# What Determines Consumer Complaining Behavior?[*]

## Devesh Raval

Federal Trade Commission

[draval@ftc.gov](draval@ftc.gov)

July 30, 2016

## Abstract

User generated reviews have increasingly become a major information source on product quality for consumers, but little is known about who chooses to review and how they compare to either consumers of the product or the general population. I examine this issue of selection in the context of a type of review – consumer complaints to the government and other third party agencies – using a large database of millions of complaints. I first find evidence of substantial differences in complaint rates across demographic communities, with higher complaint rates in more heavily black, more educated, and more urban communities and lower complaint rates in more heavily Hispanic and higher household size communities. I then utilize a set of law enforcement actions for which I can compare complaining consumers to consumer databases of victims, and find much lower complaint rates in heavily black and heavily Hispanic areas and much higher complaint rates in areas with more college graduates. This evidence demonstrates the importance of understanding the selection of who chooses to review or complain about products in the marketplace.

# 1   Introduction

With the rise of the Internet, user generated reviews have become a major source of information for consumers on the quality of products. Consumers now judge books using reviews posted on Amazon.com, restaurants using Yelp, hotels using TripAdvisor, and doctors using HealthGrades. Not surprisingly, economists and marketers have found that such reviews have substantial effects on product demand.[1]

Despite the ubiquity of user reviews, very little is known about the characteristics of the users who choose to review, and how they compare to either the consumers of the product reviewed or the general population. If certain types of consumers are more likely to review than others, the self-selection of reviewers could affect both which products are reviewed and the assessment of quality provided by the reviews.

I examine the issue of selection in the context of negative reviews – consumer complaints – using data from the Consumer Sentinel Network. The Consumer Sentinel database includes millions of consumer complaints received every year by federal and state government agencies as well as private actors such as the Better Business Bureaus (BBBs).[2] Like reviews, complaints provide details on the author's experience with a company, although all complaints are negative experiences. Consumers can file complaints online, just as they can with reviews, although, unlike reviews, they can also be filed over the phone. Just as Yelp and Amazon provide average star ratings of businesses, the BBBs publically release company grades based on the complaints they receive and the company's response to them, and some BBBs will display complaint details online.

Unlike publically available data on online reviews, however, most consumers provide their address. I use this information to connect consumer complaints with zip code level

---

[1]For a non-exhaustive list of such evidence, see Chevalier and Mayzlin (2006) on Amazon.com, Luca (2011) on Yelp, Resnick et al. (2006) and Cabral and Hortacsu (2010) on eBay, and Balat et al. (2016) on HIV drugs. Others have examined reviews by experts, such as Jin and Leslie (2003) on restaurant quality grades and Reinstein and Snyder (2005) on movies.

[2]See https://www.ftc.gov/enforcement/consumer-sentinel-network for more details on the Consumer Sentinel Network. The Consumer Sentinel Network is a secure online database available only to law enforcement. Law enforcement members can use the information in the database to enhance and coordinate investigations. Non-governmental organizations that contribute data cannot see Consumer Sentinel complaints. The data in Consumer Sentinel are unverified complaints reported by consumers.

demographics. I first examine how the "consumer base" complaining to Consumer Sentinel compares to the demographics of the general population. I find substantial differences in complaint rates between communities with different demographics. Heavily black areas complain at higher rates than areas with a small black population, while heavily Hispanic areas complain at lower rates. Zip codes with a large share of college graduates complain at much higher rates. Complaint rates are also lower in areas with low unemployment and higher household size and in less urban areas. I find fairly small differences in complaint rates between areas with different median household incomes and median ages. I find similar patterns between complaint rates and demographics for complaints to the Federal Trade Commission (FTC) and the overall Sentinel database, but fairly different patterns for the Consumer Finance Protection Bureau (CFPB). These findings on the demographics of complaints contrast to evidence on victimization that has found much higher rates of victimization for both blacks and Hispanics compared to the general population (Anderson, 2007, 2013). I thus obtain data on both victims and complaints from four law enforcement actions ("cases") in order to examine how the characteristics of complaining consumers compare to those of victims. In all cases, the number of complaints is several orders of magnitude below the number of victims, indicating severe selection in who chooses to complain. Nevertheless, I do find that areas with higher victimization also have higher complaint rates; the complaint rate rises between 13 to 17 percent with a one standard deviation rise in the victimization rate across all of the cases.

I then examine how the demographics of complaints in these cases compare to the demographics of victims. For all four cases, I find that heavily black and heavily Hispanic areas complain at much lower rates. I also find that areas with a large share of college graduates complain at much higher rates. These differences in complaint rates across communities with different demographics can affect the types of complaints received if the types of problems experienced by consumers differ by community. I find that the share of complaints about banks and lenders, debt collection, and auto related problems are much higher in heavily black areas, for example.

As its name suggests, the purpose of the Consumer Sentinel database is to alert law enforcement organizations of emerging problems in the marketplace and to aid in the remedy

of such problems. If the characteristics of consumers that provide complaints are very different from either the general population or fraud victims, policymakers at such organizations may not learn about problems that affect communities that do not complain. My evidence suggests substantial selection in who complains, and that such selection can affect the types of complaints received.

A small literature has also examined either the determinants of a consumer's decision to review or the implications of differences in reviewing behavior across consumers. Mayzlin et al. (2014) show evidence of strategic reviewing behavior, as firms place negative reviews of their competitors. Dai et al. (2014) examine how to construct an optimal quality ranking when reviewers vary in the bias and precision of their reviews. In addition, Hu et al. (2008) and Ghose and Ipeirotis (2011) examine how the characteristics of reviews and reviewers affect consumer demand.

This paper also relates to a large body of work examining consumer complaining behavior.[3] Hirschman (1970) sparked a theoretical literature on the determinants of consumer complaints. While subsequent empirical work has examined how demographics affect complaint behavior, it has been limited by access to small samples obtained either from surveys (Singh, 1989) or from complaints from one BBB (Oster, 1980; Garrett and Toumanoff, 2010). Given the small samples and limited scope of most of these studies, it is unsurprising that Garrett and Toumanoff (2010) finds that the literature is divided on how demographics such as age, income, education, and race affect the likelihood of consumer complaint. Ayres et al. (2013) also studies the determinants of 2011 mortgage complaints to the CFPB, although, as noted earlier, complaint patterns for CFPB complaints are fairly different than the overall Consumer Sentinel database.

The paper proceeds as follows. Section 2 describes the data used in this study. Section 3 examines the demographic determinants of complaints to the Consumer Sentinel database, while Section 4 examines how the demographics of complaining consumers compare to victims in a set of law enforcement actions. Section 5 then concludes.

---

[3]There is even a journal, *Journal of Consumer Satisfaction, Dissatisfaction, and Complaint Behavior*, devoted to the analysis of complaint behavior.

4

# 2 Data

## 2.1 Consumer Sentinel Network

The Consumer Sentinel Network collects data on complaints from several sources – federal government agencies such as the Federal Trade Commission (FTC) and Consumer Finance Protection Bureau (CFPB), private actors such as the Better Business Bureaus (BBBs), and state and local government agencies.[4] For the BBBs, the complaints provided to Consumer Sentinel are selected by the FTC to be those of national interest; for example, complaints about major national corporations would be more likely to be included than about local housepainters. In the 2012-2015 period, about 44 percent of BBB US and Canada complaints were reported to the Consumer Sentinel database. Since Consumer Sentinel has a five year data retention policy, I limit the sample of complaints to complaints filed from 2012 through 2015. I also exclude identity theft complaints, and complaints from a large mobile app contributor which does not provide geographic information. The remaining dataset contains over 6.7 million complaints.

The main variable of interest is the complaining consumer's zip code. Not all consumers providing complaints either live in a US state or provide a zip code. I exclude complaints without a zip code, as well as zip codes belonging to PO Boxes and Unique Organizations (such as businesses or universities that have their own zip code) and zip codes with a population of less than 100 in a given year. In addition, I only include one complaint for each unique full name, zip code, contributor, and year, in order to prevent biases in zip code complaint rates from serial complainants. The resulting dataset has about 5.2 million complaints, or about 78 percent of the original dataset.[5] Of these complaints, 1.7 million or 32.5 percent are contributed by the FTC, 1.4 million or 26.5 percent by the BBBs, 650,000 or 12.3 percent by the CFPB, and 313,000 or 6.0 percent by state and local agencies. Together, these four groups provide over three-quarters of the complaints in the dataset.

---

[4]See https://www.ftc.gov/enforcement/consumer-sentinel-network/reports for the Consumer Sentinel Data Book, which contains further detail on the Consumer Sentinel as well as a wealth of statistics on the complaints included in it.

[5]I also exclude zip codes missing the Census demographic variables described below.

## 2.2 Census Demographics

I match complaint data from the Consumer Sentinel database with demographics from the 2008-2012 American Community Survey (ACS) at the zip code level. The Census has created the Zip Code Tabulation Area (ZCTA) in order to connect Census demographics to zip codes from addresses, because the zip code is not a traditional Census geography. The boundaries of zip codes and ZCTAs do not always perfectly line up, but the exclusion of zip codes for PO Boxes and Unique Organizations should help reduce differences between the two.

I examine several demographic factors that likely proxy for cultural and economic factors that could affect the likelihood that a consumer complains. I first examine ethnic demographics, including the fraction of the zip code population that is black, that is of Hispanic origin, and that is Asian. Second, I examine information on the economic and family situation of the zip code, including the median household income, median household size, median age, the unemployment rate, and the fraction of the zip code population that is college educated. Finally, I include a measure of urbanization developed by the Economic Research Service of the Department of Agriculture.[6]

In order to examine potential non-linear effects of each variable, I divide each variable into several categories, each of which is detailed in Table I. When possible, I try to include categories for both the lowest quantiles of values and highest quantiles of values. For example, for both college education and median household income, the categories selected are very close to the bottom 5 percent, the 5th to 25th quantile, 25th to 75th quantile, 75th to 95th quantile, and top 5 percent zip codes. Some variables are skewed, such as ethnic composition or urbanization, as most zip codes are Metropolitan and contain a small share of blacks and Hispanics. For these skewed variables, a large fraction of zip codes are in the first category (such as Metropolitan areas and areas with less than 5 percent black or Hispanic population). Appendix A examines the quantiles of the variables below weighted by 2010 Census population.

---

[6]This measure is the Rural Urban Commuting Area (RUCA). It was originally created at the census tract level, and was then aggregated to the zip code level. See http://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx and https://ruralhealth.und.edu/ruca for more information.

Table I Variable Definitions

| Variable | Categories |
|---|---|
| Percent Black | 0-5, 5-25, 25-50, 50-75, 75-100 |
| Percent Hispanic | 0-5, 5-25, 25-50, 50-75, 75-100 |
| Percent Asian | 0-25, 25-100 |
| Median Age | 0-30, 30-40, 40-45, 45-50, > 50 |
| Median Household Size | 0-2, 2-2.5, 2.5-2, 3-3.5, > 3.5 |
| Unemployment Rate | 0-3, 3-5, 5-7.5, 7.5-10, >10 |
| Urbanization | Metropolitan, Micropolitan, Small Town, Rural |
| Median Household Income (thousands) | 0-30, 30-40, 40-70, 70-100, > 100 |
| Percent College Educated | 0-10, 10-20, 20-40, 40-60, > 60 |

**Note:** All datasets are as described in the text. Categories include the upper threshold in general (so a zip code that is 5 percent black is in the 0-5% category, not the 5-25% category). Zip codes are defined as Metropolitan if the USDA assigns the zip code a Rural Urban Commuting Area (RUCA) score of 1 to 3, Micropolitan with a RUCA score of 4 to 6, Small Town with a RUCA score of 7 to 9, and Rural with a RUCA score of 10.

# 3   Who complains?

Different demographics groups might complain about fraud and other issues at different rates for several reasons. First, sending a complaint takes time, and so consumers with a higher cost of time – such as those with a higher income, who are employed, or who have kids – might be less likely to complain. Another reason for different complaint rates could be knowledge of the appropriate authorities to complain to; college educated consumers might be more likely to be informed about the FTC, the BBBs, or the CFPB as authorities that receive consumer complaints and seek to remedy problems. Consumers might also be less willing to complain for several other reasons; as the FTC states in a recent report to Congress (Federal Trade Commission, 2016) about consumer complaints from black and Hispanic communities:

> In the FTC's workshops and conferences, however, many have observed a general reluctance and embarrassment to report fraud. Further, despite the higher prevalence of fraud, some have stated that African American and Latino consumers may distrust the government, may not know where to complain, may believe their complaints will not make a difference, or may have concerns about encountering the government because of their immigration status.

Consumer willingness to complain may be affected by language barriers for consumers with a limited proficiency in English as well.[7] I examine the demographics of the consumers that complain to Consumer Sentinel by examining how zip code level complaint rates vary with zip code demographics. Because the demographic information is at the zip code level, any inferences on demographics are best thought of reflecting differences between different types of American communities.

## 3.1  What is the "Customer Base" of Consumer Sentinel?

A simple way to examine how complaint rates vary by different communities is to plot how the average complaint rate varies with community demographics. In Figure 1, I do this for communities with different concentrations of blacks and Hispanics for complaints in 2015; the red and blue lines depict the average complaint rate for communities defined by their share of population that is black and Hispanic, respectively. The estimates are based upon a nonparametric loess regression, with the grey area surrounding each graph representing the 95% confidence interval.

While the average complaint rates are not monotonic, Figure 1 demonstrates that the average complaint rate tends to be lower in areas with a greater share of Hispanic residents. After a small rise in complaint rates from areas that are close to 0% Hispanic to areas that are 15% Hispanic, the complaint rate steadily falls as areas become more Hispanic. Communities that are close to 100% Hispanic have about half the complaint rate of areas that are 0% Hispanic. For black communities, by contrast, the complaint rate is much more constant with respect to the share of the population that is black. Thus, communities that are almost 100% black have about the same complaint rate as communities that are 0% black.

While Figure 1 shows that different demographic communities have different complaint rates, these communities differ on several other factors, including median age, median income, education, and urbanization. In order to disentangle the effects of several different

---

[7]While consumers can file complaints to Consumer Sentinel in Spanish and French as well as English, some consumers may not be aware that they can do so.
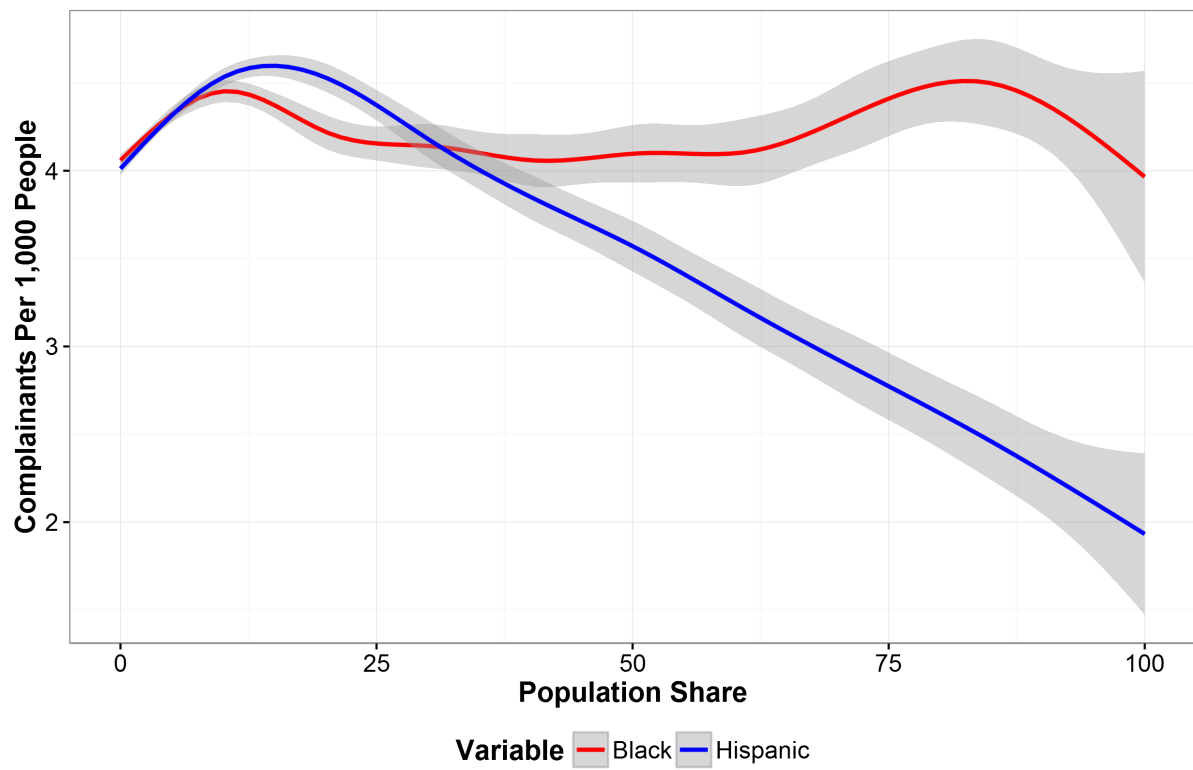
**Figure 1** Complaint Rates For Black and Hispanic Communities in 2015

demographic factors, I estimate the following regression specification:

$$\log(E(y_{szt})) = \beta D_{sz} + \eta \log Population + \gamma_t + \delta_s \qquad (1)$$

where $z$ is the zipcode, $s$ the state, and $t$ the year. The dependent variable $y_{szt}$ is the average complaint rate at the zip code, year level. In this specification, the log of the expected value of the complaint rate is linear, so all estimates of the demographic effects can easily be translated into percent changes compared to the baseline group, holding all other variables fixed. $D_{sz}$ are the zip code level demographics binned as in Table I, $\gamma_t$ are year fixed effects, and $\delta_s$ state fixed effects. I include state fixed effects because many of the contributors are state agencies, which could skew state-level differences in complaint rates. The effects of demographics are identified by variation within states. I also control for the log of the population of the zip code.

I estimate equation (1) for all complaints in the Consumer Sentinel database and for two of the largest data contributors – the the FTC and the CFPB – separately. The CFPB specializes in complaints about financial services, and so its complaints may have very different demographics than the BBBs or the FTC that include a broad spectrum of topic areas including consumer fraud.

Figure 2 - Figure 5 display the results of these regressions for each variable; each graph plots the point estimate and 95% confidence interval for the percent change in complaint rate for each category of the variable compared to the baseline category. The results for all complaints are in red, for the CFPB in green, and for the FTC in blue.

For all complaints, the complaint rate steadily rises as the population share of blacks in a zip code increases. Communities with a 25-50% black population have about a 11% higher complaint rate than those with a 0-5% black population, holding all other variables fixed, and communities with a greater than 75% black population have a 25% higher complaint rate. Communities with a higher Hispanic share of the population have lower complaint rates; areas with a greater than 75% Hispanic population have a 22% lower complaint rate than areas with less than 5% Hispanics. The confidence intervals around these estimates are fairly tight, so all of these estimates are statistically significantly different from zero.
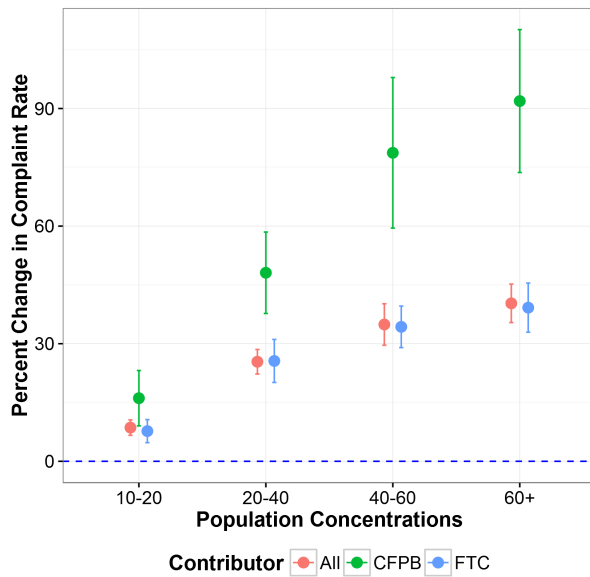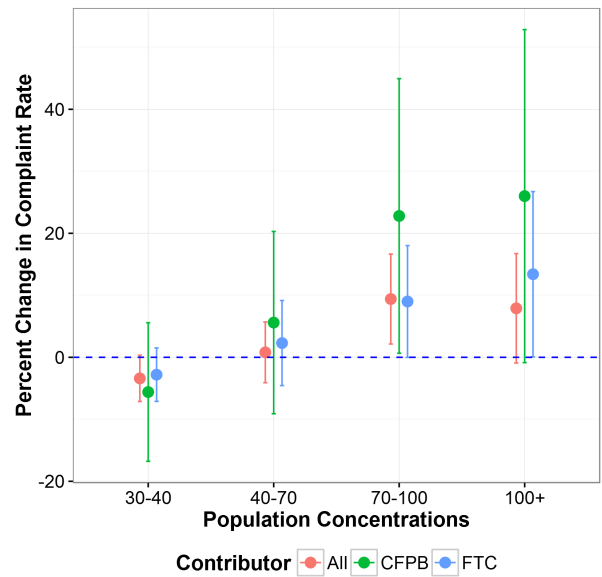
10
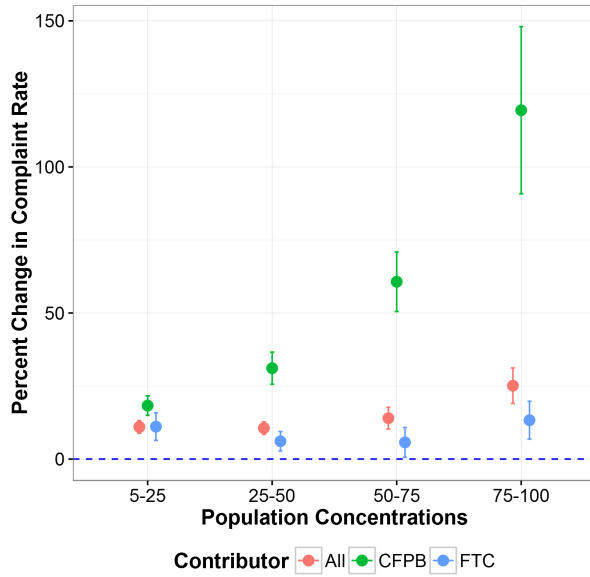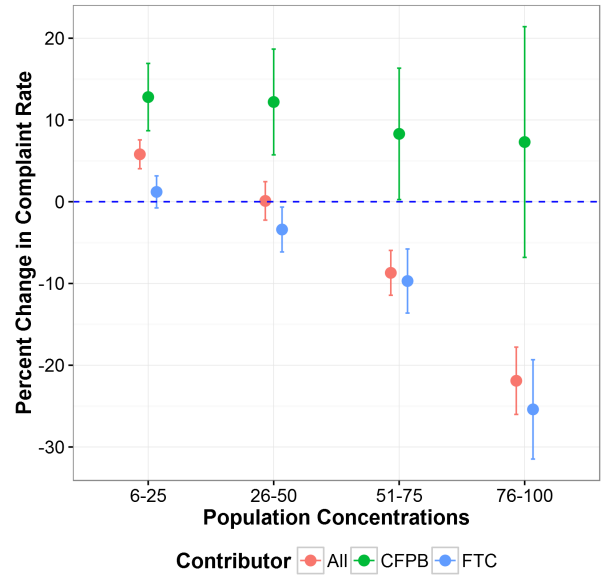
(a) Black



(b) Hispanic

**Figure 2** Percent Change in Complaint Rate by Racial Demographics

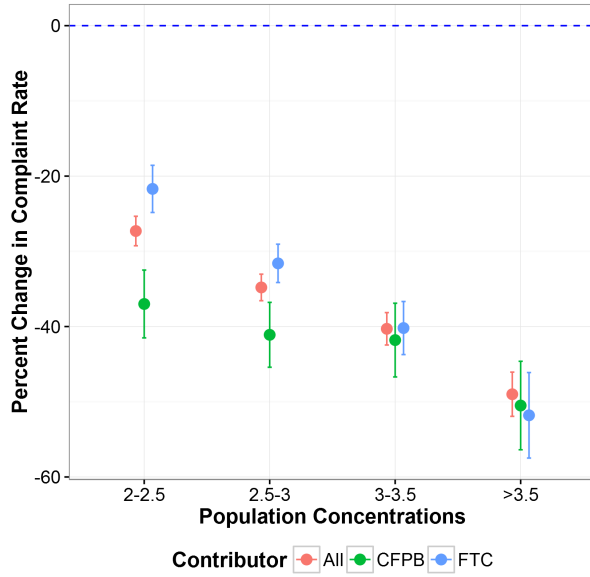**Note:** The baseline, omitted category is 0-5% for percentage black and 0-5% for percentage Hispanic.
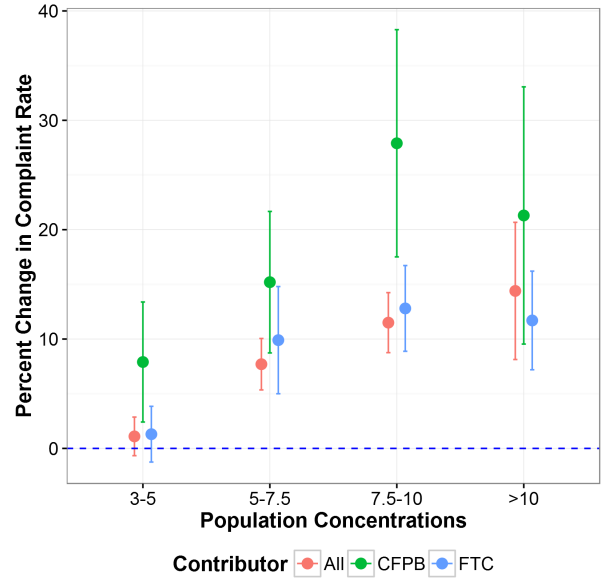


(a) College



(b) Income

**Figure 3** Percent Change in Complaint Rate by Education and Income Demographics

**Note:** The baseline, omitted category is 0-10% for percent college educated and less than 30,000 dollars for median household income.
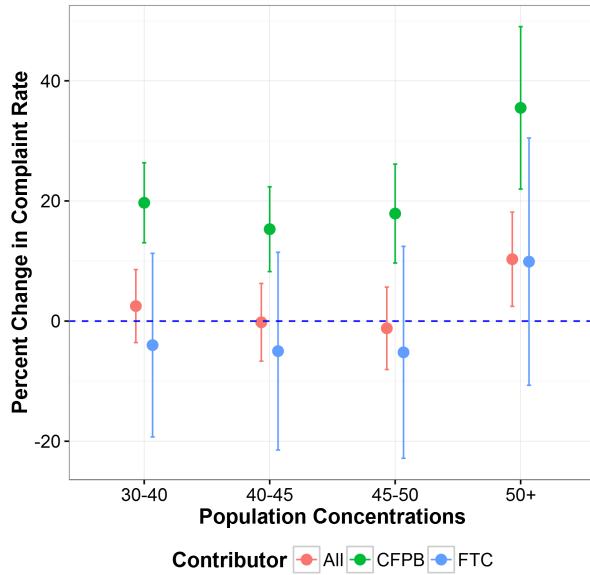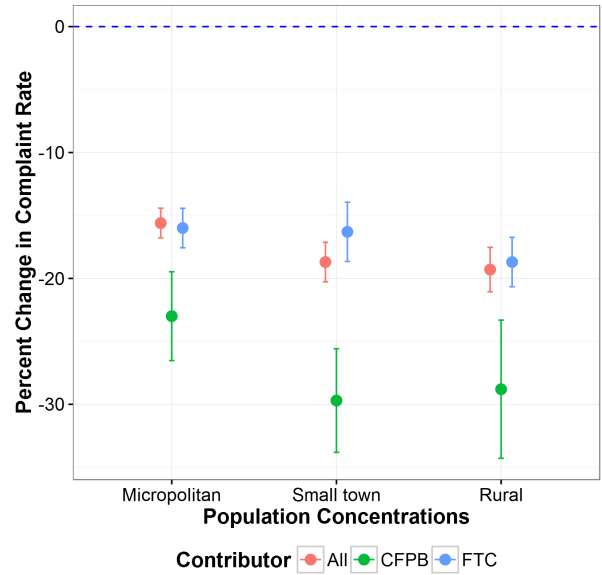
(a) Household Size

(b) Unemployment Rate

**Figure 4** Percent Change in Complaint Rate by Household Size and Unemployment Rate

**Note:** The baseline, omitted category is less than 2 people for median household size and less than 3% for the unemployment rate.



(a) Age

(b) Area Density

**Figure 5** Percent Change in Complaint Rate by Age and Density Demographics

**Note:** The baseline, omitted category is less than 30 for median age and Metropolitan Area for area density.

Areas with a larger share of college graduates complain at much higher rates. Communities with a 20-40% college educated population have about a 25% higher complaint rate than areas where less than 10 percent of the population is college educated population; the increase is 35% for communities with a 40-60% college educated population and 40% for communities with a 60% or higher college population. All of these estimates are tightly estimated and statistically significantly different from zero. By contrast, the estimates for income are much smaller and often not statistically different from zero. For example, the estimates imply that areas with greater than a hundred thousand dollars in median household income have only a 8% greater complaint rate than areas with less than thirty thousand dollars in median household income.

Areas with greater household size have lower complaint rates. For example, areas with an average household size of 2 to 2.5 have 27 percent lower complaint rates than areas with an average household size less than 2, and areas with an average household size greater than 3.5 have a 49 percent lower complaint rate. Areas with higher unemployment rates also have higher complaint rates, with areas with 5% to 7.5% unemployment, 7.5% to 10% unemployment, and greater than 10% unemployment having 8%, 12%, and 14% higher unemployment rates than areas with an unemployment rate of less than 3%.

More rural areas also have lower complaint rates. Compared to Metropolitan areas, Micropolitan areas have 16% lower complaint rates and Small Towns and Rural areas have 19% lower complaint rates. These estimates are precisely estimated. For median age, the estimates are much smaller, with most of the age groups having estimates similar to areas with a median age less than 30. Areas with a median age greater than 50 have only about a 10% higher complaint rate than areas with a median age less than 30.

I also look at how complaint rates vary with demographics for the FTC and CFPB separately. The FTC tends to have fairly similar patterns to the overall Consumer Sentinel database; the confidence intervals of the estimates for the entire database and the FTC alone typically overlap. One exception is for percentage black; areas with 75-100% black residents have 25% more complaints for the overall database but only 13% more complaints from the FTC, relative to areas with few blacks.

The CFPB, however, often has very different patterns with respect to demographics.

Although the estimates for the CFPB are noisier than for the overall database, heavily black and Hispanic areas have a much greater increase in CFPB complaints than areas with few black or Hispanic residents compared to the overall Sentinel database. Areas with a 50-75% share of blacks have a 61% higher complaint rate than areas with less than 5% blacks, and areas with a greater than 75% share of blacks have a 119% increase in complaints. Heavily Hispanic areas have 7 to 8% higher complaint rates to the CFPB, compared to declines in complaints for heavily Hispanic areas in the overall sample. Areas with more college educated residents also complain at higher rates to the CFPB, even compared to the overall sample; the increase compared to communities with less than a 10% college educated population is 79% for communities with 40-60% college educated population and 92% for communities with a 60% or higher college population.

Non-metropolitan areas also have substantially lower complaint rates, and older areas have higher complaint rates, compared to the full Consumer Sentinel dataset. I also find higher complaint rates in areas with high income and areas with high unemployment compared to the overall sample, although these estimates are often not statistically different than the estimates for the overall sample. These differences between the CFPB and other contributors could be due to the newness of the agency and the attention it has gathered, or they could be due to differences between consumers complaining about financial services and consumers with other types of complaints.[8]

To sum up, for the overall Consumer Sentinel dataset, I find higher complaint rates in heavily black areas, more college educated areas, and areas with greater unemployment. I find lower complaint rates in heavily Hispanic areas, more rural areas, and areas with greater household size. I find relatively small differences across areas with differences in median household income and median age. While the FTC has similar complaint rates to the overall database, the CFPB has very different patterns with respect to demographics.

---

[8]For example, a large fraction of the CFPB complaints are about banks. Consumers may be more likely to complain about an company with which they have a supposed trusted relationship, resulting in different demographic patterns than for other types of companies.

## 3.2 How Different are Complaints from Different Communities?

While the above analysis demonstrated that complaint rates vary considerably across communities with different demographics, it is unclear whether the issues reflected in the complaints provided by different communities differ. I examine this question using the classification of complaints into different categories in Consumer Sentinel. Each complaint is given one of 123 codes, which are then aggregated into 30 categories for published reports on complaints in the Consumer Sentinel database. I examine the seven largest categories (excluding complaints categorized as Other): Auto-Related Complaints ("Auto"), Banks and Lenders ("Bank"), Shop-at-Home and Catalog Sales ("Catalog"), Debt Collection ("Debt Col"), Impostor Scams ("Imposter") , Prizes, Sweepstakes and Lotteries ("Prize"), and Telephone and Mobile Services ("Telecom"). For each category, I calculate the share of complaints in a zip code from that category and then estimate the same specification as equation (1) replacing the complaint rate with the category's share of complaints.

Figure 6 depicts how the share of complaints from each category changes with communities with different racial demographics. Heavily black areas have much higher complaint rates about Bank, Auto, and Debt Collection complaints. Areas with a 75 to 100% black population have a 25% higher share of Debt Collection complaints, a 38% higher share of Auto complaints, and a 64% higher share of Bank complaints compared to areas in which blacks are less than 5% of the population. The other categories examined have lower shares in heavily black areas; for example, areas with a 75 to 100% black population have a 31% lower share of Imposter complaints compared to areas with a small black population.

Heavily Hispanic areas have much smaller differences in the share of complaints from these categories compared to areas with few Hispanics, the largest difference is on complaints about Prizes, Sweepstakes, and Lotteries, for which heavily Hispanic areas have a 38% lower share of complaints than areas with few Hispanics. These differences across categories are not just a product of differences between contributors. For example, looking only at FTC complaints, heavily black areas have a 35% higher share of Debt Collection complaints, a 48% higher share of Auto complaints, and a 92% higher share of Bank complaints compared to areas with few blacks. Appendix B contains figures similar to Figure 6 for other demographic
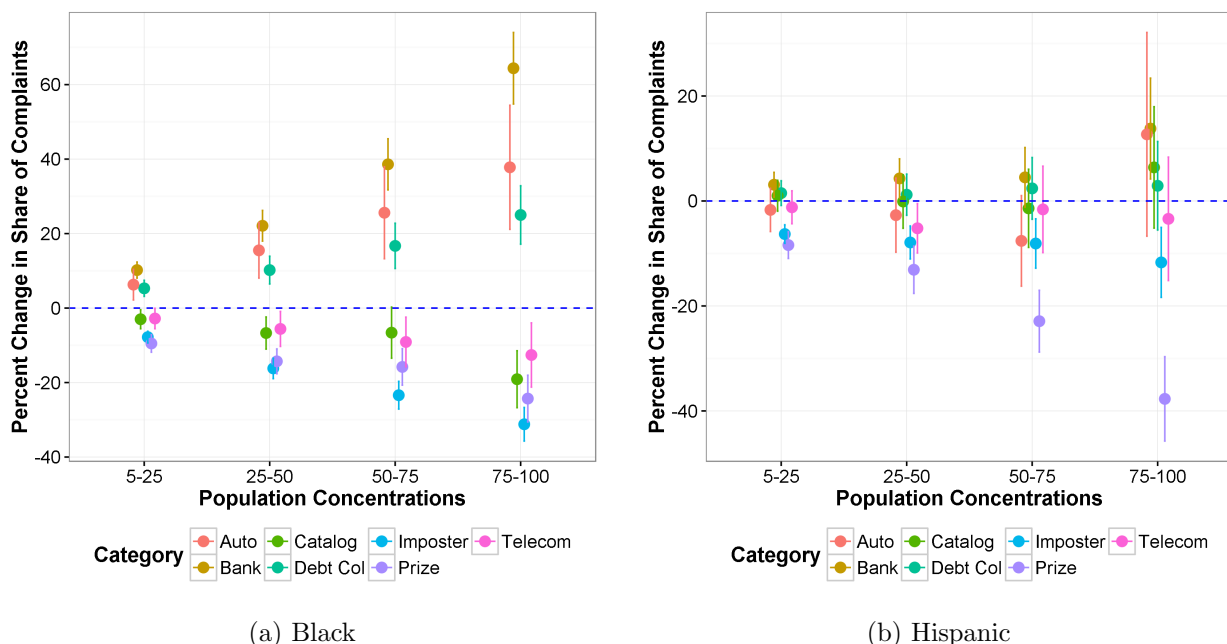
15

(a) Black

(b) Hispanic

**Figure 6** Percent Change in Share of Complaints for Complaint Categories by Racial Demographics

**Note:** The baseline, omitted category is 0-5% for percentage black and 0-5% for percentage Hispanic.

variables and for all variables looking at FTC only complaints.

# 4    How Do Complaints Compare to Victims?

The previous section showed that heavily black areas complain about 25% more than areas with a small proportion of blacks, and heavily Hispanic areas complain about 22% less than areas with a small fraction of Hispanics. The most recent FTC fraud survey (Anderson, 2013) finds, however, that blacks and Hispanics are much more likely to report being victims of fraud; blacks are about 90% more likely to be a victim of fraud than non-Hispanic whites, and Hispanics about 50% more likely. One possible way to reconcile these facts is that consumers in black and Hispanic communities are much less likely to complain after being victims of fraud than those in other communities. In this section, I examine this issue using data from a series of law enforcement actions ("cases") for which I can match consumer databases of victims to complaints in Consumer Sentinel.

16

## 4.1 Cases

I was able to obtain data from four such cases, including data on victims from consumer databases and available complaint data.[9]

I call the first case "Case B", as I am unable to disclose the company's name or details on its industry, other than that it has been successfully sued by a state or federal agency for consumer protection violations and that its industry is different than those of the other cases. Case B has the largest number of complaints and victims, with over 12 million victims and four thousand complaints.

The second case, Ideal Financial, involved a company that bought consumer payday loan applications and then used the bank account details in the applications to withdraw money from the consumers' bank accounts without their consent. The FTC sued Ideal Financial and won summary judgment, with a $43 million judgment against the defendants (two additional defendants settled for a $25 million judgment).[10] For this case, I have about two million victims and 1,400 complaints.

The third case, Platinum Trust, also involved payday loan applications as well as telemarketing. The company allegedly called online payday loan applicants and offered them credit cards with heavily deceptive terms; for example, the cards could only be used at the defendant's online store, rather than at any store accepting Visa, Mastercard, or American Express as promised. The FTC sued Platinum Trust and eventually settled the charges, with a judgment of over $7.4 million that was returned to consumers via refunds.[11] For this case, I have about 500 complaints and 70,000 victims.

The last case, WinFixer, involved a company that falsely claimed that security scans had discovered malware on consumers' computers. The company then sold computer security software that would "fix" the problems identified. The FTC sued the companies and indi-

---

[9]These complaints were typically obtained as part of the law enforcement action, and may include complaints from Consumer Sentinel on identity theft or the "Do Not Call" list.

[10]See https://www.ftc.gov/enforcement/cases-proceedings/1123211-x130044/ideal-financial-solutions-inc-et-al and https://www.consumer.ftc.gov/blog/ftc-takes-down-ideal-financials-fraud-network for additional details on this case.

[11]See https://www.ftc.gov/enforcement/cases-proceedings/1123212/apogee-one-enterprises-llc-also-dba-apogee-enterprises-llc and https://www.ftc.gov/news-events/press-releases/2013/01/ftc-sends-74-million-refunds-consumers-harmed-scheme-sold for more details.

viduals involved in the scam; most settled with multi-million dollar judgments, while the defendant that went to trial was found liable for more than $163 million.[12] For this case, I have about 300,000 victims and one thousand complaints.

**Table II** Cases with Victim Lists

| Case | Number of Victims | Number of Complaints |
|------|------------------:|---------------------:|
| Case B | 12,311,307 | 4,271 |
| Ideal Financial | 2,010,169 | 1,403 |
| Platinum Trust | 69,576 | 510 |
| WinFixer | 304,493 | 1,062 |

**Note:** The number of victims and number of complaints reflects all victims and complaints that can be matched to zip codes in Section 2.2, after duplicate entries were removed.

The amount of selection in who complains varies considerably across cases, with differences in the number of complaints per victim varying by orders of magnitude across cases in Table II. For example, for the WinFixer case there are about 0.35 complaints per 1,000 victims, compared to 7.3 complaints per 1,000 victims for Platinum Trust and 0.07 complaints per 1,000 victims for Ideal Financial. In general, the overall complaint rate will likely depend upon the characteristics and conduct of the company.

## 4.2    Are Complaint and Victim Rates Correlated?

A basic test of whether there is any relation between the complaints and victims for each case is to examine whether complaint rates and victim rates are correlated. I define the complaint rate and victim rate at the zip code level. Since there are only a few complaints for each case, for any given zip code the number of complaints is likely to be zero or one. I examine how the zip code complaint rate increases with an increase in the victim rate by estimating the following specification for each case:

$$\log(E(\text{complaint rate})_z) = \alpha(\text{victim rate})_z \tag{2}$$

where $z$ is a given zip code. I scale the victim rate by its standard deviation for each case. Thus, the estimates indicate the percentage increase in the complaint rate from a one

---

[12]See https://www.ftc.gov/enforcement/cases-proceedings/072-3137/innovative-marketing-inc-et-al and https://www.ftc.gov/news-events/blogs/business-blog/2014/02/court-appeals-upholds-win-consumers-winfixer-case for more details.

standard deviation increase in the victim rate. Table III contains these estimates. For all of the cases, the estimates of this correlation are fairly similar. A one standard deviation increase in the victim rate increases the complaint rate by 13 to 17 percent and is strongly statistically significant for all of the cases.

**Table III** Percentage Change in the Complaint Rate from a One Standard Deviation Increase in the Victim Rate

| Case | Effect |
|---|---|
| Case B | 0.125 |
|  | (0.033) |
| Ideal Financial | 0.151 |
|  | (0.026) |
| Platinum Trust | 0.155 |
|  | (0.023) |
| WinFixer | 0.173 |
|  | (0.036) |

**Note:** Each estimate is based upon estimating a specification for which the expectation of the log of the complaint rate is linear in the victim rate. Standard errors are robust.

## 4.3 Are the Demographics of Complainants and Victims Different?

The demographics of the victims themselves vary considerably across cases. In Figure 7, I plot how the per-capita number of victims changes with the percentage black and percentage Hispanic of each zip code through a non-parametric loess regression. Because these cases have very different numbers of victims, I normalize the per capita number of victims by the average across zip codes for each case. The red line in Figure 7 indicates Case B, the blue line Ideal Financial, the green line Platinum Trust, and the purple line WinFixer. For Ideal Financial and Platinum Trust, the relationship between the fraction of the zip code that is black and the per capita victim rate is similar; the victim rate is 300-500% higher in zip codes that are 100% black compared to zip codes that are close to 0% black. This pattern may reflect the demographics of consumers of payday loans, since both cases involve payday loans. For Case B, the victim rate also rises with the fraction of the zip code that is black, but much less dramatically. Heavily black zip codes have victim rates 80 to 100% higher
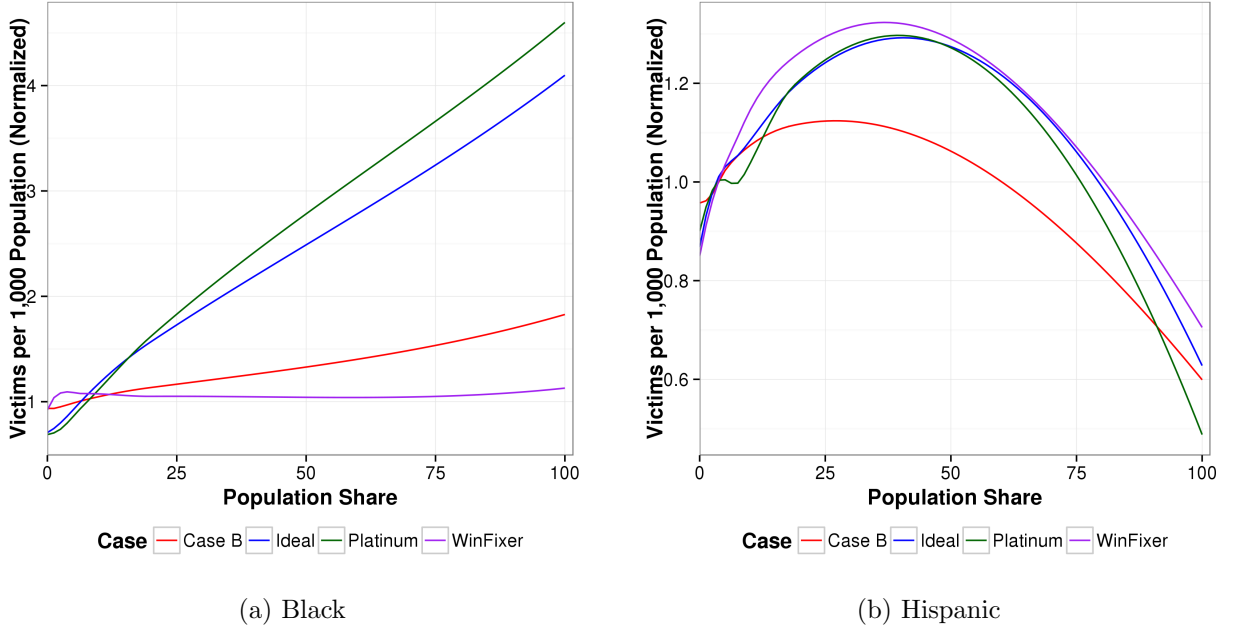
(a) Black           (b) Hispanic

**Figure 7** Percent Change in Victim Rate by Racial Demographics

**Note:**

than areas with few blacks. For WinFixer, the victim rate is roughly constant with the share of the zip code that is black.

The differences in victim rates between cases vary much less dramatically for areas with different concentrations of Hispanics. The pattern for Hispanics resembles an inverse U for all cases, with victim rates initially rising slightly as an area becomes more Hispanic but then falling. Areas that are 100% Hispanic have victim rates that are 25 to 50% lower than areas that are 0% Hispanic.

I then look at the ratio of complaints to victims for each zip code in order to understand how complaining behavior compares to underlying victim rates. Despite the fact that the relationship between victim rates and racial demographics varies with different cases, especially for the black concentration of an area, I find similar patterns for how the ratio of complaints to victims varies with racial demographics across all complaints. In Figure 8, I plot the relationship between the ratio of complaints to victims and the percentage black and percentage Hispanic of each zip code using a non-parametric loess regression. I again normalize the ratio by the average across zip codes for each case. In all cases, there are less complaints relative to victims in heavily black and heavily Hispanic areas, with a reduc-
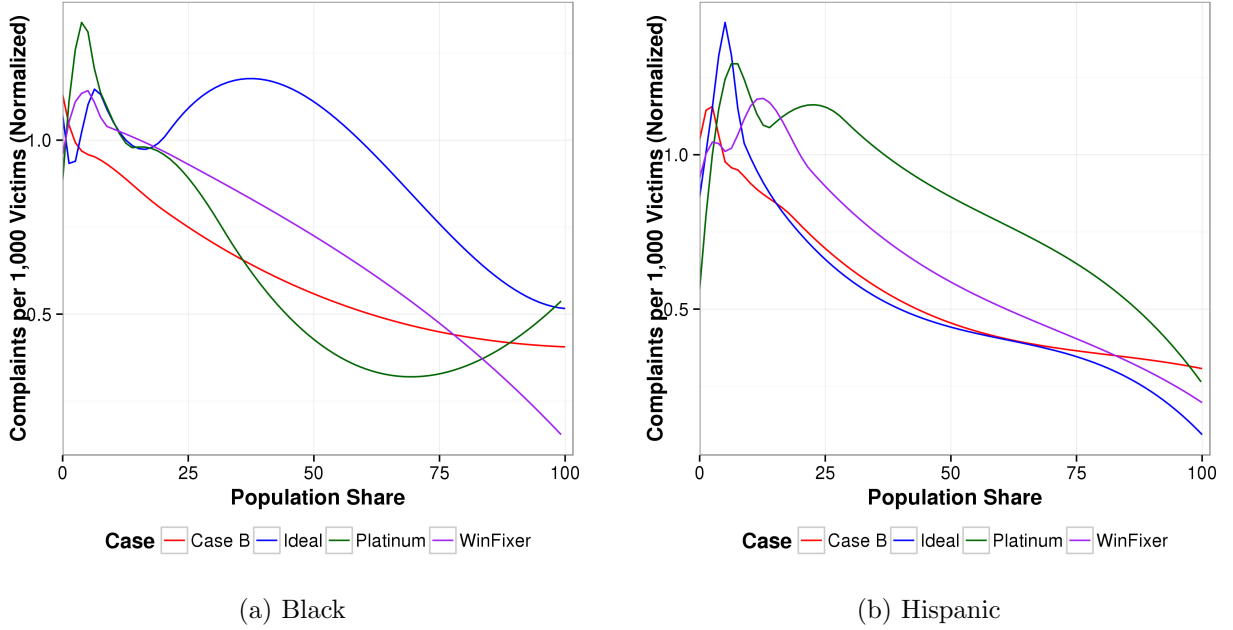
20

(a) Black

(b) Hispanic

**Figure 8** Percent Change in Ratio of Complaints to Victims by Racial Demographics

**Note:**

tion of about 40 to 80 percent for the most heavily black areas compared to areas with few blacks and 50 to 90 percent for the most heavily Hispanic areas compared to areas with few Hispanics.

While the declines in Figure 8 provide suggestive evidence that complaint rates are lower in heavily black and Hispanic communities, the declines are not statistically significant in the nonparametric regressions given that almost all zip codes have no complaints at all in these cases. I thus focus on how the individuals in the complaint datasets compare to the victim datasets. In order to examine how the demographics of complaints compare to victims while controlling for other demographic factors, I estimate the following ordered probit model for each case and for each demographic variable $y_{sz}$:

$$y_{sz}^* = \alpha 1\{complaint\} + \beta D_{sz} + \eta \log(Population) \tag{3}$$

where $y_{sz}^*$ is the latent variable for the demographic variable, $1\{complaint\}$ is an indicator that the observation came from a complaint rather than a victim, and $D_{sz}$ includes all other demographic variables. I also control for the population of the zip code.
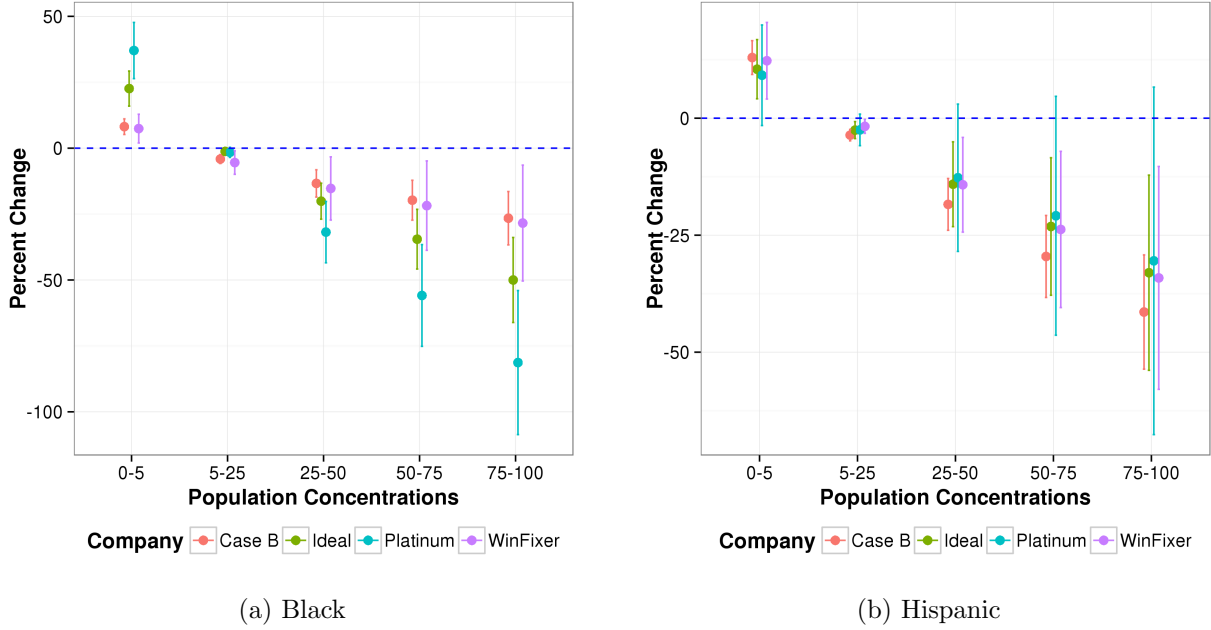
21

(a) Black                                    (b) Hispanic

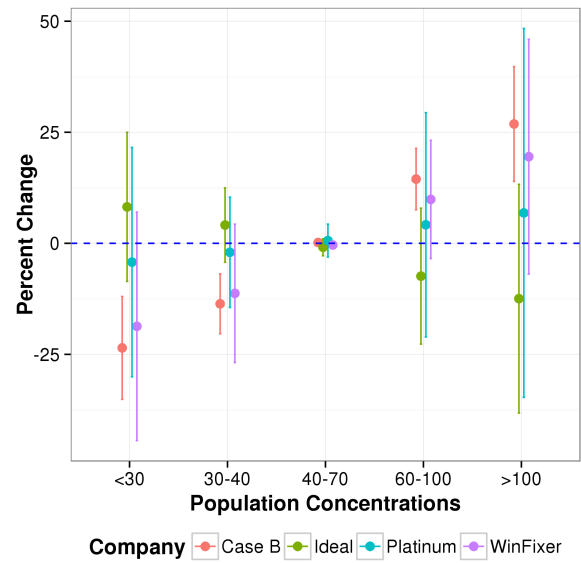**Figure 9** Percent Change in Complaint Rate by Racial Demographics

**Note:**

I use the ordered probit model in order to predict how the share of every category of each demographic variable changes between the complaint dataset and the victim dataset. The main restriction that the ordered probit model makes is that the effect of being in the complaint dataset is monotonic across categories. For example, if the coefficient on $1\{complaint\}$ is negative in the specification with the black categorization as the dependent variable, then the number of complaints relative to the number of victims falls as the fraction of black residents in the zip code rises.

Figure 9 - Figure 12 contain the marginal effects from this ordered logit model. Communities with higher shares of black residents complain at significantly lower rates relative to their number of victims. As the share of black residents rises, the number of complaints falls compared to the number of victims for all four cases. The differences are significantly different from zero in all four cases as well. For communities with 75-100 percent black residents, the estimates indicate a 26 percent lower complaint rate for Case B, 50 percent lower for Ideal, 81 percent lower for Platinum, and 28 percent lower for WinFixer, compared to areas with less than 5 percent black residents.

Communities with higher shares of Hispanic residents also complain at lower rates for all
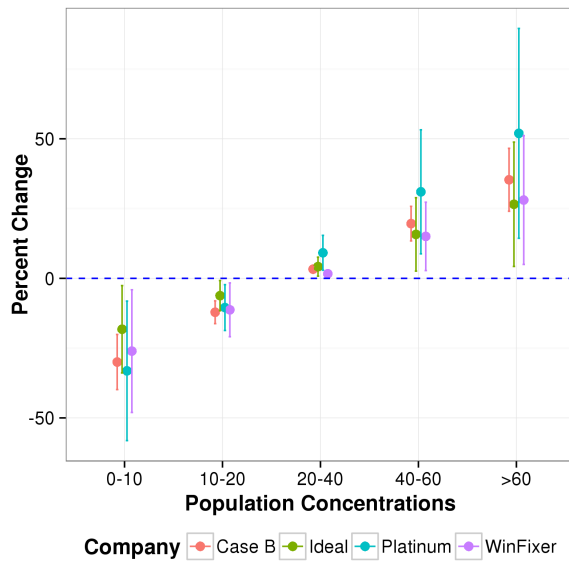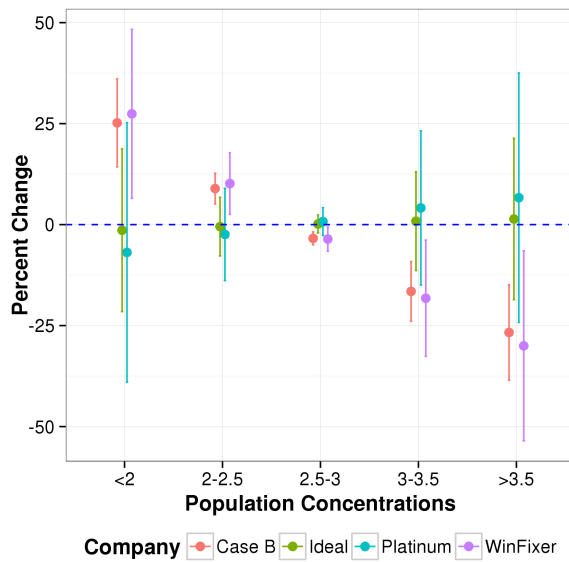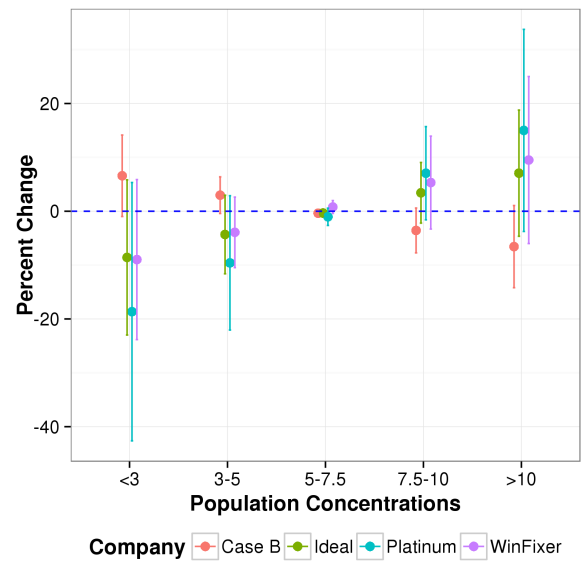
(a) College



(b) Income

**Figure 10** Percent Change in Complaint Rate by Education and Income Demographics

**Note:**



(a) Household Size



(b) Unemployment Rate

**Figure 11** Percent Change in Complaint Rate by Household Size and Unemployment Rate
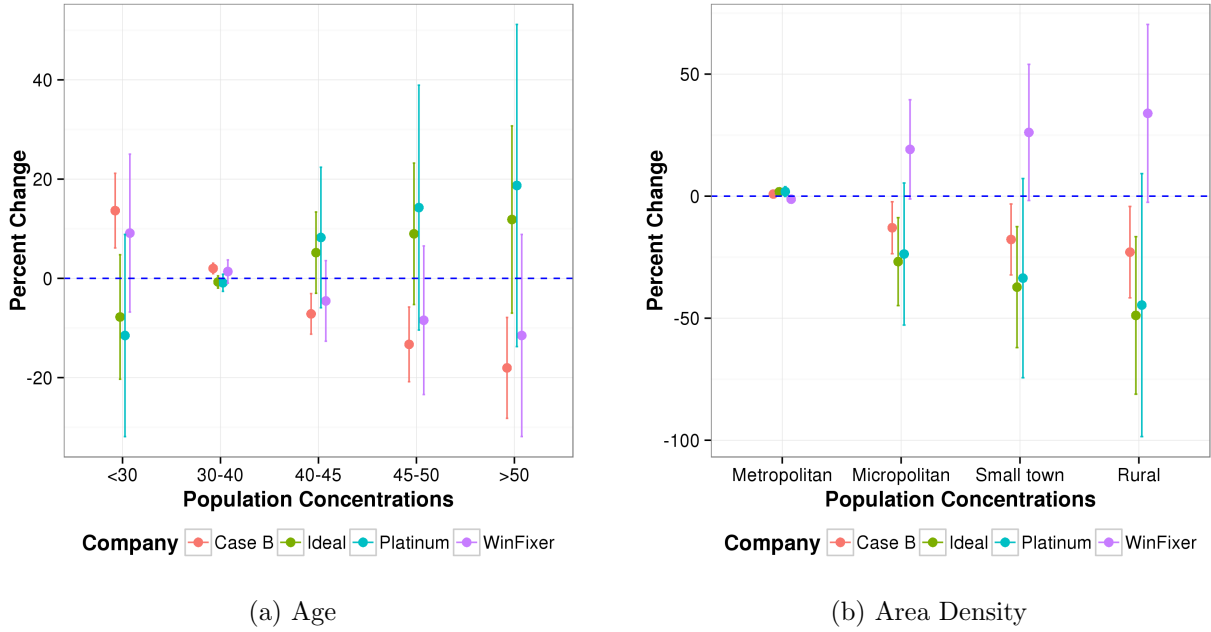
**Note:**

(a) Age



(b) Area Density

**Figure 12** Percent Change in Complaint Rate by Age and Density Demographics

**Note:**

four cases; the differences are statistically significant from zero for all cases except Platinum. For communities with 75-100 percent Hispanic residents, the complaint rate is 41 percent lower for Case B, 33 percent lower for Ideal, 30 percent lower for Platinum, and 34 percent lower for WinFixer compared to areas with less than 5 percent Hispanic residents.

Communities with a higher share of college educated residents complain at substantially higher rates. In all four cases, the number of complaints compared to victims rises with the share of college educated residents. For communities with greater than 60 percent college graduates, the complaint rate is 35 percent higher for Case B, 26 percent higher for Ideal, 52 percent higher for Platinum, and 28 percent higher for WinFixer. These increases are strongly statistically significant.

For the other demographic variables, the estimates are both often not statistically significantly different from zero, and vary in sign across cases. For income, complaint rates increase with higher median household incomes for Case B; the highest income areas have about a 25 percent higher rate of complaints, and the effect is statistically significant. For the other cases, the effect is not statistically different from zero, with a negative estimate for Ideal and slightly positive estimates for Platinum and WinFixer. For household size, areas

24

with the largest household size have a 27 percent lower complaint rate for Case B and 30 percent lower complaint rate for WinFixer compared to areas with the smallest household size. The effects are smaller and insignificant for the other two cases. For the unemployment rate, I find relatively small and insignificant effects for all cases. For age, two cases have about a 10 to 20 percent fall in the complaint rate for the oldest areas compared to the youngest, while two cases have a 10 to 20 percent rise in the complaint rate. Finally, for area density, three cases have substantial falls in the complaint rate for less dense areas, two of which are statistically significant, while for WinFixer the complaint rate rises in less dense areas, although the effect is not statistically significant.

To sum up, I find strong evidence that heavily black and heavily Hispanic areas complain at lower rates compared to their level of victimization; these effects are consistent across cases and are generally statistically significant. Areas with more college educated residents also complain at higher rates relative to their rate of victimization. I find mixed evidence for the other demographic variables.

# 5    Conclusion

In this study, I have found substantial evidence of selection in consumer complaints using a massive dataset of millions of consumer complaints. Communities that are heavily black, more college educated, and have greater unemployment have higher complaint rates, while heavily Hispanic areas, more rural areas, and areas with greater household size have lower complaint rates. These differences in complaint rates do affect the types of complaints received; heavily black areas have much greater complaint rates on finance related issues.

I have then exploited a set of law enforcement actions for which I have access to databases of victims as well as complaints for each case, which has allowed me to examine how complaints compare to victimization. I have found that heavily black and heavily Hispanic communities have much fewer complaints compared to their level of victimization, while areas with more college graduates have more complaints relative to their level of victimization. These results demonstrate the importance of controlling for victimization; heavily black areas have more complaints relative to the general population, but fewer complaints relative

to their level of victimization for the four cases I study.

While my results make clear that selection is a major issue in user generated content, it is far from clear how to account for such selection. A statistical approach to doing so would be to weight complaints based on how the complaint rate of their community compares to the general population, or to the degree of victimization; such weights would overweight communities with lower complaint rates and thus surface their complaints. A policy approach to doing so would be to conduct outreach campaigns in communities with low complaint rates. Such outreach could both provide information on how to complain and help policymakers learn why complaint rates are low in those communities.

# References

**Anderson, Keith B.**, *Consumer Fraud in the United States: The Second FTC Survey*, Federal Trade Commission, 2007.

_ , *Consumer Fraud in the United States, 2011: The Third FTC Survey*, Federal Trade Commission, 2013.

**Ayres, Ian, Jeff Lingwall, and Sonia Steinway**, "Skeletons in the Database: An Early Analysis of the CFPB's Consumer Complaints," *Fordham J. Corp. & Fin. L.*, 2013, *19*, 343.

**Balat, Jorge, Nicholas W. Papageorge, and Shaiza Qayyum**, "Positively Aware? Expert Reviews, Demand for Pharmaceuticals, and Health Outcomes," Technical Report 2016.

**Cabral, Luis and Ali Hortacsu**, "The Dynamics of Seller Reputation: Evidence From eBay," *The Journal of Industrial Economics*, 2010, *58* (1), 54–78.

**Chevalier, Judith A. and Dina Mayzlin**, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 2006, *43* (3), 345–354.

**Dai, Weijia, Ginger Jin, Jungmin Lee, and Michael Luca**, "Optimal Aggregation of Consumer Ratings: An Application to Yelp.com," Technical Report, Harvard Business School 2014.

**Federal Trade Commission**, "Combating Fraud in African American and Latino Communities: The FTC's Comprehensive Strategic Plan," Technical Report 2016.

**Garrett, Dennis E. and Peter G. Toumanoff**, "Are Consumers Disadvantaged or Vulnerable? An Examination of Consumer Complaints to the Better Business Bureau," *Journal of Consumer Affairs*, 2010, *44* (1), 3–23.

**Ghose, Anindya and Panagiotis G. Ipeirotis**, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering*, 2011, *23* (10), 1498–1512.

**Hirschman, Albert O.**, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Vol. 25, Harvard University Press, 1970.

**Hu, Nan, Ling Liu, and Jie Jennifer Zhang**, "Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects," *Information Technology and Management*, 2008, *9* (3), 201–214.

**Jin, Ginger Zhe and Phillip Leslie**, "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *The Quarterly Journal of Economics*, 2003, pp. 409–451.

**Luca, Michael**, "Reviews, Reputation, and Revenue: The Case of Yelp. com," 2011.

**Mayzlin, Dina, Yaniv Dover, and Judith A. Chevalier**, "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *The American Economic Review*, 2014, *104* (8), 2421–2455.

**Oster, Sharon**, "The Determinants of Consumer Complaints," *The Review of Economics and Statistics*, 1980, pp. 603–609.

**Reinstein, David A. and Christopher M. Snyder**, "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics," *The Journal of Industrial Economics*, 2005, *53* (1), 27–51.

**Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood**, "The Value of Reputation on eBay: A Controlled Experiment," *Experimental Economics*, 2006, *9* (2), 79–101.

**Singh, Jagdip**, "Determinants of Consumers Decisions to Seek Third Party Redress: An Empirical Study of Dissatisfied Patients," *Journal of Consumer Affairs*, 1989, *23* (2), 329–363.

# A   Demographics

Table A-1 contains the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentile quantiles of each variable across zip codes. The quantiles are estimated after weighting each zip code by its 2010 population. All of the ethnic demographics are heavily skewed – half of the American population lives in zip codes whose population is less than 5 percent black, less than 8 percent Hispanic, and less than 2 percent Asian. On the other hand, majority black and majority Hispanic zip codes each comprise more than 5 percent of population weighted zip codes. The measure of urbanization is similarly skewed. Values between 1 to 3 indicate metropolitan areas; about 83.5 percent of zip code residents in our dataset live in metropolitan areas, compared to 9 percent that live in micropolitan areas (values between 4 and 6), 4.5 percent in small towns (values between 7 and 9), and 3 percent in rural areas (values between 10 and 11).[13]

   The other variables are somewhat less skewed. The median age for the median zip code is 37.5, with the bottom 5 percent of zip codes with a median age below 28 and the top 5 percent of zip codes with a median age above 47. The median household size is 2.6 for the median zip code, compared to below 2.1 for the bottom 5 percent of zip codes and above 3.5 for the top 5 percent of zip codes. The unemployment rate for the median zip code is 5.6 percent; the bottom 5 percent of zip codes have an unemployment rate below 2.7 percent while the top 5 percent of zip codes have an unemployment rate above 10.5 percent. For the median zip code, the median household income is 52 thousand dollars; the bottom 5 percent have a median income below 29 thousand dollars and the top 5 percent have a median income above 100 thousand dollars. Lastly, tin the median zip code about 24 percent of the 25 year old and above population have completed college, compared to less than 8.6 percent for the bottom 5 percent of zip codes and above 61.2 percent for the top 5 percent of zip codes.

# B   Changes in the Share of Complaint Categories by Demographics

---

[13]Because we exclude PO Boxes, we likely miss some of the population living in rural areas, who are more likely to use PO Boxes. Compared to the overall distribution, we have a slightly higher fraction of the population in metropolitan areas and a slightly lower fraction of the population living in micropolitan, small town, and especially rural areas.
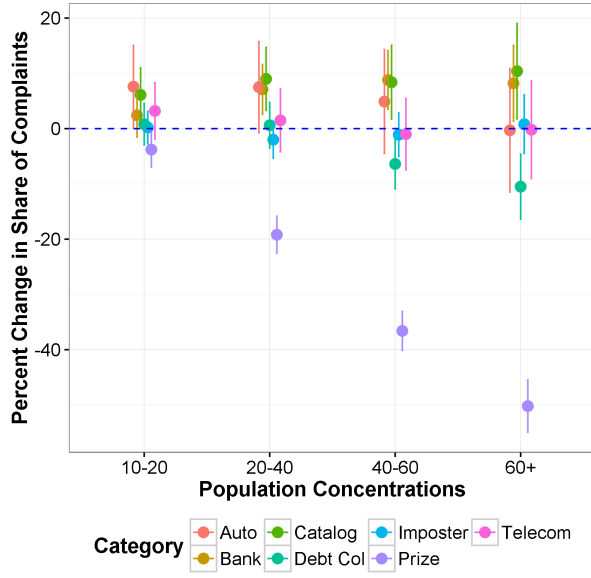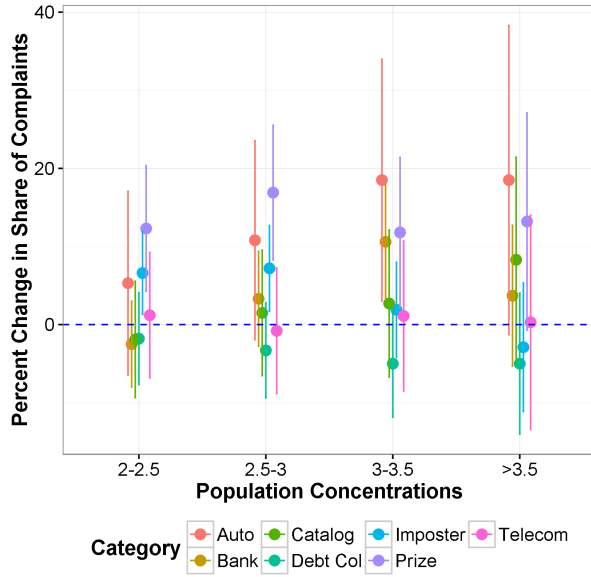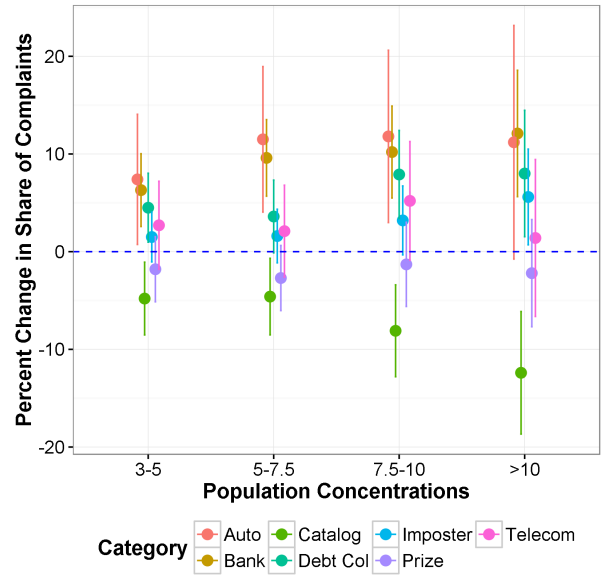
(a) College
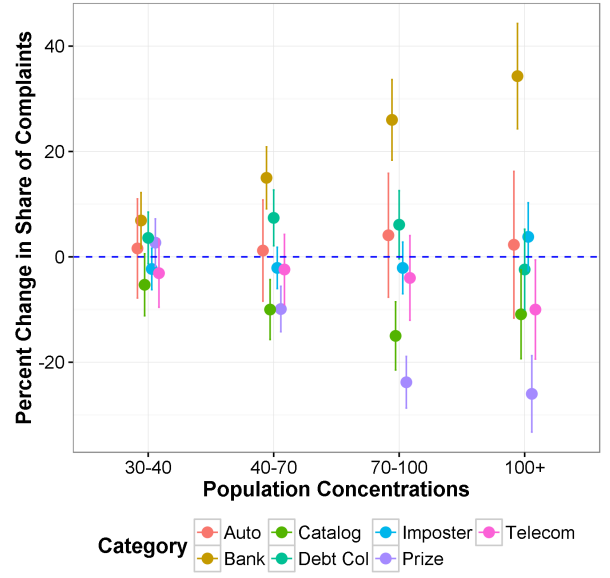


(b) Income

**Figure 13** Percent Change in Share of Complaints for Complaint Categories by Education and Income Demographics

**Note:** The baseline, omitted category is 0-10% for percent college educated and less than 30,000 dollars for median household income.



(a) Household Size



(b) Unemployment Rate

**Figure 14** Percent Change in Share of Complaints for Complaint Categories by Household Size and Unemployment Rate

**Note:** The baseline, omitted category is less than 2 people for median household size and less than 3% for the unemployment rate.

**Table A-1** Quantiles of Demographic Variables

| Variable | Quantiles | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 25% | 50% | 75% | 90% | 95% | 99% |
| Percent Black | 0 | 0.1 | 0.4 | 1.4 | 4.7 | 14.5 | 34.9 | 54.6 | 87.6 |
| Percent Hispanic | 0 | 0.7 | 1.3 | 3 | 7.7 | 20.8 | 46.9 | 65.3 | 90.8 |
| Percent Asian | 0 | 0 | 0.1 | 0.6 | 2 | 5.2 | 12 | 19.1 | 43.7 |
| Median Age | 23.5 | 28.3 | 30.2 | 33.7 | 37.5 | 41.2 | 44.6 | 47.1 | 54.8 |
| Household Size | 1.8 | 2.1 | 2.2 | 2.4 | 2.6 | 2.9 | 3.2 | 3.5 | 4.1 |
| Unemployment Rate | 1.5 | 2.7 | 3.3 | 4.3 | 5.6 | 7.3 | 9.2 | 10.5 | 13.3 |
| Urbanization (RUCA) Score | 1 | 1 | 1 | 1 | 1 | 2 | 4.1 | 7.1 | 10.2 |
| Median Household Income (thousands) | 23 | 29 | 33 | 41 | 52 | 68 | 88 | 101 | 130 |
| Percent College Educated | 5.1 | 8.6 | 10.9 | 15.8 | 24.1 | 37.4 | 52.4 | 61.2 | 75.5 |

**Note:** The 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentile quantiles of each variable across zip codes are included in the table, where the quantiles are estimated after weighting each zipcode by its 2010 population.
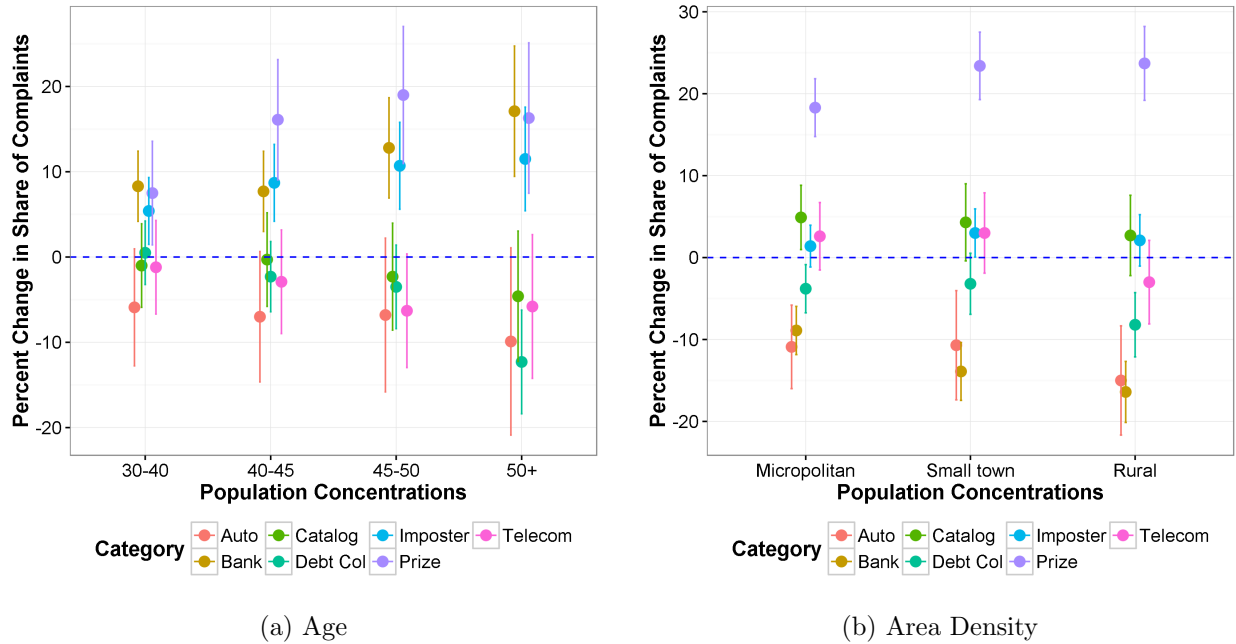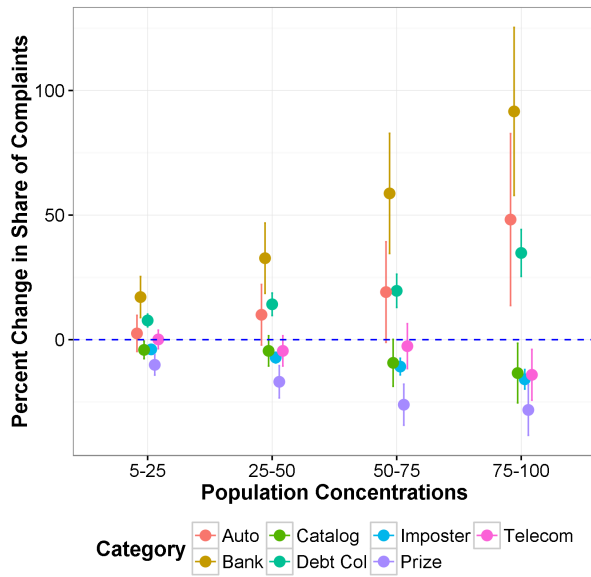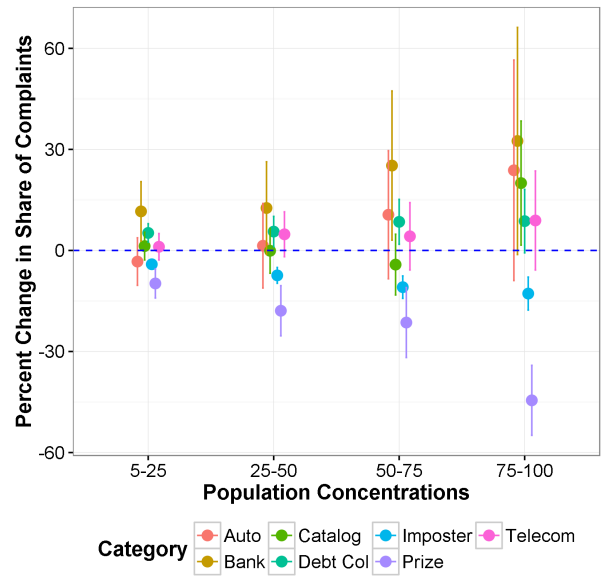


(a) Age

(b) Area Density

**Figure 15** Percent Change in Share of Complaints for Complaint Categories by Age and Density Demographics

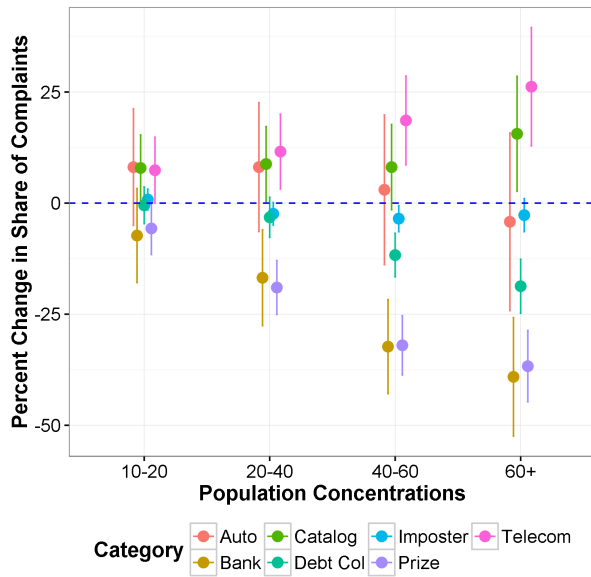**Note:** The baseline, omitted category is less than 30 for median age and Metropolitan Area for area density.
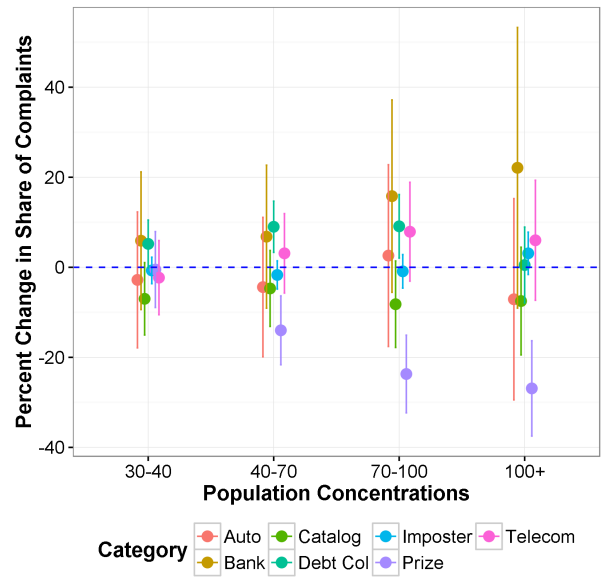
(a) Black

(b) Hispanic

**Figure 16** Percent Change in Share of Complaints for Complaint Categories by Racial Demographics, FTC Complaints Only

**Note:** The baseline, omitted category is 0-5% for percentage black and 0-5% for percentage Hispanic.
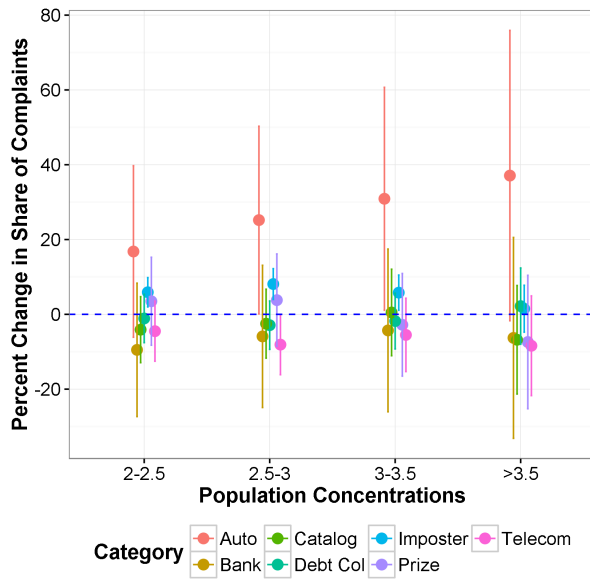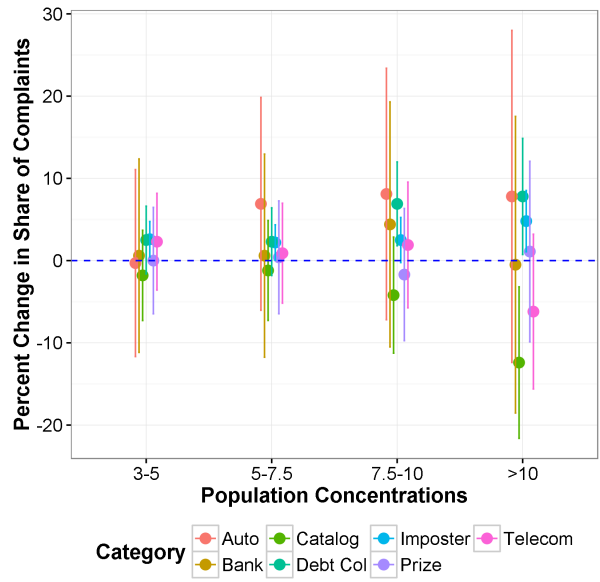


(a) College

(b) Income

**Figure 17** Percent Change in Share of Complaints for Complaint Categories by Education and Income Demographics, FTC Complaints Only

**Note:** The baseline, omitted category is 0-10% for percent college educated and less than 30,000 dollars for median household income.

(a) Household Size

(b) Unemployment Rate

**Figure 18** Percent Change in Share of Complaints for Complaint Categories by Household Size and Unemployment Rate, FTC Complaints Only

**Note:** The baseline, omitted category is less than 2 people for median household size and less than 3% for the unemployment rate.
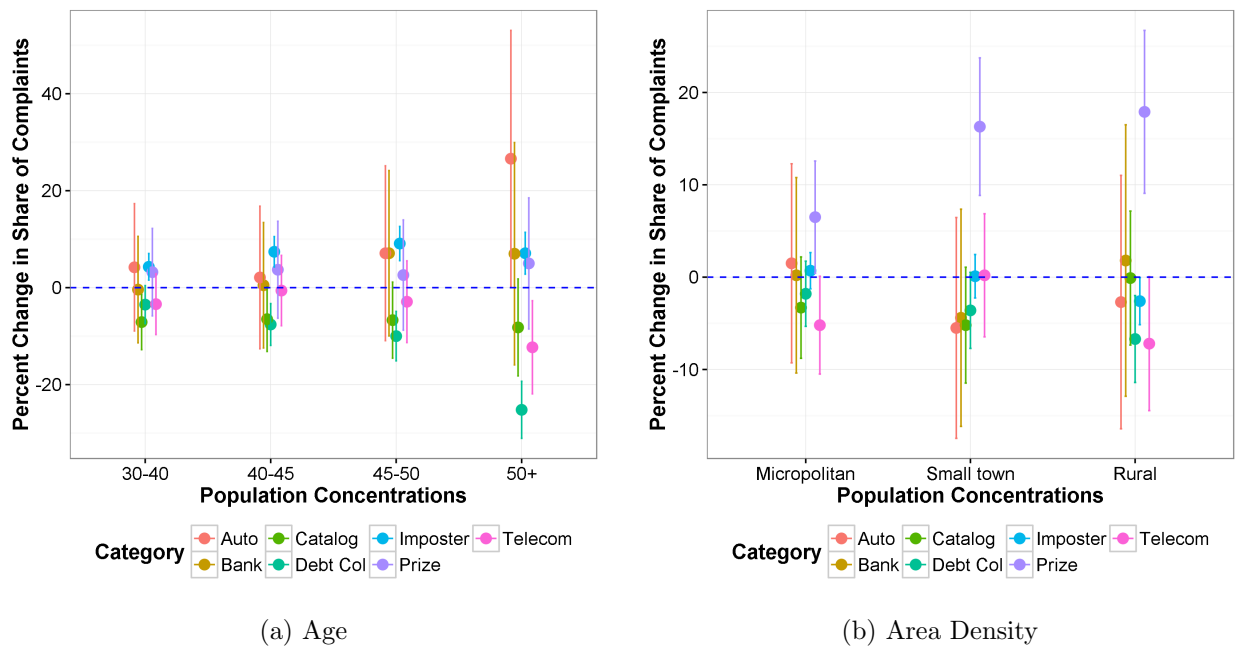
(a) Age

(b) Area Density

**Figure 19** Percent Change in Share of Complaints for Complaint Categories by Age and Density Demographics, FTC Complaints Only

**Note:** The baseline, omitted category is less than 30 for median age and Metropolitan Area for area density.