Article

Exploring Industry-Level Fairness of Auto Insurance Premiums by Statistical Modeling of Automobile Rate and Classification Data

Shengkun Xie [1,]*[ ], Rebecca Luo [1] and Yuanshun Li [2]

[1] Global Management Studies, Ted Rogers School of Management, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada
[2] School of Accounting and Finance, Ted Rogers School of Management, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada
* Correspondence: shengkun.xie@ryerson.ca; Tel.: +1-416-979-5000

Abstract: The study of actuarial fairness in auto insurance has been an important issue in the decision making of rate regulation. Risk classification and estimating risk relativities through statistical modeling become essential to help achieve fairness in premium rates. However, because of minor adjustments to risk relativities allowed by regulation rules, the rates charged eventually may not align with the empirical risk relativities calculated from insurance loss data. Therefore, investigating the relationship between the premium rates and loss costs at different risk factor levels becomes important for studying insurance fairness, particularly from rate regulation perspectives. This work applies statistical models to rate and classification data from the automobile statistical plan to investigate the disparities between insurance premiums and loss costs. The focus is on major risk factors used in the rate regulation, as our goal is to address fairness at the industry level. Various statistical models have been constructed to validate the suitableness of the proposed methods that determine a fixed effect. The fixed effect caused by the disparity of loss cost and premium rates is estimated by those statistical models. Using Canadian data, we found that there are no significant excessive premiums charged at the industry level, but the disparity between loss cost and premiums is high for urban drivers at the industry level. This study will help better understand the extent of auto insurance fairness at the industry level across different insured groups characterized by risk factor levels. The proposed fixed-effect models can also reveal the overall average loss ratio, which can tell us the fairness at the industry level when compared to loss ratios by the regulation rules.

Keywords: insurance risk classification; fairness; rate regulation; generalized linear models; fixed effect models

1. Introduction

In auto insurance rate making, to avoid the potential effect of adverse selection in the insurance market (Liu et al. 2022), the pricing of insurance contracts is based on the sophisticated evaluation of risk factors. Here, risk factors discriminate the risk level associated with different drivers and drive the underlying insurance costs. The statistical and actuarial soundness in estimating the relativities of risk factors through data modeling enhances the actuarial fairness of insurance price (Meyers and Van Hoyweghen 2018), which is one of the critical components of the auto insurance rate regulation. The risk relativity is a measure relative to a selected base level for a given risk factor. When calculated using loss costs, they are referred to empirical risk relativities. Often, they are derived by statistical modeling of loss costs. However, risk relativities used by insurance companies to determine the insurance price may be adjusted due to the economic optimization process within insurance companies (Arora and Arora 2014). This optimization process, where the minimization of bias between pure premium and average losses is targeted, is often

driven by a bias–variance tradeoff that aims at smoothing the parameter estimates (Dugas et al. 2003). The flexibility of designing their optimization process according to the desired objective function may cause higher bias for some rating groups, particularly those with smaller risk exposures. Given that pricing is based on the predictive modeling of loss severity and claim frequency, Canadian insurance companies must consider how to keep a balance between satisfying the regulatory rules and optimally adjusting the insurance rates when the desired property of risk relativity is imposed. For instance, the relativities of driving records are expected to decrease, but the empirical patterns for some other risk classifications may not be the case, and the pattern could be reversed. Therefore, from a regulatory perspective, insurance companies must explain the techniques and the rationales for making the adjustments in detail. In addition, when adjusting the relativities of risk factors, the plus–minus 10% rule needs to be ensured; credibility-weighted relativity can be made by using information from either major competitors or the benchmark values from regulators.

Although the basic principle of insurance is that the cost of risk transfer must be equal to the price of the insurance policy, adjustment to the premium charged is possible, especially in a competitive insurance market, because of the asymmetric information between insurance companies and policyholders (Chiappori et al. 2006; Cohen 2005; Saito 2006). Insurance companies tend to have more information than policyholders about their financial situation, including the return on equity, insurance loss by lines of business, and expenses. In addition, insurance companies may prefer writing policies in one area and try to avoid writing them in another if they believe it is profitable. To achieve this, insurance companies may adjust the premium for their favorite groups to lower while making the prices higher than what they should be for the unfavored groups (Regan et al. 2008). This is so-called insurance discrimination, which often occurs in the insurance market that is not regulated or lightly regulated (Avraham 2017). A recent study in Berry-Stolzlei and Born (2012) shows a significant effect of de-regulation on insurance pricing in causing fairness. Insurance discrimination is possible even in a country where auto insurance is heavily regulated. For instance, in the auto insurance regulation of Canada, regulators allow the relativities of risk factors applied to the pricing algorithm to be adjusted within a specific interval. This adjustment is often conducted by credibility weighting procedures (Ronka-Chmielowiec and Poprawska 2005; Zahi 2021). Up to ±10% of estimated risk relativity may be allowed for each factor as long as the adjustment is justifiable. There are many factors used in insurance pricing. If most factors are adjusted in the same direction, premium levels may significantly differ from loss cost levels. All of these tell us that it is arguable whether the insurance premium charged by insurance companies represents a "fair price" for an insured (Frees and Huang 2021; Hanafy and Ming 2021).

In auto insurance, the premiums are calculated by multiplying a set of risk relativities to a base rate that is further multiplied by a multiplier where a loading factor is assumed. The loading is an additional cost built into the insurance policy to cover expenses used to settle the insurance claims and costs for business operation. From the actuarial perspective, the risk relativities used in the premium charges must align with those obtained from statistical models (Frezal and Barry 2020; Landes 2015). However, as mentioned above, a minor adjustment may be made to each risk factor level. This may cause a significant difference in relativities between the premiums and loss costs at the insured level after accounting for the insurance loading factor. To estimate this loading factor, one may determine the loss ratio based on the premium and loss data. However, regulators must ensure that the premiums charged are not excessive when it comes to the loading factor (Cummins and Weiss 1992; Grabowski et al. 1989). This depends on the regulation rule on what loss ratio level is considered justifiable, fair, and exact. The new regulation rules may be applied when excessive premiums are charged. That is to say, by controlling the profit level and business expenses that the insurance companies may claim, fairness at the industry level may be addressed. In addition, because of the flexibility of the loading factor, political parties often use it for election campaigns to advertise a reduction of auto

insurance premiums to the public, which often happens in Canada. Therefore, the study on fairness at an industry level may be used to justify whether or not such an insurance premium reduction is possible. If there is a possibility, how much can it be reduced when the government implements a new regulation rule for auto insurance? This is based on the assumption that lowering the benefit coverages is not permitted. Therefore, the loss cost level is considered to be not reducible. To further investigate the fairness and whether or not the premium level is excessive at the industry level, one can compare this fixed effect to the loss ratio benchmark, around 65%, to see if they coincide.

In insurance, a risk factor is classified into different levels so that the premium rates can be computed based on the levels of given risk factors (Abraham 1985). However, when it comes to fairness, the definition of fairness is not clear, as the formal definition associated with the problem is lacking. On the other hand, there is an intrinsic difference in fairness between the insured and the insurer, which is mainly due to technologies used in insurance pricing, including financial and insurance technologies empowered by data science and analytical tools (Cao et al. 2021). The perspectives of insureds and insurance companies can be very different, so the interpretation of fairness can go completely the opposite way. Often, the insured aims for individual fairness, while the insurance pricing is completed mainly on group fairness and then adjusted by individual risk level. In Frezal and Barry (2020), some limitations and misinterpretations of actuarial fairness were discussed and studied. It has re-centered the debate around insurance fairness by going beyond the actuarial consideration. In Xin and Huang (2022), insurance discrimination was discussed and reviewed from law and regulation perspectives to address the fairness criteria. Various anti-discrimination pricing methods, such as Generalized Linear Models (GLMs), were proposed to eliminate discrimination based on fairness criteria for insurance pricing. It was revealed that to avoid insurance discrimination, there is a need to avoid adverse selection, which has been part of the results of the application of machine learning and artificial intelligence that build predictive models based on the loss data at the individual level, unlike the case in rate regulation, which uses the group-based modeling. To further classify the insurance risk, Usage-Based Insurance (UBI) has been used to address the concern of the amount of driving in pricing, which provides another dimension of risk classification. However, the study in Ferreira and Minikel (2012) shows the importance of measuring auto insurance risk using major risk factors such as class and territory. They further found that the major risk factors cannot be replaced by the annual mileage driven.

The auto insurance premium rates are not always aligned with the major risk factor relativities calculated from the industry-level insurance loss data. This is because the regulators and the insurance companies have different definitions of fairness (Thiery and Van Schoubroeck 2006). The regulators focus more on the overall reasonableness at the industry level and whether the premium rates the insurance companies charge are justifiable based on their designed risk groups. On the other hand, insurance companies focus more on fairness within their own companies. To minimize bias, they normally compute the average pure premium based on their company-level data and average claim amounts. To evaluate fairness, regulators use the major risk factors to derive the risk relativities using industry-level data. Insurance companies use these major risk factor relativities as benchmarks for the credibility-weighting purpose. Insurance companies aim to ensure actuarial fairness among the subpopulation of their customers, while regulators aim to ensure that individual companies are not charging excessive premiums. In addition, insurance companies' major risk factor relativities must comply with the regulatory rules and benchmarks, which is consistent with the traditional actuarial fairness notion that similar risks should be charged similarly. For instance, in Canada, when proposing rate changes in risk factor levels, an insurance company must submit an application for rate changes to the regulator for approval, along with justification from actuarial and statistical perspectives to support the appropriateness and reasonableness of the rate changes.

Recently, the focus on fairness has shifted from collective risk-based to individual risk-based. The pricing uses predictive modeling to determine the individual loss cost

rather than average loss costs estimated on the aggregate loss data by different risk levels (Barry 2020). These changes may lead to a further adjustment of insurance premiums from one group to the other. Therefore, the high-risk group may not be charged as high as the one determined by the relativities of risk factors. Furthermore, regulation efforts have been made to improve fairness by restricting the use of some risk factors, such as gender, eliminating gender discrimination (Ryan 1986; Schmeiser et al. 2014). However, since the fairness of the classification mechanism is part of the standard regulatory goals, charging an excessive premium for high-risk drivers may conflict with the fairness objective (Isotupa et al. 2019). In Isotupa et al. (2019), two experience-rating mechanisms were used to rate a group of drivers, where findings conclude that high driving record (DR) classes cause the highest risk drivers to pay unsustainable excessive premiums. The driving record in auto insurance refers to the total number of years of no accidents, and it is often capped at a maximum level, for example, six years.

Traditional actuarial fairness is based on company-level data to ensure the insurance rate charges are fair. The fairness criterion is often based on the minimization of mean square errors between average pure premium and average incurred loss amount in a given subpopulation across all subpopulations Dugas et al. (2003). This depends on the definition of subpopulation and the risk factors used to determine the pure premium. In this paper, actuarial fairness is built on industry-level data that aim to address fairness from the regulation perspective. Because of this, the risk factors used to address fairness are only the major risk factors, while the pricing completed by insurance companies uses many other risk factors to further classify the insurance. From a regulatory perspective, the concept of actuarial fairness at the industry level is relatively new; not much research has focused on this. A similar study was completed in Charpentier et al. (2022) to examine models for flood losses and the disparity by premiums they entail. Although offering wide coverage for moderate premiums to all may not be fair to low-risk insurers, social solidarity will be significantly improved. This is similar to the purpose of auto insurance regulation, which aims at collective actuarial fairness rather than individual fairness. However, the methodology that can be used for investigating actuarial fairness at an industry level is currently lacking.

This work applies statistical models to rate and classification data from the automobile statistical plan to investigate the disparities between insurance premiums and loss costs using major risk factors used in rate regulation as predictors. The automobile statistical plan provides aggregate loss, risk exposures, and premiums at the industry level. This information allows us to examine the relationship between the loss costs and premiums, potentially indicating overall fairness when risk classification is considered. We focus on information summarized by types of use (i.e., CLASS), driving records, regions, and major insurance coverage. We aim to demonstrate the existence and degree of auto insurance fairness by examining actuarial rates and actual insurance premiums among different levels of major risk factors. We investigate fairness by measuring the discrepancy of relativities between premiums and loss costs. We first estimate the risk relativities of the major insurance factors using Generalized Linear Models (GLM) through a combination of loss costs and premiums information. This approach differs from the traditional method, where relativities are measured based on loss costs only. Within this risk relativity estimate, we consider the major risk factors as explanatory variables: coverages, regions, types of use, and driving records. We capture the model fixed effect caused by adding premium information to the response variable in the model. We also investigate the effect on the fairness results from using different statistical models, suggesting a careful use of a statistical model in addressing fairness under different modeling objectives. The novelty of this work is the modeling techniques used to simultaneously estimate the relativities of major risk factors and the difference caused by loss costs and premiums information, which can be used to derive the intrinsic industry-level loss ratio.

The rest of this paper is organized as follows. In Section 2, the data and its basic processing are briefly introduced. The proposed methods using Generalized Linear Models

to estimate the risk relativities and the fixed effect are discussed. In Section 3, the summary of the results is presented. Finally, we conclude our findings and provide further remarks in Section 4.

2. Materials and Methods

This section mainly discusses modeling approaches that estimate fixed effects, which are defined as the difference caused by the disparity of premiums and loss costs. These estimated fixed effects from the modeling are related to the loss ratios as the fixed effect captures the difference when shifting the focus from loss costs to premiums. We analyze the loss costs and premiums both separately and jointly. The first approach is based on the similarity measured by the Euclidean distance between the relativities obtained from loss costs and premiums. This approach provides limited information to reveal the relationship between loss costs and premiums. The second approach is to jointly model the fixed effect by estimating the common risk relativities using loss costs and premiums as the model input. Within the second approach, we develop different statistical models to evaluate the fairness of insurance premiums under different objectives. Before we discuss the proposed method, we first briefly introduce the background of auto insurance regulation and the statistical plan data used in this work.

2.1. Background and Data

In many countries, including Canada, the purchase of an auto insurance policy is mandatory and regulated by insurance law. In insurance pricing, regulation rules and policies are often incorporated to ensure insurance premiums' overall fairness, non-discrimination and reasonableness. For instance, in the European Union, gender cannot be a factor used to determine insurance premiums Schmeiser et al. (2014). In Canada, a similar rule is used in auto insurance pricing. Therefore, the Type of Use pricing factor that combines other drivers' characteristics is created to avoid direct gender discrimination. In a recent study in Medders et al. (2021), the authors claimed that gender rating is considered unfairly discriminatory. In Canada, the Financial Services Regulatory Authority (FSRA) regulates auto insurance, which provides dispute resolution services and insurance rate fillings. The regulation includes both commercial vehicles and personal cars. The data used in this work are for personal vehicles only. From an actuarial perspective, rate regulation mainly involves filings and reviewing rate changes. The regulation process ensures that the proposed insurance rates and rate changes are statistically sound, fair and exact. The regulators are also responsible for providing the benchmark of key actuarial metrics, such as loss trend, loss development factors, and reform impact, to name a few. The regulation process ensures that the insurance charge is fair and not excessive at the industry level. The decision making of rate filings and rate changes review within this process are data-driven. Therefore, the data sources and their credibility are important for regulation.

Automobile statistical plan data are published by the General Insurance Statistical Agency of Canada. The statistical plan is an ongoing data collection, data reporting and data management process that provides a source of support for auto insurance rate making and rate regulation for both the industry and the government. The automobile statistical plan summarizes insurance loss and other information related to drivers, vehicles and insurance coverage at the industry level to provide high-level statistics. This work uses rates and classifications statistical plan data, separated by type of use and driving record, for different coverages, accident years, and territories. The loss cost is defined as the ratio of aggregate losses to the total number of risk exposures, which can be interpreted as the average pure premium. The aggregate losses include the claim losses and associated expenses to settle the losses. The data are transactional at a company level and are aggregated by coverages, accident years and territories to become industry-level data. These data are collected, aggregated and reported yearly, and they are from about 60 auto insurance companies in Ontario, Canada. The total number of risk exposures (i.e., vehicle years) is around 4.5 million. To illustrate the data we used to carry out our research, in Table 1, we
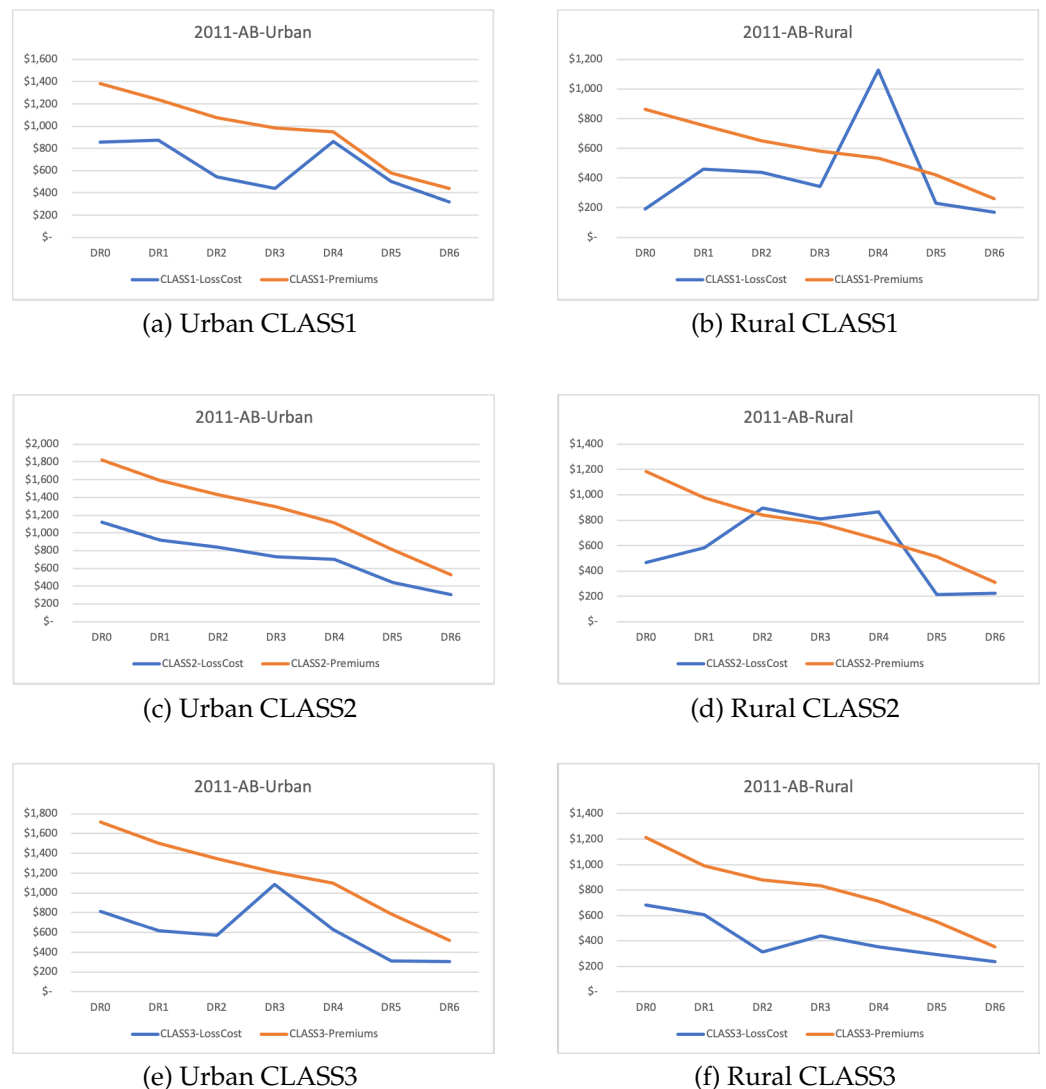
present the loss costs and premiums information for the accident year 2011, AB coverage, separated by territory. These aggregated data are formed into data matrices and taken as an input for proposed statistical models to estimate relativities of risk factors, i.e., Type of Use (i.e., CLASS) and driving records (DR). For the Type of Use, there are 14 different categorical levels, and for the driving record, there are seven different numerical levels. The coverage contains accident benefit (AB), third party liability (TPL) and collision (COL). The territory includes Urban and Rural. In addition, there were three accident years, respectively, 2009, 2010 and 2011. Due to the heterogeneity nature of loss cost, we did not consider including the coverage as a model variable. Instead, we separate our analysis by coverage.

Table 1. The comparison of loss costs and premiums (in Canadian dollars), separated by territory for the accident year 2011 and AB coverage, calculated based on a pool of total 4.5 millions risk exposures (i.e., vehicle years).

	Loss Costs							Average Premiums						
Urban														
	DR0	DR1	DR2	DR3	DR4	DR5	DR6	DR0	DR1	DR2	DR3	DR4	DR5	DR6
CLASS1	$858	$873	$546	$440	$863	$502	$318	$1381	$1237	$1074	$985	$949	$580	$440
CLASS2	$1119	$915	$840	$732	$703	$444	$305	$1824	$1591	$1428	$1295	$1115	$810	$526
CLASS3	$813	$615	$568	$1085	$630	$309	$305	$1714	$1502	$1342	$1207	$1097	$787	$521
CLASS5	$56	$68	$61	$162	$104	$56	$88	$193	$162	$147	$130	$114	$93	$80
CLASS6	$62	$49	$76	$59	$517	$75	$208	$389	$332	$317	$313	$258	$199	$199
CLASS7	$270	$389	$237	$650	$537	$305	$243	$1410	$1571	$1323	$1510	$1102	$678	$503
CLASS8	$3966	$ -	$413	$634	$ -	$ -	NA	$1600	$1258	$1364	$1153	$877	$1001	
CLASS9	$1678	$2036	$1495	$2082	$213	$350	$389	$1673	$1482	$1445	$1301	$1117	$887	$658
CLASS10	$397	$1049	$465	$1138	$311	NA	NA	$1701	$1363	$1158	$1184	$1937		
CLASS11	$6096	$467	$590	$317	$140	$150	NA	$1767	$1540	$1425	$1120	$909	$1298	
CLASS12	$383	$545	$823	$641	$204	$119	$1412	$1795	$1475	$1357	$1175	$1046	$818	$721
CLASS13	$974	$846	$797	$725	$251	$360	$217	$1826	$1600	$1343	$1233	$1078	$897	$680
CLASS18	$886	$558	$465	$522	$367	$168	NA	$1429	$1239	$1118	$989	$865	$879	
CLASS19	$750	$742	$724	$462	$441	$303	$496	$1549	$1369	$1207	$1034	$950	$762	$651
	Loss Costs							Average Premiums						
Rural														
	DR0	DR1	DR2	DR3	DR4	DR5	DR6	DR0	DR1	DR2	DR3	DR4	DR5	DR6
CLASS1	$193	$460	$440	$342	$1129	$229	$169	$865	$756	$649	$582	$533	$422	$259
CLASS2	$466	$582	$896	$811	$868	$211	$222	$1183	$975	$843	$776	$650	$512	$307
CLASS3	$680	$604	$313	$436	$353	$292	$238	$1213	$990	$878	$835	$713	$551	$352
CLASS5	$54	$59	$238	$29	$2	$0	$69	$142	$116	$100	$80	$67	$56	$47
CLASS6	$27	$9	$33	$314	$ -	$6	$54	$347	$291	$253	$229	$187	$149	$137
CLASS7	$14	$25	$188	$11	$537	$146	$111	$995	$1018	$836	$855	$716	$515	$342
CLASS8	$ -	$ -	$ -	$91	$ -	$ -	NA	$1570	$1153	$1072	$840	$700	$753	
CLASS9	$621	$ -	$366	$192	$253	$44	$371	$1319	$1042	$935	$829	$731	$572	$436
CLASS10	$852	$99	$119	$196	$ -	NA	NA	$1491	$1120	$1023	$838	$1535		
CLASS11	$510	$102	$2149	$1036	$61	$50	NA	$1797	$1429	$1241	$833	$691	$866	
CLASS12	$607	$169	$237	$3255	$139	$1362	$196	$1702	$1345	$1165	$898	$825	$595	$509
CLASS13	$404	$22	$554	$11	$273	$77	$145	$1711	$1245	$1055	$913	$796	$688	$469
CLASS18	$4215	$300	$137	$336	$136	$109	NA	$1206	$1052	$935	$770	$698	$705	
CLASS19	$586	$684	$365	$1317	$385	$449	$313	$1167	$966	$833	$707	$660	$561	$454

This industry-level data provide us with loss information and the premiums collected by insurance companies. Due to the aggregate nature of this data, we use it to illustrate how we can evaluate fairness by investigating the relationship between the loss cost and premiums. Note that the premiums are defined by the loss cost multiplied by a loading factor. This loading factor may not be the same for different coverages, different classes or different insurance companies. Therefore, even though the overall premium level is not excessive, it may be unfair for certain groups of drivers if the loading factor is too

high. For instance, in Table 1, for Rural, the loss cost associated with CLASS1 and DR0 is $193, while the average premium for this combination is $865, which is highly excessive and overcharged for this group of drivers when focusing on these two risk factors. This disparity is much more considerable for drivers with a small number of DR than those with better driving records. In Figure 1, where loss costs and premiums as a function of DR for some selected classes are displayed, Rural CLASS1 and CLASS2 do not appear excessive overall, but there are a significant amount of local adjustments. Other cases are questionable and are required further investigation. In addition, from Figure 1, we can see that premiums charged are a monotonic function of DR, but the actual loss costs from observations may not reflect this expected premium pattern. The monotonic DR pattern is desired when estimating DR relativities using loss costs data. Because of this requirement, the relativity as a function of DR will be smoothed out, and relativity for certain levels of a given risk factor can be significantly off. Due to the volatility of empirical loss cost patterns and the natural disparity of loss costs and premiums, our objective is to investigate how we can measure fairness, at an industry level, through statistical modeling approaches.

(a) Urban CLASS1

(b) Rural CLASS1

(c) Urban CLASS2

(d) Rural CLASS2

(e) Urban CLASS3

(f) Rural CLASS3

Figure 1. Comparison of loss costs and premiums as a function of DR for the first three classes in the accident year 2011 and AB coverage.

2.2. Risk Relativity Estimates by Generalized Linear Models

Using a similarity-based approach, we first compare the risk relativities between loss costs and premiums at factor levels. To do this, we apply Generalized Linear Models (GLM) to loss costs and premium data separately to derive their corresponding relativities, which can be obtained by log-transforming the GLM coefficients. Within this process, the data are also separated by territory, accident year and coverage. More specifically, we have the following two GLMs. The first model is used to derive risk relativity for the ith DR and the jth CLASS which is given as follows:

$$LossCost_{ij}^k = \alpha_0^k + \alpha_{1i}^k DR_i + \alpha_{2j}^k CLASS_j + \epsilon_{ij}^k, \tag{1}$$

for the kth combination of territory, accident year and coverage. Since we consider two territories, three different coverages and three different accident years, $k = 1, 2, \ldots, 18$. In addition, the DR has 7 levels and CLASS has 14 levels; therefore, $i = 1, 2, \ldots, 7$. $j = 1, 2, \ldots, 14$. For instance, when $k = 1$ and $i = 1$, $\alpha_{1,1}^1$ represents the risk relativity of DR_1, which is calculated from the data corresponding to the first combination (i.e., $k = 1$) of territory, accident year and coverage. Here, subscript 1 means that the duration of having no accidents is greater than one year but less than two years. The second model is used to derive the relativity of premium charged under the same combination of risk factors, i.e., DR and CLASS, and it has the same setting for i, j and k. That is,

$$Premium_{ij}^k = \beta_0^k + \beta_{1i}^k DR_i + \beta_{2j}^k CLASS_j + \epsilon_{ij}^k, \tag{2}$$

for the kth combination of territory, accident year and coverage. In GLM, the model also uses risk exposure as a weight value to indicate the importance of each combination of DR and CLASS, since their associated standard deviations are inflated linearly by the risk exposures, which are denoted by E_{ij}. The error functions used in this work are Gaussian, Inverse Gaussian, Gamma and Poisson distributions due to their popularity in the rate-making using GLM. In addition, logarithmic transformation is used as a link function for all cases. Although the different error functions used have their own canonical link functions, we use log link functions for all cases for ease of comparison and its potential in dealing with data heterogeneity. Finally, the intercepts α_0^k and β_0^k provide an estimate of the base level, and they are reflective of the basis of the loss cost or premium level across all cases.

Ideally, the relativities derived from the model will behave similarly for loss costs and premiums, so the premium levels charged are fair and exact across different levels of risk classification. However, due to the potential adjustment and the credibility weighting applied to the risk relativities obtained from loss cost at a company level, the relativities derived from loss cost may not coincide with the ones from premiums, at least for some of the factor levels.

2.3. Distance Measures for Comparing Premiums and Loss Costs

After we have obtained the relativities for the risk factor levels based on loss costs and premiums, we compare them using distance metrics. In data science, cosine similarity, Manhattan distance, Euclidean distance, Minkowski distance, and Jaccard similarity are popular metrics for measuring the similarity of two data sets (Sud et al. 2020). These distance measures are applicable for our study, but we focus on the Euclidean distance for illustration purposes. We consider the average distance per factor level based on Euclidean distance. The Euclidean distance is used because the magnitude of the value is important to reflect the risk factor level, and it is easier to interpret. This measure is defined as follows:

$$D(\alpha_{li}^k, \beta_{li}^k) = \frac{1}{n_l} \sqrt{\sum_{i=1}^{n_l} (\alpha_{li}^k - \beta_{li}^k)^2}, \tag{3}$$

for $l = 1, 2$, which corresponds to the first subscript of the model coefficients. The Euclidean distances are computed for DR and CLASS by different combinations of territories, coverages and accident years.

2.4. Modeling Fixed Effect Using Loss Costs and Premiums

The distance-based metric for comparing loss cost and premium to determine if the premium charged is fair from actuarial rate-making perspectives suffers limitations. Using distance as a metric, we cannot see if the premium is significant over the loss cost or under the loss cost. It only shows the average distance between the premium and loss cost per factor level. There is no clear cut-off value to determine if the difference is statistically significant. To overcome this difficulty, we propose the second approach by simultaneously analyzing loss cost and premium using GLM.

2.4.1. Comparing Loss Costs and Premiums within Same Territory

The risk characteristics often share some commonalities within the same territory in auto insurance. Therefore, we first analyze the combined risk relativities of loss cost and premium within the same territory to investigate the difference between loss costs and premiums. The following model estimates DR and CLASS relativities separately but uses the loss cost and premium information to see if there is a significant fixed effect on the risk relativity estimate caused by this mixed information.

$$Y_i^k = \gamma_{0i}^k + \gamma_{1i}^k DR_i + \alpha_1^k I + \epsilon_i^k, \tag{4}$$

$$Y_j^k = \gamma_{0j}^k + \gamma_{2j}^k CLASS_j + \alpha_2^k I + \epsilon_j^k, \tag{5}$$

where $Y_i^k = [LossCost_i^k, Premium_i^k]^\top$, consisting of loss costs and premiums associated with Driving Record, and $Y_j^k = [LossCost_j^k, Premium_j^k]^\top$, consisting of loss costs and premiums associated with CLASS. The input data for the GLM are by either DR or CLASS for different combinations of accident year, coverage and territory. This implies that $k = 1, 2, \ldots, 18$, the same as before. The variable I is the dummy variable. When the data correspond to premiums, variable I takes zero; otherwise, it is 1. The coefficient captures the difference caused by mixed information of loss costs and premiums. We also further investigate the case for combining different coverages and territories. In this case, we estimate the relativity by taking the average values from the combined three accident years. The two models above are referred to as a one-way model to analyze DR and CLASS. A second approach is jointly estimating DR and CLASS from the following two-way GLM.

$$Y_{ij}^k = \gamma_{0ij}^k + \gamma_{1i}^k DR_i + \gamma_{2j}^k CLASS_j + \alpha^k I + \epsilon_{ij}^k, \tag{6}$$

where $Y_{ij}^k = [LossCost_{ij}^k, Premium_{ij}^k]^\top$, consisting of loss costs and premiums from a two-way classification of DR and CLASS.

Both one-way and two-ways GLM are models for capturing the fixed effect through the estimate of coefficients, which are α_1^k and α_2^k for the one-way model and α^k for the two-way model. The final fixed effect estimate is the exponential transformation of the obtained coefficients, since the logarithmic transformation is used as a link function in GLM. The relativity of either DR or CLASS level is also the exponential transformation of the corresponding coefficient in GLM. The estimates of γ_{0i}^k, γ_{0j}^k, and γ_{0ij}^k correspond to the base rate of the premiums. Note that the estimate of the relativity uses both loss costs and premiums information, leading to an overall more reasonable estimate because multiple input data sources are used.

2.4.2. Comparing Loss Costs and Premiums between Territories

Examining variables one at a time can lead to a double penalizing effect which leads to inaccurate and higher relativity for riskier levels. For example, young drivers usually have a low DR because they are new and are deemed riskier and charged more premiums.

Additionally, it is observed that a young driver's CLASS is also deemed riskier. This "double counting" effect can be overcome by the two-way analysis where interaction between the two variables DR and CLASS is reflected. Building on top of the single variate method leads to the multivariate method, which can help identify the influence of one variable on the set of other variables. In this case, we further extend our analysis by including the Territory to undercover the relationship between DR and CLASS. We use a three-way GLM to simultaneously estimate DR, CLASS, and Territory, where Territory takes either Urban or Rural. The model is specified as follows:

$$Y_{ijl}^k = \gamma_{0ijl}^k + \gamma_{1i}^k DR_i + \gamma_{2j}^k CLASS_j + \gamma_{3l}^k Territory_l + \alpha^k I + \epsilon_{ijl}^k, \tag{7}$$

where $Y_{ijl}^k = [Loss_{ijl}^k, Premium_{ijl}^k]^\top$, consisting of loss costs and premiums from a three-way classification of DR, CLASS and Territory.

In this work, we also consider the model that takes three years of combined data as the model input to investigate an overall pattern of the DR, CLASS and Territory. In this case, the values of k take 1, 2 and 3 in the model, representing a different coverage. The combined data help improve the stability of the estimate since the overall losses and the risk exposures are three years. As a result, the estimate of loss cost becomes more accurate than the single accident year data. In addition, the premium is the average premium over the three-year period. Therefore, the fixed effect estimate from the combined data reflects the overall pattern and is more reliable from the statistical estimation perspective.

The fixed effects estimates depend on models and how the data are organized as an input for the models. The values tell the relative relationship between the loss cost and the premium charged. Since we take premium as a basis when the value of the fixed effect is significantly greater than one, we conclude that overall, the premiums are overcharged; when the value of the fixed effect is significantly smaller than one, it tells us that the premiums are undercharged instead. From this, we can see that the proposed modeling approach that captures the fixed effect between the loss cost and the premium is better than the distance-based approaches that measure the similarity between them. It estimates the relative difference between the loss cost and the premium and indicates the direction to tell whether it is overcharged or undercharged.

3. Results

We first address the limitation of using the distance-based method for comparing the relativities obtained from premiums and loss costs separately, using each accident year data. The results are presented in Table 2. Overall, the relativity difference between premiums and loss costs tends to be high for Rural for both CLASS and DR factors. This may suggest that the disparity between loss costs and premiums is high for Rural areas, which may be due to the adjustment of relativity on loss costs. Because of this significant difference in relativity, fairness may be of greater concern for Rural drivers than Urban drivers. In addition, the results in Table 2 tell that adjustment is linked to the accident year, as we can observe that the adjustment seems higher for the accident year 2009 for coverages AB and TPL, which are two major coverages. However, this distance-based approach cannot tell the direction of adjustment of relativities. We further illustrate this disparity between premium and loss cost relativities, which are separated by Urban and Rural for the CLASS factor with TPL and COL coverages. From the results presented in Figures 2 and 3, we realize that the CLASS with a small number of exposures is mainly affected, and the adjustment of relativity is much higher than other levels of CLASS, where the number of exposures is high. We also can see that relativities are adjusted much heavier for Rural than Urban areas for the groups with a smaller number of exposures. This may suggest that the fairness concern is for the minority rather than the majority groups, and it is more likely to happen for Rural drivers. Note that the obtained results are from a one-way analysis, implying a separate estimate for loss costs and premiums. In this case, the estimate of relativities for

either loss costs or premiums is subject to a potential double penalty, which is the primary concern for a one-way analysis.
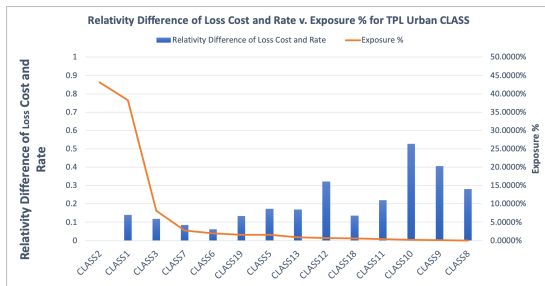
(a) Urban
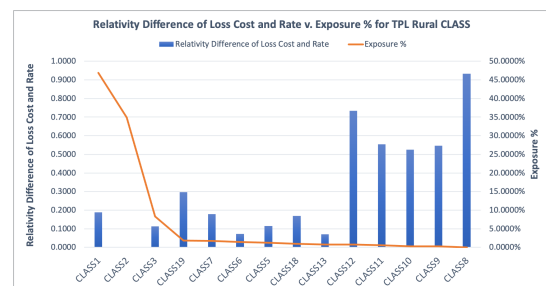
(b) Rural

Figure 2. The relativity differences between loss costs and premiums for different levels of CLASS, and the percentage of risk exposures of the corresponding levels of CLASS for TPL coverage.

Table 2. The average Euclidean distance of relativities between loss cost and premium, for DR and CLASS, by different coverages, territories and accident years.

|	2009		2010		2011	
	Gamma	**Poisson**	**Gamma**	**Poisson**	**Gamma**	**Poisson**
AB						
Rural						
DR	0.30	0.12	0.06	0.06	0.12	0.17
CLASS	0.18	0.16	0.16	0.16	0.17	0.16
Urban						
DR	0.05	0.05	0.06	0.04	0.08	0.04
CLASS	0.23	0.21	0.21	0.17	0.08	0.07
TPL						
Rural						
DR	0.26	0.22	0.12	0.11	0.12	0.11
CLASS	0.11	0.11	0.10	0.11	0.16	0.16
Urban						
DR	0.02	0.02	0.12	0.11	0.06	0.02
CLASS	0.09	0.09	0.07	0.06	0.05	0.05
COL						
Rural						
DR	0.05	0.06	0.04	0.03	0.08	0.02
CLASS	0.07	0.07	0.20	0.18	0.09	0.06
Urban						
DR	0.05	0.04	0.05	0.05	0.03	0.01
CLASS	0.19	0.19	0.13	0.10	0.08	0.06
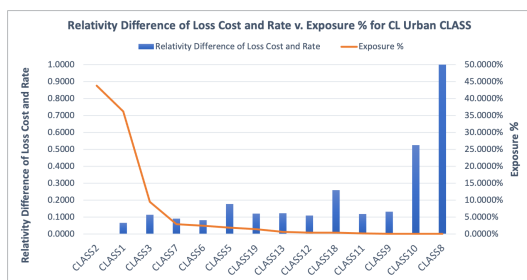
(a) Urban
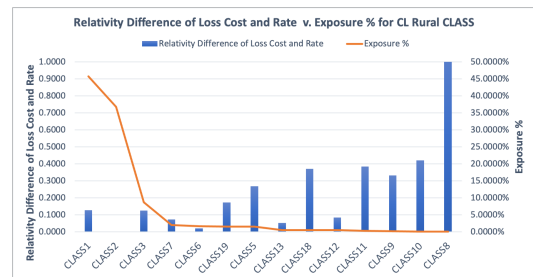
(b) Rural

Figure 3. The relativity differences between loss costs and premiums for different levels of CLASS, and the percentage of risk exposures of the corresponding levels of CLASS, for COL coverage.

We next analyze and compare the pattern of DR relativities obtained from various proposed models: one-way, two-way and three-way GLMs. Note that all three models take loss costs and premiums together as the model response, which implies the one-way model differs from the one-way analysis conducted for distance-based measures mentioned above. We find that the one-way relativity estimates for DR obtained using premium and loss cost information are much higher than those obtained from the two-way models. This may result from de-coupling the interaction between DR and CLASS, leading to less impact on premiums caused by the double risk penalty. The results, separated by Urban and Rural and by accident year, are reported in Figures 4 and 5. We also observe that the obtained results from the three-way models are similar to those from the two-way and one-way models for the case when the average loss data are used. This implies that multiple-way analysis is more appropriate for estimating the relativity of major risk factors such as DR and CLASS; therefore, it is a more suitable method for analyzing loss data to address the fairness issue. The detailed comparison among different coverages for the relativities estimates for DR, CLASS and fixed effect is reported in Tables 3–5. The fixed effect reflects the disparity of the premiums and loss costs and, therefore, can be used for measuring the fairness of premiums. The obtained fixed effect estimated for one-way, two-way and three-way models, respectively, are summarized in Tables 6–8. We observe that the estimates from different uses of the error function do not cause a significant difference, suggesting a robust estimate of such fixed effects using GLM. In all methods, collision coverage tends to have a much higher fixed effect for all accident years. This may imply that adjustment has been made more on collision coverage, which makes sense, as this coverage is directly linked to the driver's accident history, indirectly reflecting the driving habit of an insured. In addition, the fixed effects estimated from the one-way analysis behave similarly for CLASS and DR, implying that the combined fixed effect through a multiple-way analysis is appropriate.
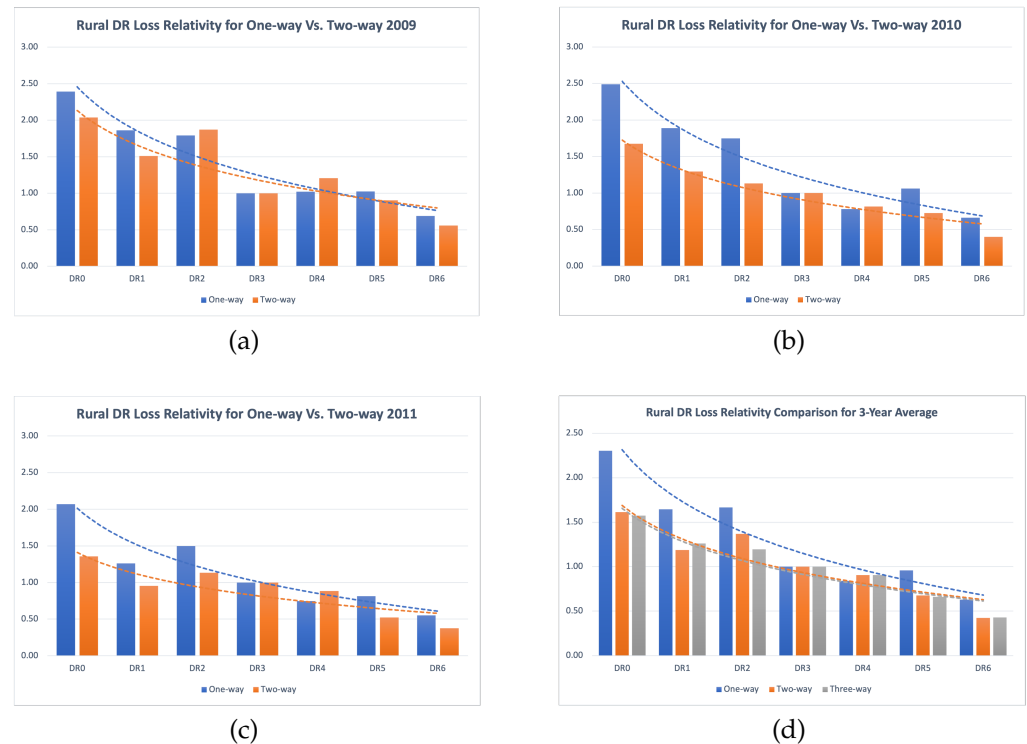

(a)

(b)

(c)

(d)
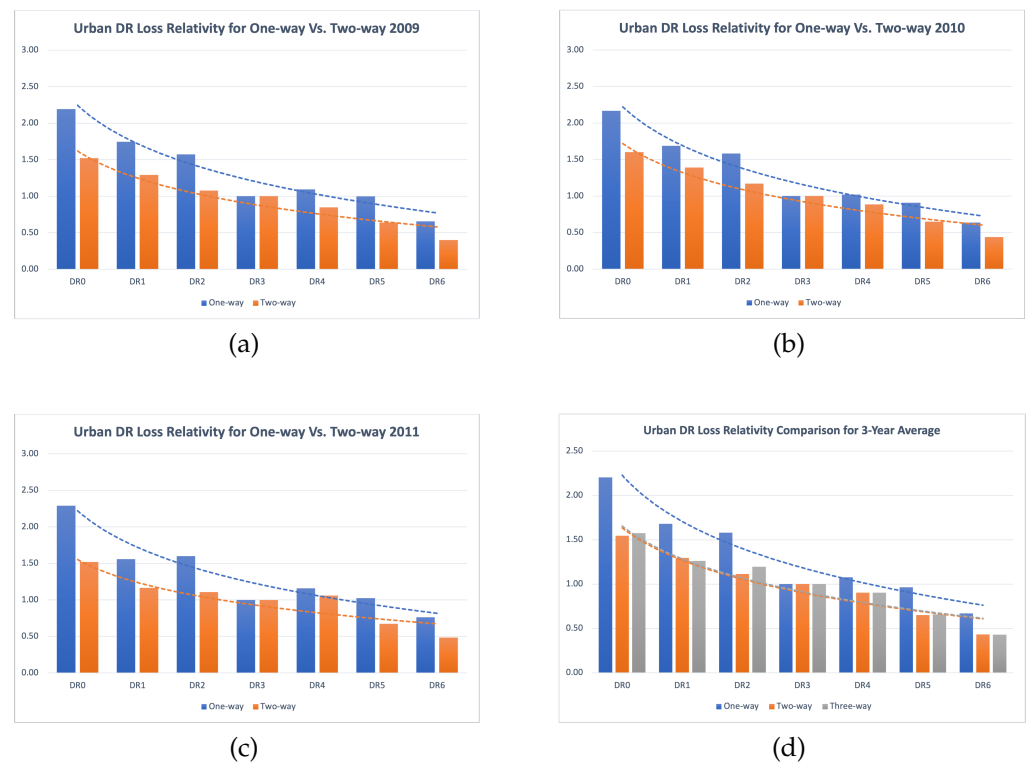
Figure 4. Comparison of DR relativity based on Urban loss data, estimated by different models including one-way, two-way and three-way models. (**a**–**c**) present the comparison between one-way and two-way analysis, respectively for year 2009, 2010 and 2011. (**d**) shows the comparisons of one way, two-way and three-way for three years average data.

In the two-way analysis, the GLM estimates the relativity associated with DR and CLASS simultaneously and the fixed effect caused by the disparity of loss costs and premiums. For within territories, the input data are separated by Territory, Rural or Urban and by the three different accident years, 2009, 2010, and 2011, respectively. In analyzing a specific example, we will examine the results obtained from Accident Benefit (AB) coverage in the Urban Territory. Loss relativities are calculated taking DR_3 and $CLASS_2$ as the bases. A fixed effect greater than one would imply that the overall premiums are higher than the loss cost, but this is what we expect, since a loading factor is applied to the loss cost to calculate the final premiums by insurance companies. Suppose the value of the fixed effect is less than one, which is the opposite of the normal condition. In that case, this may imply that the premiums charged have been significantly adjusted among coverages, accident years, territories, DR or CLASS or their combinations. In Table 7, when examining the coverage by separate accident years, we see that in 2009 and 2010, the fixed effects for AB Urban are significantly less than one. Thus, premium rates were further recovered in 2011, leading to a significant adjustment of premium rates over three years. We also observe that the fixed effects for Rural are significantly higher than those for Urban for all coverages and accident years that we consider. This evidences the fairness issue for Rural drivers, which implies that Rural drivers are overcharged for the premiums when we analyze the loss cost and premium separated by Urban and Rural. However, this two-way analysis does not consider the effect of Territory, which may significantly impact the results. This within-territory analysis is useful for the objective that aims at examining the territory-dependent fairness level based on DR and CLASS. To analyze the degree of the disparities between the premium and loss costs, we take a benchmark loss ratio of 70% (it is typical to have a loss ratio between 65% and 70% for auto insurance), which would lead to a fixed effect of 1.43. This means that we should allow a premium loading factor of 43% at an industry level and that any number smaller than 1.43 would mean a loss ratio bigger than 70%. Observing the fixed effects obtained from the two-way analysis of AB Urban, the annual average fixed effects all fall under this loss ratio. This may suggest that the overall average fixed effect obtained from a two-way model as an industry-level fixed effect is underestimated. Therefore, they are considered to be inappropriate. A fixed effect lower than 1.43 presented in the two-way analysis means that auto insurance companies would have an insufficient profit margin.

In the three-way analysis, the GLM estimates the relativities simultaneously, taking the additional explanatory variable of Territory along with DR and CLASS. The input data take Rural as the base. The data are still separated by the accident years 2009, 2010 and 2011. Loss relativities are calculated taking DR_3 and $CLASS_2$ as the base. The results are shown in Table 8. When examining fixed effects for the three-way model, we see an opposite effect to that of the two-way analysis. Overall, the Urban Territory has a higher fixed effect than the Rural Territory. The average fixed effect considering exposures shows a fixed effect higher than 1.43, which is the assumed benchmark loss ratio of 70%. This result obtained from the three-way model is more reasonable than the ones from either the one-way or two-way models. The main reason behind this is that the Territory is a major determinant of auto insurance risk and is a crucial rating variable. We can see that the Territory is a statistically significant variable in each coverage, with a *p*-value of almost zero. So, missing this critical variable from the statistical model can lead to significant bias in the fixed effect estimate and therefore an incorrect conclusion resulting from an inaccurate parameter estimate.

In Table 9, we further examine the fixed effects using the three-year average loss data, which provide a more accurate estimate of the overall fixed effects. The estimated average fixed effect amongst the error distributions for the three years is relatively stable and varied only by a small amount due to the three-year period, which leads to a more stable and reliable estimate of fixed effects. We calculate the fixed effect's weighted average for all three combined years using exposures as weight values. In Table 7, for the years 2009 and 2010, we see that the fixed effect is 1.23, meaning that the premium is being charged 23%

higher than the loss cost. This jumps significantly to 40% in 2011 (fixed effect of 1.4). Further examination of the other coverages indicates that the highest fixed effect corresponds to the COL coverage. The fixed effect is slightly over 1.8 for 2009 and 2010 and 1.66 for 2011 when looking at Rural Territory. However, when examining the fixed effect considering the 3-year average for all three coverages to see the overall difference between the premium and lost cost, the other two coverages mitigate the high fixed effect from Collision. By taking the weighted average using exposure as the weight, we can calculate a more meaningful fixed effect, as it would be misleading to compare fixed effects between coverages individually due to varying exposure amounts.

(a)

(b)

(c)

(d)

Figure 5. Comparison of DR relativity based on "Rural" loss data, estimated by different models including one-way, two-way and three-way models. (**a–c**) present the comparison between one-way and two-way analysis, respectively for year 2009, 2010 and 2011. (**d**) shows the comparisons of one way, two-way and three-way for three years average data.

Table 3. Risk relativity estimates of DR and CLASS for AB coverage using the combined data of three accident years under the three-way model.

Term	Gaussian					Poisson					Gamma				
	Estimate	Standard Error	Statistic	*p*-Value	Loss Relativity	Estimate	Standard Error	Statistic	*p*-Value	Loss Relativity	Estimate	Standard Error	Statistic	*p*-Value	Loss Relativity
(Intercept)	6.79	0.05	147.23	0.00		6.79	0.00	57,577.37	0.00		6.75	0.11	61.10	0.00	
DR0	0.00				1.54	0.00				1.53	0.00				1.57
DR1	−0.21	0.05	−4.33	0.00	1.25	−0.21	0.00	−1412.51	0.00	1.25	−0.22	0.14	−1.62	0.11	1.26
DR2	−0.36	0.05	−7.53	0.00	1.07	−0.35	0.00	−2490.09	0.00	1.09	−0.27	0.13	−2.13	0.03	1.20
DR3	−0.43	0.04	−9.82	0.00	1.00	−0.43	0.00	−3359.13	0.00	1.00	−0.45	0.12	−3.86	0.00	1.00
DR4	−0.60	0.05	−11.73	0.00	0.85	−0.58	0.00	−4188.70	0.00	0.86	−0.55	0.13	−4.41	0.00	0.91
DR5	−0.93	0.04	−20.80	0.00	0.61	−0.91	0.00	−7455.19	0.00	0.62	−0.87	0.11	−7.61	0.00	0.66
DR6	−1.37	0.04	−36.68	0.00	0.39	−1.35	0.00	−12,133.80	0.00	0.40	−1.30	0.11	−11.91	0.00	0.43
CLASS1	0.00				0.85	0.00				0.86	0.00				0.86
CLASS2	0.16	0.02	7.48	0.00	1.00	0.16	0.00	4200.70	0.00	1.00	0.15	0.02	7.47	0.00	1.00
CLASS3	0.00	0.04	0.07	0.94	0.85	0.05	0.00	810.88	0.00	0.90	0.09	0.04	2.56	0.01	0.94
CLASS5	−2.17	0.35	−6.23	0.00	0.10	−2.11	0.00	−7428.64	0.00	0.10	−2.01	0.09	−23.33	0.00	0.12
CLASS6	−1.49	0.16	−9.37	0.00	0.19	−1.42	0.00	−7674.17	0.00	0.21	−1.29	0.08	−16.34	0.00	0.24
CLASS7	−0.12	0.07	−1.63	0.10	0.76	−0.07	0.00	−629.56	0.00	0.80	−0.05	0.06	−0.86	0.39	0.81
CLASS8	0.79	0.21	3.85	0.00	1.88	0.68	0.00	871.21	0.00	1.70	0.59	0.87	0.67	0.50	1.54
CLASS9	0.47	0.11	4.45	0.00	1.36	0.43	0.00	1497.78	0.00	1.32	0.40	0.23	1.80	0.07	1.29
CLASS10	0.16	0.09	1.75	0.08	1.00	0.18	0.00	746.47	0.00	1.02	0.25	0.21	1.24	0.22	1.11
CLASS11	0.14	0.07	2.01	0.05	0.98	0.16	0.00	900.32	0.00	1.00	0.22	0.15	1.51	0.13	1.07
CLASS12	0.20	0.06	3.26	0.00	1.04	0.27	0.00	1891.16	0.00	1.12	0.41	0.11	3.61	0.00	1.29
CLASS13	0.21	0.06	3.32	0.00	1.05	0.25	0.00	1719.57	0.00	1.10	0.28	0.10	2.74	0.01	1.14
CLASS18	−0.11	0.07	−1.53	0.13	0.76	−0.06	0.00	−385.79	0.00	0.81	0.03	0.12	0.24	0.81	0.88
CLASS19	0.04	0.05	0.69	0.49	0.88	0.10	0.00	933.66	0.00	0.95	0.20	0.07	2.67	0.01	1.05
Fixed Effect	0.03	0.02	1.85	0.07		−0.03	0.00	−789.06	0.00		−0.08	0.02	−4.48	0.00	
Territory	0.61	0.03	23.33	0.00		0.62	0.00	15,838.14	0.00		0.64	0.02	33.16	0.00	

Table 4. Risk relativity estimates of DR and CLASS for TPL coverage using the combined data of three accident years under the three-way model.

Term	Gaussian					Poisson					Gamma				
	Estimate	Standard Error	Statistic	p-Value	Loss Relativity	Estimate	Standard Error	Statistic	p-Value	Loss Relativity	Estimate	Standard Error	Statistic	p-Value	Loss Relativity
(Intercept)	5.97	0.04	152.22	0.00		5.94	0.00	53,997.66	0.00		5.92	0.08	78.42	0.00	
DR0	0.00				1.49	0.00				1.49	0.00				1.51
DR1	−0.15	0.04	−3.29	0.00	1.29	−0.14	0.00	−1049.44	0.00	1.29	−0.15	0.09	−1.62	0.11	1.29
DR2	−0.21	0.04	−4.87	0.00	1.22	−0.21	0.00	−1641.89	0.00	1.20	−0.23	0.09	−2.61	0.01	1.19
DR3	−0.40	0.04	−10.06	0.00	1.00	−0.40	0.00	−3338.02	0.00	1.00	−0.41	0.08	−5.08	0.00	1.00
DR4	−0.55	0.05	−11.52	0.00	0.87	−0.55	0.00	−4140.26	0.00	0.86	−0.55	0.09	−6.41	0.00	0.87
DR5	−0.76	0.04	−18.39	0.00	0.70	−0.76	0.00	−6575.99	0.00	0.70	−0.77	0.08	−9.80	0.00	0.70
DR6	−1.06	0.03	−30.24	0.00	0.52	−1.05	0.00	−10,009.82	0.00	0.52	−1.05	0.07	−14.18	0.00	0.53
CLASS1	0.00				0.89	0.00				0.92	0.00				0.93
CLASS2	0.11	0.02	6.19	0.00	1.00	0.09	0.00	2582.05	0.00	1.00	0.07	0.01	4.69	0.00	1.00
CLASS3	0.14	0.03	4.88	0.00	1.03	0.15	0.00	2710.72	0.00	1.06	0.16	0.02	6.45	0.00	1.10
CLASS5	−1.42	0.17	−8.18	0.00	0.22	−1.39	0.00	−6840.65	0.00	0.23	−1.36	0.06	−22.13	0.00	0.24
CLASS6	−0.72	0.08	−9.08	0.00	0.43	−0.71	0.00	−5235.60	0.00	0.45	−0.65	0.06	−11.63	0.00	0.49
CLASS7	0.26	0.04	5.87	0.00	1.15	0.25	0.00	2805.29	0.00	1.18	0.24	0.04	5.56	0.00	1.19
CLASS8	0.42	0.29	1.43	0.15	1.35	0.39	0.00	439.09	0.00	1.35	0.36	0.61	0.59	0.55	1.34
CLASS9	0.50	0.10	5.10	0.00	1.46	0.48	0.00	1823.17	0.00	1.48	0.46	0.16	2.89	0.00	1.48
CLASS10	0.84	0.05	16.12	0.00	2.06	0.85	0.00	4870.39	0.00	2.14	0.87	0.14	6.02	0.00	2.23
CLASS11	0.63	0.05	14.03	0.00	1.68	0.64	0.00	4665.35	0.00	1.74	0.67	0.10	6.55	0.00	1.83
CLASS12	0.61	0.04	14.54	0.00	1.64	0.65	0.00	5553.23	0.00	1.75	0.72	0.08	8.95	0.00	1.92
CLASS13	0.50	0.05	10.79	0.00	1.48	0.50	0.00	4169.58	0.00	1.52	0.51	0.07	7.06	0.00	1.56
CLASS18	0.25	0.05	4.99	0.00	1.15	0.27	0.00	1989.33	0.00	1.20	0.30	0.08	3.53	0.00	1.26
CLASS19	0.22	0.04	5.13	0.00	1.11	0.23	0.00	2371.56	0.00	1.15	0.26	0.05	4.92	0.00	1.21
Fixed Effect	−0.22	0.02	−14.49	0.00		−0.21	0.00	−7273.30	0.00		−0.21	0.01	−16.43	0.00	
Territory	0.29	0.02	16.63	0.00		0.34	0.00	10,279.17	0.00		0.37	0.01	27.07	0.00	

Table 5. Risk relativity estimates of DR and CLASS for COL coverage using the combined data of three accident years under the three-way model.

	Gaussian					Poisson					Gamma				
Term	Estimate	Standard Error	Statistic	p-Value	Loss Relativity	Estimate	Standard Error	Statistic	p-Value	Loss Relativity	Estimate	Standard Error	Statistic	p-Value	Loss Relativity
(Intercept)	6.12	0.03	224.61	0.00		6.11	0.00	21,740.80	0.00		6.10	0.07	88.68	0.00	
DR0	0.00				1.32	0.00				1.31	0.00				1.34
DR1	−0.08	0.03	−2.46	0.01	1.22	−0.07	0.00	−192.63	0.00	1.23	−0.05	0.09	−0.57	0.57	1.27
DR2	−0.23	0.03	−7.10	0.00	1.05	−0.24	0.00	−706.65	0.00	1.03	−0.26	0.08	−3.19	0.00	1.03
DR3	−0.28	0.03	−9.74	0.00	1.00	−0.27	0.00	−890.90	0.00	1.00	−0.29	0.07	−3.96	0.00	1.00
DR4	−0.41	0.03	−12.55	0.00	0.88	−0.41	0.00	−1242.49	0.00	0.87	−0.44	0.08	−5.57	0.00	0.86
DR5	−0.53	0.03	−19.13	0.00	0.78	−0.53	0.00	−1840.77	0.00	0.77	−0.53	0.07	−7.48	0.00	0.79
DR6	−0.90	0.03	−34.84	0.00	0.54	−0.91	0.00	−3301.03	0.00	0.53	−0.90	0.07	−13.29	0.00	0.54
CLASS1	0.00				0.90	0.00				0.91	0.00				0.92
CLASS2	0.11	0.01	10.06	0.00	1.00	0.09	0.00	1348.34	0.00	1.00	0.08	0.01	7.25	0.00	1.00
CLASS3	0.17	0.02	10.77	0.00	1.06	0.18	0.00	1620.74	0.00	1.09	0.19	0.02	9.96	0.00	1.11
CLASS5	−1.33	0.08	−15.89	0.00	0.24	−1.27	0.00	−3424.25	0.00	0.26	−1.16	0.05	−25.49	0.00	0.29
CLASS6	−0.63	0.04	−16.00	0.00	0.48	−0.64	0.00	−2476.69	0.00	0.48	−0.61	0.04	−14.44	0.00	0.50
CLASS7	0.28	0.02	11.95	0.00	1.20	0.27	0.00	1505.80	0.00	1.19	0.26	0.03	7.85	0.00	1.19
CLASS8	0.82	0.17	4.92	0.00	2.04	0.84	0.00	390.00	0.00	2.10	0.88	0.69	1.27	0.21	2.21
CLASS9	0.59	0.06	9.79	0.00	1.63	0.61	0.00	1011.07	0.00	1.67	0.63	0.14	4.35	0.00	1.72
CLASS10	1.00	0.04	23.46	0.00	2.46	1.00	0.00	1706.97	0.00	2.47	1.02	0.20	5.02	0.00	2.56
CLASS11	0.84	0.03	27.23	0.00	2.09	0.83	0.00	2175.64	0.00	2.08	0.84	0.12	7.09	0.00	2.13
CLASS12	0.74	0.03	27.23	0.00	1.88	0.74	0.00	2520.96	0.00	1.91	0.77	0.08	9.59	0.00	1.98
CLASS13	0.73	0.03	27.27	0.00	1.86	0.74	0.00	2750.60	0.00	1.91	0.75	0.07	11.28	0.00	1.96
CLASS18	0.48	0.03	15.61	0.00	1.46	0.49	0.00	1520.96	0.00	1.48	0.52	0.08	6.25	0.00	1.55
CLASS19	0.44	0.02	19.86	0.00	1.40	0.47	0.00	2365.80	0.00	1.45	0.50	0.04	11.49	0.00	1.52
Fixed Effect	−0.50	0.01	−52.27	0.00		−0.50	0.00	−7957.49	0.00		−0.50	0.01	−48.62	0.00	
Territory	0.11	0.01	12.03	0.00		0.13	0.00	2033.71	0.00		0.15	0.01	14.22	0.00	

Table 6. The fixed effect estimates for DR and CLASS, separated by year, coverages and region, obtained from the one-way GLM.

	2009				2010				2011			
	Gaussian	Poisson	Gamma	Inverse Gaussian	Gaussian	Poisson	Gamma	Inverse Gaussian	Gaussian	Poisson	Gamma	Inverse Gaussian
AB												
Rural												
DR	1.24	1.22	1.21	1.20	1.44	1.45	1.45	1.45	1.41	1.42	1.44	1.45
CLASS	1.32	1.24	1.20	1.22	1.51	1.46	1.46	1.64	1.40	1.45	1.49	1.52
Urban												
DR	0.70	0.71	0.71	0.73	0.82	0.83	0.84	0.86	1.65	1.63	1.62	1.62
CLASS	0.69	0.70	0.70	0.71	0.80	0.82	0.83	0.83	1.62	1.60	1.59	1.58
TPL												
Rural												
DR	1.40	1.44	1.46	1.48	1.32	1.32	1.32	1.33	1.28	1.30	1.32	1.33
CLASS	1.45	1.46	1.44	1.42	1.35	1.32	1.31	1.29	1.26	1.33	1.35	1.36
Urban												
DR	1.40	1.44	1.46	1.48	1.15	1.16	1.17	1.17	1.33	1.29	1.27	1.26
CLASS	1.14	1.13	1.12	1.10	1.15	1.14	1.13	1.12	1.32	1.29	1.27	1.25
COL												
Rural												
DR	1.82	1.83	1.84	1.85	1.82	1.84	1.86	1.86	1.68	1.69	1.70	1.70
CLASS	1.84	1.82	1.80	1.78	1.84	1.83	1.82	1.81	1.66	1.67	1.67	1.66
Urban												
DR	1.65	1.66	1.66	1.67	1.61	1.62	1.62	1.63	1.56	1.55	1.55	1.54
CLASS	1.64	1.64	1.63	1.61	1.61	1.60	1.59	1.58	1.54	1.53	1.53	1.51

Table 7. The combined fixed effect estimates for DR and CLASS, separated by year, coverages and region, obtained from the two-way GLM.

	2009				2010				2011			
	Gaussian	Poisson	Gamma	Inverse Gaussian	Gaussian	Poisson	Gamma	Inverse Gaussian	Gaussian	Poisson	Gamma	Inverse Gaussian
AB												
Rural												
Fixed Effect	1.30	1.23	1.21	NA	1.43	1.45	1.48	NA	1.36	1.44	1.49	NA
Urban												
Fixed Effect	0.67	0.70	0.72	0.74	0.78	0.82	0.84	0.84	1.65	1.62	1.60	1.59
TPL												
Rural												
Fixed Effect	1.40	1.45	1.47	1.46	1.34	1.32	1.31	1.31	1.23	1.32	1.36	1.38
Urban												
Fixed Effect	1.14	1.14	1.12	1.10	1.14	1.15	1.15	1.13	1.35	1.29	1.26	1.24
COL												
Rural												
Fixed Effect	1.83	1.82	1.81	1.80	1.81	1.82	1.83	1.82	1.64	1.67	1.67	NA
Urban												
Fixed Effect	1.63	1.64	1.64	1.62	1.59	1.60	1.60	1.58	1.54	1.54	1.52	1.50
Average Effect	1.23	1.24	1.24	NA	1.23	1.24	1.24	NA	1.40	1.40	1.39	NA

Table 8. The fixed effect estimates for DR and CLASS, separated by year, coverages and region, obtained from the three-way GLM.

| | **2009** | | | | **2010** | | | | **2011** | | | |
	Gaussian	Poisson	Gamma	Inverse Gaussian	Gaussian	Poisson	Gamma	Inverse Gaussian	Gaussian	Poisson	Gamma	Inverse Gaussian
AB												
Rural Fixed Effect	0.72	0.79	0.86	NA	0.84	0.93	1.02	NA	1.61	1.57	1.55	NA
Urban Fixed Effect	1.54	1.60	1.71	NA	1.71	1.85	2.08	NA	2.44	2.45	2.47	NA
TPL												
Rural Fixed Effect	1.20	1.21	1.23	0.98	1.18	1.20	1.20	1.20	1.33	1.30	1.29	1.28
Urban Fixed Effect	1.70	1.79	1.89	1.00	1.62	1.68	1.71	1.74	1.66	1.74	1.80	1.84
COL												
Rural Fixed Effect	1.69	1.70	1.70	1.69	1.66	1.67	1.68	1.68	1.57	1.58	1.57	1.56
Urban Fixed Effect	1.89	1.94	1.97	1.98	1.86	1.93	1.98	2.02	1.75	1.79	1.82	1.83
Average Effect	1.50	1.56	1..63	NA	1.50	1.55	1.61	NA	1.69	1.72	1.75	NA

Table 9. The comparison of fixed effect between two-way and three-way analysis for three years of combined loss data, separated by coverages and region.

| | **Two-Way Combined** | | | **Three-Way Combined** | | |
	Gaussian	Poisson	Gamma	Gaussian	Poisson	Gamma
AB						
Rural Fixed Effect (CLASS&DR)	1.38	1.37	1.37	0.97	1.03	1.09
Urban Fixed Effect (CLASS&DR)	0.91	0.94	0.95	1.78	1.90	2.06
TPL						
Rural Fixed Effect (CLASS&DR)	1.34	1.36	1.37	1.24	1.24	1.24
Urban Fixed Effect (CLASS&DR)	1.21	1.19	1.17	1.66	1.73	1.79
COL						
Rural Fixed Effect (CLASS&DR)	1.76	1.77	1.77	1.64	1.65	1.65
Urban Fixed Effect (CLASS&DR)	1.59	1.59	1.58	1.83	1.88	1.92
Average Fixed Effect	1.28	1.28	1.27	1.54	1.60	1.65

4. Conclusions

The study of actuarial fairness in auto insurance is an important issue in the decision making of rate regulation. However, given that insurance companies have certain flexibility for adjusting the risk relativity to control the operation process and management better, the relativities of risk factors may not coincide with the ones from the regulator. The rate-making methodologies may differ from regulators to insurance companies as their rate-making goals tend to differ. This will suggest that analysis and modeling results from companies may be justifiable and considered statistically sound even though a significant difference exists among their estimates. In addition, determining the relationship between

the premium charged and loss cost and under what condition one can say the overall premium charged is considered fair and exact is a complex problem. This work has focused on measuring the disparity of loss costs and premiums using various statistical models, including one-way, two-way and three-way GLM. The shortcoming of the one-way analysis is that it assumes the variables we analyze are independent of each other. This means that modeling the single variable approach does not consider the effect of the other variables. Although the two-way analysis of DR and CLASS may achieve a de-coupling of the potential interaction, the missing territory variable completely changed the direction of the fixed effect. In the two-way approach, the fixed effect corresponding to Rural areas is deemed higher than that in Urban, but the opposite result was obtained in the three-way approach. This can be due to the disproportionate exposure amounts, where Urban has much higher exposure. However, a multiple-way is subject to its own limitation, and it is not useful if there are not enough risk exposures for the classification of risk. For example, adding more rating variables may decrease the number of exposures, which would mean the number of risks would be too low to estimate the costs accurately. This leads to a lack of credibility. Therefore, more rating variables included in the model could lead to less credibility. Therefore, we can conclude that the three-way analysis is the most appropriate model.

The results suggest a significant disparity between loss costs and premiums charged when focusing on particular groups of drivers. AB and TPL coverages for Rural drivers seem to have a small fixed effect, which implies that the loading applied to them is much smaller than other groups. However, the loading applied to Urban drivers in these two coverages seems excessive. For COL coverage, both Rural and Urban drivers have a high fixed effect, which means the loading is high, particularly for Urban drivers. From this study, we observe that Urban drivers tend to have higher loading than Rural drivers, but overall, the study using this Canadian data set does not reveal a significant excessive loading. The innovation of this work is the modeling techniques used to simultaneously estimate the relativities of major risk factors and the difference caused by loss cost and premium information. The impact of this work is to provide a statistical framework that insurance regulators can use to measure the fairness of insurance premiums across different groups of insureds. Although beyond the scope of this report, the estimated fixed effect can be useful for policymakers to decide how much to lower auto insurance premiums by identifying how much room there is between the premiums and loss costs.

Author Contributions: Conceptualization, S.X.; methodology, S.X.; software, R.L.; validation, S.X. and R.L.; formal analysis, S.X. and R.L.; investigation, S.X., R.L. and Y.L.; resources, S.X. and R.L.; data curation, S.X. and R.L.; writing—original draft preparation, S.X. and Y.L.; writing—review and editing, S.X. and Y.L.; visualization, S.X. and R.L.; supervision, S.X.; project administration, S.X. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

Abraham, Kenneth S. 1985. Efficiency and fairness in insurance risk classification. *Virginia Law Review* 403–51. [CrossRef]

Arora, Nidhi, and Poonam Arora. 2014. Insurance Premium Optimization: Perspective of Insurance Seeker and Insurance Provider. *Journal of Management and Science* 4: 43–53. [CrossRef]

Avraham, Ronen. 2017. Discrimination and Insurance. The Routledge Handbook to Discrimination Lippert-Rasmussen Ed, University of Texas Law, Law and Econ Research Paper No. E574. Available online: https://ssrn.com/abstract=3089946 or http://dx.doi.org/10.2139/ssrn.3089946 (accessed on 4 July 2022).

Barry, Laurence. 2020. Insurance, big data and changing conceptions of fairness. *European Journal of Sociology/Archives Européennes de Sociologie* 61: 159–84. [CrossRef]

Berry-Stölzle, Thomas R., and Patricia Born. 2012. The effect of regulation on insurance pricing: the case of Germany. *Journal of Risk and Insurance* 79: 129–64. [CrossRef]

Cao, Longbing, Qiang Yang, and Philip S. Yu. 2021. Data science and AI in FinTech: An overview. *International Journal of Data Science and Analytics* 12: 81–99. [CrossRef]

Charpentier, Arthur, Laurence Barry, and Molly R. James. 2022. Insurance against natural catastrophes: balancing actuarial fairness and social solidarity. *The Geneva Papers on Risk and Insurance-Issues and Practice* 47: 50–78. [CrossRef]

Chiappori, Pierre-André, Bruno Jullien, Bernard Salanié, and François Salanié. 2006. Asymmetric information in insurance: General testable implications. *The RAND Journal of Economics* 37: 783–98. [CrossRef]

Cohen, Alma. 2005. Asymmetric information and learning: Evidence from the automobile insurance market. *Review of Economics and Statistics* 87: 197–207. [CrossRef]

Cummins, J. David, and Mary A. Weiss. 1992. Regulation and the Automobile Insurance Crisis. *Regulation* 15: 48.

Dugas, Charles, Yoshua Bengio, Nicolas Chapados, Pascal Vincent, Germain Denoncourt, and Christian Fournier. 2003. Statistical learning algorithms applied to automobile insurance ratemaking. In *CAS Forum*. Arlington: Casualty Actuarial Society, Vol. 1, pp. 179–214.

Ferreira, Joseph, Jr., and Eric Minikel. 2012. Measuring per mile risk for pay-as-you-drive automobile insurance. *Transportation Research Record* 2297: 97–103. [CrossRef]

Frees, Edward W., and Fei Huang. 2021. The discriminating (pricing) actuary. *North American Actuarial Journal* 1–23. [CrossRef]

Frezal, Sylvestre, and Laurence Barry. 2020. Fairness in uncertainty: Some limits and misinterpretations of actuarial fairness. *Journal of Business Ethics* 167: 127–36. [CrossRef]

Grabowski, Henry, W. Kip Viscusi, and William N. Evans. 1989. Price and availability tradeoffs of automobile insurance regulation. *Journal of Risk and Insurance* 56: 275–99. [CrossRef]

Hanafy, Mohamed, and Ruixing Ming. 2021. Machine learning approaches for auto insurance big data. *Risks* 9: 42. [CrossRef]

Isotupa, K. P. Sapna, Mary Kelly, and Anna Kleffner. 2019. Experience-rating mechanisms in auto insurance: Implications for high-risk, low-risk, and novice drivers. *North American Actuarial Journal* 23: 395–411. [CrossRef]

Landes, Xavier. 2015. How fair is actuarial fairness? *Journal of Business Ethics* 128: 519–33. [CrossRef]

Liu, Lixin, Wenzhuo Li, Wu He, and Justin Zuopeng Zhang. 2022. Improve enterprise knowledge management with internet of things: a case study from auto insurance industry. *Knowledge Management Research & Practice* 20: 58–72.

Medders, Lorilee A., Jamie A. Parson, and Matthew Thomas-Reid. 2021. Gender X and Auto Insurance: Is Gender Rating Unfairly Discriminatory? *Journal of Insurance Regulation* 40: 1–31. [CrossRef]

Meyers, Gert, and Ine Van Hoyweghen. 2018. Enacting actuarial fairness in insurance: From fair discrimination to behaviour-based fairness. *Science as Culture* 27: 413–38. [CrossRef]

Regan, Laureen, Sharon Tennyson, and Mary Weiss. 2008. The Relationship Between Auto Insurance Rate Regulation and Insured Loss Costs: An Empirical Analysis. *Journal of Insurance Regulation* 27.

Ronka-Chmielowiec, Wanda, and Ewa Poprawska. 2005. Selected Methods of Credibility Theory and its Application to Calculating Insurance Premium in Heterogeneous Insurance Portfolios. In *Innovations in Classification, Data Science, and Information Systems*. Berlin and Heidelberg: Springer, pp. 490–97.

Ryan, Stephen R. 1986. Elimination of Gender Discrimination in Insurance Pricing: Does Automobile Insurance Rate Without Sex. *Notre Dame L. Rev.* 61: 748.

Saito, Kuniyoshi. 2006. Testing for asymmetric information in the automobile insurance market under rate regulation. *Journal of Risk and Insurance* 73: 335–56. [CrossRef]

Schmeiser, Hato, Tina Störmer, and Joël Wagner. 2014. Unisex insurance pricing: Consumers' perception and market implications. *The Geneva Papers on Risk and Insurance-Issues and Practice* 39: 322–50. [CrossRef]

Sud, Keshav, Pakize Erdogmus, and Seifedine Kadry, eds. 2020. *Introduction to Data Science and Machine Learning*. Norderstedt: BoD–Books on Demand.

Thiery, Yves, and Caroline Van Schoubroeck. 2006. Fairness and equality in insurance classification. *The Geneva Papers on Risk and Insurance-Issues and Practice* 31: 190–211. [CrossRef]

Xin, Xi, and Fei Huang. 2022. Anti-Discrimination Insurance Pricing: Regulations, Fairness Criteria, and Models. *Fairness Criteria, and Models.* Available online: https://ssrn.com/abstract=3850420 (accessed on 4 July 2022). [CrossRef]

Zahi, Jamal. 2021. Non-life insurance ratemaking techniques. *International Journal of Accounting, Finance, Auditing, Management and Economics* 2: 344–61.