Pricing risk: Analysis of Irish Car Insurance Premiums

Adrian Byrne[1,2]

[1] CeADAR, NexusUCD, University College Dublin, Belfield Office Park, Unit 9, Clonskeagh, Dublin 4, Ireland
[2] Idiro Analytics, Clarendon House, 39 Clarendon Street, Dublin 2, Ireland

Abstract
With the increasing prevalence of artificial intelligence assisting with decision-making and forthcoming EU legislation attempting to ensure it does no harm to its citizens, this paper evaluates what factors influence the cost of car insurance for individual drivers. The principle is clear: the cost of an insurance premium is determined by the perceived risk of the policyholder. But what specific elements are considered in assessing a driver's risk? To demystify this model, this study developed an automated process to gather quote data based on various factors, such as the driver's gender, age (as a proxy for driving experience), geographical location, occupation, and driving history. We conducted an audit of pricing algorithms employed by insurance companies in the Irish car insurance industry, by gathering quotes through online car insurance websites available in Ireland. This research provides insights into some of the factors influencing car insurance premiums in Ireland, highlighting some of the intricacies behind the complex, algorithmic calculations of car insurance quotations. While acknowledging the complexity of the industry, we find evidence of several potentially problematic issues. We show that place of residence and occupation have a direct and sizeable impact on the prices quoted to drivers.

Keywords
Car insurance premiums, bias detection, explainable AI

1. Introduction

Companies are developing increasingly complex machine learning-based models for car insurance pricing [1, 2]. As such, forthcoming EU Artificial Intelligence (AI) Act legislation[2] classifies insurance as 'high-risk' via the access to essential services domain thereby compelling all insurers within the EU who use 'AI' as part of their operation to comply with the forthcoming legislation. Consequently, compliance means undertaking bias detection to mitigate any harmful effects arising from the use of AI (this aspect is covered within section 5 of Article 10 – Data and Data Governance). The motivation for this paper stemmed from trying to understand the potential for bias in AI systems that need to comply with the forthcoming EU AI Act legislation. Moreover, insurers do not normally provide customers with much information on why they have been offered a certain price, and they do not publish data showing the prices offered to different groups. This study aimed to shine a light on these inconspicuous aspects of car insurance premiums.

At the intersection of these considerations, this study delves into the pricing practices of ten insurance companies in Ireland. By scrutinising the factors and their respective magnitudes that contribute to car insurance quotes, this research aimed to unravel the complexities of the business model. Moreover, this study sought to contribute to a comprehensive understanding of the dynamics of car insurance premiums and foster greater transparency within the industry. We hope that this endeavour will encourage insurers to consider undertaking and publishing independent algorithmic bias audits on their website. Similar to the requirement stipulated by the New York City Local Law 144 of 2021 on employment decisions assisted by AI[3], this practice will go towards enhancing transparency and trust.

EWAF'24: European Workshop on Algorithmic Fairness, July 01–03, 2024, Mainz, Germany
EMAIL: Adrian.byrne@ucd.ie (A. 1)
ORCID: 0000-0003-4887-2572 (A. 1)

[2] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
[3] See https://rules.cityofnewyork.us/wp-content/uploads/2023/04/DCWP-NOA-for-Use-of-Automated-Employment-Decisionmaking-Tools-2.pdf

To the best of our knowledge, no study to date has analysed the contextual influence of certain features on car insurance pricing in Ireland. Additionally, we are unaware of any study that has examined the influence of socio-demographic features on pricing.

Research questions. In this paper, we exploited nearly 40,000 quotes gathered via an automated retrieval process.

RQ1: What are the factors that play a major role in setting Irish car insurance premiums?
RQ2: Do gender and ethnicity inferred by name directly influence quoted premiums?
RQ3: Are riskier driver profiles uniformly discriminated against regardless of location and occupation?

Using an experimental design, we gathered data by varying some features of the 'applicant' while keeping the rest of the features constant. This allowed us to conduct an algorithmic audit to uncover any connections between inputs and outputs. This study builds upon the work of Fabris et al. (2021) [3] and Cook et al. (2022) [4] whose efforts inspired this research.

2. Background and related work

Since 2012, the use of gender in the Irish car insurance industry has been regulated. The European Union has adopted legislation which prohibits the direct use of gender for setting insurance premiums [5, 6]. The principle of gender equality is enshrined by Articles 21 and 23 of the Charter of Fundamental Rights of the European Union [7,8]. Gender equality has been explicitly operationalised in the context of insurance, with Article 5(1) of Council Directive 2004/113/EC [9], stating that no difference in individuals' premiums can result from the use of gender as an explicit factor, and fully confirmed in a 2011 judgement by the European Court of Justice [6]. Official guidelines on the application of the ruling [5] explicitly mention motor insurance, clarifying that indirect discrimination remains possible where justifiable: "For example, price differentiation based on the size of a car engine in the field of motor insurance should remain possible, even if statistically men drive cars with more powerful engines". Moreover, information about gender may still be collected, stored, and used, e.g. to monitor portfolio mix or for the purposes of reinsurance.

In terms of geographic location, which is a publicly acknowledged but often underestimated factor, it emerges as a substantial determinant of car insurance premiums. An Irish insurance broker, Chill, has released research on the price differences in car insurance based on the driver's location within Ireland [10]. The study by the car insurance broker found that Longford had the highest quotes among all locations (average €738). While the rationale behind this practice often hinges on the increased risks associated with higher-crime areas (for example), questions arise about the equity of imposing a 'penalty' on drivers based solely on their residential neighbourhood.

Despite this focus on creating an individual price for each customer, there is evidence that some groups are likely to pay more for insurance than others. Previous research has indicated that people of colour could be experiencing worse outcomes - including higher prices - in insurance markets compared to white consumers. This paper follows a range of investigations looking at unequal outcomes in the insurance market. In November 2015, the Consumer Federation of America found that price of car insurance offered to drivers increased where the proportion of African Americans living in a community increased [11]. In July 2016, research by Webber Phillips found a postcode-based ethnicity penalty for motor insurance customers in the UK, affecting 12 million people [12].

To the best of our knowledge, no such investigation has been conducted for the Irish car insurance market. Our aim was therefore to close a transparency gap between insurers and customers in Ireland and help develop best practice when it comes to bias detection which forms part of the forthcoming EU AI Act.

3. Data and methods[4]

Design of "mystery shopper" experiment

Firstly, the Competition and Markets Authority (CMA) has suggested mystery shopping as an appropriate technique for investigating potential harm caused by algorithmic bias [13]. Secondly, to conduct our research, we crafted individual scripts tailored to each insurer used in this study. The scripts were written using the set of questions provided on each insurer's website. A total of ten insurance companies were included in this study. This approach facilitated the automated retrieval of insurance quotes by referencing a spreadsheet containing the profiles of individual drivers and their corresponding responses. Driver names were sourced from online databases categorising names by gender and ethnicity. For standardisation, we set driver ages at 25, 40, and 60, linking age to years of driving experience and policy in their own name. We established the relationship between the driver's age and driving experience: age - 18 = years of driving experience, and age - 19 = years of policy in their own name. Consequently, age was a perfect proxy for driving experience and years of policy in own name in our study. We also strategically selected 20 locations (i.e. 'locales') to capture diverse contrasts, encompassing urban versus rural settings, variations in house prices, levels of deprivation, ethnic diversity, and per capita statistics for crime, road traffic accidents, and penalty points. These 20 locations were drawn from the following six counties: Dublin, Wicklow, Cork, Longford, Roscommon, and Donegal.

We employed multivariate log-linear regression models to derive mean predicted quotes, transforming the quotes on a log base 10 scale to address positive skewness. The predictor variable sets varied depending on the like-with-like comparison under consideration, but the model predicted mean quotes all accounted for age, gender, locale, insurer, and time of quote (hour/day/month). To illustrate feature importance within each model setup, we used SHAP (SHapley Additive exPlanations) plots after running random forest regression models [14]. We also complemented this approach with AVTS (Absolute Value Test Statistic) numbers derived from the log-linear regression models [15].

4. Results[5]

We contextualised the quotes retrieved according to characteristics attributable to either the locale or county level and we present these results in Figure 1 below. These characteristics include ethnic makeup, house prices, deprivation score, crime, road traffic accidents, and penalty points.

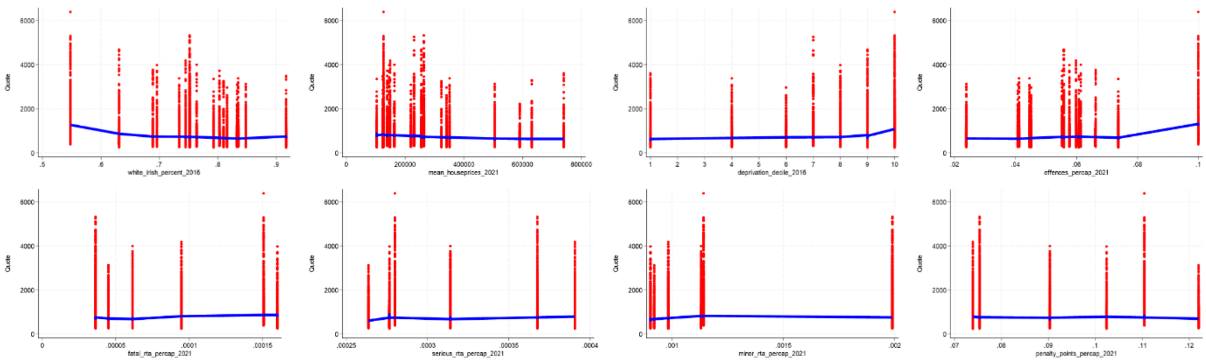

Figure 1: Effect of county and locale contextual data on quote amount

In Figure 1, the top four plots displaying proportion of white Irish, mean house prices, level of deprivation (10 = highest), and criminal offences per capita appear to have a bigger impact on quote amount than the bottom four plots displaying fatal/serious/minor road traffic accidents and penalty points per capita.

[4] Section reduced due to space constraints. Full list of assumptions will be presented at EWAF'24.

[5] It is not possible to present all our results in this extended abstract due to space constraints. Full results will be presented at EWAF'24.

5. Discussion and conclusion

Our research revealed several noteworthy findings. We found no disparity between male and female drivers' insurance quotes, which shows that the EU legislation banning gender discrimination in car insurance premiums is working well in Ireland. Next to NCB/days since claim and age, which was a proxy for driving experience, locale was the third most important variable affecting car insurance premiums. The locations with higher crime rates, more deprivation, more ethnic diversity, and more fatal road accidents directly correlated with higher car insurance premiums. Additionally, our investigation revealed that drivers from higher-premium locations were disproportionately penalised for having zero NCB, making claims, or receiving penalty points compared to their counterparts in lower-premium areas.

Combining the contextual data with our study results gives us a grasp of how insurance pricing works. It is essential to consider the contextual factors associated with specific locations and the socioeconomic dynamics of communities when seeking to comprehend the determinants influencing car insurance costs. This approach significantly contributes to a comprehensive understanding of the overarching implications for individuals with insurance in varied geographic settings. Furthermore, occupation, specifically retail, significantly influenced car insurance premiums. This impact also varied across locations, highlighting geographical disparities.

The interplay of location, demographics, and occupation adds depth to the industry's complexities. These insights extend beyond immediate premium considerations, offering a foundation for future research and policy considerations. As innovative technologies, including AI, continue to advance, it is inevitable that car insurance companies will integrate these tools into their processes. Therefore, a comprehensive understanding of the sophisticated dynamics involved in determining car insurance premiums is paramount for stakeholders and policymakers. It is imperative to conduct a thorough assessment of the decision-making mechanisms employed by car insurance companies' models to mitigate the risk of potential biases and perceived unfairness.

Limitations and future work. Despite retrieving nearly 40,000 quotes, our final dataset cannot be considered fully representative of the Irish driving population at large. Moreover, we were only able to examine a subset of the relevant features, which does not fully characterise the performance of the pricing algorithm. Fabris et al. (2021) ran into a similar limitation when studying the Italian car insurance market [3]. Also, like Cook et al. (2022)[6], our exploratory research allowed us to test the outcomes of pricing mechanisms, but we cannot explain why the outcomes we identified occurred [4]. While our experiments showed possible discrimination regarding differential quotes, we did not attempt to quantify the impact of this discrimination for all of Ireland. We leave this large and complex task as an interesting endeavour for future work.

6. Acknowledgements

This research, jointly produced by CeADAR, Ireland's Centre for Artificial Intelligence, and Idiro Analytics, has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 847402. The content of this paper reflects only the authors' views. Neither Enterprise Ireland nor the European Commission are responsible for any use that may be made of the information disclosed.

7. References

[1] Nicolas Chapados, Yoshua Bengio, Pascal Vincent, Joumana Ghosn, Charles Dugas, Ichiro Takeuchi, and Linyan Meng. 2001. Estimating Car Insurance Premia: A Case Study in High-Dimensional Data Inference. In *Proceedings of the 14th International Conference on Neural Information Processing Systems:*

[6] Our study established similar results in terms of no significant difference in premiums charged to people with different names in the same location, but higher average quotes in areas where non-white ethnicities make up a large proportion of the population.

Natural and Synthetic (Vancouver, British Columbia, Canada) *(NIPS'01)*. MIT Press, Cambridge, MA, USA, 1369–1376.

[2] Chun Yan, Xindong Wang, Xinhong Liu, W. Liu, and J. Liu. 2020. Research on the UBI Car Insurance Rate Determination Model Based on the CNN-HVSVM Algorithm. *IEEE Access* 8 (2020), 160762–160773.

[3] Alessandro Fabris, Alan Mishler, Stefano Gottardi, Mattia Carletti, Matteo Daicampi, Gian Antonio Susto, and Gianmaria Silvello. 2021. Algorithmic Audit of Italian Car Insurance: Evidence of Unfairness in Access and Pricing. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021, Virtual Event, USA. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3461702.3462569

[4] Tilly Cook, Aiden Greenall, Emer Sheehy. Discriminatory pricing: Exploring the 'ethnicity penalty' in the insurance market. Citizens Advice (2022).

[5] EC - European Commission. 2012. Guidelines on the application of Council Directive 2004/113/EC to insurance, in the light of the judgment of the Court of Justice of the European Union in Case C-236/09 (Test-Achats) C-11/1. https: //eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XC0113(01).

[6] ECJ - European Court of Justice. 2011. Association belge des Consommateurs Test-Achats ASBL v Conseil des ministres (2011) C-236/09. https://eur-lex.europa. eu/legal-content/EN/ALL/?uri=CELEX:62009CJ0236.

[7] EU - European Union. 2000. Charter of Fundamental Rights of the European Union C-364/01. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX% 3A32000X1218%2801%29.

[8] EU - European Union. 2012. Charter of Fundamental Rights of the European Union C-326/391. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX: 12012P/TXT.

[9] Council of the EU. 2004. Implementing the principle of equal treatment between men and women in the access to and supply of goods and services L-373-37. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32004L0113.

[10] https://lovin.ie/uncategorized/study-reveals-the-irish-counties-that-pay-the-most-for-car-insurance, last accessed 2024/03/12.

[11] https://consumerfed.org/wp-content/uploads/2015/11/151118_insuranceinpredominantlyafricanamericancommunities_CFA.pdf, last accessed 2024/03/12.

[12] https://www.thompsons.law/media/1839/ethnic-penalties.pdf, last accessed 2024/03/12.

[13] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/954331/Algorithms_++.pdf, last accessed 2024/03/12.

[14] https://shap.readthedocs.io/en/latest/, last accessed 2024/03/12.

[15] Altman, D. G., & Bland, J. M.: How to obtain the P value from a confidence interval. Bmj, 343 (2011).