in financing small and medium enterprises in the UK

Citation for published version:
Lu, X & Calabrese, R 2023, 'The Cohort Shapley value to measure fairness in financing small and medium enterprises in the UK', *Finance Research Letters*, vol. 58, 104542. https://doi.org/10.1016/j.frl.2023.104542

Digital Object Identifier (DOI):
10.1016/j.frl.2023.104542

Journal Pre-proof

The Cohort Shapley value to measure fairness in financing small and medium enterprises in the UK

Xuefei Lu, Raffaella Calabrese

PII: S1544-6123(23)00914-5
DOI: https://doi.org/10.1016/j.frl.2023.104542
Reference: FRL 104542

To appear in: *Finance Research Letters*

Received date : 31 May 2023
Revised date : 29 August 2023
Accepted date : 30 September 2023

Please cite this article as: X. Lu and R. Calabrese, The Cohort Shapley value to measure fairness in financing small and medium enterprises in the UK. *Finance Research Letters* (2023), doi: https://doi.org/10.1016/j.frl.2023.104542.

The Cohort Shapley Value to Measure Fairness in Financing Small and Medium Enterprises in the UK

Abstract

Banks are relying on machine learning techniques to support their decisions in financing small and medium enterprises (SMEs). As regulators require that credit decisions are transparent, there is a need to develop methods to measure fairness. We propose a weighted average of the Cohort Shapley value, which removes impossible feature combinations, and a relative fairness score for assessing the fairness level within sub-populations. Based on our knowledge, this is the first paper that investigates the fairness of UK financial institutions in providing funding to SMEs. Our findings reveal discrimination against start-up, micro, women-led companies, and owners of Asian ethnic backgrounds.

Keywords: Explainable AI, Shapley value, fairness, small and medium enterprises

1. Introduction

To protect consumers, credit decisions taken by financial institutions need to be transparent (Bracke et al., 2019). For example, if a credit application has been rejected, the applicant has the right to ask for explanations about the credit decision. Machine learning (ML) techniques, known for their black-box nature, are increasingly being utilized by banks for credit scoring due to their high accuracy. This has led to a growing body of literature focusing on the need for explainability (Bussmann et al., 2021) and fairness (Kozodoi et al., 2022) in credit assessment processes. As banks are adopting ML techniques for their high accuracy, and the complexity of the ML techniques makes them black-box models, there is an extended literature on explainability (Bussmann et al., 2021; Bueff et al., 2022) and fairness (Kozodoi et al., 2022) for credit scoring models. These previous studies analyze accepted credit applications and assess the fairness of credit default.

As most of the literature studies the customer classification based on default records, however, to the best of our knowledge, no analysis has previously focused on fairness for access to finance for SMEs. This gap in the literature is mainly due to the lack of data availability on the outcomes of SMEs' credit applications. To fill this gap, we analyze the SME Finance Monitor survey, the largest survey in the UK on finance for SMEs. This survey provides data not only on accepted applicants but also on those that have been rejected.

Explainable Artificial Intelligence (XAI) methods have been developed to open up black-box models and provide insights into the mechanisms of ML algorithms.

Within XAI, the study of feature importance has garnered significant attention. A feature is deemed important either because changing it has a causal effect, or because changing it leads to a significant change in the predictions, or because leaving it out of a model reduces the model's prediction accuracy. Shapley value is one of the most widely used techniques for explaining feature importance in individual predictions. Shapley value was initially introduced in game theory to define a reward allocation to a team and the importance of each player in the team Shapley (1953). In the XAI field, Štrumbelj and Kononenko (2014) and Lundberg et al. (2017) adapt the concept of Shapley value, known as SHapley Additive exPlanations (SHAP), to explain the importance of features in individual predictions, where the game players are considered as the features, and the rewards are defined as the expected predictions. However, a limitation of the classical Shapley value-based explanation methods is that its calculation involves feature value combinations that may be infeasible, known as the extrapolation issue (Rudin, 2019; Kumar et al., 2020; Hooker et al., 2021; Mase et al., 2022).

In this analysis, we consider the Cohort Shapley value proposed by Mase et al. (2019, 2022) to address this issue. As the survey data are weighted to make the observations representative of the population, we further suggest a Weighted average of the Cohort Shapley (WCS) value. Furthermore, we propose a Relative Fairness Score (RFS) as a function of the WCS values. We apply our proposals to data collected by the SME Finance Monitor survey, which represents the largest survey on SME finance in the UK. We consider $1,102$ SMEs that sought a bank loan in the last 12 months from the second quarter 2018 to the third quarter 2020. In line with the literature on access to finance for SMEs, we focus on some firms' characteristics, such as owner's gender and ethnicity, number of employees, age, and turnover (Calabrese et al., 2022). We emphasize that previous studies on discrimination in accessing external finance for SMEs are mainly based on regression approaches that do not allow to assess the fairness of the credit decisions.

Coherently with previous results (Hewa-Wellalage et al., 2022), we obtain that loan applications from women-led companies face a mild level of disadvantage. Micro companies and start-ups are usually considered unreliable debtors for their lack of funding resources and company history (Department for Business, Innovation and Skills, 2014). For these reasons, micro and start-up companies face significant disadvantages in securing external funding. There is an extended literature on the disadvantages faced by companies whose owners are from ethnic backgrounds (Cowling et al., 2023; Fraser, 2009; Smallbone et al., 2003). We obtain that owners with a Asian background are significantly discriminated against, while the Black ethnic group is highly advantaged in the credit decision process.

The paper is organized as follows. Section 2 illustrates the methodology on the Cohort Shapley value. Section 3 describes SME data set, and Section 4 provides the main empirical results. Finally, Section 5 presents the conclusions.

2. Methodology

2.1. The Cohort Shapley value

Consider a data set of n SMEs with each observation characterized by d SME features. Each observation is denoted as $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})$, $i = 1 \ldots n$. The outcome of the SME credit application is denoted as y_i, where $y_i = 1$ if the loan is granted, and 0 otherwise. In practice, a financial institution often uses a ML model $f(\cdot)$ to provide predictions $\hat{y}_i$ based on the values $\mathbf{x}_i$, i.e., $\hat{y}_i = f(\mathbf{x}_i)$. $f(\cdot)$ is often a black-box model, which can be challenging to interpret. As financial institutions need to be able to explain their credit decision outcome to the applicants, XAI methods such as Shapley value have been widely used to explain the default importance for the credit application outcome $\hat{y}_i$.

Specifically, the definition of Shapley value assumes that d features produce a value $val(1:d)$. For a subset u of d features, denoted as $u \subseteq \{1:d\}$, the produced value is $val(u)$. To explain the prediction $f(\mathbf{x}_t)$ of a target observation $\mathbf{x}_t$, $t \in \{1...n\}$, we consider the Shapley value $\phi_{t,j}$ is the marginal contribution of features j, defined as:

$$\phi_{t,j} = \frac{1}{d} \sum_{u \subseteq \{1:d\} \setminus j} \binom{d-1}{|u|}^{-1} \left(val_t(u+j) - val_t(u)\right), j \in \{1:d\}, \tag{1}$$

where $|\cdot|$ is the cardinal of a set, and $val(\emptyset) = 0$ by convention.

Traditional explainable methods, such as SHAP, use the value function

$$val_t(u) = \mathbb{E}(f(\cdot)|\mathbf{x}_{tu}) - \mathbb{E}(f(\cdot)), \tag{2}$$

where $\mathbf{x}_{tu}$ is a sub-vector of $\mathbf{x}_t$ contains only feature values in u, and $\mathbb{E}(f(\cdot)|\mathbf{x}_{tu})$ is a conditional expectation function of ML model f with features in u fixed at values of the target observation. The calculation of $val_t(u)$ in equation (2) requires the evaluation of f on data points that might do not exist in the given data set, which can result in using combinations of feature values that are unlikely or even infeasible in reality, known as the extrapolation issue (Kumar et al., 2020).

The Cohort Shapley method proposed by Mase et al. (2019, 2022) circumvents the need for accessing $f(\cdot)$ on infeasible feature combinations by defining cohort sets of the target observation, where the cohort set includes observations that are deemed 'similar' to the target observation based on a subset of predictors. The notion of similarity is established using target-specific similarity functions z_{tj}, where $z_{tj}(x_{ij}) = 1$ indicates that observation i is similar to target x_{tj} based on feature j, whereas dissimilarity is represented by a value of 0. For categorical features, a straightforward approach to identifying observations similar to the target with respect to this feature is by employing an identity function. For instance, two observations can be deemed similar in terms of the company leader's gender if they are both women-led. On the other hand, for real-valued features, a distance-based similarity function can be employed.

Then for each target observation $\mathbf{x}_t$, the cohort set for a subset u of features is defined as

$$C_{t,u} = \{i \in \{1:n\} | z_{tj}(x_{ij}) = 1, \text{for all } j \in u\},$$

with $C_{t,\emptyset} = \{1:n\}$ by convention. $C_{t,u}$ is the cohort set of observations that are similar to the target for all features in u but not necessarily features in $\{1:d\} \setminus u$. Given the cohort set, the cohort averages are defined as:

$$\bar{y}_{t,u} = \frac{1}{|C_{t,u}|} \sum_{i \in C_{t,u}} y_i. \tag{3}$$

As in our case, the SME observations are associated with survey weights, we use a weighted average to calculate equation (3). Then the corresponding Cohort Shapley value function is given by:

$$val_{CS}(u) = \bar{y}_{t,u} - \bar{y}_{t,\emptyset} = \bar{y}_{t,u} - \bar{y}, \tag{4}$$

where $\bar{y} = (1/n)\sum_i^n y_i$. Following equation (1), the *Cohort Shapley value* for the j-th feature of target observation $\mathbf{x}_t$ is given by:

$$\phi_{t,j}^{CS} = \frac{1}{d} \sum_{u \subseteq \{1:d\} \setminus j} \binom{d-1}{|u|}^{-1} (\bar{y}_{t,u+j} - \bar{y}_{t,u}). \tag{5}$$

A positive value of Cohort Shapley $\phi_{t,j}^{CS}$ implies that the corresponding feature j has a positive impact on the outcome y, while a negative value implies a negative impact. Notably, traditional explainable methods such as SHAP necessitate the existence of an ML model, for example, computing equation 2 using $\hat{y}_i = f(\mathbf{x}_i)$. In contrast, by equation (5), Cohort Shapley values can be computed from the observed response values $\{y_i\}_{i=1}^n$, enabling its direct application to a provided dataset even without a built ML model. We apply the Cohort Shapley value to the observed outcomes of the credit applications and not to the ML techniques as the survey data are provided by several UK financial institutions that adopt different ML techniques. Furthermore, using real observations can avoid that the results are dependent on the particular ML technique Rudin (2019).

However, when an existing ML model is in use, Cohort Shapley operates similarly to other XAI methods like SHAP by replacing y_i with $\hat{y}_i$ in Equations (3) to (5), while effectively addressing the extrapolation issue. Moreover, Cohort Shapley can also be used to explore the feature importance concerning the residuals $y_i - \hat{y}_i$.

2.2. A relative fairness score

Each categorical feature j may divide the population into several groups, to summarize the degree of fairness for the k-th group with respect to a feature j, we calculate the Weighted average $WCS_{k,j}$ of the Cohort Shapley values $\phi_{t,j}^{CS}$ for all observations in the group:

$$WCS_{k,j} = \frac{\sum_{i \in \text{group}k} w_i \phi_{i,j}^{CS}}{\sum_{i \in \text{group}k} w_i}, \tag{6}$$

where w_i is the survey weight associated with each observation in the SMEs Finance Monitor survey analyzed in this paper. $WCS_{k,j} < 0$ indicates that group k is, on average, disadvantaged in the system, while $WCS_{k,j} > 0$ implies an advantage.

To further gauge the severity of the disadvantage, we normalize the weighted averages by the minimum and maximum values that a group could obtain given the dataset. For a disadvantaged group, we create a biased system where observations in that group receive no financial support, while all observations in the privileged group receive financial support. We then calculate the Cohort Shapley values under this biased system to obtain the average minimum score $WCS_{k,j}^{min}$ that the group can obtain. Similarly, we can define another biased system where the privileged and disadvantaged groups are swapped to obtain the maximum score $WCS_{k,j}^{max}$ that the group can achieve.

Then, a *relative fairness score* (RFS) can be calculated as:

$$RFS_{k,j} = \frac{WCS_{k,j} - WCS_{k,j}^{min}}{|WCS_{k,j}^{max} - WCS_{k,j}^{min}|} = \frac{WCS_{k,j} - WCS_{k,j}^{min}}{|2 \cdot WCS_{k,j}^{min}|}. \tag{7}$$

The last equation in (7) results from the observation that $WCS_{k,j}^{max}$ is equivalent to $-WCS_{k,j}^{min}$, owing to the exchange of groups between the two biased systems. The RFS score takes values within $[0, 1]$. A low $RFS_{k,j}$ value less than 0.5 suggests that group k is at a disadvantageous position in the system concerning feature j. Conversely, a high $RFS_{k,j}$ value greater than 0.5 indicates a privileged situation.

Fairness measures based on Cohort Shapley show the following advantages: first, they bypass the extrapolation issue mentioned in Section 2.1; second, compared to other fairness measures based on false/negative errors (Babii et al., 2020; Chouldechova, 2017), they are independent of the specific classification threshold and less affected by the class imbalance (Chen et al., 2024).

3. Data

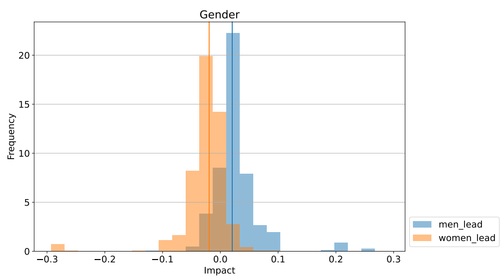
We analyze data gathered from the SME Finance Monitor survey over the period from the second quarter 2018 to the third quarter 2020. This is the largest survey on SME finance in the UK, with 4,500 companies interviewed every quarter. Coherently with the criteria established by the European Commission, SMEs are eligible for the SME Finance Monitor interview if they meet the following conditions: (i) not 50%+ owned by another company; (ii) not run as a social enterprise or as a not-for-profit organization; (iii) turnover of less than £25m and (iv) number of employees lower than 250. Quotas are set by size, sector and region to have a representative sample.

Initially, we consider only the SMEs that sought a bank loan in the last 12 months, given by $1,620$ companies. We exclude records with missing or unknown values, resulting in a dataset with $n = 1,102$ SMEs.

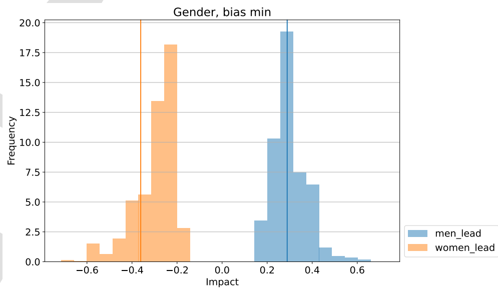
4. Empirical results

To evaluate the fairness of funding allocation to SMEs, we consider some firms' characteristics based on the literature on access to finance for SMEs (Cowling et al., 2023; Hewa-Wellalage et al., 2022; Department for Business, Innovation and Skills, 2014). Particularly, we examine ethnicity, start-up, women leading, and enterprise size following the classification of the European Commission[1] based on turnover and number of employees. When computing the Cohort Shapley values, we use the given observations directly and no ML model is built, i.e., use y_i in equation (5), as the several financial institutions represented in the survey data adopt different ML techniques.

Figures 1-5 display the distributions of the Cohort Shapley values for gender, ethnicity, number of employees, turnover and firms' age, respectively, while Figure 6 shows the RFS results for the above features. The group sample sizes are reported in Appendix A.



(a) True system
$WCS_{Women} = -0.025$
$WCS_{Men} = 0.027$



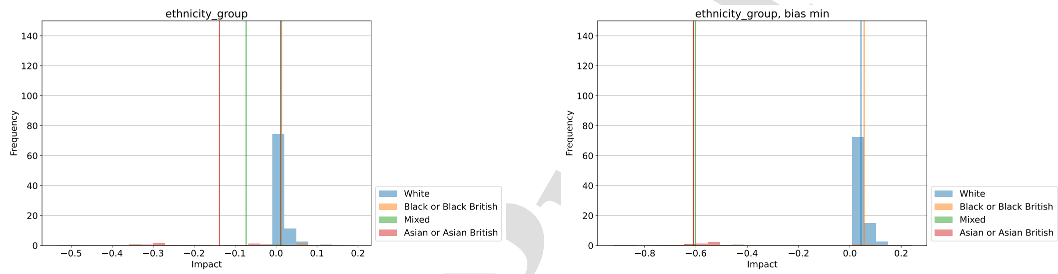
(b) Biased system
$WCS_{Women} = -0.298$
$WCS_{Men} = 0.300$

Figure 1: The Cohort Shapley results for feature 'women lead'.

Plot (a) of Figure 1 illustrates the histogram of Cohort Shapley values for each observation obtained from the dataset, therefore, the 'true' system. For companies with women leaders, the WCS value is -0.025, represented by the orange vertical line. Based on equation (6), the negative value of WCS indicates that women-led companies face, on average, disadvantages in their loan applications. This is because most of the cohort Shapley values $\phi_{t,j}^{CS}$ defined in equation (5) are negative so being a women-led company has a negative impact on the chance of having the loan application approved. Coherently with the expectations and the results obtained in the literature (Hewa-Wellalage et al., 2022), men-led companies exhibit, on average, privilege, with $WCS_{Men} = 0.027$, represented by the blue vertical line. We observe that the histograms for men and women-led companies have similar central tendencies.

[1] https://single-market-economy.ec.europa.eu/smes/sme-definition_en

To assess the severity of unfairness, we construct a highly 'biased' system, wherein all women-led companies have their loan applications rejected while all men-led companies are granted loans. Subsequently, we recompute the WCS values for the two groups of companies using these artificial outcomes, as depicted in Plot (b). In this extremely biased system based on gender, we obtain $WCS^{min}_{Women} = -0.298$ and $WCS^{min}_{Men} = -0.300$. Consequently, the estimated range for WCS_{Women} (or WCS_{Men}) lies within the interval $[-0.298, 0.298]$. Comparing this range to the WCS values in the true system, it is evident that the difference between men and women-led companies (i.e., the advantage and disadvantage groups) is not obvious. Indeed, the computed RFS for women-led companies is $RFS_{Women} = \frac{-0.025-(-0.298)}{|2\times(-0.298)|} = 0.46$, and for men-led companies, $RFS_{Men} = \frac{0.027-(-0.300)}{|2\times(-0.300)|} = 0.54$, both of which are close to the RFS threshold of 0.5 (See Figure 6).



(a) True system
$WCS_{White} = 0.012$
$WCS_{Black} = 0.027$
$WCS_{Mixed} = -0.142$
$WCS_{Asian} = -0.173$

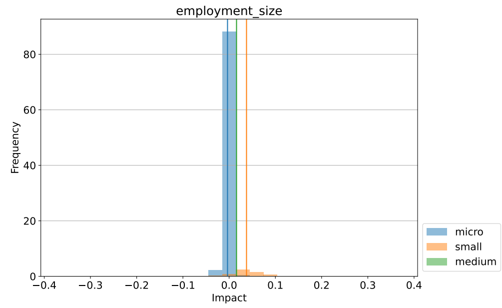
(b) Biased system
$WCS_{White} = 0.044$
$WCS_{Black} = 0.046$
$WCS_{Mixed} = -0.567$
$WCS_{Asian} = -0.582$

Figure 2: The Cohort Shapley results for feature 'Ethnicity group'.

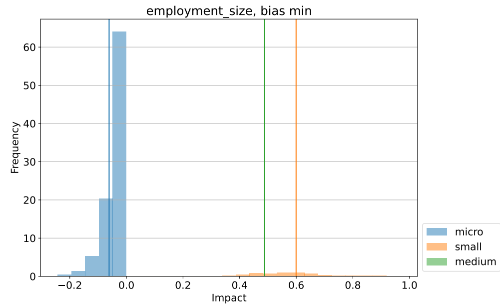
Figure 2 illustrates the results for the 'Ethnicity group'. Within the dataset, the Asian (red) and mixed-ethnic groups (green) are identified as disadvantaged, with respective WCS values of $WCS_{Asian} = -0.173$, and $WCS_{Mixed} = -0.142$. Conversely, the white and black ethnic groups are observed to be privileged. In the biased system (Plot (b) of Figure 2), notable changes in WCS values are observed for the Mixed and Asian groups.

For company size, we consider the classification of the European Commission [2] based on turnover and the number of employees. For turnover, a company is defined as micro if the turnover is lower than 2 million euros, small if the turnover is lower than 10 million euros and medium if the turnover is lower than 43 million euros. For employees, a company is classified as micro if the number of employees is lower than 10, small if it is lower than 50 and medium if it is lower than 250. Figures 3 and 4

[2] https://single-market-economy.ec.europa.eu/smes/sme-definition_en

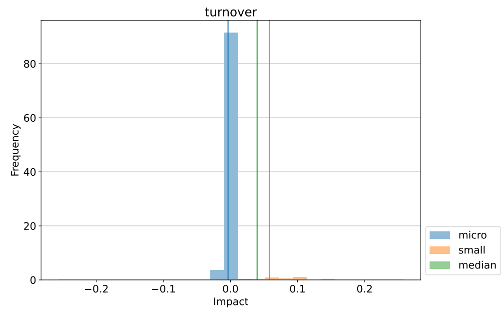


(a) True system
$WCS_{Micro} = -0.003$
$WCS_{Small} = 0.042$
$WCS_{Medium} = 0.018$

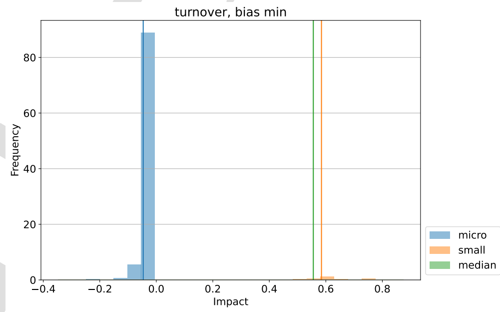


(b) Biased system
$WCS_{Micro} = -0.044$
$WCS_{Small} = 0.591$
$WCS_{Medium} = 0.481$

Figure 3: The Cohort Shapley results for feature 'Employment size'.



(a) True system
$WCS_{Micro} = -0.002$
$WCS_{Small} = 0.074$
$WCS_{Medium} = 0.026$



(b) Biased system
$WCS_{Micro} = -0.021$
$WCS_{Small} = 0.625$
$WCS_{Medium} = 0.581$

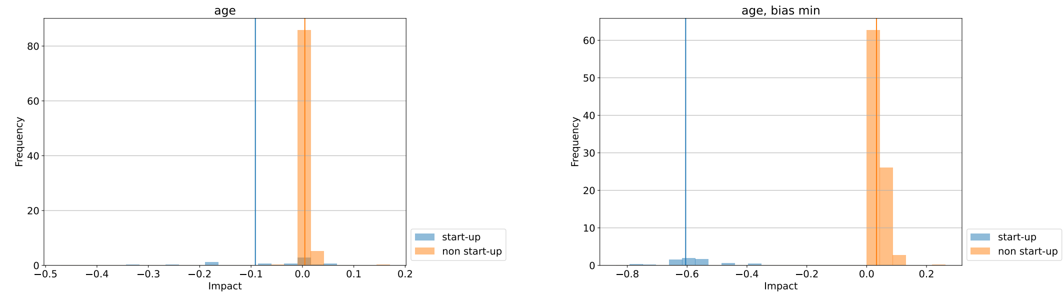
Figure 4: The Cohort Shapley results for feature 'Turnover'.

reveal the disadvantages faced by micro companies according to both the definitions based on the number of employees and turnover.

The results depicted in Figure 5 indicate that start-up companies, defined as companies with a duration of less than 12 months, face a disadvantage in loan applications, in line with the results in the literature (Cowling et al., 2018).

We also conduct two-tail weighted Kolmogorov–Smirnov tests on Cohort Shapley values among groups in the true system. Appendix B presents the corresponding test statistics and p-values. The results reveal statistically significant differences (at the 5% significance level) in Cohort Shapley values among various group divisions based on feature categories. The only exception is the Cohort Shapley values for employment size small and medium groups.

Figure 6 visually summarizes the RFS values for all the groups analyzed in the previous sections. The bars in the figure are centered at 0.5, representing the RFS



(a) True system
$WCS_{Startup} = -0.064$
$WCS_{NonStartup} = 0.005$

(b) Biased system
$WCS_{Startup} = -0.580$
$WCS_{NonStartup} = 0.042$

Figure 5: The Cohort Shapley results for feature 'Age'.

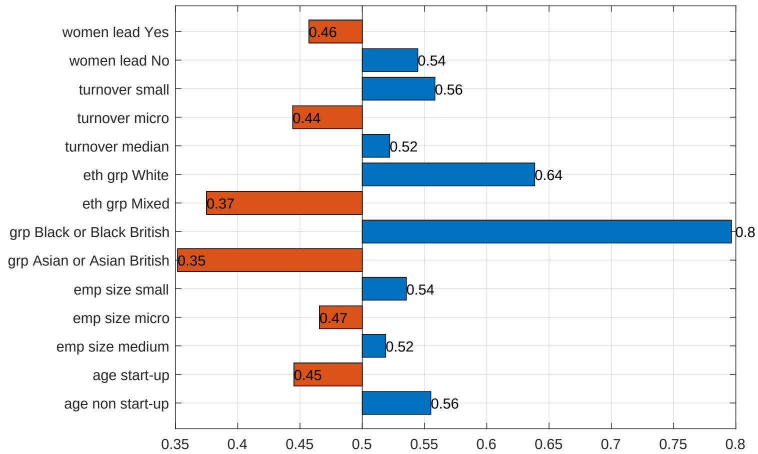


Figure 6: The RFS values.

threshold, and the length of each bar reflects the distance of the corresponding RFS value from this threshold. A longer bar indicates a higher level of advantage (shown in blue) or disadvantage (shown in red) for the respective group.

Analyzing the results, it is evident that women-led companies experience a relatively mild level of disadvantage, as indicated by the RFS values. On the other hand, micro and start-up companies, as characterized by turnover, employment size, and company age, face notable disadvantages in loan applications. Among the different ethnic groups, Asian companies demonstrate the greatest level of disadvantage, while the Black ethnic group exhibits the highest level of advantage in terms of loan application outcomes. The latter result contrasts with previous US studies (Blanchard et al., 2008) showing that African Americans are mostly disadvantaged.

To check the robustness of our results, we randomly extract 80% of the observations and we obtain similar results by performing the same analysis [3].

5. Conclusion

Based on our knowledge, this is the first paper that assesses the fairness of SMEs in accessing external finance. The main contribution of our analysis is that we consider not only accepted credit applications but also those that have been rejected. To achieve this goal, we analyze the data collected by the SME Finance Monitor survey, the largest survey on SME finance in the UK.

To evaluate fairness, we consider the Cohort Shapley value because it bypasses the extrapolation issue in the Shapley value calculation, avoiding model evaluation on the infeasible combinations. We suggest the relative fairness score based on the Cohort Shapley value to measure the extent of privilege or disadvantage experienced by different groups based on factors such as gender, turnover, employment size, company age, and ethnicity.

Our findings reveal that women-led companies in the UK face a relatively mild level of disadvantage in loan applications. Micro and start-up companies, characterized by smaller turnovers, employment sizes, and younger ages, encounter notable disadvantages in securing loans. Moreover, the analysis highlights the significant disadvantage faced by Asian companies in the loan application process. Our results show that the Black ethnic group in British exhibits a higher level of advantage. Financial institutions can adopt the approaches suggested in this study to assess the fairness of their credit decisions and be compliant with the regulations on the adoption of ML techniques.

[3]The code of our analysis is available at `https://github.com/LuXuefei/CohortShapleyforFairnessSMESurvey`

Appendix A. Dataset

Table A.1: Sample sizes and RFS for the SME Finance Monitor Survey data.

Feature	Group	Sample Size	RFS
Women lead	No	665	0.545
	Yes	437	0.457
Turnover	Micro	929	0.444
	Small	145	0.558
	Medium	28	0.522
Ethnicity group	White	1031	0.639
	Black or Black British	12	0.796
	Mixed	14	0.375
	Asian or Asian British	45	0.352
Employment size	Micro	543	0.465
	Small	447	0.535
	Medium	112	0.519
Age	Start-up	31	0.445
	Non start-up	1071	0.555

Appendix B. Kolmogorov–Smirnov Tests for Cohort Shapley Values

The computed two-tailed weighted Kolmogorov–Smirnov (KS) test statistic for Cohort Shapley values for women-led vs. men-led companies is 0.72 (p-value$= 1.25 \times 10^{-138}$). The KS test statistic for start-up vs. non start-up companies is 0.52 (p-value$= 5.53 \times 10^{-8}$). Table B.2 provides the KS test statistics and their corresponding p-values for the remaining features, including ethnicity group, employment size, and turnover.

Table B.2: Two-tail Weighted Kolmogorov–Smirnov Tests for Cohort Shapley Values with p-values in the parentheses. If the p-value is lower than the significance level (5%), we reject the null hypothesis that the two samples are coming from the same distribution.

Ethnicity group	White	Black or Black British	Mixed	Asian or Asian British
White	0 (*1*)	0.87 (4.11×10^{-11})	0.90 (2.71×10^{-14})	0.90 (6.33×10^{-44})
Black or Black British		0 (*1*)	0.89 (3.49×10^{-6})	0.90 (2.70×10^{-9})
Mixed			0 (*1*)	0.41 (*0.032*)
Asian or Asian British				0 (*1*)
Employment size	Micro	Small	Medium	
Micro	0 (*1*)	0.84 (8.92×10^{-196})	0.80 (6.50×10^{-65})	
Small		0 (*1*)	0.11 (0.18)	
Medium			0 (*1*)	
Turnover	Micro	Small	Medium	
Micro	0 (*1*)	0.92 (7.11×10^{-140})	0.89 (1.38×10^{-26})	
Small		0 (*1*)	0.87 (4.11×10^{-11})	
Medium			0 (*1*)	

References

Bracke, P., Datta, A., Jung, C., Shayak, S., 2019. Machine learning explainability in finance: An application to default risk analysis. URL: `https://www.bankofengland.co.uk/working-paper/2019/machine-learning-explainability-in-finance-an-application-to-default-risk-analysis`

Bueff, A.C., Cytryński, M., Calabrese, R., Jones, M., Roberts, J., Moore, J., Brown, I., 2022. Machine learning interpretability for a stress scenario generation in credit scoring based on counterfactuals. Expert Systems with Applications 202, ISSN 0957–4174.

Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J., 2021. Explainable machine learning in credit risk management. Computational Economics 57, 203–216.

Calabrese, R., Cowling, M., Liu, W., 2022. Understanding the dynamics of UK COVID-19 SME financing. British Journal of Management 33, 657–677.

Babii, A., Chen, X., Ghysels, E., Kumar, R., 2020. Binary choice with asymmetric loss in a data-rich environment: Theory and an application to racial justice. arXiv preprint arXiv:2010.08463 .

Blanchard, L., Zhao, B., Yinger, J., 2008. Do lenders discriminate against minority and woman entrepreneurs? Journal of Urban Economics 63 (2), 467–497.

Chen, Y., Calabrese, R., Martin-Barragan, B., 2024. Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research 312 (1), 357–372.

Chouldechova, A., 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big data 5, 153–163.

Cowling, M., Liu, W., Conway, E., 2023. Ethnicity and bank lending before and during COVID-19. International Journal of Entrepreneurial Behavior & Research 29 (3), 614–642.

Cowling, M., Liu, W., Zhang, N., 2018. Did firm age, experience, and access to finance count? SME performance after the global financial crisis. Journal of Evolutionary Economics 28, 77–100.

Department for Business, Innovation and Skills, 2014. Understanding localised policy interventions - startup loans in England's south west. BIS Research paper number 160 , 1–26.

Fraser, S., 2009. Is there ethnic discrimination in the UK market for small business credit? International Small Business Journal 27, 583–607.

Hewa-Wellalage, N., Boubaker, S., Hunjra, A.I., Verhoeven, P., 2022. The gender gap in access to finance: Evidence from the COVID-19 pandemic. Finance Research Letters 46(A), 102329.

Hooker, G., Mentch, L., Zhou, S., 2021. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. Statistics and Computing 31, 1–16.

Kozodoi, N., Jacob, J., Lessmann, S., 2022. Fairness in credit scoring: Assessment, implementation and profit implications. European Journal of Operational Research 297 (3), 1083–1094.

Kumar, I.E., Venkatasubramanian, S., Scheidegger, C., Friedler, S.A., 2020. Problems with Shapley-value-based explanations as feature importance measures.

Lundberg, S.M., Allen, P.G., Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30, 4765–4774.

Mase, M., Owen, A.B., Seiler, B., 2019. Explaining black box decisions by Shapley cohort refinement .

Mase, M., Owen, A.B., Seiler, B.B., 2022. Variable importance without impossible data. arXiv preprint arXiv:2205.15750 .

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence 1, 206–215.

Shapley, L.S., 1953. A value for n-person games. URL: https://apps.dtic.mil/sti/pdfs/AD0604084.pdf.

Smallbone, D., Ram, M., Deakins, D., Aldock, R.B., 2003. Access to finance by ethnic minority businesses in the UK. International Small Business Journal 21, 291–314.

Štrumbelj, E., Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems 41, 647–665. doi:10.1007/s10115-013-0679-x.

The Cohort Shapley Value to Measure Fairness in Financing Small and Medium Enterprises in the UK

Xuefei Lu[a], Raffaella Calabrese[b]

[a]SKEMA Business School, Université Côte d'Azur, 5 Quai Marcel Dassault, Paris, 92150, France
[b] University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh, EH8 9JS, United Kingdom

---

Abstract

Banks are relying on machine learning techniques to support their decisions in financing small and medium enterprises (SMEs). As regulators require that credit decisions are transparent, there is a need to develop methods to measure fairness. We propose a weighted average of the Cohort Shapley value, which removes impossible feature combinations, and a relative fairness score for assessing the fairness level within sub-populations. Based on our knowledge, this is the first paper that investigates the fairness of UK financial institutions in providing funding to SMEs. Our findings reveal discrimination against start-up, micro, women-led companies, and owners of Asian ethnic backgrounds.

Funding Information

Raffaella Calabrese is grateful to ESRC for financial support (ESRC Award Number ES/W010259/1).

Highlights

The Cohort Shapley Value to Measure Fairness in Financing Small and Medium Enterprises in the UK

- We analyze the fairness of financing small and medium enterprises in the UK.
- We propose a relative fairness score based on the Cohort Shapley value.
- Women-led and start-up companies face a limited disadvantage.
- Micro companies and owners with Asian backgrounds are highly discriminated against.

Credit Author Statement

Xuefei Lu: Conceptualization, Methodology, Software, Validation , Formal analysis, Writing - Original Draft, Writing - Review & Editing, Data Curation, Visualization
Raffaella Calabrese: Conceptualization, Methodology, Software, Validation , Formal analysis, Writing - Original Draft, Writing - Review & Editing, Data Curation, Visualization, Funding acquisition