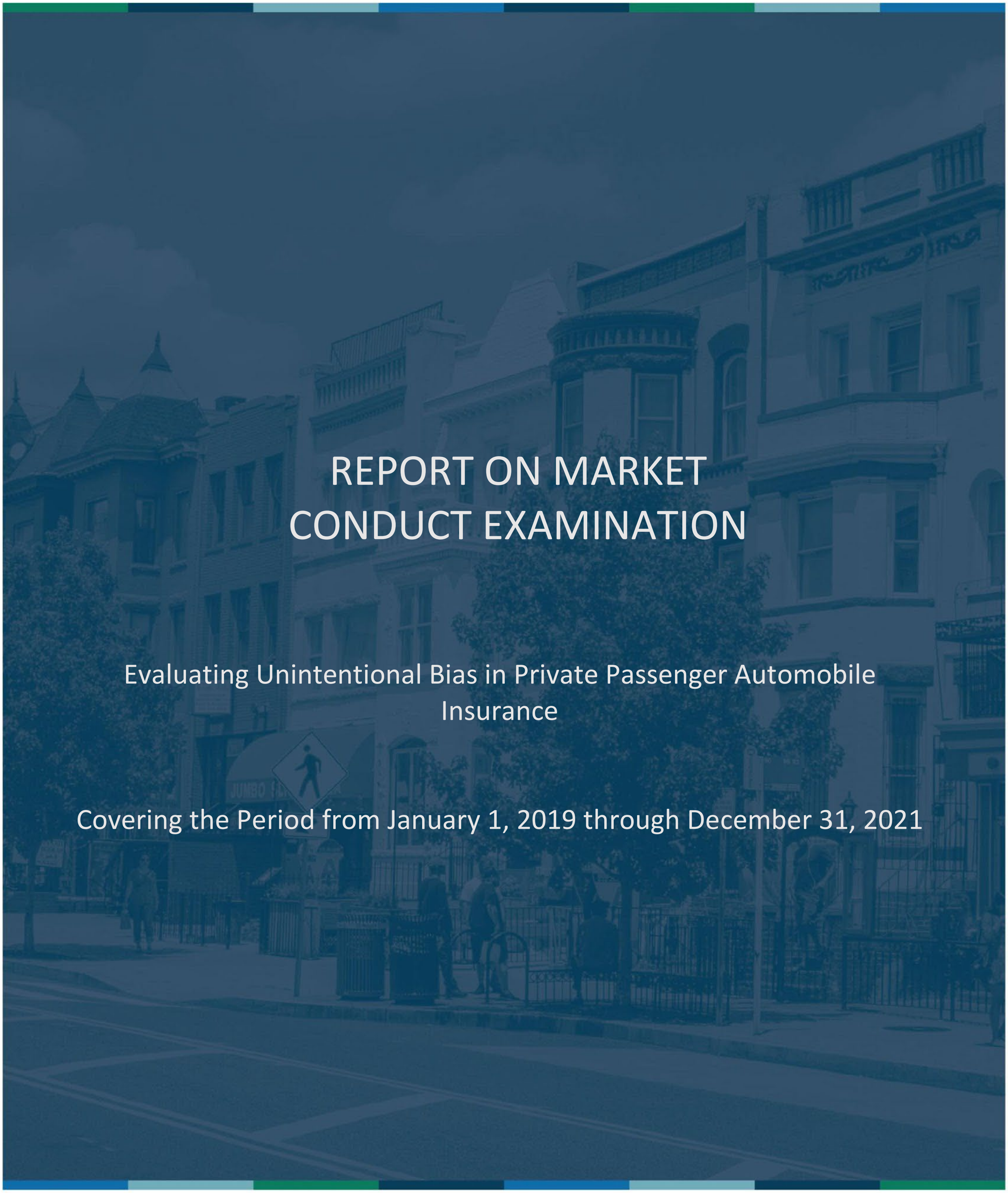
DC DEPARTMENT OF INSURANCE, SECURITIES AND BANKING



REPORT ON MARKET CONDUCT EXAMINATION

Evaluating Unintentional Bias in Private Passenger Automobile Insurance

Covering the Period from January 1, 2019 through December 31, 2021

DISB | GOVERNMENT OF THE DISTRICT OF COLUMBIA
MURIEL BOWSER, MAYOR

Contents

Executive Summary

Black and Hispanic drivers who have obtained private passenger auto insurance ("insureds") are paying higher premiums, and we have identified several actual or potential reasons why that is happening, but fundamentally, we do not believe that race is an attribute of increased risk. There may be systemic issues that lead to higher losses that are correlated with race and there may also be rating or other policy characteristics that result in higher premiums for Black and Hispanic insureds because they are correlated with race. This investigation uncovered differences about how much particular racial groups paid in premiums along with the discrepancies in insurance claims among racial groups. In essence, there may be an unintentional bias that causes Black and Hispanic drivers to pay more in premiums than white and API drivers, and there may be specific reasons why Black drivers represented more claims on average than white, Hispanic or API drivers.

During this review, DISB received information and comments about issues that affect the Black/white average premium gap, but that may not be resolved by insurance rating reform. DISB is exploring other ways to address these issues. Most notably, regarding the significantly higher losses for Black and Hispanic insureds, both in dollar terms and relative to premiums paid than for white insureds, additional study is warranted on the types and causes of claims by Black and Hispanic drivers to see if infrastructure or other changes may help reduce the claim differential.

Additionally Black and Hispanic drivers are more likely to have driving infractions, but it is not clear as to why these infractions occur. Driving record is a practical and reasonable criterion to use in car insurance rating, because not only does it have a rational explanation for the risk it is addressing, but it also provides potential risk mitigation benefits. This is another area that may benefit from additional study.

What explains the average premium difference by race? There is no evidence that Black and Hispanic drivers are riskier while operating a vehicle or more accident prone. However, there are a variety of factors that may explain this premium disparity including credit-based insurance scores and homeownership discounts. Additionally, some insurers price auto premiums using predictive models. Since premiums are determined mechanically, there may be some factors and/or steps in those models that correlate with race in a way that produces different average premiums by race for DC drivers.

For example, credit-based insurance scores, used by some auto insurance companies in the District, may be correlated with race.[1] Similarly, discounts, like homeownership discounts, may be correlated with race.[2] Because data on these characteristics was not collected for this review, this study cannot conclude how much of the gap could be explained by them. Whether DISB would allow such factors to explain the average premium gap, given their correlation with race, is a separate question – one that

[1] See Figures 8 and 9 in Credit-Based Insurance Scores: Impacts on consumers of automobile insurance, FTC 2007, which display the distribution of ChoicePoint credit-based insurance scores by race, and show that African Americans and Hispanics are overrepresented (compared to Whites and API) at the lower end of the distribution. Note that this fact is completely independent of the issue of how credit-based insurance scores relate to expected risk.

[2] E.g., "Homeownership is profoundly stratified by race and ethnicity, as is health. Approximately 74% of White households own their homes, compared to only 45% of Black and Latino households (U.S. Census Bureau 2013)." from Finnigan R. Racial and ethnic stratification in the relationship between homeownership and self-rated health. Soc Sci Med. 2014 Aug; 115:72-81. doi: 10.1016/j.socscimed.2014.06.019. Epub 2014 Jun 13. PMID: 24953499; PMCID: PMC4301401.

could be addressed by a "balancing test," as outlined in the Analytic Approach section.

One such opportunity for further analysis: at a recent Congressional hearing,[3] Mr. Robert Gordon, Senior Vice President, American Property and Casualty Insurance Association (APCIA) indicated the APCIA has studies showing a direct connection between credit scores and actual hard braking and hard acceleration driving behavior. Gordon also indicated they have studies showing credit scores are not discriminatory, but DISB believes the referenced study does not look at the racial impact of credit scores on insurance premiums. DISB believes that hard braking and hard acceleration provide a rational explanation for expected losses and would allow them as factors to explain the average premium gap. However, when a proxy for hard braking and hard acceleration is used, the insurer should also be required to consider a balancing test like that mentioned above.

The DISB examination, explained in detail in this report, found that inferred Black drivers pay 1.46 times as much as inferred white drivers, whereas inferred Hispanic drivers pay 1.20 times as much, and inferred API drivers pay 1.02 times as much. In dollar terms, that translates into the average annual premium of $705 for white drivers, $1,031 for Black drivers, $849 for Hispanic drivers, and $722 for API drivers. This will be referred to as a "Black/white average premium gap" of $326.

The examination also looked into cumulative paid losses by race, and found that Black drivers, as a group, represented more claims on average than white, Hispanic, or API drivers. In fact, Black drivers' average losses were 2.38 times that of white drivers. From the actuarial perspective, this means Black drivers are more costly as a group than the other groups, because although their premiums are high (relative to white drivers), their losses are even higher (relative to white drivers).

In undertaking this analysis, however, DISB did not expect the exact same premiums for different race groups. Rather, DISB, with input from stakeholders, sought to identify factors that could explain the average premium gaps, and specifically the Black/white average premium gap. For example, DISB looked at age, type of policy, driving record, claim history, gender, and many others factors to determine if and how those factors impact premiums. When accounting for all these factors, however, there was still an average premium gap of $271 between Black and white drivers.

Because we have no data on the driving skills of residents based on race of the drivers, a commonsense approach to understanding the results of the study is for DISB to take further action, including, but not limited to, conducting further study and deeper analysis to determine if there are social and systemic reasons for this result.

[3] The Factors Influencing the High Cost of Insurance for Consumers, Thursday, November 2, 2023 2:00 PM, Subcommittee on Housing and Insurance at approximately 1:51 (available at: financialservices.house.gov/calendar/eventsingle.aspx?EventID=409012:).

Background

In 2020, Karima Woods, Commissioner for the District of Columbia Department of Insurance, Securities and Banking (DISB), directed the creation of DISB's first Diversity, Equity, and Inclusion Committee to engage in a wide-ranging review of financial equity and inclusion and to make recommendations to remove barriers to accessing financial services. DISB staff developed draft initiatives, including an initiative related to insurers' use of factors such as credit scores, education, occupation, home ownership and marital status in underwriting and ratemaking. Stakeholder feedback on this draft initiative resulted in DISB concluding that data was necessary to properly address this initiative. DISB staff conducted research and contacted subject matter experts before determining that relevant data was not generally available.

DISB undertook this project to collect the relevant data. Through the data call, insurers provided data on car insurance premiums in Washington, DC for one-vehicle/one-driver policies, inferring race with the BIFSG methodology[4]. For reference, one-vehicle/one-driver policies represent 64% of all auto insurance policies in the District and 54% of aggregate premium.[5]

DISB determined this initiative will be deliberative and transparent to ensure the resultant data would address the issue of unintentional bias. DISB also decided to focus initially on private passenger automobile insurance, as that line of insurance affects many District consumers, and DISB's staff have questioned insurers' use of non-driving factors in their rating and underwriting practices.

For this project looking at the potential for unintentional bias in auto insurance, DISB conducted a review of auto insurers' rating and underwriting methodologies. As a first step, DISB held a public hearing on Wednesday, June 29, 2022, to gather stakeholder input on the review plan. DISB engaged the services of O'Neil Risk Consulting and Algorithmic Auditing (ORCAA) to assist and provide subject matter expertise. Additionally, DISB held follow up meetings with industry and consumer representatives to address any items requiring additional information and provided exposures for public comment. Finally, DISB staff reached out to subject matter experts during the conduct of this examination to address specific issues. All details of meetings, exposures and comments are on our website — disb.dc.gov/page/evaluating-unintentional-bias-private-passenger-automobile-insurance.

Motivation

DISB wanted to explore whether the use of certain information by auto insurers in the application and underwriting process may cause unintentional harm to those who are Black, indigenous, people of color, or belong to another protected class of Washington, DC consumers. Examples of such information that may be considered proxies for race include: credit scores, education, home ownership, occupation, and marital status.

This analysis builds off earlier work undertaken by DISB. Previous work included a market conduct examination that looked at specific insurers' models used for auto insurance. From that examination,

---

[4] See Appendix A for a discussion of this methodology.

[5] We did a preliminary review of premiums for multi-vehicle/multi-driver policies. Average premiums were about 50% higher overall than for one-vehicle/one-driver policies ($1,313 vs $884), which makes sense given that these policies have more exposures. Importantly, the distribution of average premium by race was similar to one-vehicle/one-driver policies.

DISB concluded the models reviewed were working as intended and did not include any factors that would directly introduce bias into the rating process. However, the review could not determine if unintentional bias was present.

All private passenger automobile insurance rates are reviewed and approved by DISB actuaries to our statutory standard that the rates are not excessive, inadequate, or unfairly discriminatory. That includes the rates analyzed as part of this examination. However, DISB is responsible for maintaining a well-functioning market for insurance in DC, not just ensuring the actuarial appropriateness of premiums. DISB has taken actions previously to address potential racial bias in auto insurance premiums through our longstanding prohibition on territorial ratings. This examination looked at premiums using methodologies that have been developed more recently to evaluate potential bias in the use of algorithms.

Analytic Approach

This section discusses the report's main analysis: what DISB measured (the average premium difference by race) and why, how the measurement accounted for certain factors, and why those factors were chosen.

This is not an actuarial analysis, nor is it intending to be. It does not seek to determine whether particular rating factors are biased; DISB's previous examination addressed that. Rather, this analysis investigates whether the aggregate effect of correlations between rating factors and race, as reflected in premiums, could amount to a bias impact. This is the potential harm: that the use of rating factors correlated with race may result in members of a protected class paying higher premiums, particularly given the imperfection inherent in rating classification systems.

As for which protected classes to consider, this is inherently a subjective determination, in this case by the regulator. For this review, because of its historical significance, perceived potential impact on outcomes and the significance for residents of the District of Columbia, DISB chose to focus on racial bias, and for simplicity, much of the discussion will address the average premium difference between Black drivers and white drivers. [6]

DISB notes that none of the insurance companies provided any indication in their answers to the qualitative questions in the data call of taking any action before using rating factors to minimize the potential bias or even establish whether the factor is correlated with race such that certain race groups are more likely to be included in high-risk categories.

[6] DISB is concerned with racial equity among all groups, not just Black and white. In this first analysis and report we focused on the Black-white premium gap because they are by far the largest racial/ethnic groups in DC ("white alone" being 39.6% of the DC population and "Black alone" being 41.4%, as of the 2020 Census), and given the history of this inquiry by DISB.

The critical question is, how can DISB test for this potential harm? Or, to frame it positively, how do we define racially equitable? We mean:

The distribution of premium is "*racially equitable*" if:
[the difference] between race groups in average premium
is [small enough],
after accounting for [allowed factors].

Below we discuss the bracketed terms in this definition.

How is the difference measured? Linear regression

This analysis measures the differences using linear regression. DISB chose this technique because it widely used, relatively simple, and understandable. DISB does not contend this is the only possible approach; the difference could be measured using other techniques, e.g. Generalized Linear Models (GLMs)[7]. Other techniques make fewer and/or different assumptions about the structure of the underlying data, yet produce results that are arguably more difficult to interpret or explain.

Nonetheless DISB is open to incorporating additional measurement approaches in future analyses of this kind, as a robustness test.

How small is 'small enough'? No threshold for now

The analysis done so far has no threshold of acceptability. The goal was simply to measure the average premium difference by race.

Looking ahead: Once a measurement approach is developed and a baseline measurement taken, DISB may consider normative questions. Does the difference need to disappear entirely? Or at least shrink to a modest enough size that it is acceptable?

Thresholds may be defined in relative terms (e.g., "Insurers for whom the difference is above the median among DC insurers, must take some action…") and may move over time (e.g., there could be an absolute limit on size of the difference, which tightens year over year).

Accounting for differences: 'allowed factors'

In the context of this report's analysis, 'allowed factors' are characteristics (of drivers, their vehicles, or their insurance policies) that are used to explain some of the variation across race groups in premium. Concretely, they are used as control variables in regressions modeling premium.

[7] Linear regression is a statistical modeling tool that predicts the relationship between unknown and known data points based on explanatory variables. A GLM is a type of linear regression that can examine a broad range of data types and relationships using error distributions and link functions.

DISB recognizes that since insurers follow detailed rating plans, for a given carrier it is possible to fully explain the premium for a given policy, simply by applying the relevant rating plan to the characteristics of that policy (including its drivers and vehicles). DISB also recognizes, as noted in "Motivation", that it reviewed and approved the rating plans from which the premiums in this analysis arise.

That notwithstanding, DISB asserts that at the marketplace level premiums may be racially inequitable. To determine whether they are, DISB seeks a way of comparing premiums among like-risks, across the marketplace. This notion of like-risks needs to be common across carriers, and meaningful to consumers as an explanation for premiums The 'allowed factors' in the main analysis in this report were chosen, with these criteria in mind, among factors suggested by stakeholders' in response to DISB's July 19, 2022 request for comments.

There are three ways a factor would be allowed by DISB:

1. Factors with a "rational explanation" for the relationship with losses. According to the NAIC Casualty Actuarial and Statistical Task Force's Regulatory Review of Predictive Models white paper:

    > A "rational explanation" refers to a plausible narrative connecting the variable and/or treatment in question with real-world circumstances or behaviors that contribute to the risk of insurance loss in a manner that is readily understandable to a consumer or other educated layperson. A "rational explanation" does not require strict proof of causality but should establish a sufficient degree of confidence that the variable and/or treatment selected are not obscure, irrelevant, or arbitrary. [8]

2. Factors for which there is no expectation of a strong correlation with race, such as gender.

    Not used in this analysis, but looking ahead:

3. Factors that pass a "balancing test," which is to say (1) the factor is correlated both with race and with claim frequency and/or severity (i.e., risk), and (2) its contribution to estimating frequency/severity outweighs its functioning as a proxy for race. It was not a goal of this analysis to conduct such a test, but designing and calibrating an appropriate balancing test for the DC auto insurance context would be a significant and worthwhile project.

Scope of analysis

DISB reviewed recent applications for auto insurance from consumers who reside in Washington, DC.The review required all carriers writing private passenger auto policies in the District of Columbia to submit data from recent applications to DISB for testing through a data call. This report is based on a review of the data in aggregate. The elements of the data call were developed based on input at a

---

[8] Casualty Actuarial and Statistical (C) Task Force Regulatory Review of Predictive Models White Paper, Adopted by the Property and Casualty Insurance (C) Committee, 12/8/20, https://content.naic.org/sites/default/files/CA-WP-20.pdf

public hearing and appropriate follow up.

Since insurance carriers do not collect information about applicants' races or ethnicities, this information was inferred for the limited purpose of the testing in this review. ORCAA used the Bayesian-improved firstname surname geocoding (BIFSG) methodology to infer the race of policyholders and applied it to all insurers in the same way. Appendix A of the report provides additional information about the BIFSG methodology.

Additional comments

DISB did not require detailed information about carriers' underwriting or pricing models, such as a description of the models' structures, lists of variables used, and their weights. The focus was on the outcomes of these models.

The analysis relied on data provided by the examined insurers. Some data was excluded if elements necessary for the analysis were missing or corrupted. Neither DISB nor ORCAA audited the underlying data for accuracy or completeness.

DISB did not have sufficient information to evaluate two of the identified rating factors (vehicle make and model) and recognize that individual companies may have additional rating factors that are not reflective of unintentional bias and companies vary in the way factors are used in developing rates. In light of this variation across insurers, DISB had to decide how to treat certain factors in this analysis so data from multiple insurers could be aggregated. For example, the driving records data insurers submitted to DISB contained many distinct coding schemes, reflecting that different insurers classify driving histories differently; to enable aggregation, in this analysis DISB structures driving history as a categorical variable with 3 values.

The Study

DISB engaged ORCAA in connection with the work of its Diversity Equity and Inclusion Committee, as described above. ORCAA's dual mission is to help define accountability for algorithms, and to keep people safe from harmful consequences of AI and automated systems. Whether it's a hiring algorithm, healthcare AI, predictive scoring system, or generative AI platform, ORCAA thinks about how it could fail, for whom, and what can be done to monitor and mitigate these risks. Cathy O'Neil, CEO of ORCAA, has been an independent data science consultant since 2012 and has worked for clients including the Illinois Attorney General's Office and Consumer Reports. She wrote the book *Doing Data Science* in 2013 and *Weapons of Math Destruction: How Big Data Increases Inequality And Threatens Democracy*, released in September 2016. ORCAA's prior work in insurance includes assisting the Colorado Division of Insurance with implementation of SB21-169 — Protecting Consumers from Unfair Discrimination in Insurance Practices.

Following input submitted by stakeholders in writing and at a June 29, 2022, public hearing, DISB published a memo[9] further explaining its plans for this review and requesting comments. The memo includes a preliminary list of outcomes to be studied (Quoted prices; Underwriting decisions; Premiums;

[9] disb.dc.gov/sites/default/files/dc/sites/disb/page_content/attachments/request-comment-unintentional-bias-auto-insurance.pdf

Loss ratio)[10] and a preliminary list of factors that DISB believes provide a rational explanation for differences across race groups in these outcomes.

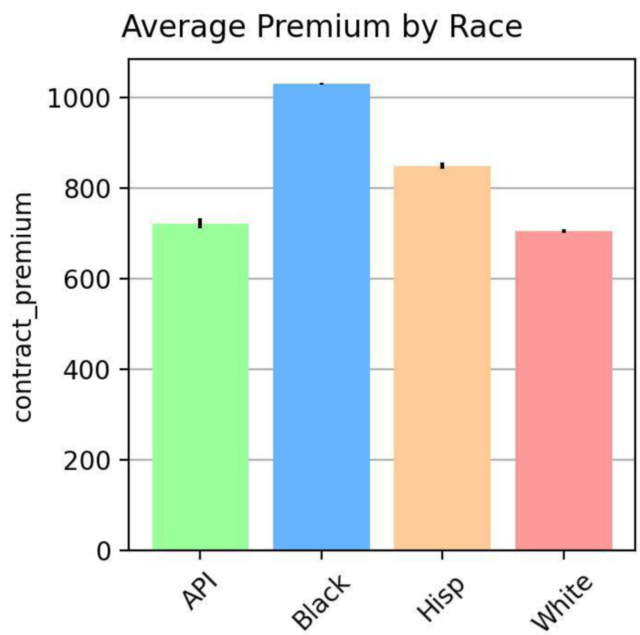
Throughout the process, a wide range of stakeholders including insurers, regulators, consumer advocates, and professional organizations provided input and answered questions that strengthened this study. We acknowledge and appreciate their assistance.

The sections below describe the work performed by ORCAA to evaluate the data. DISB and ORCAA staff met weekly throughout the analysis and DISB staff provided input into the analysis process throughout the review.

1. Premium Gap by Race

After narrowing the dataset to one-vehicle/one-driver policies, and applying BIFSG to infer the "most likely race" of the insured, ORCAA studied average premiums by race and made the following graph:

CHART 1



The average annual premium is $705 for white drivers, $1,031 for Black drivers, $849 for Hispanic drivers, and $722 for API drivers. This shows a "Black/white average premium gap" of $326.

[10] This report highlights the analysis of Premium, as price paid is a highly salient outcome for consumers, and the data provided on this outcome was relatively complete and commensurable across carriers. Data on Losses was also incorporated, as discussed in Sections 3 and 5. The data received on Quotes was noisier, but generally consistent with Premiums; it is discussed in Appendix B. While data on Underwriting decisions was collected in the data call, DISB determined the information was covered by the Quote and Premium analysis, in the sense that ratings and classifications are reflected in the prices consumers see and pay.

2. Premium Compared to Losses, by Race[11]

ORCAA computed the overall losses by race and noted that on average, Black policyholders pay more in premium compared to white policyholders — a factor of 1.44 (=1024.42 / 709.96) — but generate 2.38 (=611.54 / 256.49) times the losses, on average. This means Black drivers are more costly as a group than the other groups, because although their premiums are high (relative to white drivers), their losses are even higher (relative to white drivers):

TABLE 1

Inferred race	Average premium ($)	Average Loss ($)	Aggregate losses / Aggregate premiums
API	734.54	279.76	0.38
Black	1024.42	611.54	0.60
Hispanic	858.82	370.32	0.43
White	709.96	256.49	0.36

3. Explanatory Factors

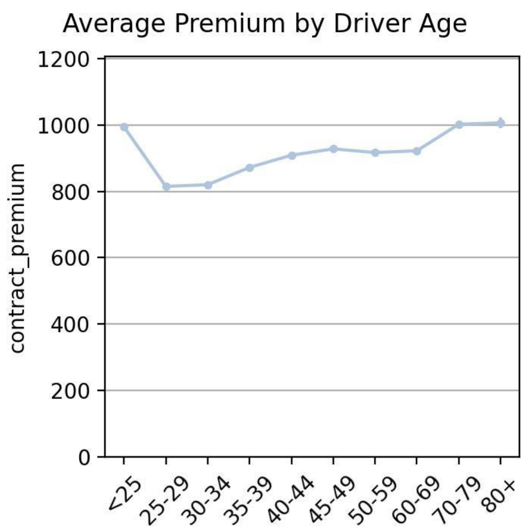
DISB and ORCAA looked for explanations for why premiums were higher for Black drivers. The method is shown here for four of the factors investigated: age of policyholder, driving record of policy holder, gender of policyholder, and age of vehicle. Appendix C includes a full list of factors investigated.[12] The list of factors was developed by DISB based on input from our July 19, 2022 request for comments.

First, ORCAA takes a given factor and groups its values (into "bins") then looks at the average premium within each bin:

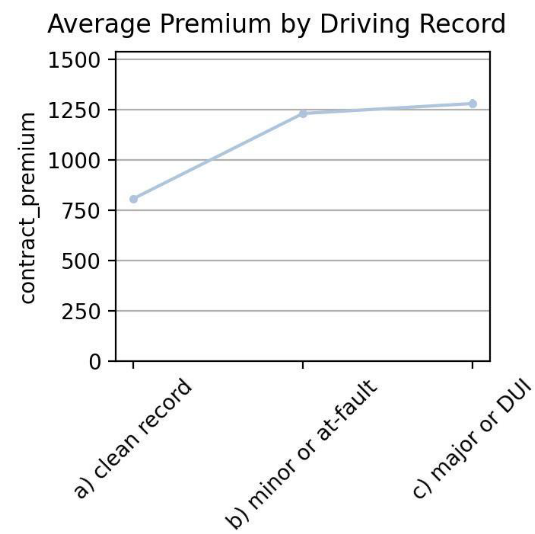
[11] This analysis could only be conducted on a subset of the data because (1) some carriers submitted data to DISB in a format that did not permit linkage of claims data with premium amounts; and (2) policies written after 2020 were excluded from the analysis since their claims experience is relatively immature. The analysis in this section comprises 237,595 policies from 2019-2020 – a subset of the policies in the previous section's analysis of premiums.

[12] DC does not allow territorial ratings; so, territory is not an explanatory factor in this analysis.

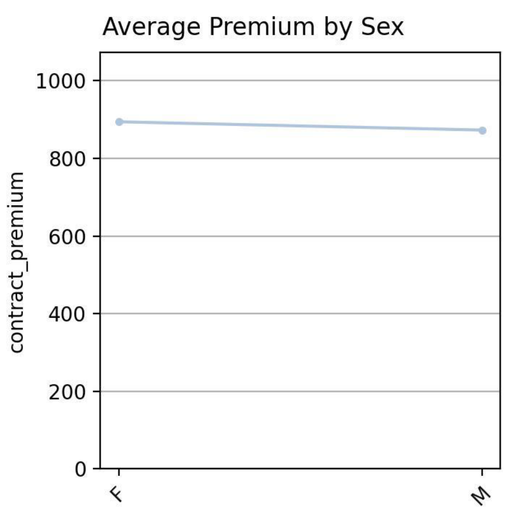
CHARTS 2-5



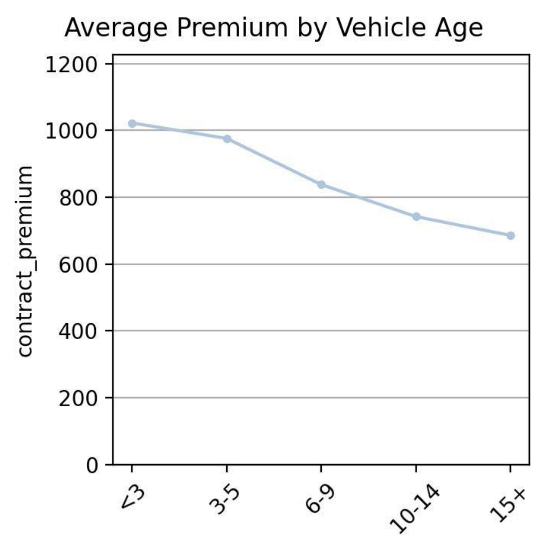
Except for the youngest drivers, younger drivers are less expensive



Prior incidents on driving records are associated with higher premiums, as expected



Premium does not vary much by sex.[13]

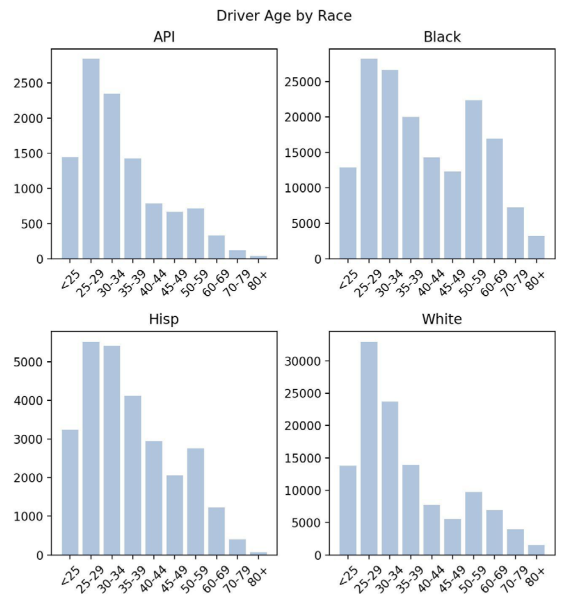


As expected, insurance costs are lower for older vehicles.

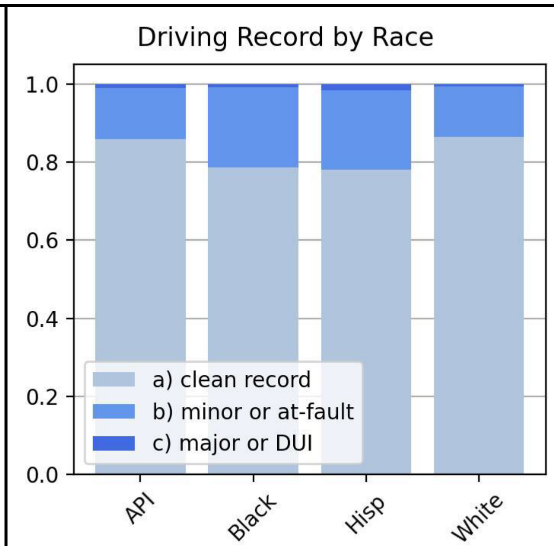
[13] F/M are the standard gender identity data points currently collected by insurers.

Next, ORCAA looked at the distribution of these factors by race:

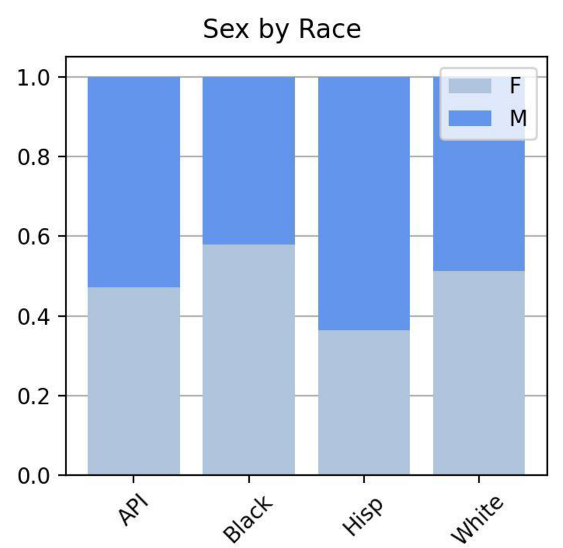
CHARTS 6-9



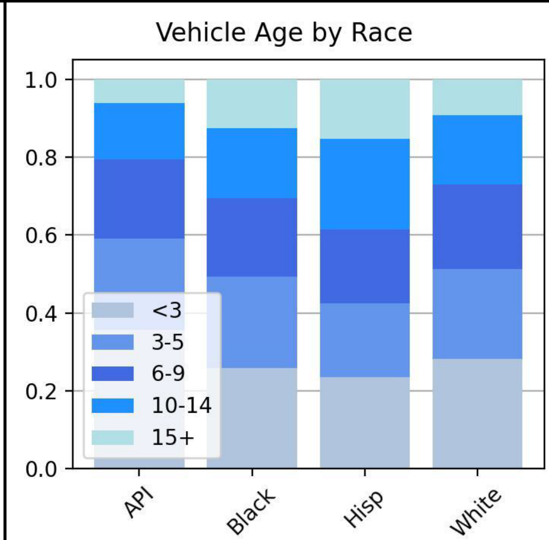
There are relatively more older Black drivers.



Black and Hispanic drivers have more minor or at-fault accidents.



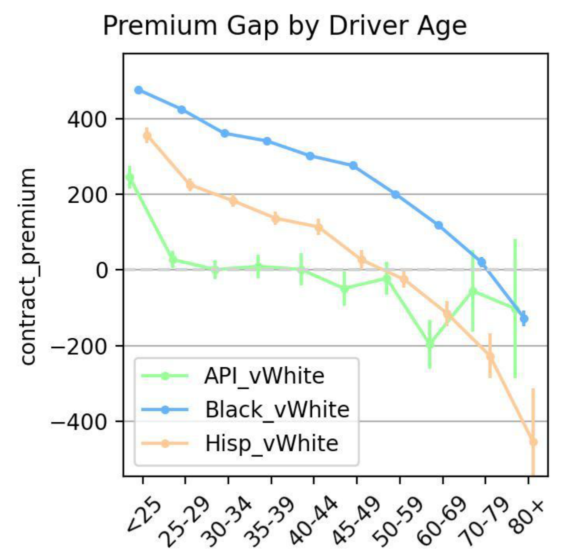
Black, white, and Asian and Pacific Islander (API) have similar M/F splits; Hispanic drivers are more male.



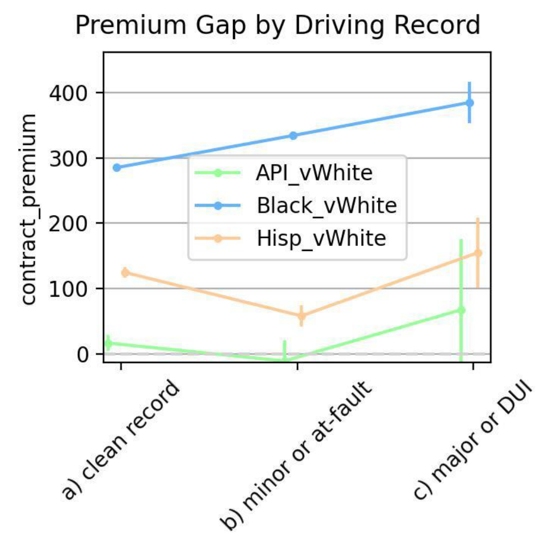
Vehicle ages are similar for Black and white, slightly older for Hispanic, and newer for API

Third, ORCAA looked at how rates vary by race across these factors individually, which is to say we see if the factors are "explaining" the average premium gap. Note that these are the premium gap relative to what white policyholders pay in the same bin, so we are seeing the amount non-white drivers pay *relative* to white drivers:

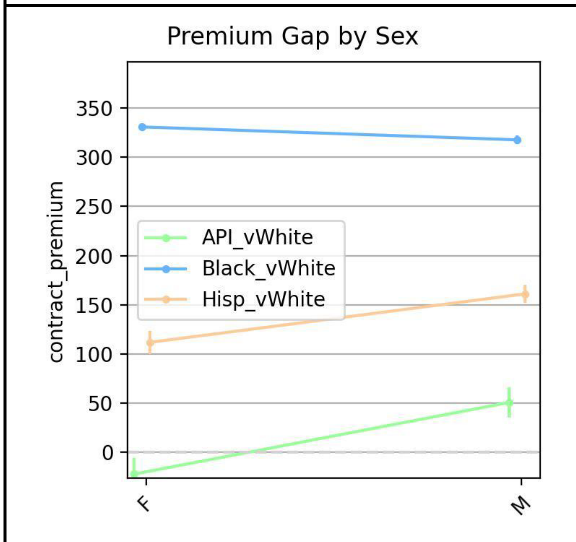
CHARTS 10-13



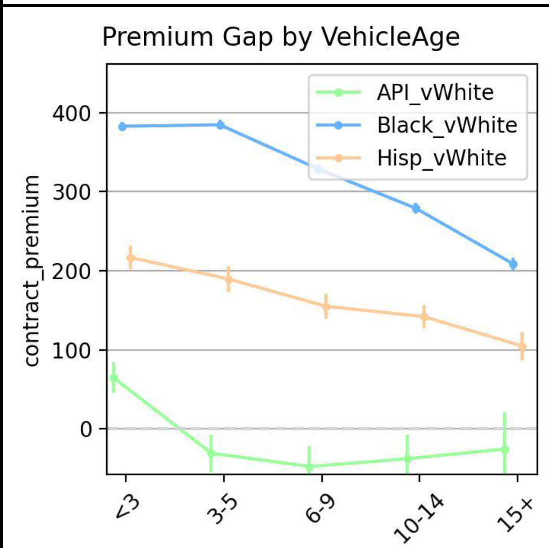
The gap disappears as people get older



The Black/white gap increases with severity of driving record



Black/white premium gap varies little across sex



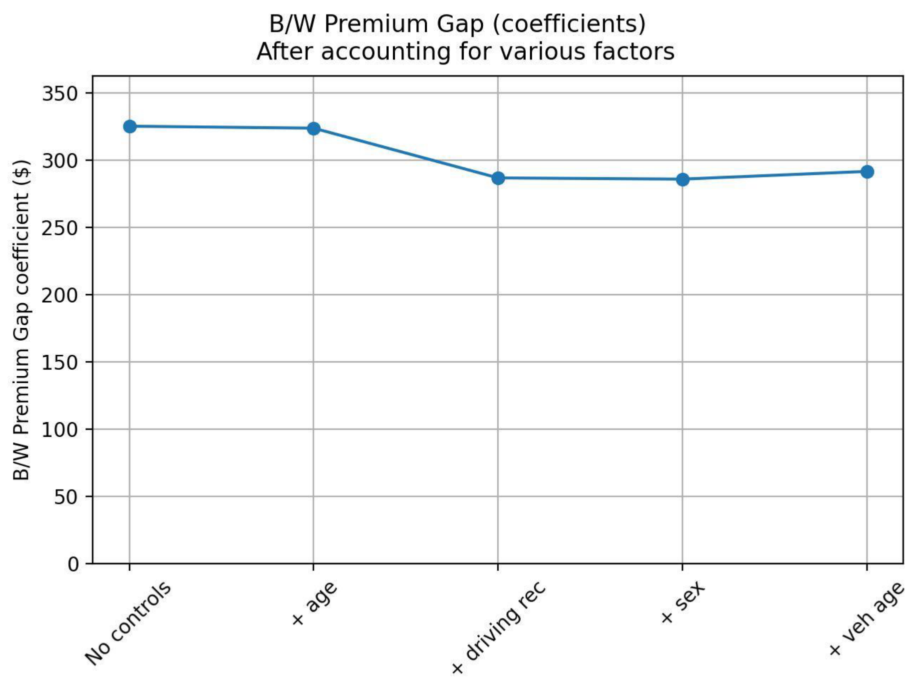
The gap decreases with vehicle age

From the above examples, we can conclude that the driving record will “explain” some of the Black/white average premium gap, because Black drivers have more at-fault accidents *and* those policies tend to be more expensive. Indeed, the gap is highest for policy holders that have a DUI in their record.

On the other hand, given the above picture we don’t expect to see much “explanation” of the Black/white average premium gap due to the age of the driver or the age of the car.

The final step is to account for the above four factors and see how much the Black/white average premium gap is diminished.[14] Therefore, consider the following graph:

CHART 14



In other words, the Black/white average premium gap has gone from $325 to $295 once we account for age of driver, driving record, payment type, and age of car.

Appendix C includes a similar graph for the full set of explanatory factors we considered (adding marital status, coverage limits, policy year, new car, and prior lapse in coverage). Note: DISB did not collect data on the value of vehicles, so it could not be accounted for. The Black/white average premium gap after accounting for all these factors is $271.

4. Relationship of Premium to Claims and Losses

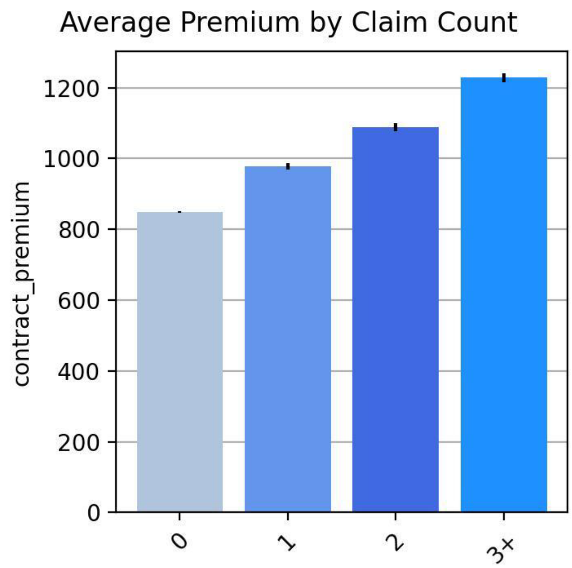
This report treats claims as an explanatory factor separately, for two reasons. First, while, ideally, DISB would have three prior years of a given driver’s claims experience leading up to a given policy premium, but unfortunately only had claims data from the same three years that there was premium data (2019-2021). Also, DISB could not track a given driver across insurers, so if a driver’s prior claims were with a

[14] See Appendix D for the methodology used here.

different insurer, DISB could not see them. Claims data is statistically very different from the other kinds of factors considered here, in that it is both sparse (85% of drivers have no claims) and fat-tailed (the claims vary in size tremendously).

Even so, DISB learned some important things by looking at the claims data available. First, DISB looked at how premiums vary by the number of claims a policyholder has (note these claims occurred *during* the policy period, i.e., after the associated contract premium was set):

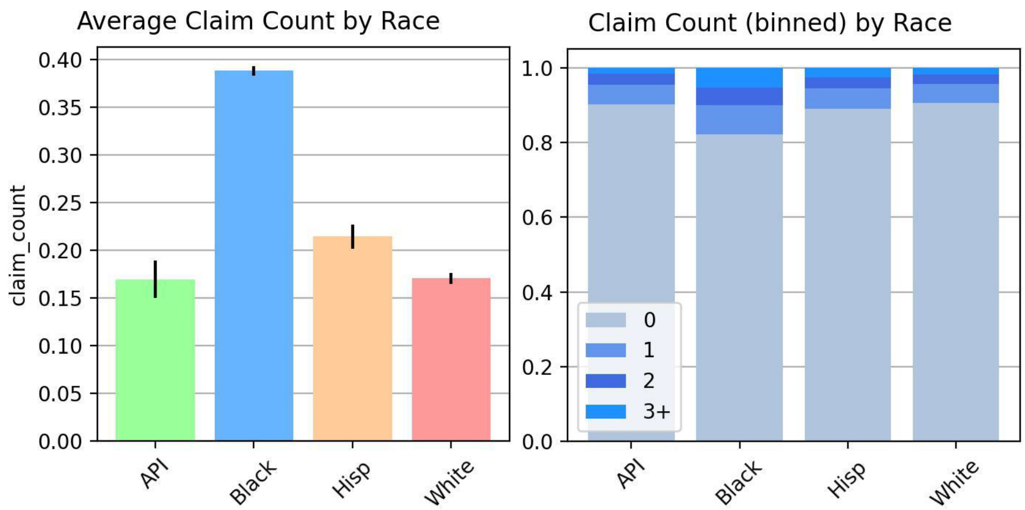
CHART 15



Customers who went on to file more claims were charged higher premiums.

Next, we look at the average number of claims by race:

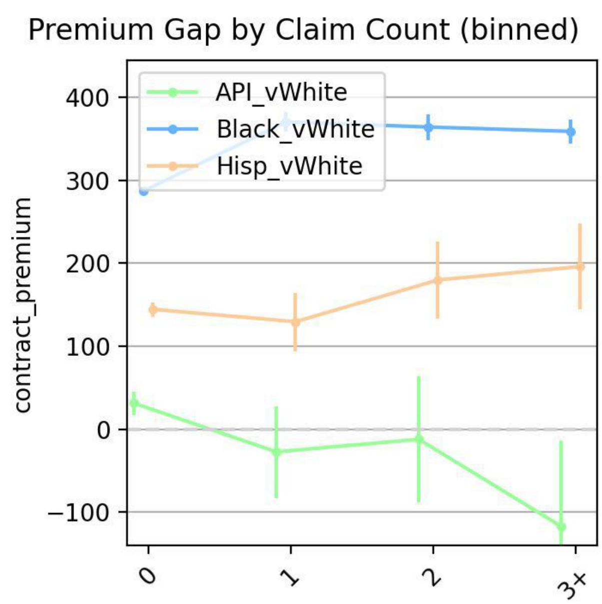
CHARTS 16-17



Black drivers have more than twice as many claims as white drivers on average (left), though most drivers of all races have no claims (right)

Third, we look at how the premium gaps vary with the number of claims:

CHART 18



The gap is bigger as soon as one claim arises.

As noted, DISB lacks sufficient information about "previous claims" for drivers, such that this report cannot conclude that claims do not explain the race gap. To make that conclusion would require building a historical model over a longer time period, allowing DISB to predict the premium next period based on the premium and the count of claims this period. In fact, DISB did this with the (limited) data on hand, and found that the expected difference in next-period premium between policies that had claims this period and policies that had none is typically about $30 per claim.[15] Even though Black policyholders have over twice as many claims on average as white policyholders, the overall incidence of claims is low: in every race group, >80% of drivers have no claims. This $30-per-claim next-period premium difference, which applies to only a small fraction of policyholders, could not explain very much of the overall Black/white average premium gap of $271.

5. What Does Explain the Black/white Average Premium Gap? Other Factors Not in Our Data

Since premiums are determined mechanically according to rate plans, conceptually every premium can be fully “explained” as the result of applying a known plan to a given applicant’s rating characteristics. Since insurers' rate plans do not vary by applicant race, if the results differ on average by race, then the rating characteristics must, too. Put another way, there must be some factors and/or steps in rating plans that correlate with race in a way that produces different average premiums by race for DC drivers. Examples of such factors, that either were not collected for this review or not analyzed, that are commonly used by DC insurers, and which could explain some of the average premium gap, include credit-based insurance scores, discounts (e.g. for homeownership, education level, or multiline), and payment modes.

Based on comments from stakeholders familiar with insurance pricing, these factors are all important differentiators for the makeup of premiums. Moreover, they tend to be multiplicative factors, which is to say they pile up and amount to even more cumulatively.

However, there are arguably no rational explanations for why they predict accident risk; and there is evidence[16] they are correlated with race in ways that would widen the Black/white average premium gap. For example, multiline discounts reward customers who have more assets to insure — and therefore correlate with race as well, again without a rational explanation. Finally, payment mode discounts (e.g. discount for payment-in-full) also correlate with wealth -- and thus race[17] -- because they rely on the ability to pay in advance. With premium rates in mind, that translates to white drivers being asked to pay on average $705 in advance, whereas Black drivers are being asked to pay on average $1,031 in advance. This is indeed once again an extra charge against those who can least afford it.

Reflections and Next Steps

As discussed in this report, reviewing premiums relative to insurance losses is important, but DISB also has an interest in ensuring premiums are racially equitable. To that end, DISB should develop a process for identifying rating factors that are correlated with losses and adversely correlated with race for which there is no rational explanation for the relationship with losses. Once identified there should be a determination of thresholds for balancing the impacts of both correlations on premiums.

---

[15] See Appendix E for more on this.

[16] For credit-based insurance scores and homeownership discounts, see Executive Summary and footnotes 5 and 6, respectively.

[17] In the United States the “racial wealth gap” describes how race correlates with wealth. See, e.g., https://www.brookings.edu/articles/black-wealth-is-increasing-but-so-is-the-racial-wealth-gap/

During this review, DISB received information and comments about issues that impact the Black/white average premium gap, but that may not be resolved by insurance rating reform. DISB should consider other ways to address these issues. Most notably, the losses for Black and Hispanic insureds were significantly higher, both in dollar terms and relative to premiums paid than for white insureds. This warrants additional study on the types and causes of claims by Black and Hispanic drivers to see if infrastructure or other changes may help reduce the claim differential.

Additionally Black and Hispanic drivers are more likely to have driving infractions, but it is not clear if this is the result of difference in enforcement rather than differences in driving practices. Driving record is a practical and reasonable criterion to use in car insurance, because not only does it have a rational explanation for the risk it is addressing, but it also provides potential risk mitigation benefits. This is another area that may benefit from additional study, but that study likely falls outside the purview of DISB.

Telematics includes a wide range of vehicle specific information that can, under the right circumstances, develop rating methodologies that have a strong rational explanation. Telematics also has the potential to exacerbate some of the concerns that led DISB to undertake the analysis in this examination. A review of how telematics currently is and potentially may be used in underwriting and rating is warranted to identify appropriate consumer protections and standards to balance telematics results when there is a correlation with race.

There are wealth-related factors that may contribute to the Black/white average premium gap. These include:

- Wealthy drivers are better able to directly pay for some insured losses without filing insurance claims.
- Wealthy drivers are more likely to have secure off-street parking for their vehicles.
- Road conditions (paving, lighting, signage, etc.) are likely to be better in wealthy neighborhoods.
- Multiple policy and other discounts likely to provide greater benefit to wealthy drivers.

DISB should consider other ways to address these impacts such as direct subsidies or refundable income tax credits.

Finally, our review showed that quote prices mirror differences by race in average premium. From this we concluded that there was not likely quote bias introduced that would be separate from premium bias. The data also shows that quotes given by agents were lower than quotes through other sales channels and Black consumers are less likely to get quotes through agents. DISB should evaluate the reasons why quotes from agents were lower than otherwise and Black drivers are less likely to get quotes from agents.

As with the work to date, all future work should be done in a deliberate and transparent manner, with a public process and involvement from stakeholders.

Appendices

Appendix A: BIFSG methodology for race inference

Background

We used the Bayesian-improved first name surname geocoding (BIFSG) methodology to infer the race of policyholders.

A forerunner of this methodology, BISG, was developed by the RAND Corporation to "help U.S. organizations produce accurate, cost-effective estimates of racial and ethnic disparities within datasets—and illuminate areas for improvement."[18] At a high level, this approach leverages U.S. Census data to guess an individual's self-reported race/ethnicity based on their name and address. The Census publishes tables on the aggregate demographics of specific geographic areas (e.g., census blocks) and common surnames. For a given address (geocoded to a census block), we can make a guess of a person's race/ethnicity, and likewise for their surname. Then Bayes' theorem[19] is applied to combine the guesses.

We note this approach generates probabilistic guesses, but a given person's self-reported race/ethnicity is a definite label. BISG will not "guess right" for every individual; but still it can be "calibrated" in the sense that if you took many individuals with a 90% probability of being Hispanic (per BISG), 90% would indeed be self-reported Hispanic.

This methodology is used by regulatory agencies in fairness analyses. For example, the CFPB has written about its use of BISG in "conducting fair lending analysis of non-mortgage credit products in both supervisory and enforcement contexts."[20]

The BIFSG methodology we used incorporates another piece of information — the individual's first name (i.e., "F" in "BIFSG") – and another application of Bayes' theorem to further refine the race/ethnicity inference. This refinement of the methodology was proposed and validated by Ioan Voicu of the U.S. Department of the Treasury, Office of the Comptroller of the Currency (OCC).[21] We follow the implementation described in his paper.

Details of our implementation

In our implementation of BIFSG, for each individual we obtained a vector with seven entries. Each entry is a decimal value between 0 and 1 (inclusive) and indicates the probability that the individual belongs to a certain race/ethnicity category used in the 2010 U.S. Census: Hispanic, non-Hispanic Black, non-Hispanic white, non-Hispanic Asian or Pacific Islander ("API"), non-Hispanic American Indian or Alaska Native, Multiracial, Other. These categories are mutually exclusive and collectively exhaustive, so the seven probabilities sum to one.

We took the highest value among the seven and applied the corresponding race/ethnicity label to that individual. This report focuses on the largest categories (API, Black, Hispanic, and white); the graphs and charts omit the categories non-Hispanic American Indian or Alaska Native, Multiracial, and Other, which together represent less than 1% of all data.[22]

[18] For background, see rand.org/health-care/tools-methods/bisg.html

[19] wikipedia.org/wiki/Bayes%27_theorem

[20] files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf

[21] Ioan Voicu (2018) Using First Name Information to Improve Race and Ethnicity Classification, Statistics and Public Policy, 5:1, 1-13, DOI: 10.1080/2330443X.2018.1427012

[22] In the policies data, these 3 race/ethnicity categories comprised 6,561 out of 1,607,113 records; in the quotes

We used the federal geocoding server[23] to find the associated Census block for each customer address provided by an insurer. Geocoding was successful — i.e., a Census block was found — for 303,260 of 322,329 addresses in the data insurers submitted.

For a given observation (insurance application) any of the three BIFSG inputs (firstname, surname, or address) could be missing or degenerate. A "degenerate" address would be one that failed geocoding; a "degenerate" firstname or surname would be one not included in the relevant table. When one or two BIFSG inputs were degenerate or missing, we made an inference with the viable inputs available, which in effect meant "skipping" one or two applications of Bayes' Theorem. When all three BIFSG inputs were missing, we made no race inference.

data, they comprised 24,351 out of 2,607,088 records. Both are less than 1%.

[23] API documentation available at: geocoding.geo.census.gov/geocoder/Geocoding_Services_API.pdf

Appendix B: Quotes analysis

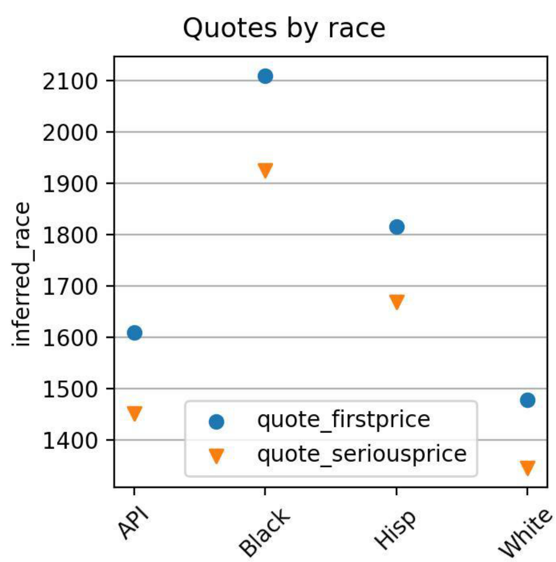
DISB's data call to carriers included a request for data on quotes for insurance coverage that were provided to consumers. Based on discussions with DISB and other stakeholders, we understand the quote process varies considerably across carriers. Different carriers may provide quotes at different junctures, through different channels, based on different information, and display the quotes in different ways.

DISB provided definitions in the data call[24] to collect comparable information from everybody. For instance, DISB defined a "serious quote" as one where the consumer's name and address were validated, a price was shown, some vendor data (e.g., motor vehicle records, credit based insurance scores (CBIS), comprehensive loss underwriting exchange (CLUE)) was pulled, and underwriting was called. Still, based on carriers' feedback we understand the data on quotes is somewhat noisy. Consequently, this analysis is preliminary.

Some key findings:

Quote prices by race mirror premium gaps
As with premiums, quotes are highest for Black consumers, lower for Hispanic consumers, and lowest for white and API consumers. For all race groups, the first price shown (quote_firstprice) is about 10% higher than the price shown when/if the quote becomes "serious" (quote_seriousprice).

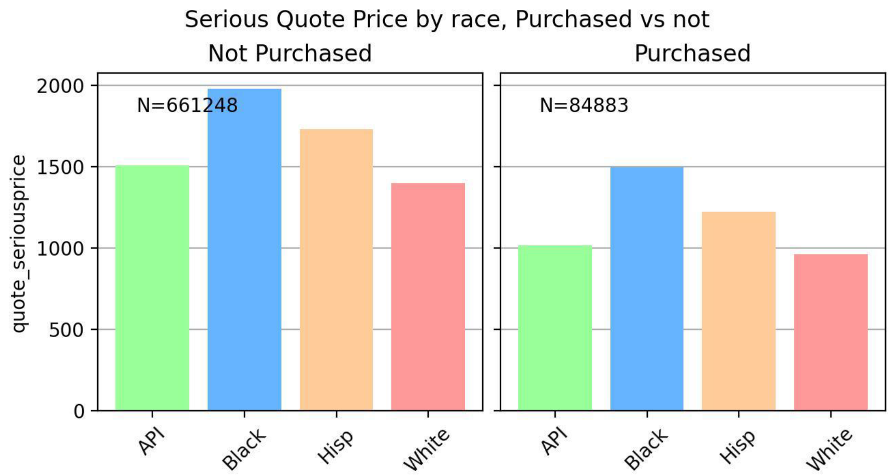
CHART A1



Consumers shop around
About 13% of serious quotes are purchased by consumers. Quotes that are purchased are lower on average, which is consistent with consumers shopping around for price. The pattern of race gaps is persistent across purchase.

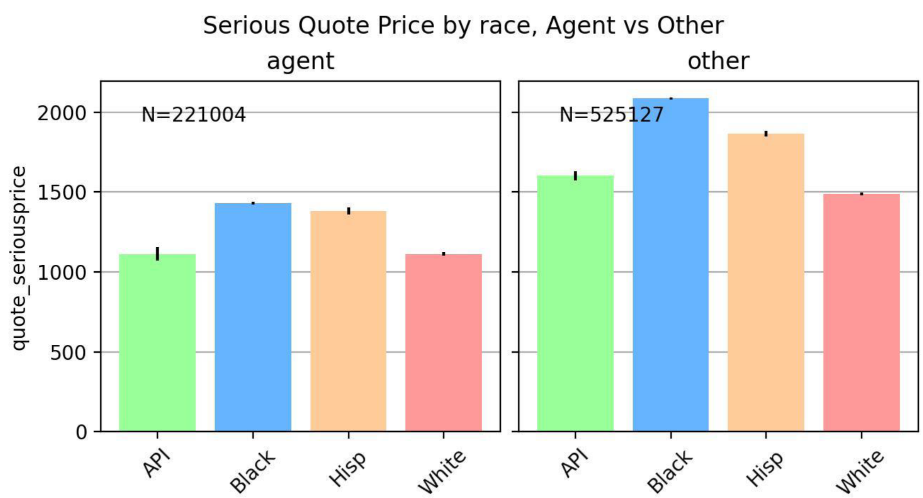
[24] See data call tab "Quotes - Scope and definitions", rows 45 ("quote_firstprice"), 45 ("quote_serious"), and 52 ("quote_finalprice")

CHARTS A2-A3



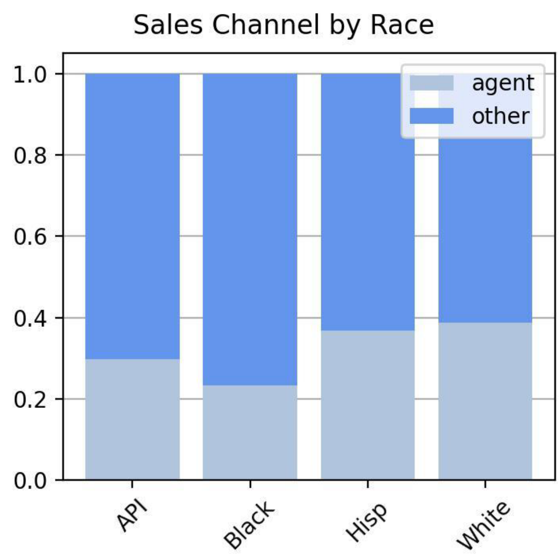
<u>Quotes given by agents are lower than through other sales channels</u>
Quotes given by agents are lower on average than quotes given through other channels (e.g., online), and the race gap is smaller for quotes given by agents.

CHARTS A4-A5



However, Black consumers are less likely than others to get quotes through agents:

CHART A6



Anecdotally, some stakeholders suggested agents could be helping customers find discounts that lead to lower premiums. Future data calls by DISB could request more information about policy-level discounts to explore this issue.

Appendix C: Detail on explanatory factors we tried

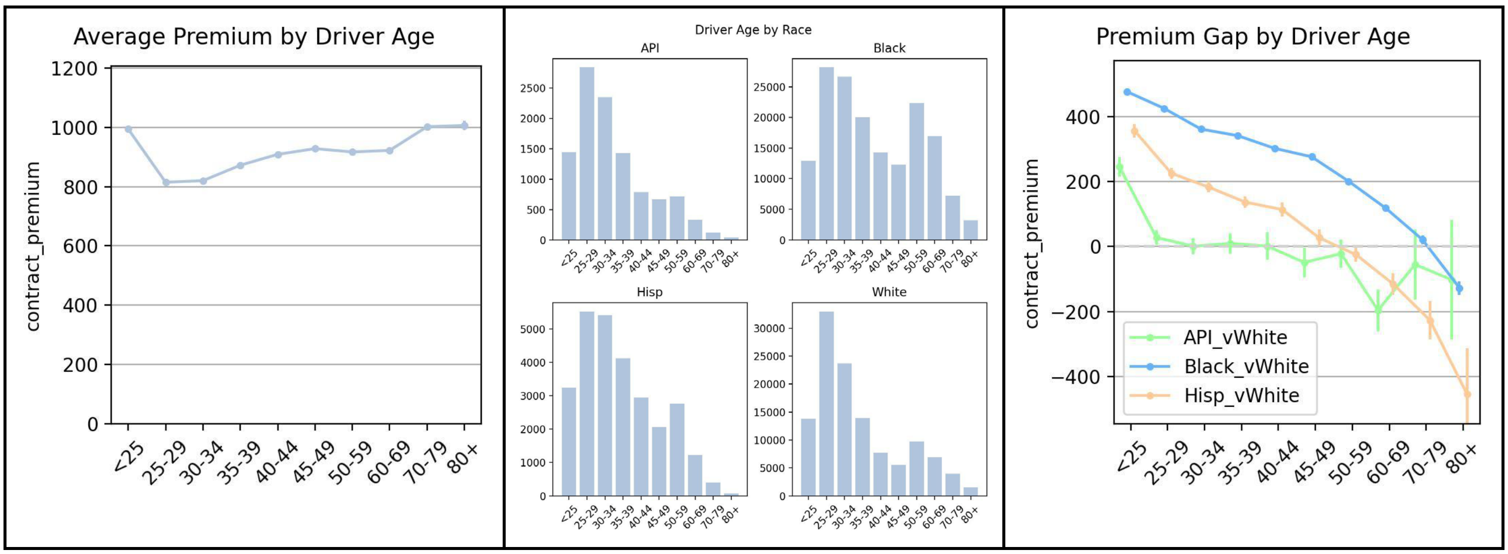
Here is the full list of factors we analyzed as potential explainers of race differences in premium, using the analysis approach described in the "Explanatory Factors" section.

Driver age

<u>Definition:</u> driver's age in years, as of the policy effective date

<u>Derivation from data call:</u> calculated as difference in years between "policy_effective_date" and "dob"

CHARTS A7-A9



Driving record

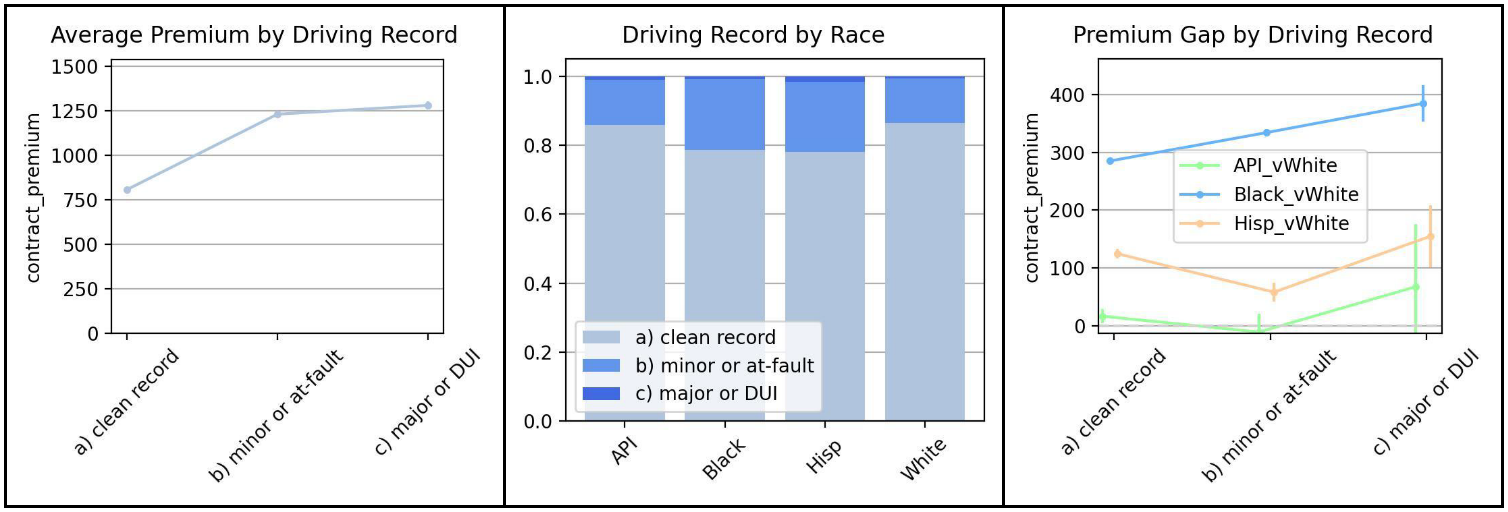
<u>Definition:</u> categorical variable

= "major or DUI" if driver has at least one major violation or DUI

= "minor or at-fault" if driver has at least one minor violation or at-fault accident but no major violation or DUI

= "clean record" if driver has no major or minor violations or at-fault accidents or DUI

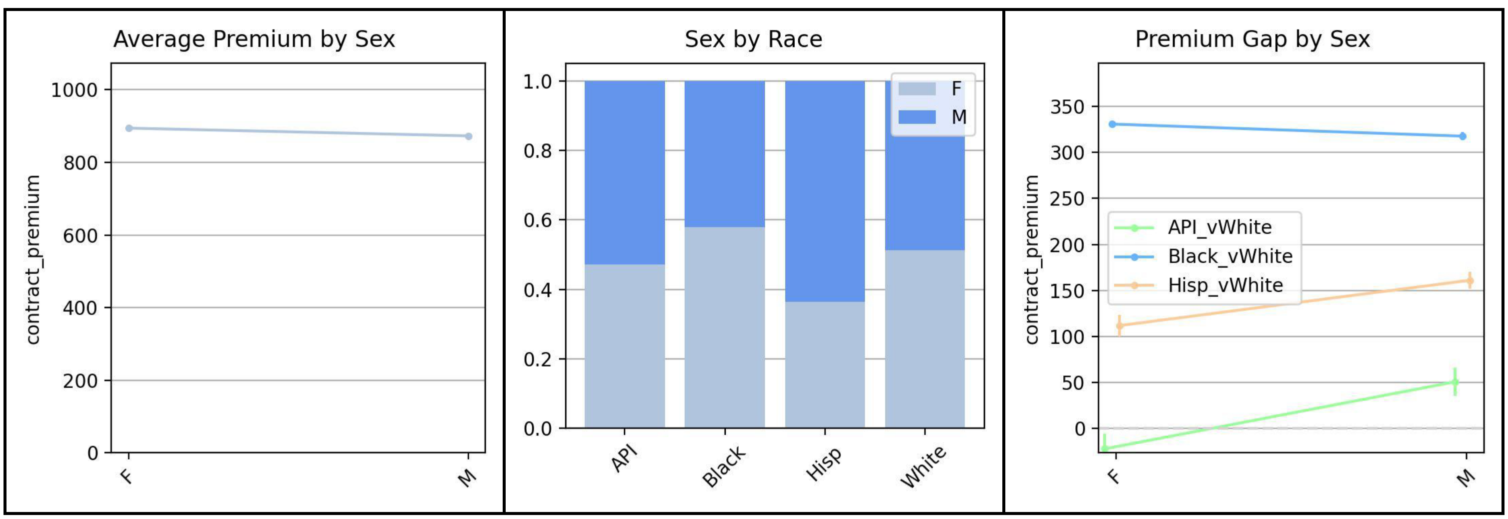
<u>Derivation from data call:</u> calculated based on "major_violations_count", "minor_violations_count", "dui_violations_count", "atfault_accident_count"



Sex

Definition: driver's sex

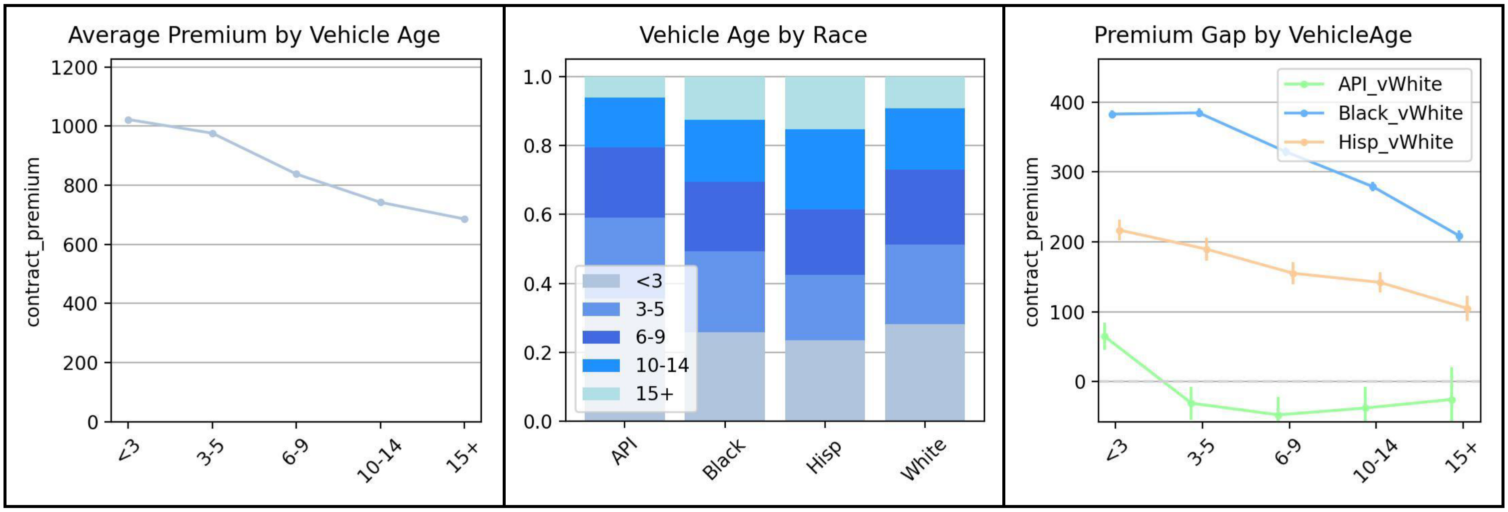
Derivation from data call: "sex"



Vehicle Age

Definition: age of the insured vehicle in years, at the time of the policy effective date, then binned into <3, 3-5, 6-9, 10-14, 15+. Bins were chosen to divide the data relatively evenly.

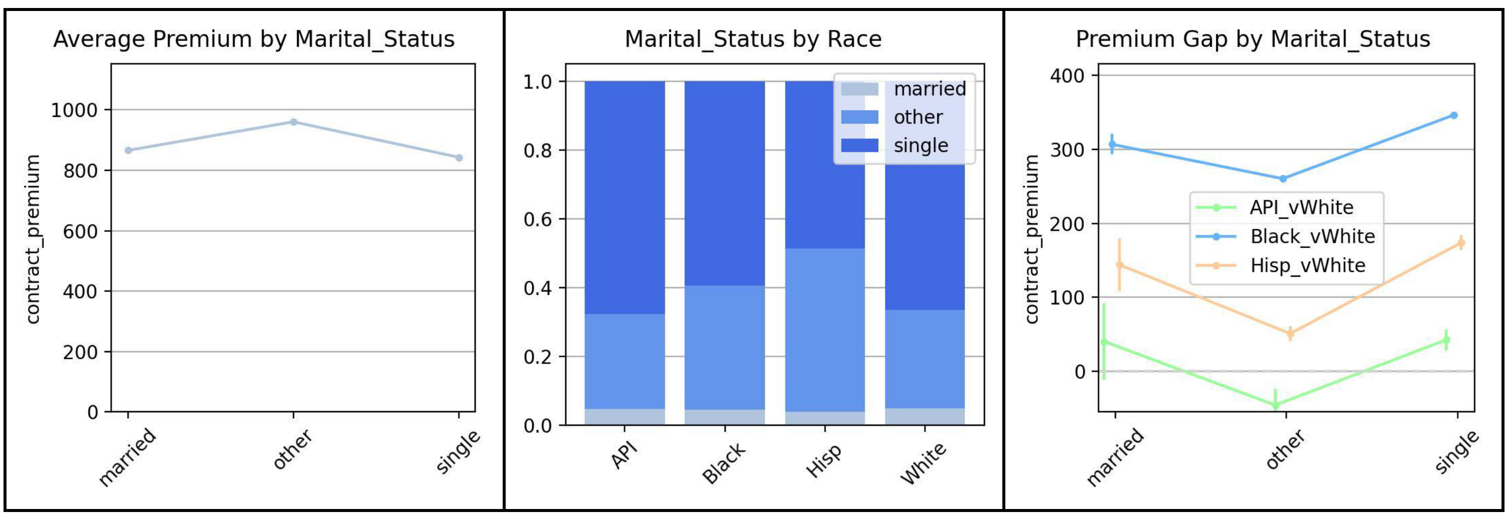
Derivation from data call: calculated as the difference between the year in "policy_effective_date" and the "vehicle_year", then binned as above.



Marital status

Definition: driver's marital status, binned as "married", "single", or "other"

Derivation from data call: based on "marital_status"



Coverage limits

Definition: categorical variable indicating common bundles of (bodily injury per-person, bodily injury per-incident, property damage per incident) liability coverage limits on the policy depicted

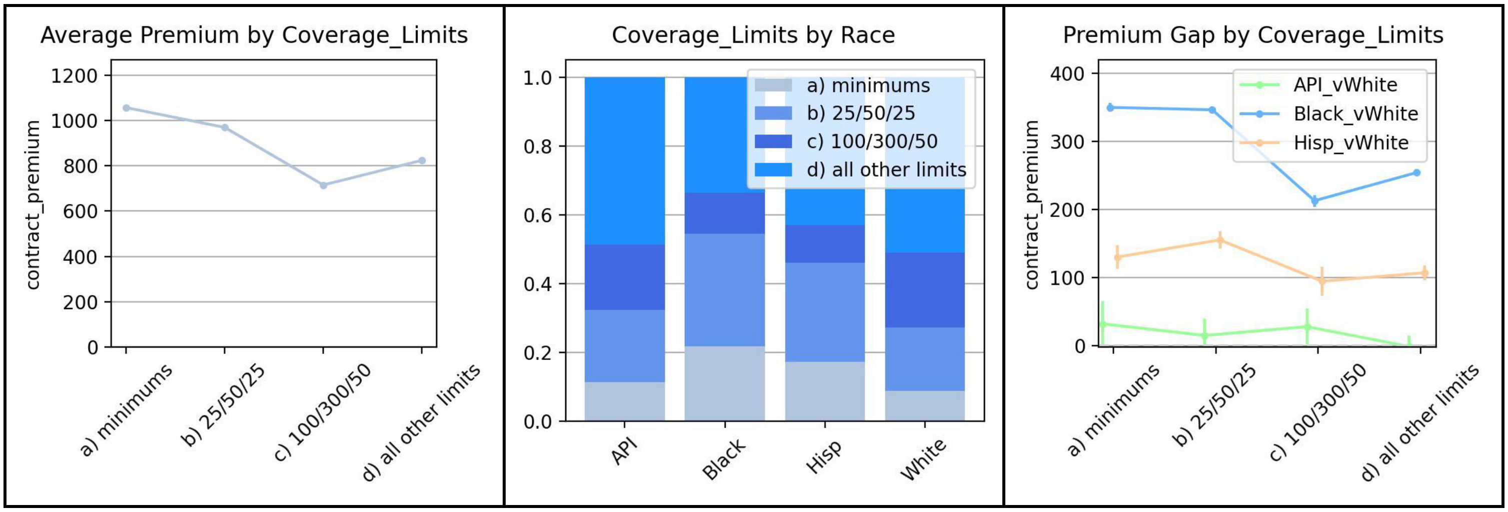
="minimums" if the limits are ($25k, $50k, $10k) [i.e., DC mandatory minimums]

="25/50/25" if the limits are ($25k, $50k, $25k)

="100/300/50" if the limits are ($100k, $300k, $50k)

="all other limits" otherwise

Derivation from data call: calculated from "BI_limit_perperson", "BI_limit_perincident", and "PD_limit"

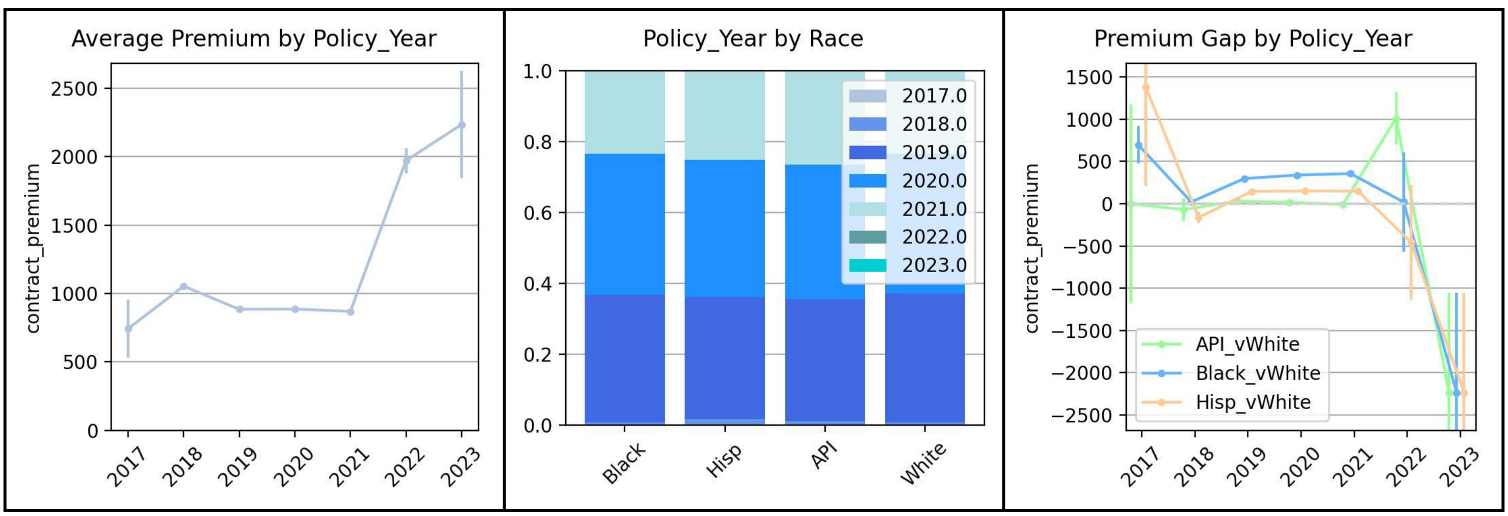


Policy year

Definition: the calendar year of the effective date of the policy

Derivation from data call: calculated from "policy_effective_date"

Note: As the middle chart below shows, >95% of our data is from policy years 2019, 2020, and 2021. The extreme values of average premium and premium gap outside those years are a reflection of small data.



New Car

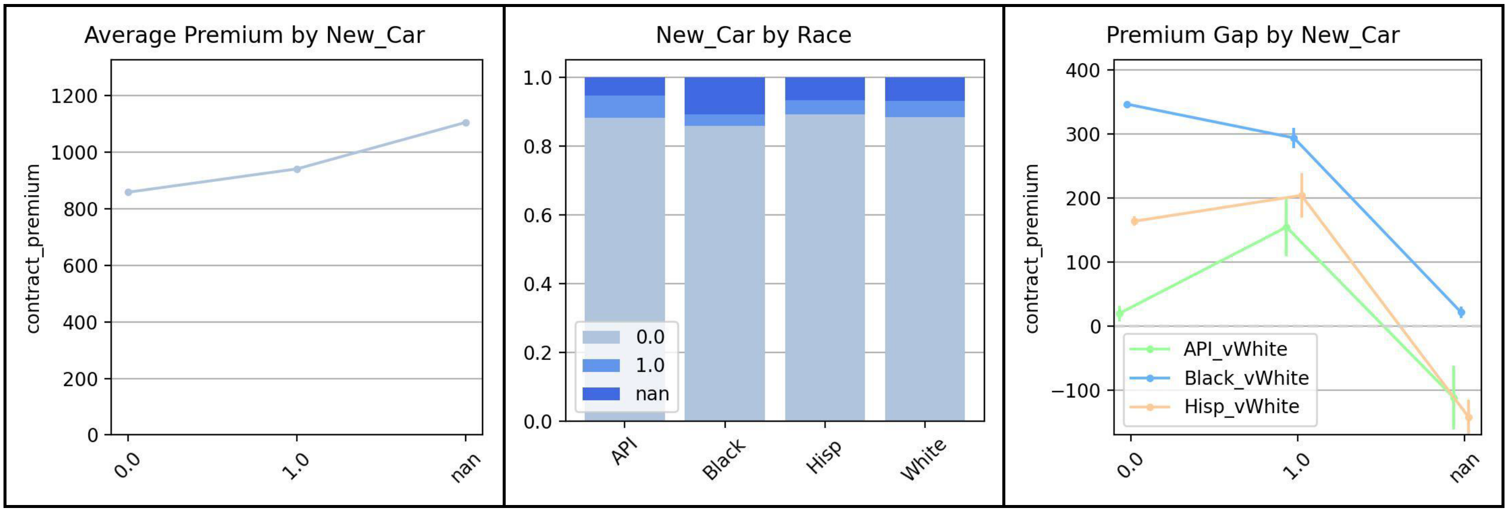
Definition: indicator variable for whether this is a new vehicle

=0.0 if the vehicleis not new

=1.0 if the vehicleis new

=nan if unknown

Derivation from data call: based on "vehicle_new"



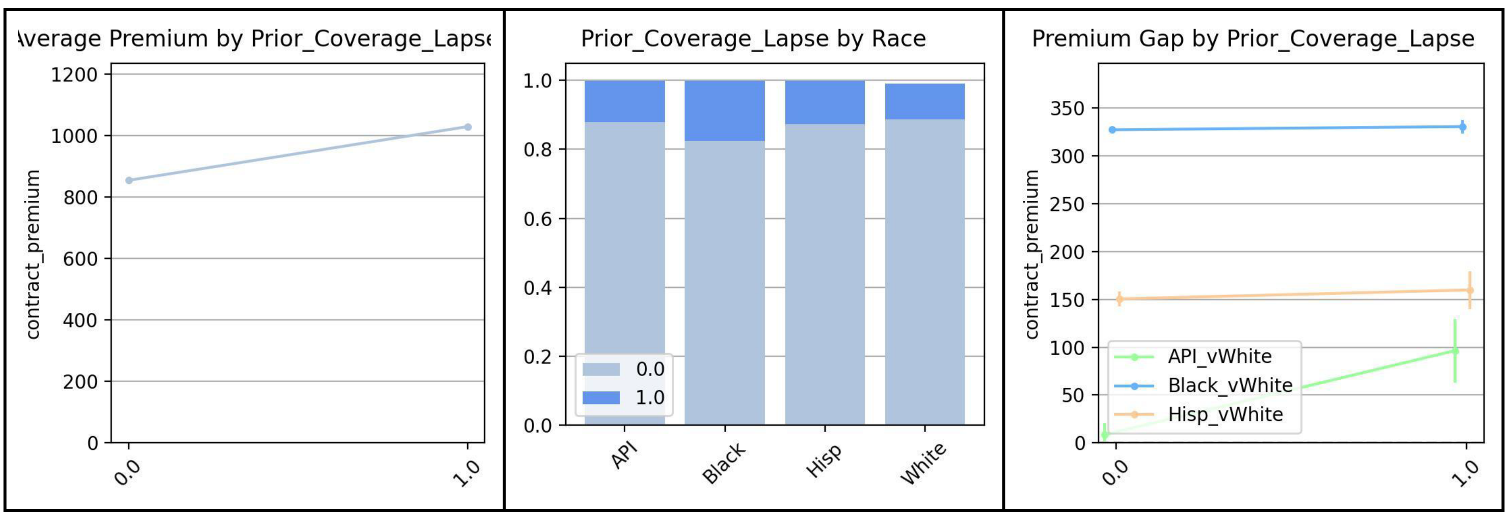
Prior Coverage Lapse

Definition: indicator variable: did this driver have a prior lapse in auto coverage?

=0 if no prior lapse

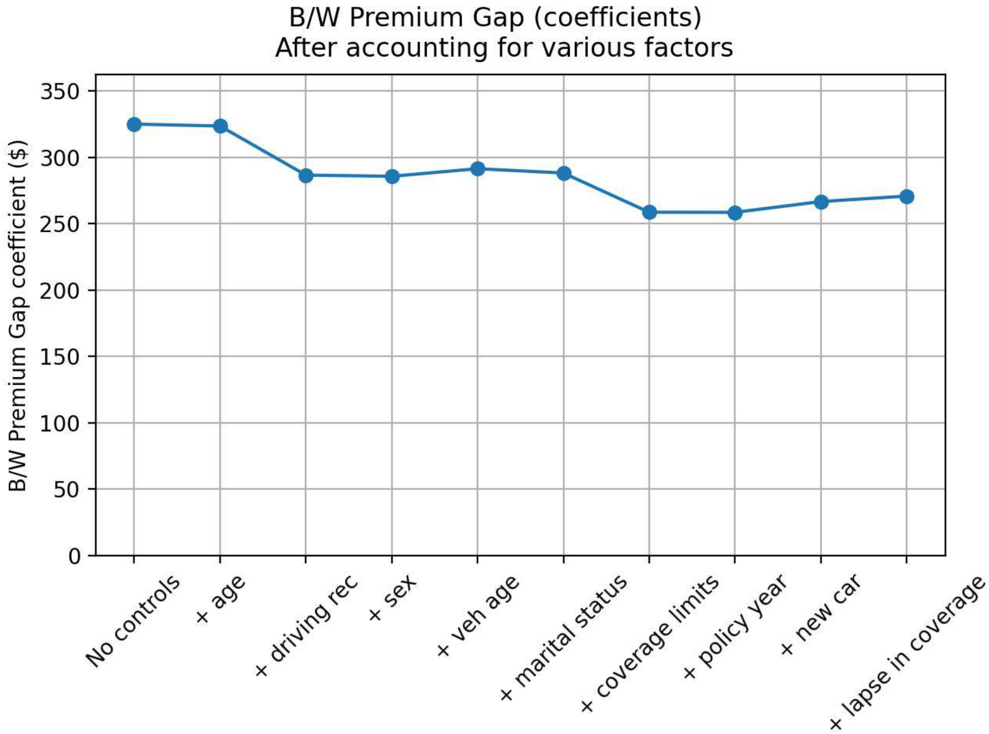
=1 if some prior lapse

Derivation from data call: based on "lapse_in_coverage"



Explaining the Black/white Premium Gap

Taking all the above explanatory factors into account, the Black/white gap in premiums goes from $325 (at left) to $271 (at right). The regression methodology is explained in Appendix D.



New vs renewal[25]

<u>Definition:</u> categorical variable

=”new” if this policy is new business

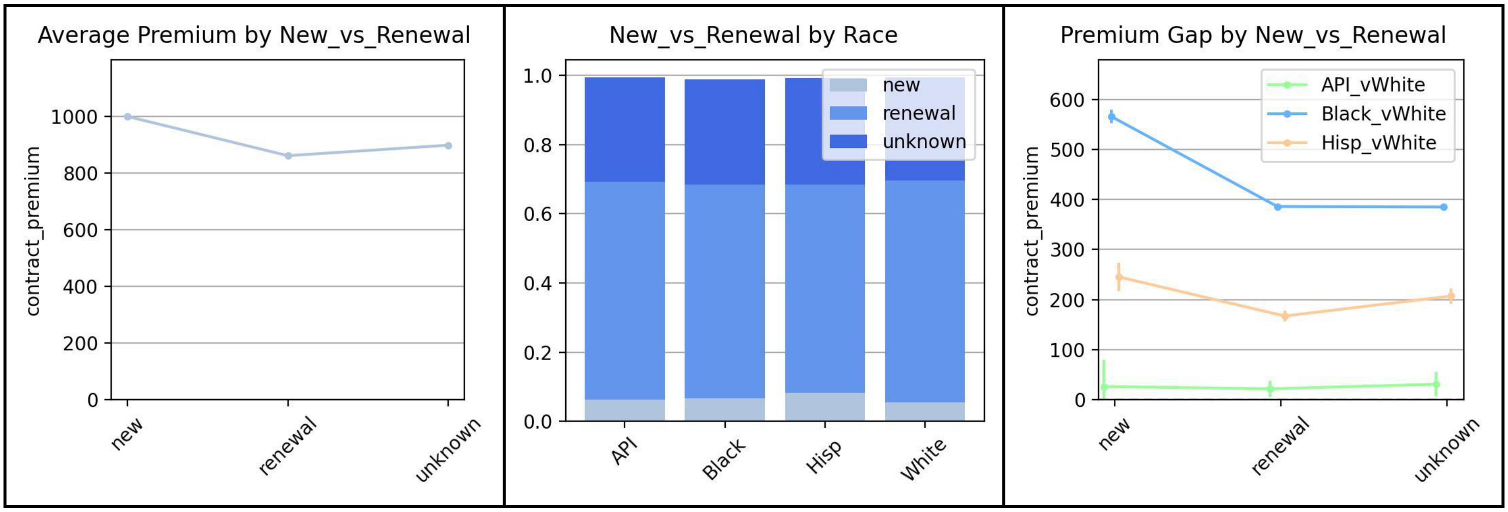
=“renewal” if this policy is a renewal

=“unknown” if we cannot tell from the data

<u>Derivation from data call:</u> this is derived based on matching “driverID” or “policyID” within a given insurer, across policy periods.

<u>Note:</u> only a subset of data submissions supported the matching of IDs across policy period based on the structure of the data, so this variable is available only for some insurers, representing about ⅔ of all policies.

[25] Note we were missing substantial coverage for this column in our data.


Average Premium by New_vs_Renewal
contract_premium
0
200
400
600
800
1000
new
renewal
unknown
New_vs_Renewal by Race
0.0
0.2
0.4
0.6
0.8
1.0
new
renewal
unknown
API
Black
Hisp
White
Premium Gap by New_vs_Renewal
contract_premium
0
100
200
300
400
500
600
API_vWhite
Black_vWhite
Hisp_vWhite
new
renewal
unknown

Appendix D: Methodology detail: Measuring the Black/white premium gap while accounting for various factors

The basic structure of this analysis is a series of linear regressions that represent nested models predicting premium. Each successive model "accounts for" an additional explanatory factor by adding one or more control variables to the regression equation.

The first, and simplest, model predicts premium (P) based on race alone. Since race is a categorical variable with four possible values in our data (API, Black, Hispanic, white), this appears in regressions as three dummy variables.[26] We fix white as the reference category.

(1) $P = \alpha + \beta_1 * API + \beta_2 * Black + \beta_3 * Hispanic + \varepsilon$

The next model accounts for driver age as an explanatory factor for premium:

(2) $P = \alpha + \beta_1 * API + \beta_2 * Black + \beta_3 * Hispanic + \beta_4 * DriverAge + \varepsilon$

At each step, we add one or more variables to the right-hand side of the equation to account for another explanatory factor.

β_2 is the coefficient of interest throughout this analysis. Its interpretation is roughly "the expected difference (in dollars) between white and Black premiums, after accounting for the other explanatory factors in the equation." For equation (1), since there are no other included explanatory factors, β_2 is simply the difference in average premiums between white and Black.

<u>Modeling premium vs log(premium) – robustness check</u>

The regressions described in this appendix model premium, stated in dollars. As a robustness check we did the same analyses modeling log(premium) instead. This was motivated by two considerations:

1. We understand that in practice insurance premiums are often constructed by taking a base rate and *multiplying* relativity factors for a given applicant. Modeling log(premium) as a sum is equivalent to modeling premium as a product;[27] so this structure might better fit the data generation process.
2. Premium has a long tail (some premiums are very large). Taking the log of premium condenses the distribution and could reduce the impact of outliers on the estimated coefficients.

The results were substantially the same modeling log(premium) instead of premium.

[26] "Dummy variables" are binary variables that indicate when an observation belongs to a certain category. E.g., API=1 if this policyholder is Asian/Pacific Islander, and =0 otherwise.

[27] Since $\log(P) = \alpha+\beta \rightarrow P = e^{\alpha+\beta} = e^{\alpha} * e^{\beta}$

Appendix E: Model of premium bump in response to claims

We merged policies and claims data then built a dataset of premium jumps as follows:

1. Group policies data by driverID
2. For driverIDs with N>1 rows of policies data (i.e. multiple policy periods of the same driverID), create N-1 "premium jump snapshots" showing premium in the previous policy period, claims in the previous policy period, and premium in this policy period
    - This approach excludes 101,456 policies that have just one row of data (i.e., we observe these driverIDs for only one policy period)
    - Overall, we get 84,565 "premium jump snapshots"

Then we do linear regressions to investigate the relationship between a given driver's premium in this policy period (P_{now}), their premium in the previous policy period (P_{last}), their claims activity in the previous policy period (C_{last}), and their race[28]:

1. $P_{now} = \alpha + \beta_1 * P_{last} + \varepsilon$
2. $P_{now} = \alpha + \beta_1 * P_{last} + \beta_2 * C_{last} + \varepsilon$

2B. $P_{now} = \alpha + \beta_1 * P_{last} + \beta_2 * C_{last} + \varepsilon$ [Black drivers only]

2W. $P_{now} = \alpha + \beta_1 * P_{last} + \beta_2 * C_{last} + \varepsilon$ [white drivers only]

3. $P_{now} = \alpha + \beta_1 * P_{last} + \beta_2 * C_{last} + \beta_3 * \text{Race} + \varepsilon$

The outputs of these regressions are below. We noted about these regressions:

- The coefficient on claim count (β_2) is positive and highly statistically significant. As expected, claims this policy period lead to higher premium next policy period. Specifically: for each claim this period, premium next policy period is about $30 higher on average.
- R-squared barely changes when adding claims to the regression (0.797 without [regression 1], vs 0.799 with claims [regression 2]). Given that claims do predict premium jumps, it is surprising that including claims information adds little explanatory power to the model. This may reflect the fact that claims are sparse.
- Running the regressions separately for Black and white drivers (regressions 2B and 2W), β_2 is positive and highly statistically significant in both cases, with similar values (28.83 in 2B and 27.52 in 2W). So, the bump in next period's premium based on a claim this period is roughly the same for both races.
- There is still a Black penalty: $35 hike after accounting for last period's premium and claims last period (regression 3)

[28] Race appears in regressions as 3 dummy variables, as described in Appendix D.

Regression 1

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             prem_after   R-squared:                       0.797
Model:                            OLS   Adj. R-squared:                  0.797
Method:                 Least Squares   F-statistic:                 3.322e+05
Date:                Tue, 29 Aug 2023   Prob (F-statistic):               0.00
Time:                        13:43:31   Log-Likelihood:            -5.9084e+05
No. Observations:               84565   AIC:                         1.182e+06
Df Residuals:                   84563   BIC:                         1.182e+06
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept      58.0150      1.643     35.301      0.000      54.794      61.236
prem_before     0.9217      0.002    576.359      0.000       0.919       0.925
==============================================================================
Omnibus:                    79070.621   Durbin-Watson:                   1.812
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         29940326.608
Skew:                           3.750   Prob(JB):                         0.00
Kurtosis:                      94.875   Cond. No.                     1.88e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.88e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Regression 2

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             prem_after   R-squared:                       0.799
Model:                            OLS   Adj. R-squared:                  0.799
Method:                 Least Squares   F-statistic:                 1.682e+05
Date:                Tue, 29 Aug 2023   Prob (F-statistic):               0.00
Time:                        13:43:31   Log-Likelihood:            -5.9042e+05
No. Observations:               84565   AIC:                         1.181e+06
Df Residuals:                   84562   BIC:                         1.181e+06
Df Model:                           2
Covariance Type:            nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept            55.2416      1.638     33.724      0.000      52.031      58.452
prem_before           0.9147      0.002    568.405      0.000       0.912       0.918
claim_count_before   30.7377      1.055     29.138      0.000      28.670      32.805
==============================================================================
Omnibus:                    78301.647   Durbin-Watson:                   1.815
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         30162303.109
Skew:                           3.678   Prob(JB):                         0.00
Kurtosis:                      95.229   Cond. No.                     1.88e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.88e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Regression 2B

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             prem_after   R-squared:                       0.757
Model:                            OLS   Adj. R-squared:                  0.757
Method:                 Least Squares   F-statistic:                 6.503e+04
Date:                Thu, 21 Sep 2023   Prob (F-statistic):               0.00
Time:                        15:05:48   Log-Likelihood:            -2.9882e+05
No. Observations:               41726   AIC:                         5.976e+05
Df Residuals:                   41723   BIC:                         5.977e+05
Df Model:                           2
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept             78.3996      3.013     26.023      0.000      72.495      84.305
prem_before            0.9088      0.003    354.032      0.000       0.904       0.914
claim_count_before    28.8300      1.499     19.229      0.000      25.891      31.769
==============================================================================
Omnibus:                    37840.124   Durbin-Watson:                   1.808
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7252842.226
Skew:                           3.761   Prob(JB):                         0.00
Kurtosis:                      67.149   Cond. No.                     2.33e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.33e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Regression 2W

OLS Regression Results

Dep. Variable:	prem_after	R-squared:	0.839
Model:	OLS	Adj. R-squared:	0.839
Method:	Least Squares	F-statistic:	8.535e+04
Date:	Thu, 21 Sep 2023	Prob (F-statistic):	0.00
Time:	15:05:48	Log-Likelihood:	-2.1681e+05
No. Observations:	32663	AIC:	4.336e+05
Df Residuals:	32660	BIC:	4.337e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	41.2867	1.833	22.528	0.000	37.695	44.879
prem_before	0.9135	0.002	410.055	0.000	0.909	0.918
claim_count_before	27.5200	1.673	16.446	0.000	24.240	30.800

Omnibus:	23946.431	Durbin-Watson:	1.819
Prob(Omnibus):	0.000	Jarque-Bera (JB):	51011362.708
Skew:	1.998	Prob(JB):	0.00
Kurtosis:	196.561	Cond. No.	1.50e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.5e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Regression 3

OLS Regression Results

Dep. Variable:	prem_after	R-squared:	0.800
Model:	OLS	Adj. R-squared:	0.800
Method:	Least Squares	F-statistic:	6.764e+04
Date:	Tue, 29 Aug 2023	Prob (F-statistic):	0.00
Time:	13:43:31	Log-Likelihood:	-5.9023e+05
No. Observations:	84565	AIC:	1.180e+06
Df Residuals:	84559	BIC:	1.181e+06
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	45.9190	1.829	25.107	0.000	42.334	49.504
C(inferred_race, Treatment(reference="White"))[T.API]	-13.8812	5.019	-2.766	0.006	-23.719	-4.044
C(inferred_race, Treatment(reference="White"))[T.Black]	34.9454	2.007	17.412	0.000	31.012	38.879
C(inferred_race, Treatment(reference="White"))[T.Hisp]	3.2631	3.383	0.965	0.335	-3.368	9.894
prem_before	0.9063	0.002	543.262	0.000	0.903	0.910
claim_count_before	29.0008	1.057	27.446	0.000	26.930	31.072

Omnibus:	78969.362	Durbin-Watson:	1.814
Prob(Omnibus):	0.000	Jarque-Bera (JB):	30559969.247
Skew:	3.734	Prob(JB):	0.00
Kurtosis:	95.829	Cond. No.	5.86e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.86e+03. This might indicate that there are strong multicollinearity or other numerical problems.