Gender-Neutral Pricing and Statistical Discrimination: Evidence from Insurance Pricing Algorithms *

Ozge Demirci†

February 16, 2022

PRELIMINARY - PLEASE DO NOT CIRCULATE

Click **here** for the most recent version

Abstract

Traditionally, car insurance companies used to charge women and men differently. As male and female drivers differ in their accident and claim statistics, insurance companies use this as a reason for gender-based pricing. Although it can be efficient for firms, it is statistical discrimination to take someone's group belonging into account while pricing them. Because of this, many policy regulations to ban the use of gender when assessing risk factors for car insurance were introduced over time in different states in the US. This paper focuses on the recent policy change in California requiring gender-neutral pricing in auto insurance. By exploiting the variation in the policy implementation over time and across states, I show that the insurance price gap between male and female drivers decreased by 6 percentage points after the policy. It decreased premiums for young male drivers, the riskiest group, but they increased for young females. Leveraging different machine learning methods, I estimate features that predict gender and show that insurance pricing algorithms started to proxy gender with other information already collected by firms. Hence, these characteristics correlated with gender gained more weight in the pricing algorithm. Drivers using specific car models associated with young males, the riskiest group, end up paying up to 10 percent more irrespective of their gender and driving records.

Keywords: Statistical discrimination, algorithmic bias

JEL Codes: J16, G22

---

*I am grateful to Manuel Bagues, Mirko Draca, and Ludovica Gazze for their invaluable mentorship and advice. I would like to thank all the Warwick internal seminars participants for their insightful comments and suggestions.

†Department of Economics, University of Warwick. Email address: o.demirci@warwick.ac.uk.

1 Introduction

Insurance companies use historical data to predict the likelihood of a claim based on information collected from each customer, and they base their prices on the expected individual cost. Hence, using all information that may reflect risk is economically efficient for companies. If they cannot differentiate between groups with different perceived risks, it may raise adverse selection and moral hazard issues (Puelz and Snow, 1994; Finkelstein and Poterba, 2004; Cohen, 2005). On the other hand, using statistical information on the group belonging to infer individual risk is discrimination (Arrow, 1973; Phelps, 1972). If one group is characteristically riskier than the other, then each individual in this group will be assumed to be riskier, although their individual predicted risk may be different. In that sense, the trade-off between efficiency and equality has created public policy debates in insurance pricing.

Auto insurers collect various information like other insurance policies, including gender, age, marital status, and driver characteristics such as car model and driver history. Young males under 25 often pay more for car insurance than young females, whereas this gender gap disappears with more experienced drivers. As companies collect more information on individual driving records and accident histories, they do not need to proxy risk with gender. The main reason for this initial gender gap is driver statistics itself. Young males are more likely to involve in more accidents, and their accidents are often more severe (Huh and Reif, 2021; Moore and Morris; 2021). In that sense, using gender information of those who do not have driving records is efficient for firms, as they cannot perfectly observe individual-level risk for the young. However, it is also statistical discrimination to price people based on group-based characteristics rather than their individual risk level. Hence, in many places worldwide, gender-neutral pricing policies have been adopted. In Europe, together with the EU Gender Initiative in 2011, the EU countries must have gender-neutral insurance policies. Also, some states in the US have banned gender-based pricing for car insurance, and recently California joined them in 2018 [1].

[1] As of 2021, Hawaii, Massachusetts, Pennsylvania, North Carolina, and Montana are gender-neutral

In this paper, I analyze the impact of a recent policy change that bans the use of gender in automobile insurance in California. Particularly, I intend to answer if this policy eliminates the gender gap, what are the distributional consequences of this ban on different demographic groups, and whether firms start to use other characteristics correlated with gender after the policy change. In order to answer these questions, I exploit two sources of variation in policy implementation. First, I focus on gender gaps in different age groups and compare the gender gap among young people and older people. By using triple difference-in-differences, I find that the 15 percent gender gap has decreased by 8 percentage points after the gender-neutral pricing policy. Second, I use the variation in the policy implementation across states and find that 4 percentage points decrease in the gender gap of young people in California compared to the young gender gap in non-treated states.

Furthermore, I investigate how firms responded to this policy change. Specifically, did they change their pricing algorithms to compensate for the information loss and use characteristics correlated with gender to proxy it? I construct a novel dataset based on insurance company filings that contain risk factors, risk scores, and pricing algorithms used in insurance pricing. Using different machine learning prediction models, I first estimate characteristics correlated with gender, such as different demographics and car types. Then, I analyze whether there is any change in the weightings of the gender-correlated features in the pricing algorithms after the policy. I find that people with young male characteristics (the riskiest group) started to pay up to 10 percent more on average, and particularly, insurance risk scores for car types associated with young males increased up to 20 percent.

Regulations on asking specific features such as age, gender, race are common in insurance markets. Many scholars focus on these regulatory reforms in different insurance markets and their impact on the economies. Finkelstein et al. (2011) theoretically show that the gender asking ban in the UK annuity market distorts the redistribution from less risky groups to more risky groups and leads to inefficient market outcomes. Other theoretical studies

states for car insurance.

focus on characteristics-based pricing in insurance and find that restrictions on including these characteristics in pricing create a loss of efficiency (Hay, 1982; Cracker and Snow, 1987; Reo, 1987; Polborn et al.;, 2006). From an empirical perspective, Huang and Sham (2018) investigate the outcomes of the gender ban in the health insurance market in Germany. They show that the unisex pricing mandate made higher risk groups (women, in this context) switch from social health insurance that never uses gender as a pricing factor to public health insurance that previously uses gender as a risk factor. Hence, the unisex mandate creates an adverse selection and inefficient equilibrium outcomes. More broadly, Pope and Sydnor (2011) focus on implementing anti-discrimination policies in statistical profiling models. They stress that banning a specific feature leads to the use of proxies of banned variables and further provides a framework to eliminate proxy effects while maintaining predictive accuracy. Another empirical study by Asservatham et al. (2011) focuses on a similar reform in the German automobile insurance market and analyzes how proxy variables may impact pricing. They find that firms' behaviors will likely be affected by proxy effects for especially young and old drivers. However, this study uses a dataset of insurance policies for a single year prior to the policy implementation year and predicts how firms will change their pricing. In that sense, my paper provides a comprehensive analysis using data from before and after the policy with actual prices paid for insurance. To the best of the author's knowledge, this work is, therefore, the first to provide an empirical investigation of the impact of gender-neutral auto insurance pricing policy in the US.

Similar to insurance markets, there are various regulations in labor markets that limit using different features of applicants by a decision-maker. For instance,' Ban the Box' policies require not asking criminal records in job applications until a late stage. These policies aim to prevent potential discrimination against ex-offenders in the labor market. However, some studies show that under these policies, employers start to use other observables such as the race to proxy for a criminal conviction (Agar and Starr, 2018; Doleac and Hansen, 2020). Hence, these policies have created unintended outcomes such as a lower likelihood of getting an interview or employment for African-American males. In another related literature, Bartik and Nelson (2016) focus on bans on credit history checks of job applicants. They find that preventing employers from accessing applicants' credit histories leads them to use other characteristics such as race to predict credit scores. Therefore, this ban reduced job finding rates significantly for black job seekers. In that sense, my study contributes a growing literature emphasizing that anti-discrimination policies that remove information about different applicant characteristics might create unintended outcomes and sometimes even more harm on other demographic groups than targeted ones.

This paper is also relevant to the current debates and ongoing work on algorithmic discrimination (Arnold et al., 2018; Lambrecht and Tucker, 2019; Cowgill and Tucker, 2020; Ukanwa and Rust, 2020; Arnold et al., 2021). Insurance companies often use prediction algorithms to assign a likelihood of making an accident or a claim to each customer. Based on these predictions, pricing algorithms offer insurance price quotes for each individual. These mechanisms rely on historical data from previous customers. Therefore, they inherently perpetuate bias placed in previous policy holders' information. Although anti-discriminatory policies in insurance aim to eliminate bias and offer fair insurance prices across different groups, algorithmic pricing can lead to further discriminatory outcomes. This study aims to contribute to the growing algorithmic bias literature by focusing on anti-discriminatory policy regulations in algorithmic insurance pricing. One genuine difficulty of studying policies targeting algorithmic decision-making is that pricing algorithms are usually considered black boxes. Therefore, researchers can observe only the outcomes after the policy regulation but not the direct mechanisms driving the results. However, the automobile insurance industry provides a salient advantage to study algorithmic bias, as companies are required to disclose their pricing rule due to regulations in most US states. In this study, I can observe pricing rules from different insurance providers and analyze the changes in these algorithms after the gender-neutral pricing policy.

The structure of the paper is as follows. Section 2 focuses on a conceptual framework about the US automobile insurance system, Section 3 discusses the theoretical background and Section 4 presents data. In Section 5, I discuss the results of empirical analysis.

2 Background on the Automobile Insurance

Auto insurance is one of the most used types of personal insurance, and it is mandatory in most states.[2] Insurers collected 162.4 billion dollars in 2020 for private automobile insurance in the US. There are two broad categories for automobile insurance: liability and property coverages. Liability insurance covers the property damage and bodily injuries to another person caused by accident with an insured's fault. Property insurance provides coverage for one's car regardless of fault. All states require liability insurance, but different states mandate different minimum level requirements for this coverage.

The insurance industry in the US is heavily regulated. Insurance companies are required to submit detailed filings to regulatory agencies in each state and explain the details of their pricing. Each state has its regulation to ensure insurance pricing is fair and affordable for all drivers and prevent excessive pricing for some demographic groups. State-level insurance laws specify which features can be used as a risk factor in pricing. Companies have to show which risk factors they use in their pricing algorithm and the risk scores attached to each risk factor.[3]

In general, risk factors can be grouped into three broad categories: driver, vehicle, and location characteristics. These characteristics include age, gender, marital status, ZIP code, vehicle make and model, driving history, etc. In addition to these factors, an insurance policy price has a base rate that varies over the state and insurance coverage choice in a way that choosing more extensive policy coverage increases the price.

[2] Although its minimum requirements vary across states, all US states except New Heaven mandate to have liability insurance.

[3] In insurance terminology, risk scores are often referred to as risk relativities. Risk relativity shows how risky a specific feature is seen relative to other features in that risk category. For instance, risk relativity (risk score as used throughout this paper) for single male drivers with no driving experience gives an expected risk of this group relative to single male drivers with driving experience.

One stream of the policy debates regarding automobile insurance in the US focuses on characteristics-based pricing. Regulation supporters often argue that using group-based characteristics rather than individual-level risk is discriminatory. In 2018, California banned the use of gender in auto insurance rating, and New York prohibited using occupation and education information in 2017. Some states such as Massachusetts, North Carolina, Pennsylvania, and Hawaii restricted using gender as a risk factor in the 1980s. The main argument for these regulations is to ensure that insurance pricing is based on factors within a driver's control but not characteristics given at birth, such as race or gender.

3 Theoretical Framework

This section lays out a brief theoretical framework that tries to explain the impacts of the gender-blind policy and adapts a signaling framework commonly used for studying statistical discrimination (Aigner and Cain 1977; Altonji and Pierret 2001; Lundberg and Startz 1983). I introduce a model where firms try to predict individual-level risk by observables they have collected. Then, it suggests some theoretical predictions of what can happen when firms' ability to monitor risk predictors is constrained.

3.1 Environment

Assume that insurance firms cannot observe an individual i's actual risk while driving, but instead, they can observe a noisy signal of the individual risk level. The risk signal is defined as:

$$y_i = x_i + \eta_i$$

where $\mathrm{y_i}$ represents the risk signal of individual i, $\mathrm{x_i}$ is the true risk of individual i, and $\mathrm{n_i}$ denotes the noise in the signal. The noise η_i is assumed to be independently and normally distributed with zero mean and finite variance implying that $\eta_\mathrm{i} \sim \mathrm{N}(0, \sigma_{\eta_\mathrm{i}})$. Moreover, firms learn from experience that the individual risk x_i is normally distributed with mean $\bar{x}$ and variance σ_{x_i}. The accuracy of the signal is independent of individual risk, $E(x_i\eta_i) = 0$

The expected risk for individual i is the weighted average of the signal y_i and the distributional prior p_i for group belonging. Here, signals can be considered as individual accident records, and distributional priors are group-based characteristics such as gender or postcode. The expected risk for individual i conditional on signal y_i is given by

$$E(x_i|y_i, z_i) = \gamma y_i + (1-\gamma)p_i$$

where $\gamma = E(x_i y_i) \ / \ E(y_i y_i) = \frac{\sigma_{\eta_i}}{\sigma_{\eta_i}+\sigma_{x_i}}$ which is the coefficient from a bivariate regression of x_i on y_i and a constant.

3.2 Restrictions on monitoring consumer characteristics

Let us define two cases depending on the firm's ability to monitor consumer i's characteristics:
Non-blind case: Firms can observe all signals and group belongings. Hence, the expected risk for individual i will be:

$$E(x_i|y_i, z_i) = \gamma y_i + (1-\gamma)p_i$$

Blind case: Firms cannot observe group belonging.Hence, firstly, they will estimate a probability of being from a particular group based on signal y_i. For simplicity, let us assume there are two groups as g_1 and g_2. The probability of being from gg_n where $n \in 1{,}2$ is given by:

$$P(i \in g_1|y_i) = \alpha_0 + \alpha_1 y_i + u_i$$

Next, the firm formulates an expected risk based on $E(p_i)$, an expected value of distributional prior:

$$\mathrm{E(x_i|y_i)} = \gamma y_i + (1-\gamma)\mathrm{E(p_i)}$$

where

$$\mathrm{E(p_i) = P(i \in g_1|y_i)p_1 + P(i \in g_2|y_i)p_2}$$

denotes the weighted average distributional priors fr different groups.

There can be different possibilities depending on firms' ability to predict the probability of being from a particular group and hence the ability to calculate the expected value of distributional prior. For simplicity, let us consider two extreme cases:

- If firms are myopic and cannot predict banned variables, then the expected value of risk will only depend on the signal:

$$\mathrm{E(x_i|y_i)} = \gamma y_i$$

- If firms can fully predict group belonging, we can infer that the expected value of distributional prior is equal to its true value, i.e., $\mathrm{E(p_i) = p_i}$. Hence, the expected value of risk conditional on the signal will be

$$\mathrm{E(x_i|y_i, z_i)} = \gamma y_i + (1-\gamma)p_i$$

3.3 Predictions

Based on this framework and following Pope and Sydnor (2011), I generate different testable predictions. For the proofs of these testable predictions, please refer to Appendix A.

Prediction 1: Expected risk between two groups will be less in blind case if firms cannot perfectly predict group belonging. Let us assume group 1 is riskier than group 2 on average without loss of generality.

$$\mathrm{E}^{\mathrm{Non-blind}}(\mathrm{x_i}|\mathrm{y_i}, \mathrm{i} \in \mathrm{g_1}) - \mathrm{E}^{\mathrm{Non-blind}}(\mathrm{x_i}|\mathrm{y_i}, \mathrm{i} \in \mathrm{g_2}) >$$
$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}, \mathrm{i} \in \mathrm{g_1}) - \mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}, \mathrm{i} \in \mathrm{g_2})$$

Prediction 2: Signals $\mathrm{y_i}$ positively correlated with group belonging will gain more weight in risk prediction.

$$\mathrm{E}^{\mathrm{Non-blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_0 y_i + (1 - \gamma_0) p_i$$

and

$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_1 y_i + (1 - \gamma_1)\mathrm{E}(\mathrm{p_i})$$

Suggesting that, $\gamma_1 \geq \gamma_0$.

4 Data

This paper uses three main datasets on (i) car insurance prices, (ii) car insurance pricing filings, and (iii) car user characteristics.

4.1 Car insurance expenditures

The data on car insurance prices are obtained through the US Consumer Expenditure Survey Microdata, a large-scale longitudinal dataset weighted to be representative of the US population. This survey provides information about US consumers' quarterly expenditures, incomes, and consumer unit characteristics. It also includes information on how much people spend on car insurance and demographics such as age, marital status, occupation, education level, living area. Furthermore, the survey presents car characteristics such as car types, car brands, and year. These characteristics are essential to conducting a policy evaluation for car insurance, as they are the main features collected by insurance firms to offer price quotes. This paper focuses on the sample of consumers who own a car and buy car insurance between 2010 and 2020, and it includes around 125k individuals.

Table B1 provides information on the main summary statistics for the Consumer Expenditure Survey sample used in this paper. Panel A and B present individual characteristics, car characteristics, and yearly insurance expenditures for the population under 25 years old and all sample, respectively. Among young under 25, 49 percent are female, 72 percent are single, 33 percent have a college degree. In contrast, in all sample, the female population is 51 percent, single people are 43 percent, and 48 percent has a college degree. The younger population also differs significantly in terms of their car insurance expenditures. Annual car insurance spending is $2142 on average for all sample, whereas people under 25 pays approximately 38 percent more. Younger people are also more likely to second-hand vehicles.

4.2 Car insurance pricing filings

Auto insurers in the US have to submit filings that show their pricing details to insurance regulatory agencies in their state. These filings include information on (i) risk factors used in pricing calculations, (ii) risk scores attached to each risk factor and category, and (iii) pricing algorithm that combines these risk factors with risk scores and translates into insurance prices for each customer. These filings often span thousands of pages to verify that firms charge fair and affordable prices, and they have to disclose these filings. Using the National Association of Insurance Commissioners' SERFF database, I construct a comprehensive dataset including filings from various insurance companies. This dataset provides insurance risk factors and risk scores associated with these factors for multiple insurance companies between 2015 and 2020. They also display pricing rules for each company. Therefore, this feature of data allows me to replicate firms' pricing algorithms by incorporating all possible combinations of risk characteristics and risk scores. This paper uses this dataset to analyze how firms changed their pricing strategy when banned to use gender in their pricing.

4.3 Car user characteristics

An important question this paper aims to answer is if firms change their pricing algorithms after the gender-neutral pricing policy and if they use gender-correlated features to compensate for their information loss. To analyze this, we need a dataset to observe the type of information that car insurance companies collect, including gender and other characteristics that may correlate with gender. To conduct this analysis, I use US National Household Travel Survey; a large-scale survey conducted every 9 years. The sample in the paper focuses on the last three waves at 2001, 2009, and 2017. It includes a stratified random sample of U. S. households and has approximately 700K individuals. This survey provides information on various individual characteristics (age, gender, education, occupation, etc.), detailed vehicle characteristics (car brand, model, features, year), and location-specific characteristics. This combined dataset broadly covers all the information insurance collect from their customers. Using this dataset, I conduct different empirical analyses to estimate features correlated with gender.

5 Empirical Analysis

This paper studies the impact of the gender-neutral car insurance policy in California introduced on January 1, 2019. First, I will investigate how the policy affected the insurance gender gap in California. By using the difference-in-differences strategy, I will exploit variation in policy implementation within California across different age groups and states. Then, I will focus on how gender-correlated characteristics were affected in the firm's pricing algorithm. I will first investigate features that predict gender by using different machine learning prediction techniques to analyze this. Then, I will estimate how much people with these characteristics start to pay after the policy.

5.1 Relation between age, gender, and insurance prices

Gender is used as a rating factor in car insurance in many states as far as it is not prohibited by law. Insurers base their rating on how risky male or female drivers are seen on average. However, one crucial aspect of gender-based pricing is its interaction with age or driving experience. For younger drivers with a few years of driving experience, the gender gap is more apparent, and young males are associated with higher risk. Hence, they usually pay higher premiums than young female drivers. However, as drivers gain more driving experience, firms can base their insurance prices on individual driving and accident histories rather than gender as a noisy signal.

Figure 1 shows the age and gender risk scores used in the firm's pricing algorithm. Drivers with less than 10 years of experience are considered two times riskier than more experienced drivers. Also, the gender gap is more salient for less experienced drivers, but it becomes negligible as drivers gain more experience. Risk scores are raw measures of a firm's perception of risk and how a specific risk group is regarded as risky compared to other risk groups. They directly enter the firm's pricing algorithm and contribute to the insurance price.

Another angle of looking at the interaction of gender and age is focusing on the premium gap for different demographic groups. Figure 2 provides information on annual insurance premiums paid by male and female drivers of different ages. Young drivers under 25 spend more than twice what older drivers pay on average. After 25 years old, insurance premiums sharply decrease, and the middle-aged group pays the lowest premiums between 40 and 70 years old. In the elder group, premiums show a slight increase. Regarding differential pay between men and women, the raw gap among drivers under 25 is about 14 percent, whereas the after 25 years old men and women pay very similarly for car insurance. This finding is in line with the prediction of insurance companies using driving records rather than gender as a risk proxy.

5.2 How gender gap has evolved after the gender-neutral insurance pricing policy?

The first question asked in this paper is how gender-blind pricing policies affect how much males and females pay for car insurance. This section will analyze if the policy eliminates the gender gap entirely and how much change occurred in premiums paid by different age and gender groups. I will start by estimating the differential impact of the policy on the gender gap among young compared to the gender gap among old drivers in California. Then in the following subsection, I focus on the exogenous variation in the policy implementation across states and analyze the impact of policy on the gender gap for young in California relative to young in other states that adopts gender-based pricing policies.

5.2.1 Variation across age groups within California

By using a triple difference-in-differences empirical strategy, this section estimates the impact of gender-neutral pricing policy on the gender gap in insurance premiums in California. I compare the gender gap among young individuals under 25 with older people. As explained in the previous section, gender is not considered as a risk factor after certain years of driving experience. The identifying assumption is that the gender gap among older people is a good counterfactual of how the gender gap among the young would have evolved in the absence of policy. As shown in Figure 3, all groups followed parallel trends before the policy supporting the assumption that they would have moved similarly without the policy.

In order to estimate the differential impact of policy on the young gender gap, I estimate the following equation:

$$\begin{aligned}\text{Log(Price)}_{it} = \alpha_0 \text{Young}_i + \alpha_1 \text{Male}_i + \alpha_2 (\text{Young x Male})_i + \beta_1 (\text{Policy x Young})_{it} \\ + \beta_2 (\text{Policy x Male})_{it} + \beta_3 (\text{Policy x Male x Young})_{it} + \gamma \text{X}_{it} + \alpha_t + \epsilon_{it} \quad (1)\end{aligned}$$

where for each individual i and year t, $\log(\mathrm{Price}_{\mathrm{it}})$ represents the logarithm of insurance premium price paid by individual i at year t, $\mathrm{Young}_{\mathrm{i}}$ and $\mathrm{Male}_{\mathrm{i}}$ represent if individual i is under 25 years and male, respectively. $\mathrm{Policy}_{\mathrm{t}}$ takes value 1 if the policy is implemented at time t, 0 otherwise. X_{it} denotes a set of control variables such as age, marital status, car brand, and year for each individual i at time t. Year fixed effects denoted by α_{t} are also included in each regression.

Table 1 reports the results from estimations of equation (1). In both specifications, the sample is California, and the treatment group is young people under 25 years old. In specification (1), the gender gap among young compared to older people decreased 4 percent. If we also account for control variables, the treatment effects become 6 percent in the specification (2).

5.2.2 Variation in young gender gap across states

This section analyzes the differential impact of policy on the gender gap among young in California compared to other states adopting gender-based insurance pricing. By considering young drivers in other states as a control group, the identifying assumption requires that young in California would have similar trends in the absence of gender-neutral policy.

Figure 4 depicts parallel trends in insurance premiums paid by these four demographic groups and the gender gaps in treatment and control states. The left panel shows the average premiums young drivers pay in California overtime. After the policy, the average premiums paid by young males decrease, and young females' premiums increase in the treated state. Hence, there is a slight convergence between young men's and women's premiums paid. On the other hand, as seen in the right panel, the gender gap among young in non-treated states is more stable than in California after the policy.

In order to estimate the causal impact of policy, I use a triple difference-in-difference strategy and compare the gender gap among young male and female drivers in California and other states with gender-based insurance pricing. The following equation is estimated:

$$\begin{aligned}\text{Log(Price)}_{it} = {} & \alpha_0 \text{California}_i + \alpha_1 \text{Male}_i + \alpha_2 (\text{California x Male})_i + \beta_1 (\text{Policy x California})_{it} \\ & + \beta_2 (\text{Policy x Male})_{it} + \beta_3 (\text{Policy x Male x California})_{it} + \gamma \text{X}_{it} + \alpha_t + \epsilon_{it} \quad (2)\end{aligned}$$

where for each individual i and year t, $\log(\text{Price}_{\text{it}})$ represents the logarithm of insurance premium price paid individual i at year t. $\text{California}_{\text{i}}$ and Male_{i} represent if individual i lives in California and is male, respectively. Policy_{t} is 1 after the policy is implemented at time t, 0 otherwise. X_{it} is a set of control variables such as age, marital status, car brand, and year for each individual i at time t. Year fixed effects denoted by α_{t} are also included in each regression. All standard errors are clustered at the state level.

Table 2 presents estimation results for equation (2). The main treatment effect leads to a 4 percent decrease in the gender gap in California compared to the gender gap in other states after accounting for a set of controls. One potential mechanism to explain this gender gap under gender-neutral pricing is the modifications in the pricing algorithm. It could be the case that to compensate their ability to charge young men and women differently, they can price other features that predict being a young male more after the policy to alleviate their profit loss in this riskiest driver group.

5.3 How do gender predictors change in pricing algorithms?

This section examines the impact of gender-neutral pricing on a firms' pricing algorithm. Specifically, I analyze whether the observables correlated with the riskiest demographic group, i.e., young males are given more weight in price calculations. In order to answer this question, I will first start estimating features that predict being a young male by using a rich dataset that includes both individual and vehicle characteristics. I will employ different machine learning prediction techniques to select features. Next, I estimate the impact of policy on risk scores of these gender predictors by using insurance firms' pricing dataset.

Lastly, I analyze how drivers with these features start to pay regardless of their age and gender after the policy.

5.3.1 Gender-correlated characteristics

Insurance companies collect much information to give a price quote to each customer. These features are individual characteristics such as age, gender, marital status, education, occupation, location-based characteristics such as ZIP code, and vehicle characteristics such as car brand, model, year, whether it is used or a new car, etc. However, together with the gender asking ban, they cannot use gender in their pricing. In that sense, using other features they have already collected as proxies is one way to compensate for their information loss. In this section, by using an extensive household travel survey that matches individual and car characteristics in the US, I will estimate features that predict gender. The focus in this analysis will be on the prediction of being a young male given other features since the gender gap is not significant for middle-aged and elderly groups.

Table 3 presents the results for the prediction of being a young male. Being a young male under 25 is regressed on a wide range of control variables in these prediction models. Each column displays features that predict being a young male according to OLS, LASSO, ridge regression, and elastic net models, respectively. A given feature is listed in the table; only if it predicts the dependent variable, i.e., being a young male; otherwise, it is not shown. Mean squared errors in each method are very similar, implying that results are robust across different prediction models.

5.3.2 Changes in the pricing algorithm

This section analyzes how firms change their behavior in response to the gender-neutral policy. This behavior can reveal itself in two ways. The first one is the adjustments in the pricing algorithm for risk factors that are estimated as correlated with gender, and the latter is how much people with these characteristics start to pay after the policy change. In order to analyze the raw changes in the risk scores, I use the pricing filings for insurance companies. These filings show the risk factors used in pricing, the risk scores attached to each risk factor, and the pricing rule. In that sense, it enables us to analyze changes in the raw risk scores for different attributes, which eventually reflects the insurance price by entering into the pricing algorithm.

The prediction of being a young male conditional on different attributes suggests that owning a particular type of car model or brand such as pickup cars is associated with young males. Hence, I created an aggregate term called young male cars by using these specific car types and brands that young males predominantly use. This variable takes the value of 1 if a given car brand predicts being a young male compared to other demographic groups. In line with the same intuition, another variable for gender-neutral cars is created. This neutral cars variable is an aggregate term for cars if a given car model is equally likely to be used by young males and other groups. Then, I compare how risk scores change over time for these two car categories.

Figure 5 shows the risk scores of these cars between 2015 and 2020. Risk scores can be interpreted as relative measures of how risky each group is compared to other groups. For instance, car models associated with young males are considered around 7 percent riskier than gender-neutral cars. However, after the policy, raw risk scores increased for young male vehicles, and the risk score gap between the two groups became around 14 percent.

Interpretation of risk scores and how they changed relative to each other for different groups provide a solid intuition of the adjustments in the insurance pricing algorithms. However, this analysis is not easy to interpret how risk scores translated into insurance price. In order to examine how each customer is affected by these algorithm adjustments, we need to analyze insurance prices. Also, each consumer's price is calculated as a combination of risk scores in different dimensions, such as risk score according to their car model, marital status, or driving histories. In that sense, how people are affected by changes in different dimensions of pricing rules depends on their unique features. To illustrate, the increase in young male car risk scores could affect anyone who uses these cars regardless of their age and gender. In that sense, people who have young male features in terms of observables can end up paying more after these adjustments. These unintended outcomes are one of the most controversial aspects of anti-discriminatory laws in this domain.

In order to estimate the changes in insurance price for young male features, I estimate the following equation:

$$\text{Log(Price)}_{it} = \alpha \text{Feature}_i + \beta(\text{Policy x Feature})_{it} + \gamma \text{X}_{it} + \alpha_t + \epsilon_{it} \quad (3)$$

where for each individual i and year t, log(Price_{it}) represents the logarithm of insurance premium price paid individual i at year t. $\text{Feature}_{\text{i}}$ represents if individual i has a given feature such as having a specific car brand. Policy_{t} is 1 after the policy is implemented at time t, 0 otherwise. X_{i}t is a set of control variables such as age, marital status for each individual i at time t. Year fixed effects denoted by α_{t} are also included in each regression.

Figure 1 shows estimation results of changes in insurance prices of different young male features. In column (1), a composite variable that captures all young male predictors is created. Estimation results show that drivers with these features pay 4 percent more on average. However, after the policy change, the price differential for these prices increase to 7 percent. The remaining specifications list estimates of insurance price change in different characteristics separately. Column 2 reports that single drivers pay 10 percent more on average. Significantly, as shown in Column 3, the insurance price for pickup car owners increase by 9 percent after the gender-neutral pricing policy. On the other hand, there is no significant change in how much people who one gender-neutral car pay.

6 Conclusion

This paper investigates the impact of gender-neutral pricing policies on firms and their pricing strategy. Specifically, I aim to answer which risk factors gain more weight in the pricing after the gender-neutral pricing policy is implemented, how women and men are affected by the policy, and if there are any heterogeneous impacts for people with different features. In order to answer these questions, I focus on the recent law change in California in 2019 requiring gender-neutral pricing in auto insurance.

By exploiting the variation in the policy implementation over time and across states, I show that the price gap between male and female drivers decreased after the gender-neutral pricing policy, as young male drivers used to pay significantly more before. Evidence suggests that after implementing this anti-discrimination policy in California, the firms' pricing algorithms started to use other features that it treats as correlated with gender to compensate for the gaps in data. Therefore, people who choose specific car models associated with the riskier demographic groups started to pay more for car insurance, irrespective of their gender. After implementing the gender-neutral policy, even if individuals have a good driving record with no history of accidents, they have to pay more if they use specific car brands. Furthermore, the policy has heterogeneous impacts on different demographic groups. It decreased premiums for young male drivers while increasing for young females and middle-aged people.

This tendency is an interesting example of how algorithmic pricing may perpetuate the bias embedded in previous customer data, even though the anti-discrimination policies aim to achieve more equity. A critical implication of this paper is understanding to which extent anti-discrimination policies achieve their aims, especially when dealing with inherent bias in algorithms. The anti-discrimination initiatives have good intentions and recognize that discrimination is a significant problem. However, these well-intended policies may not necessarily bring equity in all settings. In that sense, policymakers need to acknowledge that discrimination may still exist in other forms and may be weaved into the algorithms designed to automate decision processes.

References

Agan, A., & Starr, S. (2018). Ban the box, criminal records, and racial discrimination: A field experiment. The Quarterly Journal of Economics, 133(1), 191-235.

Aigner, D. J., & Cain, G. G. (1977). Statistical Theories of Discrimination in the Labor Market. Industrial and Labor Relations Review, 175–187.

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. The Quarterly Journal of Economics, 116(1), 313-350.

Arrow, K. J., Ashenfelter, O., & Rees, A. (2015). The theory of discrimination (pp. 1-33). Princeton University Press.

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. The Quarterly Journal of Economics, 133(4), 1885-1932.

Arnold, D., Dobbie, W., & Hull, P. (2021). Measuring racial discrimination in algorithms. In AEA Papers and Proceedings (Vol. 111, pp. 49-54).

Aseervatham, V., Lex, C., & Spindler, M. (2016). How do unisex rating regulations affect gender differences in insurance premiums?. The Geneva Papers on Risk and Insurance-Issues and Practice, 41(1), 128-160.

Bartik, A., & Nelson, S. (2016). Deleting a signal: Evidence from pre-employment credit checks.

Cohen, A. (2005). Asymmetric information and learning: Evidence from the automobile insurance market. Review of Economics and statistics, 87(2), 197-207.

Cowgill, B., & Tucker, C. E. (2020). Algorithmic fairness and economics. Columbia Business School Research Paper.

Crocker, K. J., & Snow, A. (1986). The efficiency effects of categorical discrimination in the insurance industry. Journal of Political Economy, 94(2), 321-344.

Doleac, J. L., & Hansen, B. (2020). The unintended consequences of “ban the box”: Statistical discrimination and employment outcomes when criminal histories are hidden. Journal of Labor Economics, 38(2), 321-374.

Finkelstein, A., & Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the UK annuity market. Journal of Political Economy, 112(1), 183-208.

Finkelstein, A., Poterba, J., & Rothschild, C. (2009). Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities. Journal of financial economics, 91(1), 38-58.

Hoy, M. (1982). Categorizing risks in the insurance industry. The Quarterly Journal of Economics, 97(2), 321-336.

Huang, S., & Salm, M. (2020). The effect of a ban on gender-based pricing on risk selection in the German health insurance market. Health economics, 29(1), 3-17.

Huh, J., & Reif, J. (2021). Teenage driving, mortality, and risky behaviors. American Economic Review: Insights, 3(4), 523-39.

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Management science, 65(7), 2966-2981.

Lundberg, S. J., & Startz, R. (1983). Private discrimination and social intervention in competitive labor market. The American Economic Review, 73(3), 340-347.

Moore, T. J., & Morris, T. (2021). Shaping the Habits of Teen Drivers (No. w28707). National Bureau of Economic Research.

Phelps, E. S. (1972). The statistical theory of racism and sexism. The American Economic Review, 62(4), 659-661.

Polborn, M. K., Hoy, M., & Sadanand, A. (2006). Advantageous effects of regulatory adverse selection in the life insurance market. The Economic Journal, 116(508), 327-354.

Pope, D. G., & Sydnor, J. R. (2011). Implementing anti-discrimination policies in statistical profiling models. American Economic Journal: Economic Policy, 3(3), 206-31.

Puelz, R., & Snow, A. (1994). Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market. Journal of Political Economy, 102(2), 236-257.

Rea, S. A. (1987). The market response to the elimination of sex-based annuities. Southern Economic Journal, 55-63.

Ukanwa, K., & Rust, R. T. (2020). Discrimination in service. Marketing Science Institute Working Paper Series Report, (18-121).

Figures

Figure 1: Risk scores by age and gender

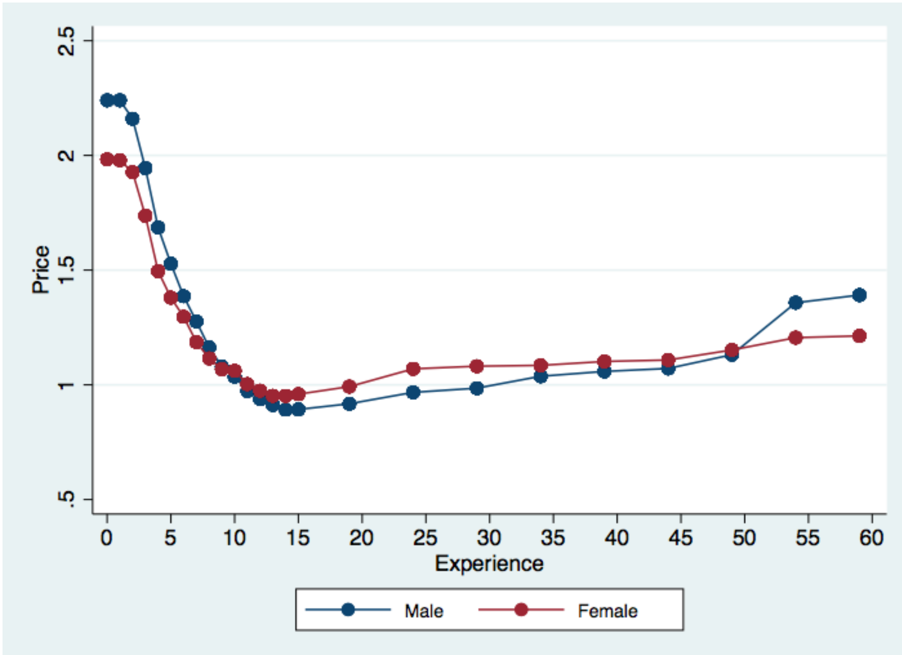


Figure 2: Insurance premiums paid by age and gender

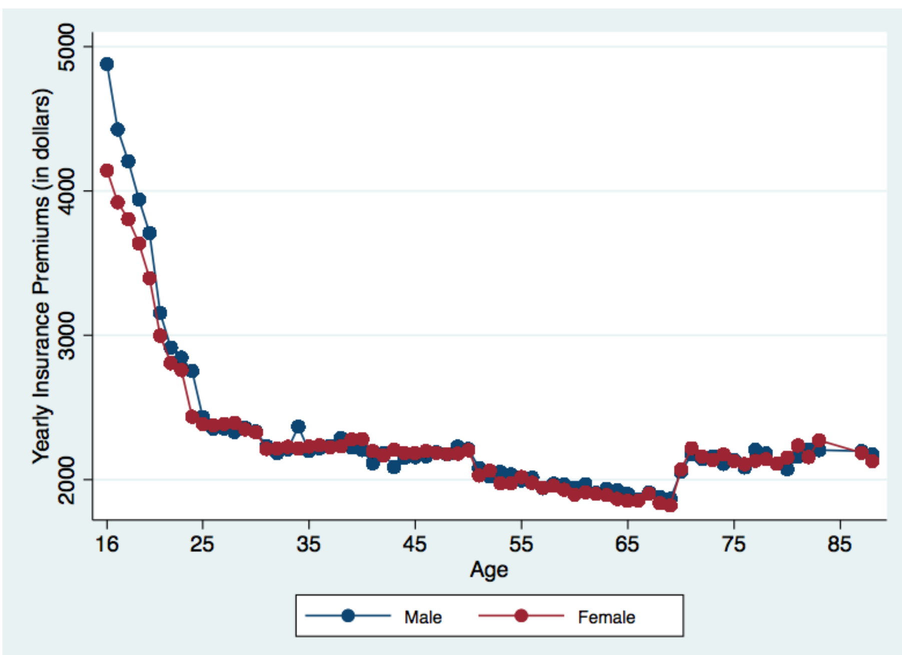


Figure 3: Gender gap among young and old people in California

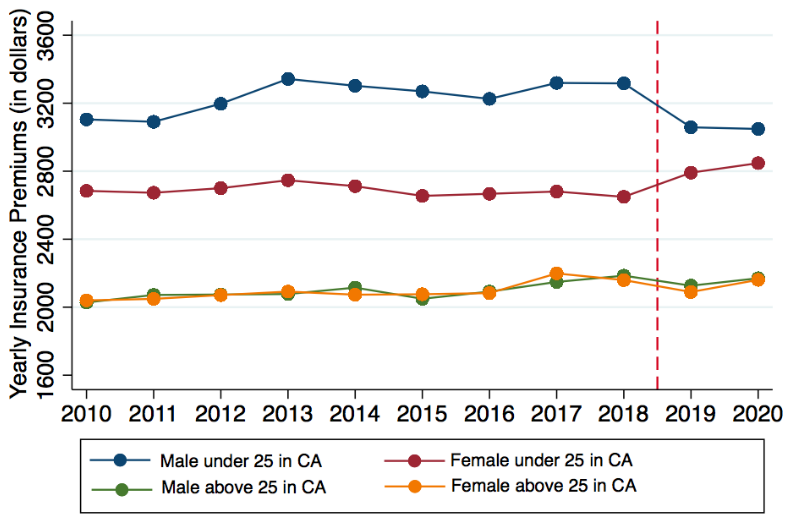
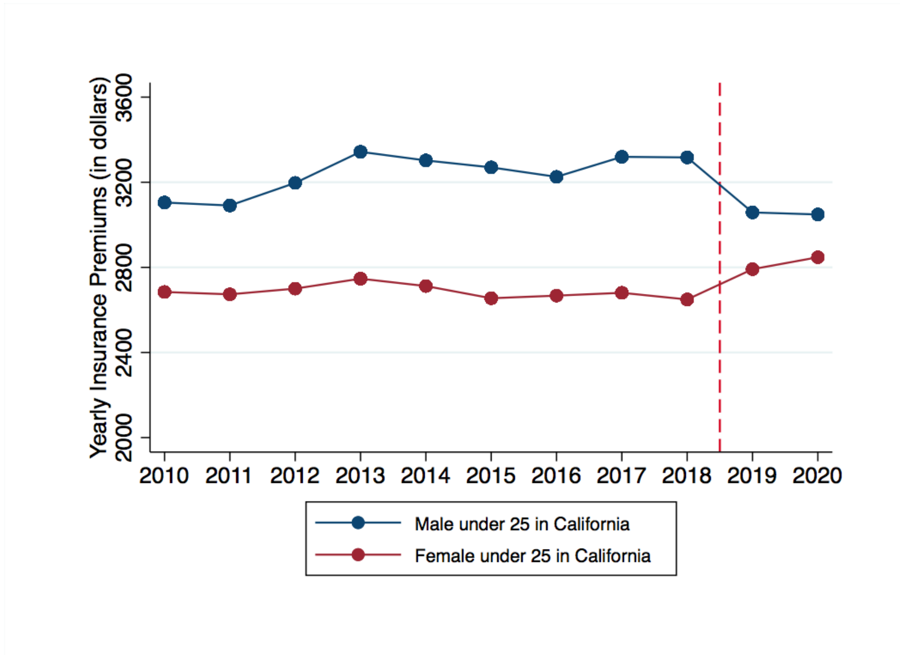
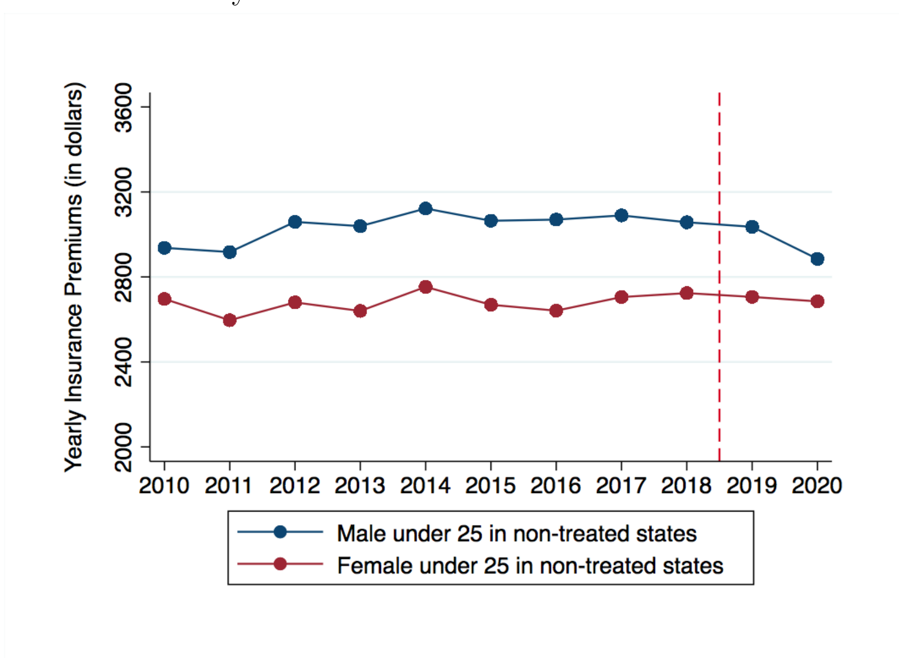


Figure 4: Gender gap among young in California vs non-treated states

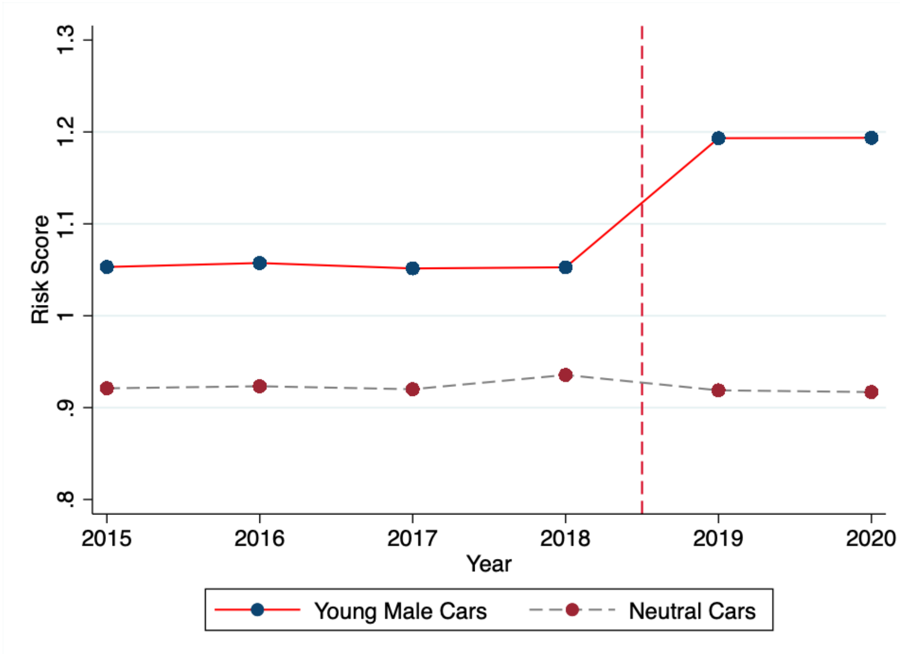


Under 25 years old in California



Under 25 years old in non-treated States

Figure 5: Changes in vehicle risk scores after the policy



Tables

Table 1: Gender gap among young and old poeple in California

Log(Price)	(1)	(2)
Policy x Young x Male	-0.048*	-0.062**
	(0.026)	(0.026)
Young (under 25)	0.197***	0.116***
	(0.015)	(0.016)
Male	0.005	0.006
	(0.005)	(0.005)
Young x Male	0.108***	0.124***
	(0.018)	(0.018)
Sample	California	California
Controls	No	Yes
Observations	15,509	15,509
R-squared	0.074	0.190

Notes: US CEX Survey 2010-2020. Controls are age, marital status, years of education, MSA dummy, race, if car is used or new, car year, car model. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Gender gap among young people in California and non-treated states

Log(Price)	(1)	(2)
Male	0.071***	0.096***
	(0.025)	(0.014)
California	0.004	0.003
	(0.008)	(0.007)
Policy x Male	-0.027	-0.049
	(0.027)	(0.033)
California x Male	0.051***	0.087**
	(0.012)	(0.012)
Policy x California x Male	-0.011	-0.041**
	(0.020)	(0.018)
Sample	Under 25 in all states	Under 25 in all states
Controls	No	Yes
Observations	4,189	4,189
R-squared	0.043	0.342

Notes: US CEX Survey 2010-2020. Controls are age, marital status, years of education, MSA dummy, race, if car is used or new, car year, car model. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Feature selection with different prediction methods

Category	Variable name	OLS	Lasso	Ridge	Elastic Net
Individual	Age	x	x	x	x
	Education Level	x	x	x	x
	Occupation	x	x	x	x
	Ethnic group	x	x	x	x
HH	HH size		x	x	x
	HH income		x	x	x
	MSA	x	x	x	x
	Homeownership		x	x	x
Car	Car category	x	x	x	x
	Annual miles	x	x	x	x
	Gas/diesel/hybrid		x	x	x
	Ford F-series pickup	x	x	x	x
	Ram pickup	x	x	x	x
	Toyota Tacoma pickup		x	x	x
	Honda Accord	x	x	x	x
	Taurus		x	x	x
	Silverado			x	
	Toyota Four Runner			x	
	Mean squared error	0.02262	0.02262	0.02263	0.02263
	R-squared	0.1642	0.1642	0.1642	0.1642
	Observations	58,805	58,805	58,805	58,805

Table 4: Changes in insurance price of young male features

Log(Price)	(1)	(2)	(3)	(4)
Prediction of being male under 25	0.045**			
	(0.035)			
Policy * Prediction of male under 25	0.029**			
	(0.014)			
Single		0.106***		
		(0.013)		
Single x Policy		0.002		
		(0.006)		
Pickup Cars			0.019*	
			(0.008)	
Pikcup Cars x Policy			0.091***	
			(0.017)	
Gender Neutral Cars				0.038
				(0.027)
Gender Neutral Cars x Policy				0.014
				(0.017)
Controls	No	Yes	Yes	Yes
Sample	All in CA	All in CA	All in CA	All in CA
Observations	15,509	15,509	15,509	15,509
R-squared	0.167	0.131	0.128	0.128

Notes: US CEX Survey 2010-2020. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Controls are age, marital status, years of education, MSA dummy, race, if car is used or new, car year, car model.

Appendix A. Proofs

This sections presents the proofs of the theoretical predictions presented in Section 3.3.

Prediction 1: Expected risk between two groups will be less in blind case if firms cannot perfectly predict group belonging. Let us assume group 1 is riskier than group 2 on average without loss of generality, i.e. $(\mathrm{p_1} > \mathrm{p_2})$.

$$\mathrm{E^{Non-blind}(x_i|y_i, i \in g_1) - E^{Non-blind}(x_i|y_i, i \in g_2) > E^{Blind}(x_i|y_i, i \in g_1) - E^{Blind}(x_i|y_i, i \in g_2)}$$

Proof:

$$\mathrm{E^{Non-blind}(x_i|y_i, i \in g_1) - E^{Non-blind}(x_i|y_i, i \in g_2) > E^{Blind}(x_i|y_i, i \in g_1) - E^{Blind}(x_i|y_i, i \in g_2)}$$

$$\gamma \mathrm{y_i} + (1-\gamma)\mathrm{p_1} - \gamma \mathrm{y_i} - (1-\gamma)\mathrm{p_2} > \gamma \mathrm{y_i} + (1-\gamma)\mathrm{E(p_1|y_i)} - \gamma \mathrm{y_i} - (1-\gamma)\mathrm{E(p_2|y_i)}$$

$$(1-\gamma)(\mathrm{p_1} - \mathrm{p_2}) > (1-\gamma)[\mathrm{E(p_1|y_i) - E(p_2|y_i)}]$$

$$(\mathrm{p_1} - \mathrm{p_2}) > \mathrm{P(i \in g_1|y_i)p_1 + P(i \in g_2|y_i)p_2 - P(i \in g_1|y_i)p_1 - P(i \in g_2|y_i)p_2}$$

$$(\mathrm{p_1} - \mathrm{p_2}) > 0 \qquad \blacksquare$$

Prediction 2: Signals $\mathrm{y_i}$ positively correlated with group belonging will gain more weight in risk prediction.

$$\mathrm{E^{Non-blind}(x_i|y_i)} = \gamma_0 y_i + (1-\gamma_0) p_i$$

and

$$\mathrm{E^{Blind}(x_i|y_i)} = \gamma_1 y_i + (1-\gamma_1)\mathrm{E(p_i)}$$

Suggesting that, $\gamma_1 \geq \gamma_0$.

Proof:

$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_1 y_i + (1-\gamma_1)\mathrm{E}(\mathrm{p_i})$$

$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_1 y_i + (1-\gamma_1)[\mathrm{P}(\mathrm{i} \in \mathrm{g_1}|\mathrm{y_i})\mathrm{p_1} + \mathrm{P}(\mathrm{i} \in \mathrm{g_2}|\mathrm{y_i})\mathrm{p_2}]$$

$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_1 y_i + (1-\gamma_1)[(\hat{\alpha_0} + \hat{\alpha_1}\mathrm{y_i})\mathrm{p_1} + (\hat{\beta}_0 + \hat{\beta}_{1\mathrm{i}})\mathrm{p_2}]$$

$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_1 y_i + (1-\gamma_1)[(\alpha_0\mathrm{p_1} + \beta_0\mathrm{p_2}) + (\alpha_1\mathrm{p_1} + \beta_1\mathrm{p_2})\mathrm{y_i}]$$

Let us call the constant term $\alpha_0\mathrm{p_1} + \beta_0\mathrm{p_2}$ as c.

$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_1 y_i + (1-\gamma_1)[\mathrm{c} + (\alpha_1\mathrm{p_1} + \beta_1\mathrm{p_2})\mathrm{y_i}]$$

$$\mathrm{E}^{\mathrm{Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \underbrace{[\gamma_1 + (1-\gamma_1)(\alpha_1\mathrm{p_1} + \beta_1\mathrm{p_2})]}_{\text{Weights given to signal}}\mathrm{y_i} + \mathrm{c}$$

Whereas the expected risk $\mathrm{x_i}$ given signal $\mathrm{y_i}$ in non-blind case is

$$\mathrm{E}^{\mathrm{Non-Blind}}(\mathrm{x_i}|\mathrm{y_i}) = \gamma_1 y_i + (1-\gamma_1)\mathrm{p_i}$$

By definition $\gamma > 0$ and $0 \leq \mathrm{p_i} \leq 1$.

Also, β_1 is non-negative by definition as it is the coefficient in the regressions of signals $\mathrm{y_i}$ that are positively correlated with group belonging.

Then, we can infer that

$$\gamma_1 + (1-\gamma_1)(\alpha_1\mathrm{p_1} + \beta_1\mathrm{p_2}) > \gamma_0 \qquad \blacksquare$$

Appendix B. Additional Tables

Table B1: Summary Statistics

Panel A. Population under 25

Individual Characteristics	N	Mean	Min	Max
Female	5,593	0.49	0	1
Age	5,593	22.89	16	25
Single	5,593	0.72	0	1
White	5,593	0.81	0	1
College	5,593	0.33	0	1
Living in a MSA	5,593	0.97	0	1
Insurance				
Yearly insurance expenditure ($)	5,593	2,951	1480	9000
Car type				
Automobile	5,593	0.76	0	1
Turck or van	5,593	0.23	0	1
Used car	5,593	0.79	0	1

Panel B. Main Sample

Individual Characteristics	N	Mean	Min	Max
Female	125,735	0.51	0	1
Age	125,735	51.28	16	88
Single	125,735	0.43	0	1
White	125,735	0.82	0	1
College	125,735	0.48	0	1
Living in a MSA	125,735	0.98	0	1
Insurance				
Yearly insurance expenditure ($)	125,735	2,142	804	6026
Car type				
Automobile	125,735	0.69	0	1
Turck or van	125,735	0.31	0	1
Used car	125,735	0.58	0	1