Digital Technologies, Customer Experience, and Decisions

by

Shuyi Yu

B.A. in Economics, Peking University, 2013
B.S. in Statistics, Peking University, 2013
A.M. in Statistics, Harvard University, 2015
S.M in Management Research, Massachusetts Institute of Technology, 2018

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2021

Author ....
Sloan School of Management
March 10, 2021

Certified by ....
Catherine Tucker
Sloan Distinguished Professor of Management
Professor of Marketing
Thesis Supervisor

Accepted by ....
Catherine Tucker
Sloan Distinguished Professor of Management
Professor of Marketing
Chair, MIT Sloan PhD Program

Digital Technologies, Customer Experience, and Decisions

by

Shuyi Yu

Submitted to the Sloan School of Management
on March 10, 2021, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY IN MANAGEMENT

Abstract

This dissertation consists of three chapters that investigate how digital technologies have changed customer experience and their decisions.

The first chapter investigates market participants' reactions to predictive algorithms and the effects of this public information source on market outcomes. In particular, I study the extent to which buyers and sellers rely on a home's Zestimate when making decisions. Using detailed property transaction data for 120,482 properties sold between May 2017 and May 2019 in the Greater Philadelphia area, I show that the sale price of a property does respond to exogenous shocks to its estimated home value. I develop a theoretical framework and provide empirical evidence to show how people use the Zestimate as a source of publicly available information that plays an important role in coordination and helping people reach an agreement. The results suggest that market participants tend to rely more on this public information source when it is harder to reach a consensus based on private information. Moreover, I show that people's reliance on the Zestimate might mitigate racial disparities in the housing market by providing less biased information.

In the second chapter, we study how consumers respond to repeated marketing campaigns driven by algorithms and how the responses vary across different algorithms. To investigate it, we collaborate with a U.S. food delivery company and conduct a field experiment where targeted coupons are sent by applying the same algorithms repeatedly. Our results show that algorithms utilizing more information perform better than simpler algorithms, and this difference only exists when the consumers have already been treated by the same algorithm-driven policy a few times. By exploring the variation in the purchase patterns, we show that those differences arise because advanced algorithms reduce the level of learning and strategic behaviors against the rules. This result also suggests that consumers may have some level of algorithm awareness, especially when algorithms are easy to learn, and are forward-looking enough to play strategically against the policies powered by those algorithms.

In the third chapter, we study how digitization has transformed customer experience in the public sector. Customers with more education may get better service after complaining, because they are better placed to advocate for themselves. It is unclear how digitization of the consumer complaint process will change this situation. To investigate this, we analyze 364,189 customer complaints to the city of Boston. Empirically, complaints that originate from areas with high levels of education are more likely to be solved quickly. However, dedicated mobile app technologies that automate the complaint process can help mitigate the advantage conferred by education. Since the adoption of digital devices is endogenous to wealth and education, we instrument their usage using granular geographic data on a proxy for cellular signal strength. This analysis again suggests that mobile applications can partially eliminate the disparity between educated and uneducated people. We present suggestive evidence that this is because mobile devices and the standardization of communication they require, eliminate potential differences in treatment of cases that arise due to differences in communication skills. This result suggests that using newer forms of automated digital communication tools enhances equality in customer service.

Thesis Supervisor: Catherine Tucker
Title: Sloan Distinguished Professor of Management
Professor of Marketing

Acknowledgments

First and foremost, I would like to express my heartfelt appreciation to my advisor Catherine Tucker, for her advice, encouragement, and support. During my PhD, Catherine has not only taught me how to conduct research hand-over-hand, but also encouraged me to explore research ideas and guided me to think independently. In addition to her tremendous guidance on research, Catherine has also been my role model that inspires me to be a better person. I also want to thank her for always being incredibly caring and understanding.

I would like to thank my committee. Duncan Simester gave me extremely constructive feedback and suggestions on my dissertation, which I truly appreciate. Birger Wernerfelt provided me amazing opportunities to learn from him as research assistant and teaching assistant. I benefited from working with Birger and our co-author, Alvin Silk. They are brilliant scholars and extremely generous people that I deeply admire.

I have had the privilege to interact with many great faculty members at the MIT Marketing group, from whom I have learned a great deal. Thanks to Sinan Aral, Sharmila Chatterjee, Dean Eckles, Renee Gosline, John Hauser, Tony Ke, John Little, Drazen Prelec, David Rand, and Juanjuan Zhang.

I would like to thank my peers and friends who made my years at MIT joyful. I enjoyed the time spent with the fellow marketing PhD students – Jenny Allen, Xinyu Cao, Yiqun Cao, Matthew Cashman, Cathy Xi Chen, Jason Du, James Duan, Graelyn Humiston, Marat Ibragimov, Madhav Kumar, Keyan Li, Xiang Song, Artem Timoshenko, Yifei Wang, Jeremy Yang, Jerry Yunhao Zhang, and Yuting Zhu. The same acknowledgment goes to my friends outside the department, including Sean Shiyao Liu, Yinying Ren, Fei Song, Sophie Liyang Sun, Shujing Wang, Duanyi Yang, and many others, for supporting each other through the years at MIT.

I would also like to thank Briana Blake, Sarah Massey, Allison McDonough, Aileen Menounos, Hillary Ross, and Davin Schnappauf, for making my days at MIT go smoother.

Last but by no means least, I want to dedicate this dissertation to my parents, Hongjing Shen and Shengyuan Yu, for their endless love and support.

Contents

List of Figures

List of Tables

Chapter 1

Algorithmic Outputs as Information Source: The Effects of Zestimates on Home Prices and Racial Bias in the Housing Market

1.1 Introduction

There has been an increasing interest in how algorithms have reshaped the economy (Bughin *et al.* , 2018). Breakthroughs in machine learning techniques make automated decision-making available for many giant players in the economy. For example, sharing economy companies such as Uber and Lyft dynamically adjust their prices based on the data-driven real-time pricing system that utilizes information from both supply and demand sides.[1] E-commerce sites go even further by adopting AI-powered demand forecasting tools to automate restocking of products.[2]

But more fundamentally and profoundly, algorithms may also affect economic outcomes via their influences on human decisions. One way algorithms can change human behaviors is by altering the information presented to decision-makers. For example, sophisticated ad targeting algorithms have tailored the information presented to consumers, which has proved crucial for consumers' ability to make good decisions (Payne *et al.* , 1991). Moreover,

[1] See `https://www.forbes.com/sites/nicolemartin1/2019/03/30/uber-charges-more-if-they-think-youre-willing-to-pay-more`

[2] See `https://www.npr.org/2018/11/21/660168325/optimized-prime-how-ai-and-anticipation-power-amazons-1-hour-deliveries`

algorithms can provide market participants with novel information sources by processing massive information and making predictions and recommendations. This use of algorithms has appeared in various domains, such as travel agencies, matchmaking service, and financial advisory service. However, it is still controversial whether people are in adherence to algorithmic forecasts (Dietvorst *et al.* , 2015, Logg *et al.* , 2019).

To answer this question empirically, I study real estate market the reaction of real estate market participants to the Zestimate in this paper. The Zestimate is Zillow's estimate of a home's market value based on its home valuation model, which incorporates data from multiple sources, taking into account home facts, location, tax information and market conditions.[3] It shows right below the current list price (the most recent sale price if the property is not on the market currently) on the property page (see Figure 1-1)[4] and it has been displayed for 97.5 million homes out of the 110 million homes found on Zillow.com, the most popular real estate website in the United States [5]. Anecdotal evidence has shown that even though Zillow is commonly used by both home sellers and buyers in the U.S., it is not clear ex-ante whether Zillow or its home price estimates affect home-buying decisions. On one hand, it provides public and easy-to-process information that may help market participants evaluate and compare home values. On the other hand, the real estate decision is a stressful major financial decision and it is unknown whether people will still trust online information sources and algorithmic estimation techniques when making this important decision.

I combine 120,482 property transaction records with the Zestimate history collected from Zillow.com. An obvious endogeneity concern in this setting is that Zestimates may reveal the unobserved quality of a property. To address this, I turn to an instrumental variables approach, where I use the number of months since the last revaluation or reassessment as a plausibly exogenous instrument. The idea here is that the time since the last reassessment should affect the severity of covariate shifts but the differences in the frequency of revaluation and reassessment across townships are jointly decided by many forces, like laws and budget

[3]See the presentation by Zillow's data scientist: `https://www.slideshare.net/NicholasMcClure1/python-datascienceatzillow/1`

[4]It has been moved to listing details after a major change came into effect in Sep 2019.

[5]`https://investors.zillowgroup.com/overview/default.aspx`

plans, which are not affected by the changes in current sale prices after controlling the current assessed value and various fixed effects. Empirical evidence proves the validity of this instrument and shows how the covariate shift problem affects the model performance. Using this approach, I find that the final transaction price tends to be higher when the estimated home value displayed on Zillow.com is higher.

Furthermore, I investigate how the algorithmic estimation changes users' information acquisition process by providing information that is available to all the market participants. The empirical results show that the reliance on Zestimates is correlated with users' costs of acquiring information from other sources but not with the perceived average accuracy of the Zestimate in the neighborhood.

Finally, I explore the heterogeneity in the effect and ask whether the algorithm helps eliminate the racial biases existing in the housing market for a long time. Due to the lingering impacts of historical "redlining", the properties located in minority or more diversified neighborhoods are usually undervalued. I find suggestive evidence that the Zestimate doesn't fully reflect this white-premium in home values. Since the Zestimate's influence on decision making doesn't vary much across neighborhoods, this effect leads to a smaller racial gap in final sale prices compared to the gap in list prices.

1.2 Literature Review

This paper is related to four streams of research. The first is the literature on information search and information sources. Early work in marketing studying the prepurchase information search and acquisition (e.g. Newman & Staelin, 1972, Claxton *et al.* , 1974, Westbrook & Fornell, 1979, Schaninger & Sciglimpaglia, 1981, Kiel & Layton, 1981, Hauser *et al.* , 1993) focuses on how different customers determine their total information-seeking effort and the allocation of effort among information sources. More recent research by Ratchford *et al.* (2003) studies how the Internet as a new information source reshapes the information acquisition process by substituting other information sources, especially the dealer/manufacturer sources. Zettelmeyer *et al.* (2006) further extends the discussion and shows that the Internet lowers the negotiated prices in car retailing markets by providing buyers more purchaserelevant information. Kuruzovich *et al.* (2010) instead looks at the seller side and shows that lower search costs facilitated by the Internet also equip sellers with the ability to search for high-valuation buyers and raises the final sale price. Other studies (e.g. Brown & Goolsbee, 2002, Jensen, 2007, Ellison & Ellison, 2009) discuss the impact of IT on market structure and efficiency. This paper contributes to the literature by investigating the impact of a specific data product on the allocation of attention and offline market outcomes.

The second is a stream of research that focuses on home prices and racial differentials in housing markets. Previous research has shown that many factors affect the final transaction prices, e.g. school quality (Black, 1999), marketing platforms (Hendel *et al.* , 2009), agent characteristics (Seagraves & Gallimore, 2013), and policy changes (Tucker *et al.* , 2013). Most importantly, significant racial disparities have been found in the US housing market. Individual black buyers tend to pay premiums for comparable units (King & Mieszkowski, 1973, Myers, 2004, Ihlanfeldt & Mayock, 2009, Bayer *et al.* , 2017) and this racial discrimination even persists in emerging online rental markets (Edelman *et al.* , 2017, Cui *et al.* , 2020). On the other hand, house values have been proven to decline in neighborhoods as the percentage of blacks increases (Berry, 1976, Chambers, 1992, Kiel & Zabel, 1996, Myers, 2004) as the consequence of racial prejudice. Results in this paper provide new evidence for racial differentials in home prices caused by prejudice and suggest that this gap can be mitigated by less biased home value estimates based on data-driven methods.

The last one is emerging literature on algorithms and biases. Even though evidence has shown that algorithms reproduce existing racial and gender disparities in various applications (Angwin *et al.* , 2016, Ali *et al.* , 2019, Lambrecht & Tucker, 2019, Obermeyer *et al.* , 2019), it is also important to compare bias between automated algorithms and human judges or other benchmarks (for a recent survey see Cowgill & Tucker, 2020). In this study, I focus on a specific prediction model and study its effects on human prejudice in decision-making.

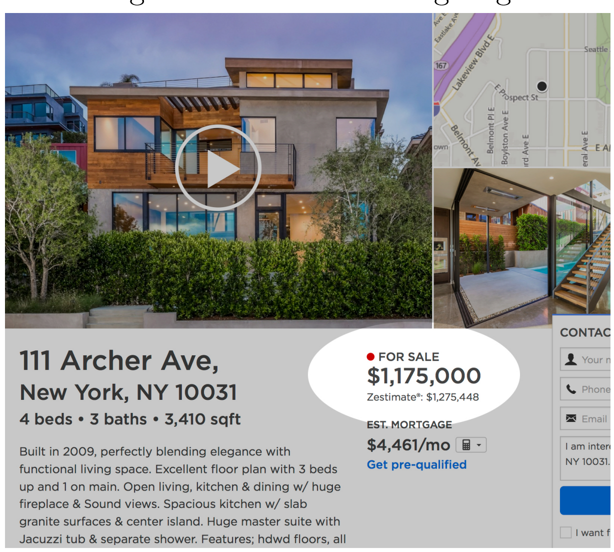
1.3 Data

I use data collected from three sources: Zillow.com, local government's property records, and social-demographic data from the Decennial Census administered by the Census Bureau.

1.3.1 Zillow Data

First of all, I use the property transaction data collected from Zillow.com. I collect detailed property transaction information for 209,016 properties (excluding lots and commercial buildings) sold between May 2017 and May 2019 in the Greater Philadelphia Area. 372 zip codes in 4 states (PA, NJ, DE, MD) are included in this sample. I am able to find the geographic location for 168,818 out of them using their addresses[6] and keep only these properties in the data set to guarantee the data accuracy. To avoid the extreme cases of predatory pricing and the potential threats of misrecorded information, I also drop the 45,141 cases where the list price is missing or equals zero. Finally, I exclude the observations where the sale-to-list is too large (greater than the 99 percentile) or too small (less than the 1 percentile). It excludes about 2,400 properties from the data set and reduces the total number of observations to 120,482.

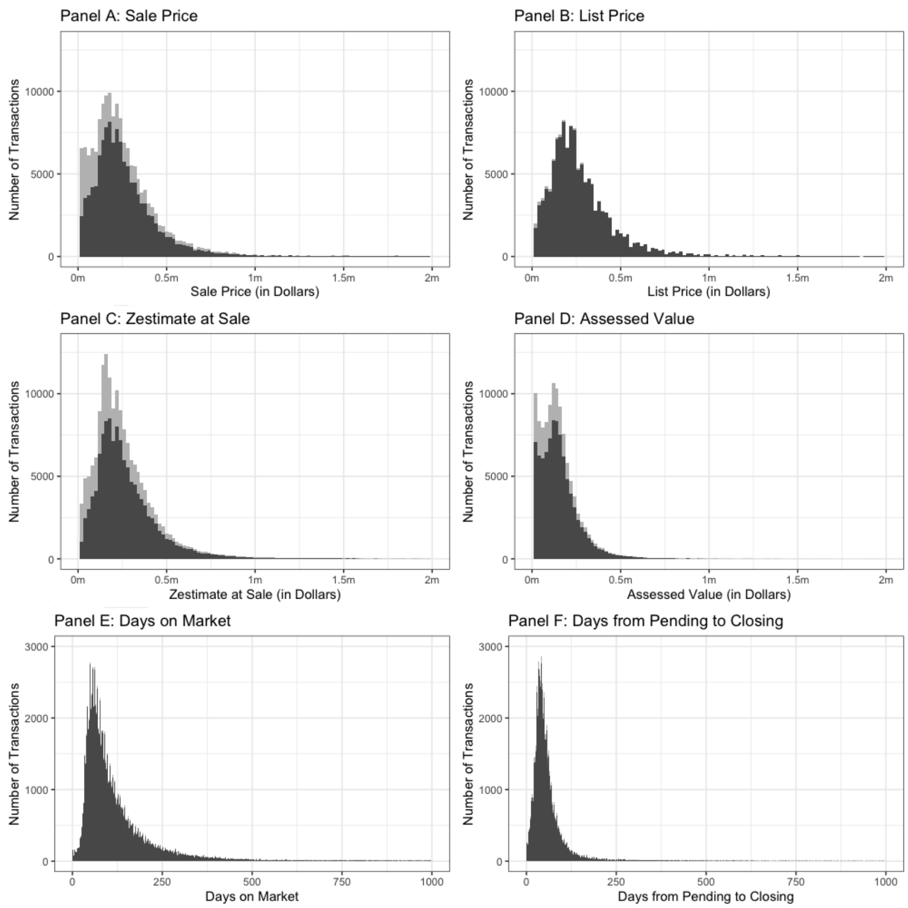
Figure 1-1: Zillow Listing Page



The transaction details collected include the address, the listing date, the list price, the assessed value, the date the seller accepts the offer, the closing date, the sale price, and any price changes that happened in between. In addition, I collect the historical Zestimates for the properties included in the data set. Monthly Zestimates in the last 5 years are

[6] See https://geocoding.geo.census.gov.

Figure 1-2: Transaction Details



Notes: The distribution of sale prices, list prices, the Zestimates at sale, and assessed values are plotted in Panel A, B, C, D, respectively. These distributions are truncated at 2M. The distribution of numbers of days on market and numbers of days from pending to closing are plotted in Panel E and F, respectively. These distributions are truncated at 1000. The light grey bars represent observed frequencies in the entire sample (with 168,818 observations) and the dark grey bars represent observed frequencies in the selected sample (with 120,482 observations) The census block groups included in the sample are shown in grey.

available on Zillow.com. Based on the information provided by Zillow, the median error of Zestimates is 1.9% and more than 1.8M homes have been included in the model in the city of Philadelphia, which is very close to the national average. And there is no evidence showing that Zestimates are more accurate in more active markets or more metropolitan areas [7]. In Figure 1-2, I plot the distribution of the sale price, the list price, the Zestimates[8],

[7] See `https://www.zillow.com/zestimate/`

[8] The Zestimate displayed one month before the sale is used for the plot.

Table 1.1: Summary of Variables

Variable	Mean	Std. Dev.	Min.	Max.	N
Independent and Dependent Variables					
Sale Price	270,474.3	213,814.5	1,000	8,000,000	120,482
List Price	286,383.3	237,352.3	750	1,000,000	120,482
Zestimate_Sold	277,646.1	252,352.5	5,390	4,470,000	120,482
Assessed Value	154,687.5	119,269.7	0	4,108,700	88,110
Log(Days on Market)	4.637	0.905	0	8.369	120,482
Log(Days from "Pending" to "Sale")	3.905	0.817	0	8.066	68,490
Instrument					
Months_Last_Update	184.116	169.186	0	562	120,482
Moderators					
#Transactions	0.007	0.007	0.000	0.103	120,482
%Deviation	0.157	0.563	0	68.093	112,747
Important Sociodemographic Variables					
#Years of Education	13.857	1.273	8.182	17.938	120,482
Log(Median Income)	11.179	0.509	8.849	12.391	118,210
%Internet Subscription	0.826	0.136	0	1	120,482
%Computer Ownership	0.890	0.099	0.148	1	120,482
%White	0.709	0.279	0	1	120,482
%White Owners	0.740	0.274	0	1	120,482

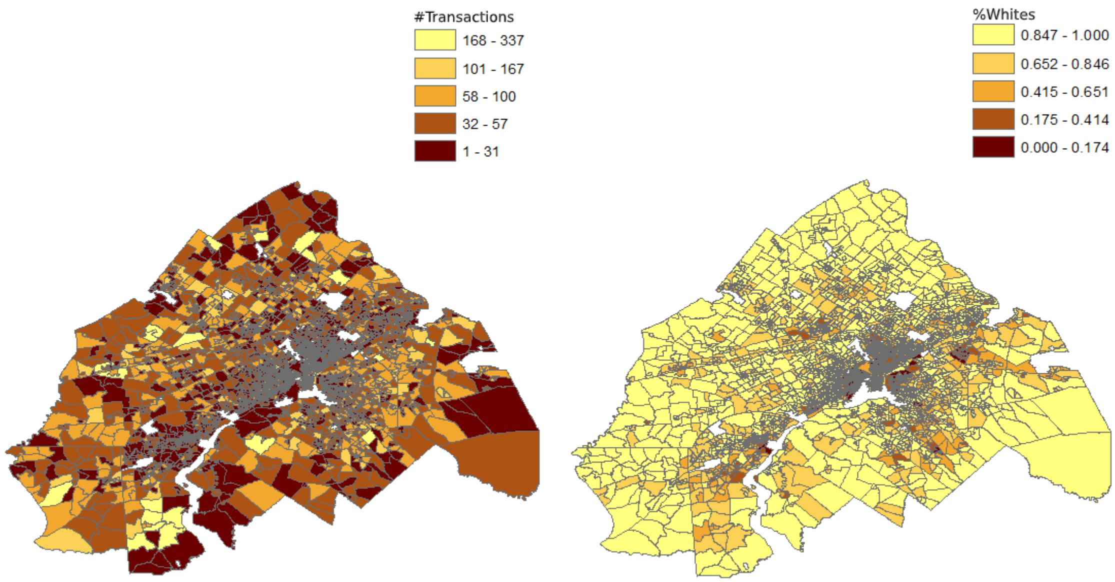
the assessed value, the number of days on market, and the number of days from pending to closing. All the distributions are right skewed and it seems that the Zestimate model is pretty accurate while assessed values don't fully reflect market values of those properties. The summary statistics are reported in Table 1.1.

The information collected from Zillow.com also includes many exterior and interior features of properties, including but not limited to the size, the property type, the number of bedrooms, the number of bathrooms, the exterior material, the exterior and interior amenities, the view, the cooling/heating conditions, the heating conditions, the appliances, and the flooring conditions (see Table 1.A1 for a full list of variables).

1.3.2 Census Data

The second data set I use is the neighborhood-level socioeconomic characteristics collected from the Decennial Census administered by the Census Bureau. The properties are matched with the census block group level data using the street address and the social-demographic characteristics are found for the 4,108 census block groups they belong to. Those block groups are highlighted on a map in Figure 1-A1.

Figure 1-3: Geospatial Distribution



(a) Transaction Frequency

(b) Share of Whites

Notes: The number of transactions for each census block are plotted in Panel A and the share of whites is plotted for each census block in Panel B.

The characteristics used in this study include but not limited to the population density, the gender distribution, the age distribution, the race and ethnicity distribution, the geographical mobility, the place of work, the commute methods, the one-way commute time, the household status, the household size, the average education level, the property status, the median household income, the income sources, the number of housing units, the race and ethnicity distribution for homeowners, the distribution of the number of bedrooms in a property, and the home type distribution (see Table 1.A2 for a full list of variables). In Figure 1.1, I plot the number of observations (Panel A) and the share of white residents (Panel B) in each block group. The graph shows the segmentation in the housing market: the suburbs are whiter than the city center. However, the transaction frequency is not aligned with the difference in racial markup and the market is active in some more diverse neighborhoods. The summary statistics for some important sociodemographic variables are reported in Table 1.1.

1.3.3 Assessment Information and Public Records

Finally, I collect the assessment information from the local government's website and newspapers. The time of the last revaluation is found for most towns (this information is missing for only 7 out of 529 towns (cities) included in the data set). This time varies from 0 months to 562 months as places like the city of Philadelphia and the state of Maryland conduct a regular reassessment every three years while in other places, such as Buck County, the assessed valuation of property has not been updated since 1970s.

The property records are also collected as a supplementary data set from the assessor's website for most of the properties located in New Jersey and Pennsylvania. Unfortunately, the property records are not available to the public in Delaware, Maryland, and Chester County in Pennsylvania. These public records are matched with the Zillow data using the street address. The most important variable I collect from this supplementary data set is the name of the current owner of a property. I only keep the most recent transaction for a property in the data set if it was traded multiple times during our time window. So I can identify the buyers for the transactions included in the data set using the owners' information. Using a prediction model that exploits the US census data, I am able to predict a buyer's race and ethnicity based on her last name ($ethnicolr$0.2.1). Gender is predicted based on the first name as well using prediction models that utilize the Social Security data sets (*gender*). If there are two owners owning the property jointly, I collect these variables for both of them. Moreover, the properties owned by firms are identified and later removed from the analysis.

1.4 The Effect of Zestimates on Home Prices

1.4.1 Model

The analysis focuses on the effect of changes in the Zestimate on the final sale price because this market outcome is a natural measure of how market participants react to the statistic.

The sale price for home i in census block k listed in month t is modeled as:

$$
\begin{aligned}
Sale\_Price_{itk} =& \alpha + \beta Zestimate_{i,t-1} + \gamma List\_Price_i + \theta Assessed\_Value_{it} + \delta Log(Days \\
& \_on\_Market)_i + \lambda Log(Days\_from\_\text{``}Pending\text{''}\_to\_\text{``}Sold\text{''})_i + \mu_t \\
& + \eta L_k + \zeta S_i + \epsilon_{ikt}, \qquad (1) \\
Zestimate_{i,t-1,k} =& \alpha' + \kappa Months\_Last\_Update_{i,t-1} + \gamma' List\_Price_i + \theta' Assessed\_Value_{it} \\
& + \delta' Log(Days\_on\_Market)_i + \lambda' Log(Days\_from\_\text{``}Pending\text{''}\_to \\
& \_\text{``}Sold\text{''})_i + \mu'_t + \eta' L_k + \zeta' S_i + \epsilon'_{ikt}. \qquad (2)
\end{aligned}
$$

$Zestimate_{i,t-1}$ is the Zestimate in month t-1. I use the lagged Zestimate here to isolate the effect of estimated market values on home prices from final sale prices' impacts on prediction outcomes, and β is the parameter of interest that shows the magnitude of this effect. I use the original list price ($List\_Price_i$), the assessed value ($Assessed\_Value_{it}$), and the number of days between listing and pending sale ($Log(Days\_on\_Market)_i$) to partially control the unobserved quality of the property and market condition. The number of days from the pending sale to closing ($Log(Days\_from\_\text{``}Pending\text{''}\_to\_\text{``}Sold\text{''})_i$) is added into the model as well to control the unobserved quality of the buyer (like whether they are making an all-cash offer or a mortgage offer and the uncertainties in eventually receiving a mortgage) and the property. The distribution of these explanatory variables is plotted in Figure 1-2. Moreover, μ_t is a vector of month indicators that control the month fixed effects. L_k is a vector of social-demographic controls at the block group level (see Table 1.A2 for a detailed list of variables) and S_i is a vector of controls related to home features (see Table 1.A1 for a detailed list of variables). ϵ_{ikt} is the idiosyncratic error term.

Given the nature of the predictive algorithm, there may be concerns over the endogeneity of Zestimates. It is because that the advanced prediction models Zillow uses may better reflect the unobserved quality of a property, even though I have controlled the important features used by Zillow in the linear model.[9]

[9]The variables used by Zillow can be found from the Kaggle competition hosted by them. See `https://www.kaggle.com/c/zillow-prize-1`.

As stated in Equation (2), I use 2SLS regressions where the lagged Zestimate is instrument with the number of months since the last revaluation or reassessment at time $t-1$ ($Months\_Last\_Update_{i,t-1}$, the summary statistics are reported in Table 1.1) to address those potential endogeneity issues. The idea is that the new observations a Zestimate is based on are more likely to be different from the data used by Zillow for the model training when a reassessment has been done recently. Therefore, the prediction errors caused by dataset shifts will be more severe if the reassessment frequency is higher. In particular, both the covariate shift in the assessed value itself and the shift in its relationship with the property's market value and other covariates such as property features will be larger if the assessed values are updated more frequently. Those shifts will exacerbate the estimation bias in the model (Kanamori & Shimodaira, 2009), and the first-stage results reported in Table 1.3 confirm this hypothesis.

Moreover, the difference in the frequency of revaluation and reassessment across townships is not correlated with the current sale prices (after controlling the assessed value) but jointly decided by many forces, like laws and budget plans. For example, in the city of Philadelphia and in Maryland, the assessed values have to be updated every three years by law while they have not been updated for more than 20 years in some other parts of Philadelphia and New Jersey. To illustrate the exclusion restriction assumption, I plot the geospatial distribution of the dependent variable (sale price), the independent variable of interest (the Zestimate) , and the instrumental variable (the number of months since the most recent reassessment) in Figure 1-A2. In particular, the darker the census block is in Panel A, the longer the time since the last assessment update is in the area. The segmentation in the time since the last reassessment caused by policy differences suggests that it is unlikely to be related to the sale price besides through the Zestimate's effect.

1.4.2 Results

The 2SLS results are reported in Panel A of Table 1.2. Column (1) in the table presents the result of a regression that includes only the Zestimate at month $t-1$ as the independent variable. The original list price, the assessed value, $Log(Days\_on\_Market)_i$,

Table 1.2: The Effect of Zestimates on Final Sale Prices

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.998***	0.507***	0.337***	0.335***	0.270***	0.231***
	(0.008)	(0.064)	(0.048)	(0.048)	(0.049)	(0.054)
List Price		0.439***	0.599***	0.600***	0.639***	0.658***
		(0.060)	(0.045)	(0.045)	(0.044)	(0.048)
Assessed Value		-0.046***	-0.024***	-0.023***	-0.023***	-0.031***
		(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
Log (Days on Market)			-12967.583***	-12778.470***	-12716.006***	-12918.462***
			(614.912)	(608.319)	(576.655)	(609.640)
Log (Days from "Pending" to "Sold")			5149.837***	5045.879***	5224.922***	5257.542***
			(364.152)	(359.637)	(362.475)	(369.564)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	120,482	88,110	52,981	52,981	52,387	52,164
Panel B: Subsample without Missing Information						
	(1)	(2)	(3)	(4)	(5)	(6)
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	1.010***	0.397***	0.344***	0.342***	0.277***	0.231***
	(0.004)	(0.050)	(0.051)	(0.051)	(0.052)	(0.054)
List Price		0.545***	0.592***	0.593***	0.633***	0.658***
		(0.047)	(0.048)	(0.048)	(0.047)	(0.048)
Assessed Value		-0.036***	-0.024***	-0.023***	-0.024***	-0.031***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Log (Days on Market)			-12828.849***	-12641.816***	-12617.347***	-12918.462***
			(624.945)	(618.303)	(595.072)	(609.640)
Log (Days from "Pending" to "Sold")			5119.504***	5016.118***	5225.440***	5257.542***
			(369.130)	(364.367)	(365.514)	(369.564)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	52,164	52,164	52,164	52,164	52,164	52,164

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The regressions reported in Panel A use the entire sample for estimation and the regressions reported in Panel B use only the observations without missing information.

$Log(Days\_from\_\text{“}Pending\text{”}\_to\_\text{“}Sold\text{”})_i$, the month fixed effects, the socio-demographic controls, and property feature controls are added into the model incrementally in Columns (2)-(6). The standard errors are clustered at city×month level to control the correlation between sales.

As we can see, the final sale price is significantly affected by the Zestimate displayed on Zillow.com – particularly, based on Column (6), on average a property is going to be sold 0.205 dollars higher if its Zestimate increases by one dollar. If we assume that the exogenous shock caused by covariate shifts is constant after controlling the variables used in the prediction model, this estimator estimates a weighted average of conditional average treatment effects. It suggests that the popular market value model does have an influence on transactions, apart from reflecting the unobserved quality and market conditions. Moreover, the final sale is more likely to be higher if the list price is higher, ceteris paribus, since the list price implies not only the quality of the property but also the seller's private information. It also makes sense the assessed value is negatively correlated to the sale price after controlling the property features and other conditions, since the property tax proportional to it is sometimes a major burden for the owner. As we can see from the table, the longer the property is on the market the lower the final sale price is: unpopular properties are more likely to be low quality and an unlucky seller who sells her house in a buyer's market has to lower the price. It is also consistent with the findings in Tucker *et al.* (2013). Reversely, buyers are more likely to receive a risk premium if the time between the pending sale and the final closing is longer.

The first stages are reported in Panel A of Table 1.3. The Wald F statistic always suggests a strong first stage and the number of months since the last update has a significant impact on the Zestimate displayed. Based on the results reported in Column (6), the Zestimate will be more than 92 dollars higher if the assessed value was updated one month earlier. In other words, Zestimates are more biased (i.e., underestimated, since the Zestimate is on average smaller than the final sale price) in those areas where assessed values are updated more frequently when the model used allows heterogeneity. It can be explained by the additional biases caused by covariate shifts as discussed before. To further illustrate it, in Figure 1-4 I show the estimated market values of two similar properties located next to each other. The first one is located in Elkins Park, PA, and the last time its assessed value got updated was in 1996 (the current assessed value is \$118,860). Even though the second home is only ten-minute away from the first one, it is located in the city of Philadelphia, which means that its assessed value is updated every three years (the current assessed value is \$148,200). Despite the second home is larger and have more bathrooms, its Zestimate is significantly lower than that of the first home. This example shows how the impact of reassessment frequency on the Zestimate.

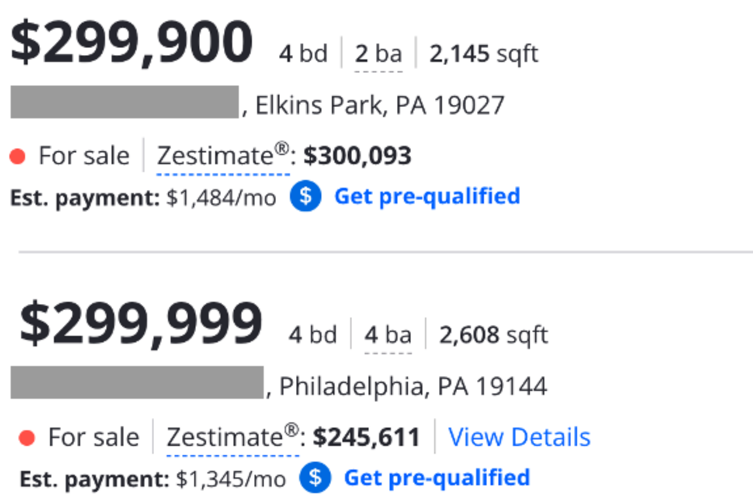
Table 1.3: First Stages

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate
Months_Last_Update	265.854***	163.572***	118.466***	118.341***	109.419***	96.786***
	(11.300)	(15.600)	(10.673)	(10.672)	(10.359)	(10.149)
List Price		0.748***	0.801***	0.800***	0.775***	0.765***
		(0.020)	(0.013)	(0.013)	(0.015)	(0.016)
Assessed Value		0.307***	0.223***	0.223***	0.199***	0.168***
		(0.031)	(0.022)	(0.022)	(0.021)	(0.021)
Log (Days on Market)			-10102.850***	-9828.774***	-8989.805***	-9043.551***
			(558.297)	(547.414)	(515.036)	(494.792)
Log (Days from “Pending" to “Sold")			3719.434***	3579.912***	3569.213***	3485.599***
			(370.919)	(370.504)	(358.610)	(350.474)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	120,482	88,110	52,981	52,981	52,387	52,164
F statistic	3939.31	980.59	2397.47	2391.47	1877.90	1398.43
R^2	0.032	0.589	0.928	0.928	0.928	0.931
Adjusted R^2	0.032	0.589	0.928	0.928	0.928	0.931
Panel B: Subsample without Missing Information						
	(1)	(2)	(3)	(4)	(5)	(6)
	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate
Months_Last_Update	288.451***	116.164***	111.237***	111.105***	104.950***	96.786***
	(13.704)	(10.935)	(10.908)	(10.910)	(10.480)	(10.149)
List Price		0.804***	0.808***	0.807***	0.781***	0.765***
		(0.013)	(0.013)	(0.013)	(0.015)	(0.016)
Assessed Value		0.204***	0.205***	0.206***	0.190***	0.168***
		(0.022)	(0.022)	(0.022)	(0.021)	(0.021)
Log (Days on Market)			-10025.264***	-9774.361***	-9000.666***	-9043.551***
			(555.432)	(544.513)	(515.616)	(494.792)
Log (Days from “Pending" to “Sold")			3754.129***	3629.701***	3625.058***	3485.599***
			(366.551)	(365.620)	(354.691)	(350.474)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	52,164	52,164	52,164	52,164	52,164	52,164
F statistic	2680.57	2293.42	2107.03	2100.93	1732.58	1398.43
R^2	0.049	0.927	0.928	0.928	0.930	0.931
Adjusted R^2	0.049	0.927	0.928	0.928	0.930	0.931

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The regressions reported in Panel A use the entire sample for estimation and the regressions reported in Panel B use only the observations without missing information.

The overall explanatory power of the first stage model represented by the R-squared is also reasonably high – the full model can explain 93% of the variability. Consistent with how Zestimates are computed, the original list price, the assessed value, the neighborhood socio-demographic features, and the property features play a significant role in explaining the variation in Zestimates. $Log(Days\_on\_Market)$ and $Log(Days\_from\_\text{“}Pending\text{”}\_to\_\text{“}Sold\text{”})$ also have significant effects because the effect of list price decays and the pending price which is recorded as the usually higher original list price is more likely to enter the model when it

Figure 1-4: First Stage: Illustrative Examples



takes a longer time to finalize the deal.

1.4.3 Robustness Checks

To check the robustness of the results, I first replicate those results using a sub-sample where all the observations with missing values are excluded from the analysis, which is the same sub-sample as the one used in Column (6) of Panel A in both Table 1.2 and 1.3. The replication results reported in Panel B of the corresponding table are very close to the ones estimated using the full sample. It suggests that the missing information doesn't reflect anything fundamental nor leads to biased results.

Another concern regarding the identification is that Zestimates displayed at the time the data was collected are different from the Zestimates displayed when buyers were making their purchase decisions. Though Zillow makes a major improvement in how they calculate their Zestimates every year or two[10], I find no evidence suggesting that they update the historical Zestimates as well. However, if it is true, the fact that the Zestimate is predicted by a model trained using the final sale price will undermine the identification results here. So I check the robustness of the model by replicating the results using only recent sales. The idea here

[10]See `https://www.forbes.com/sites/johnwake/2019/06/30/new-zillow-zestimate-accuracy/\#5548c2a28a07`

is that the more recent the sale was the less likely it will be included in the training data set. I present the results for properties sold less than one month before the data collection in Panel A of Table 1.A3 and for those properties sold less than three months before the data collection in Panel B of Table 1.A3. Despite the data sparsity, we can still observe a positive effect and the effect size is not different from the one reported in Table 1.2 at a significance level of 0.05. The first stages reported in 1.A4 are also similar to the ones for the main model.

Finally, I check the model specifications by replacing the social demographic controls with city fixed effects. So the variation (in both reassessment frequency and market prices) across cities is fully captured by the fixed effects. The results reported in Appendix (Tables 1.A5 an 1.A6) are close to those from the main model. It also further validates the exclusion restriction assumption.

1.5 Mechanism: Zestimates as a Public Source of Information

The Zestimate home valuation, as a summary statistic from a popular online real estate database, affects how real estate market participants acquire and process information. The sellers (buyers) often need to search market information before making the sell (purchase) decision and there are various information sources available to them, like advertising, word-of-mouth, internet, and even a trial sale (purchase). Particularly, there is an enormous need for information when making a major financial decision such as home sale and purchase and people spend tremendous time and effort acquiring and processing necessary information. For example, sellers and buyers usually hire professional agents for advice and register in the MLS system to receive notifications about new listings. In addition, buyers also go to open houses to see properties in person, search-related information (like crime maps and recently sold properties) online for their reference, and ask friends and colleagues for suggestions. In this section, I extend the basic model presented in the previous section and study how the market value estimated using algorithms differ from other information sources.

I find that people are more likely to rely on the Zestimate home valuation when there is a larger set of private signals. I suggest that it is because users view the estimated market value as a public source of information that other parties can also observe, and relying on this summary is more cost-efficient when it becomes harder to reach a consensus with private information sources.[11] A theoretical model that explains the underlying information acquisition process is presented in Appendix.

1.5.1 Subgroup Analysis

Comparable sales are one of most important private signals in the housing market. Usually agents will go through all the comparable sales and compare them with the focal property before they make price suggestions for their clients. Those signals are private because different people have different opinions on the same set of properties and it may lead to different estimated values for the focal property.

In Table 1.4, for each property. its Zestimate is interacted with the transaction frequency of comparable properties sold within 6 months prior to the sale (in Panel A, or within 12 months prior to the sale in Panel B). I divide the properties into 5 segments based on their final sale prices: the properties sold at a price lower than 135k, the properties sold at a price between 135k and 196k, the properties sold at a price between 265k and 375k, the properties sold at a price higher than 375k. I only count the number of sales in the same segment because users usually only consider comparable homes when they are evaluating properties. To control the size of the neighborhood and thus that of the buyer consideration set, I standardize the number of properties sold in the same segments and divide it by either the population of the area (Columns (1), (2), (5), and (6) in each panel) or the total number of housing units in the area (Columns (3), (4), (7), and (8) in each panel). The summary statistics for the moderator used in Column (1) of Panel A are reported in Table 1.1. The Zestimate is interacted with the housing market activeness at both block group (Columns (1)-(4) in each panel) and city (Columns (5)-(8) in each panel) levels. As we can see, the number of recent sales has a negative effect on the final sale price, which is consistent with the hypothesis that a market with higher supply is less likely to be a seller market where

[11] Because people are less likely to choose to observe the same set of private signals when the choice set becomes larger if there is a constraint on the number of signals they can observe.

Table 1.4: Mechanism: Interaction with Number of Comparable Sales

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Interaction with Frequency of Similar Transactions Nearby ($t-6$)								
Area	Block Group				City			
Measurement	per Capita		per Housing Unit		per Capita		per Housing Unit	
Zestimate_Sold	0.345***	0.240***	0.349***	0.239***	0.364***	0.242***	0.351***	0.240***
	(0.049)	(0.054)	(0.049)	(0.055)	(0.053)	(0.055)	(0.048)	(0.054)
# Transactions	-1275060.638***	-436734.618*	-548333.146***	-164017.904	-1365560.718*	-370443.009***	-386168.813***	-129545.705***
	(291592.872)	(244943.577)	(151556.395)	(111914.450)	(716143.083)	(140184.685)	(112902.358)	(44622.119)
Zestimate_Sold	3.755***	2.424***	1.702***	0.948**	3.499*	1.076***	0.996***	0.390***
× # Transactions	(1.018)	(0.895)	(0.486)	(0.375)	(1.823)	(0.345)	(0.283)	(0.110)
List Price	0.585***	0.645***	0.580***	0.645***	0.565***	0.646***	0.579***	0.648***
	(0.046)	(0.048)	(0.046)	(0.048)	(0.051)	(0.048)	(0.045)	(0.048)
Assessed Value	-0.024***	-0.030***	-0.024***	-0.030***	-0.029***	-0.033***	-0.027***	-0.033***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Log (Days on Market)	-12825.156***	-12813.274***	-12809.992***	-12823.744***	-12320.859***	-12751.286***	-12565.889***	-12780.412***
	(626.918)	(610.874)	(625.494)	(610.132)	(771.537)	(619.890)	(645.662)	(614.987)
Log (Days from "Pending" to "Sold")	5118.994***	5229.373***	5138.132***	5239.319***	5035.742***	5219.117***	5088.805***	5229.582***
	(371.941)	(372.352)	(373.685)	(372.502)	(378.147)	(369.121)	(364.850)	(368.241)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	52,981	52,164	52,981	52,164	52,981	52,164	52,981	52,164
Panel B: Interaction with Frequency of Similar Transactions Nearby ($t-12$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Area	Block Group				City			
Measurement	per Capita		per Housing Unit		per Capita		per Housing Unit	
Zestimate_Sold	0.347***	0.241***	0.349***	0.240***	0.365***	0.246***	0.354***	0.243***
	(0.049)	(0.054)	(0.049)	(0.054)	(0.052)	(0.055)	(0.048)	(0.054)
# Transactions	-796409.968***	-298056.195*	-328162.582***	-107087.095	-745732.650**	-252819.535***	-221249.713***	-88025.446***
	(182748.518)	(157134.731)	(94141.464)	(73426.032)	(365159.708)	(80605.890)	(62574.799)	(25756.795)
Zestimate_Sold	2.399***	1.629***	1.039***	0.629***	1.904**	0.724***	0.573***	0.262***
× # Transactions	(0.629)	(0.565)	(0.298)	(0.241)	(0.916)	(0.194)	(0.156)	(0.062)
List Price	0.584***	0.644***	0.580***	0.645***	0.566***	0.643***	0.577***	0.645***
	(0.045)	(0.047)	(0.045)	(0.047)	(0.050)	(0.048)	(0.045)	(0.048)
Assessed Value	-0.023***	-0.030***	-0.024***	-0.030***	-0.028***	-0.033***	-0.027***	-0.033***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Log (Days on Market)	-12847.774***	-12825.006***	-12837.488***	-12835.724***	-12386.913***	-12717.164***	-12577.214***	-12748.203***
	(622.250)	(608.236)	(622.280)	(608.759)	(726.864)	(617.871)	(638.630)	(613.655)
Log (Days from "Pending" to "Sold")	5125.016***	5218.766***	5144.814***	5229.976***	5060.833***	5205.761***	5100.364***	5217.095***
	(369.134)	(370.178)	(371.661)	(371.068)	(371.365)	(368.280)	(363.435)	(367.640)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	52,981	52,164	52,981	52,164	52,981	52,164	52,981	52,164

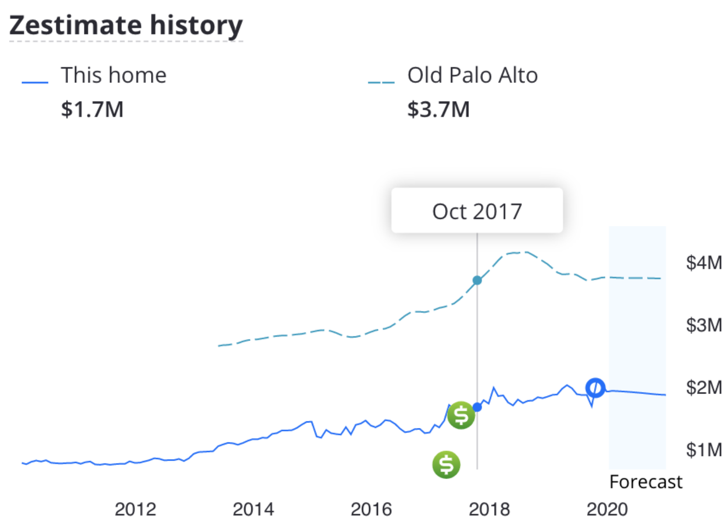
Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The number of transactions (adjusted for the neighborhood size) is mean centered to allow easy interpretation of the main effects.

buyers usually overbid for the properties. However, the effect of the Zestimate on the sale price increases with the number of recent sales in the neighborhood: based on the results reported in Column (2), the effect size will become about 10 times larger if there is one more recent transaction per Capita. It suggests that users rely more on the statistics that everyone can easily observe rather than focusing on analyzing the comparable sales and getting the private signals when the private signals they can from the comparable sales are noisier.

Robustness checks reported in Appendix (Table 1.A7) replicate the results but exclude those sales with a logged sale-to-list ratio that is too high or too low, which may not be used as a reliable reference for the home's market value. The results are consistent with the conclusion drawn from Table 1.4.

1.5.2 Alternative Explanation

Figure 1-5: Zestimate Details



Another potential explanation for the increasing usage of Zestimates when there are more comparable sales is that real estate market participants may use the richness of data as a proxy for the summary statistic's accuracy. Here I provide evidence to eliminate this possibility by interacting the Zestimate with the average accuracy of Zestimates for the comparable properties sold in the surrounding area recently (sold within 6 months prior to

Table 1.5: Mechanism: Interaction with Information Accuracy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Interaction with Average Deviation for Similar Transactions Nearby ($t-6$)								
Area	Block Group				City			
Measurement	$Zestimate_t$		$Zestimate_{t-1}$		$Zestimate_t$		$Zestimate_{t-1}$	
Zestimate_Sold	0.345***	0.244***	0.344***	0.244***	0.340***	0.240***	0.338***	0.240***
	(0.050)	(0.058)	(0.050)	(0.059)	(0.048)	(0.053)	(0.048)	(0.053)
% Deviation	-2511.465	-3280.874	-2834.807	-3447.803	-17722.572	-9872.767	-17722.429	-8375.687
	(4186.535)	(2948.192)	(4052.144)	(2857.283)	(11046.479)	(6830.298)	(12888.196)	(7716.131)
Zestimate_Sold	0.002	0.006	0.003	0.007	0.039	0.023	0.040	0.020
× % Deviation	(0.009)	(0.006)	(0.009)	(0.006)	(0.025)	(0.016)	(0.030)	(0.018)
List Price	0.592***	0.645***	0.592***	0.645***	0.598***	0.651***	0.599***	0.651***
	(0.047)	(0.051)	(0.047)	(0.051)	(0.045)	(0.047)	(0.045)	(0.047)
Assessed Value	-0.028***	-0.034***	-0.028***	-0.034***	-0.023***	-0.032***	-0.024***	-0.032***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)
Log (Days on Market)	-12991.715***	-12763.784***	-12953.288***	-12726.558***	-12716.353***	-12713.136***	-12696.927***	-12686.735***
	(646.046)	(660.328)	(645.423)	(660.545)	(617.036)	(603.497)	(615.141)	(602.462)
Log (Days from "Pending" to "Sold")	5381.650***	5461.936***	5369.044***	5449.588***	5142.910***	5246.551***	5140.017***	5234.109***
	(379.645)	(389.125)	(379.913)	(389.371)	(363.891)	(367.774)	(363.869)	(367.439)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	48,193	47,501	48,195	47,503	52,508	51,711	52,510	51,713
Panel B: Interaction with Average Deviation for Similar Transactions Nearby ($t-12$)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Area	Block Group				City			
Measurement	$Zestimate_t$		$Zestimate_{t-1}$		$Zestimate_t$		$Zestimate_{t-1}$	
Zestimate_Sold	0.335***	0.237***	0.335***	0.237***	0.324***	0.224**	0.287	0.197
	(0.049)	(0.056)	(0.049)	(0.056)	(0.102)	(0.100)	(0.463)	(0.421)
% Deviation	-9327.655	-6291.070	-8940.396	-5986.132	-43440.374	-30206.358	-92556.569	-58276.898
	(6417.173)	(4345.996)	(5874.783)	(3970.178)	(166395.047)	(122284.624)	(704970.950)	(504676.677)
Zestimate_Sold	0.015	0.011	0.015	0.011	0.093	0.066	0.204	0.129
×% Deviation	(0.012)	(0.008)	(0.011)	(0.008)	(0.356)	(0.261)	(1.548)	(1.107)
List Price	0.600***	0.651***	0.600***	0.651***	0.613***	0.665***	0.650	0.689*
	(0.046)	(0.049)	(0.046)	(0.049)	(0.101)	(0.088)	(0.463)	(0.371)
] Assessed Value	-0.026***	-0.033***	-0.026***	-0.033***	-0.023***	-0.032***	-0.023***	-0.033*
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)	(0.008)	(0.018)
Log (Days on Market)	-13065.105***	-12890.977***	-13030.870***	-12854.729***	-12689.190***	-12772.371***	-12619.408***	-12849.709***
	(631.133)	(633.741)	(631.390)	(634.396)	(720.845)	(646.015)	(1205.348)	(1581.054)
Log (Days from "Pending" to "Sold")	5298.621***	5368.578***	5287.444***	5356.610***	5247.413***	5350.865***	5441.910**	5495.004**
	(368.005)	(377.875)	(368.708)	(378.272)	(632.851)	(648.783)	(2639.005)	(2515.343)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	49,305	48,587	49,307	48,589	52,585	51,782	52,587	51,784

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The accuracy of the Zestimate for a recent sale is calculated as $|Zestimate - Sale\_Price|/Zestimate$ and the average accuracy is mean centered to allow easy interpretation of the main effects.

the sale for the results reported in Panel A of Table 1.5 and sold within 12 months prior to the sale for the results reported in Panel B of the same table). The idea here is that if users attempt to infer the reliability of the Zestimate of their focal interest from the recent sales, they will be very likely to also look at more direct evidence, for example, the realized difference between the Zestimate and the final sale price for recent sales.[12]

Similar to Table 1.4, I interact the Zestimate with the average difference for properties sold in the same price segment at both block group (Columns (1)-(4) in each panel) and city (Columns (5)-(8) in each panel) levels. I also measure the average difference in terms of both the percent deviation of the final sale price from the Zestimate reported during the month of sale (Columns (1), (2), (5), and (6) in each panel) and that from the Zestimate reported one month before the sale (Columns (3), (4), (7), and (8) in each panel). The former one is a natural comparison because both numbers have the same y-value on Zillow's trend graph. The later one further entrenches the conclusion by considering the possibility that users are sophisticated so that they are able to realize that the potential changes in the Zestimate between the time they are looking at it and the closing. The summary statistics for the moderator used in Column (1) of Panel A are reported in Table 1.1. The results reported in Table 1.5 suggest that the average deviation level doesn't have any material impact on how much customers rely on the Zestimate to make their purchase decision. It implies that either users don't update their belief about the Zestimate's accuracy using recent sales or the inferred accuracy is not the main factor that determines their reliance on the Zestimate. Unlike how users construct their reference sets, it is possible that customers will infer the accuracy of the summary statistic from how it performed for properties in other price segments because they may still browse them just out of curiosity. So I include the properties from other segments and replicate Table 1.5. The results are reported in 1.A8 and they lead to confirm the finding that it is unlikely that users increase their usage of Zestimates because of the alleviation of accuracy concerns.

[12]As shown in Figure 1-5, this difference has been visualized on Zillow.com.

1.6 Heterogeneity in Effect Size and How it Moderates Racial Biases in Housing Market

In this section, I further investigate how the influence of the Zestimate varies across neighborhoods and how it potentially affects the racial biases in the housing market. I first show how the effect changes with demographic variables in Section 1.6.1. Then I examine the racial biases in the housing market caused by the redlining policy in section 1.6.2. Finally, in Section 1.6.3 I provide evidence of a smaller bias in Zestimate and combine the results from the previous sections I discuss how it leads to a moderated racial bias in final sale price.

1.6.1 Influence of Demographics on Zestimate's Effect

In this subsection, I inspect what socioeconomic variables affect the market participants' reliance on the Zestimate. First, I focus on how the dependence on the summary statistic changes with education and income levels. The answer to this question is ambiguous. On one hand, previous studies (Newman & Staelin, 1972, Claxton *et al.* , 1974, Schaninger & Sciglimpaglia, 1981) have shown that the depth and breadth of information search before a purchase are usually higher among buyers with higher education and higher income. So those privileged people are likely to search for more information and be less reliant on a single information source. Also, as shown in Table 1.4, they are also well-trained to process the raw data. On the other hand, due to the well-known digital divide, low-income populations may not have access to the Internet and thus Zillow.com.

To study it, I interact the Zestimate with the average education level and the median income separately. I take a log transformation of the median income since its distribution is highly right-skewed and we use the average number of years of education received to measure the average education level. The results are reported in Table 1.6: The Zestimate is interacted with the average number of years of education for results reported in Columns (1)-(3) while it is interacted with the logged median income for results reported in Columns (4)-(6). Control variables are added into the model incrementally. The results show that the areas with a better-educated population and higher median income are more likely to be a

Table 1.6: Heterogeneous Effect: Influence of Education and Income Levels

	(1)	(2)	(3)	(4)	(5)	(6)
Moderator	#Years of Education			Log(Median Income)		
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.281***	0.239***	0.238***	0.304***	0.263***	0.236***
	(0.046)	(0.052)	(0.054)	(0.047)	(0.053)	(0.054)
Moderator	8116.035***	7649.521***	7833.780***	16851.218***	13647.872***	11603.691***
	(957.874)	(939.070)	(1277.983)	(2057.227)	(1922.090)	(3510.055)
Zestimate_Sold	-0.015***	-0.014***	-0.021***	-0.020**	-0.019**	-0.032***
× Moderator	(0.004)	(0.004)	(0.005)	(0.009)	(0.008)	(0.011)
List Price	0.654***	0.678***	0.682***	0.627***	0.649***	0.670***
	(0.040)	(0.044)	(0.046)	(0.042)	(0.045)	(0.045)
Assessed Value	-0.017***	-0.025***	-0.028***	-0.026***	-0.032***	-0.028***
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)
Log (Days on Market)	-13055.079***	-13084.741***	-13089.605***	-12891.582***	-12927.252***	-12991.159***
	(570.259)	(606.991)	(591.699)	(573.050)	(601.435)	(594.689)
Log (Days from “Pending" to “Sold")	5391.538***	5326.206***	5307.059***	5201.606***	5181.476***	5290.396***
	(356.011)	(358.672)	(362.335)	(354.627)	(357.244)	(365.609)
Month Fixed Effects	No	Yes	Yes	No	Yes	Yes
Socio-Demographic Controls	No	No	Yes	No	No	Yes
Property Feature Controls	No	Yes	Yes	No	Yes	Yes
Observations	52,981	52,757	52,164	52,515	52,292	52,164

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The average number of years of education (the moderator used in Columns (1)-(3)) and the logged median income (the moderator used in Columns (4)-(6)) in the census block are mean centered to allow easy interpretation of the main effects.

"seller market", in which the final sale prices are higher due to the excess demand. However, users participating in those markets are less likely to rely on the summary statistics provided on Zillow.com to make their purchase decisions. This finding is consistent with our previous conclusion that the Zestimate serves as the summary of one of the information sources and its effect will diminish when 1) people's ability to process the raw data is higher 2) other information sources are more accessible. I have shown that more educated people are better at processing detailed transaction information. Moreover, other information sources are more accessible to privileged people, for example, they can pay a premium and hire experienced buyer agents or they are more likely to have friends or colleagues who have participated in local markets before and know the local market very well.

To further investigate this difference, I replace the average education level and the median income with measures of information access and search cost. The two measures I use here are the percentage of households that have an Internet subscription and the percentage of households that have one or more types of computing devices. The results are reported in Table 1.7: Zestimate is interacted with the percentage of households that have an Internet subscription for results reported in Columns (1)-(3) while it is interacted with the percentage

Table 1.7: Heterogeneous Effect: Influence of Internet and Computer Access

	(1)	(2)	(3)	(4)	(5)	(6)
Moderator	%Internet Subscription			%Computer Ownership		
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.324^{***}	0.272^{***}	0.237^{***}	0.330^{***}	0.276^{***}	0.223^{***}
	(0.048)	(0.053)	(0.055)	(0.047)	(0.051)	(0.054)
Moderator	38017.777^{***}	29607.734^{***}	18377.405^{*}	55582.731^{***}	46708.265^{***}	28989.568^{*}
	(7609.364)	(6984.501)	(9649.062)	(10542.538)	(9915.323)	(14801.575)
Zestimate_Sold × Moderator	-0.035	-0.036	-0.079^{*}	-0.114^{**}	-0.117^{***}	-0.116^{*}
	(0.037)	(0.033)	(0.044)	(0.049)	(0.045)	(0.063)
List Price	0.609^{***}	0.639^{***}	0.659^{***}	0.608^{***}	0.640^{***}	0.673^{***}
	(0.044)	(0.046)	(0.047)	(0.043)	(0.046)	(0.046)
Assessed Value	-0.024^{***}	-0.032^{***}	-0.031^{***}	-0.024^{***}	-0.032^{***}	-0.031^{***}
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
Log (Days on Market)	-12857.756^{***}	-12953.386^{***}	-12908.936^{***}	-12876.829^{***}	-12972.860^{***}	-12952.265^{***}
	(607.520)	(632.734)	(607.453)	(604.060)	(628.033)	(605.820)
Log (Days from "Pending" to "Sold")	5143.732^{***}	5129.542^{***}	5252.362^{***}	5139.963^{***}	5133.594^{***}	5209.047^{***}
	(360.727)	(363.317)	(368.700)	(358.690)	(361.118)	(363.482)
Month Fixed Effects	No	Yes	Yes	No	Yes	Yes
Socio-Demographic Controls	No	No	Yes	No	No	Yes
Property Feature Controls	No	Yes	Yes	No	Yes	Yes
Observations	52,981	52,757	52,164	52,981	52,757	52,164

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The percentage of households that have an Internet subscription (the moderator used in Columns (1)-(3)) and the percentage of households that have one or more types of computing devices (the moderator used in Columns (4)-(6)) in the census block are mean centered to allow easy interpretation of the main effects.

of households that have one or more types of computing devices for results reported in Columns (4)-(6). Control variables are added into the model incrementally. I find that even though that people who don't have stable access to the Internet or computing devices are less likely to utilize the Zestimate, the Internet subscription and computing device ownership do slightly reduce buyers' reliance on Zestimate. It suggests that people are more likely to rely on accessible and simple information sources such as Zestimate and other summary statistics when it is hard for them to acquire and search for other information.

As I have already shown in Table 1.6, users' reliance on the summary statistic depends on some socio-demographic characteristics of the neighborhood. Here I investigate whether there is a difference between "whiter" neighborhoods and more diverse neighborhoods after controlling the socio-demographic status. In Table 1.8, I interact the Zestimate with two measures of racial makeup – the percentage of white residents (reported in Columns (1)-(3)) and the percentage of white homeowners (reported in Columns (4)-(6)). The share of white residents is the most common measure of neighborhood racial diversity and I introduce the percentage of white homeowners here as well to better approximate the race of the average market participant in the neighborhood. Those two measures are close to each other: the

Table 1.8: Heterogeneous Effect: Influence of Racial Makeup

	(1)	(2)	(3)	(4)	(5)	(6)
Moderator	%White			%White Owners		
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.323***	0.264***	0.229***	0.323***	0.264***	0.232***
	(0.048)	(0.052)	(0.054)	(0.048)	(0.053)	(0.054)
Moderator	7308.393*	4495.520	3259.813	9986.045**	6684.631*	-3589.206
	(4109.398)	(3740.966)	(4259.760)	(4436.761)	(3930.080)	(6772.646)
Zestimate_Sold	0.012	0.019	0.010	-0.001	0.009	-0.004
× Moderator	(0.022)	(0.019)	(0.020)	(0.025)	(0.022)	(0.026)
List Price	0.607***	0.640***	0.658***	0.608***	0.641***	0.658***
	(0.045)	(0.047)	(0.048)	(0.044)	(0.047)	(0.048)
Assessed Value	-0.022***	-0.031***	-0.031***	-0.022***	-0.031***	-0.031***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Log (Days on Market)	-12983.688***	-13083.290***	-12928.026***	-12983.307***	-13083.137***	-12917.784***
	(611.579)	(636.319)	(611.637)	(611.800)	(637.583)	(611.125)
Log (Days from "Pending" to "Sold")	5243.095***	5222.780***	5263.553***	5253.606***	5231.249***	5256.706***
	(368.246)	(372.687)	(372.891)	(367.849)	(372.907)	(372.971)
Month Fixed Effects	No	Yes	Yes	No	Yes	Yes
Socio-Demographic Controls	No	No	Yes	No	No	Yes
Property Feature Controls	No	Yes	Yes	No	Yes	Yes
Observations	52,981	52,757	52,164	52,981	52,757	52,164

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The share of white residents (the moderator used in Columns (1)-(3)) and the share of white home owners (the moderator used in Columns (4)-(6)) in the census block are mean centered to allow easy interpretation of the main effects.

average neighborhood is about 79 percent white in the data set while the share of white is 2 percent higher in terms of home-ownership. The heavy tails show that the residential segregation between whites and minorities. If minorities are inferior in terms of information acquisition and processing even after controlling the education and wealth inequality, it is likely that we will see heavier use of Zestimates in those less-white neighborhoods. This interaction effect exists in our data but it is not significant: In Table 1.8, the sale price is less likely to be influenced by the Zestimate when the percentage of white homeowners in the neighborhood increases, even though the interaction term is not significant.

To double-check this difference, I split the observations into two subgroups based on the share of white residents in the neighborhoods and run the 2SLS regressions separately for each group. The results for those properties that sold in a neighborhood where the white population share is greater than the median (0.852) is reported in Columns (1)-(3) of Panel A in Table 1.9 and the results for the properties that sold in a neighborhood that the white population share is less or equal to the median is reported in Columns (4)-(6) in the same panel.[13] Similar to the results reported with interactions, the coefficient of the

[13]The average white population share is 0.927 for the first group and 0.653 for the second group.

Table 1.9: Subgroup Analysis: Influence of Racial Makeup

Panel A: %White Residents in Census Block Group						
	(1)	(2)	(3)	(4)	(5)	(6)
	%White > Median			%White ≤ Median		
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.300***	0.269***	0.222***	0.374***	0.276**	0.235**
	(0.049)	(0.061)	(0.067)	(0.105)	(0.111)	(0.110)
List Price	0.627***	0.640***	0.664***	0.568***	0.637***	0.657***
	(0.045)	(0.051)	(0.054)	(0.099)	(0.102)	(0.099)
Assessed Value	-0.024***	-0.028***	-0.028***	-0.021***	-0.034***	-0.036***
	(0.004)	(0.006)	(0.006)	(0.005)	(0.006)	(0.008)
Log (Days on Market)	-14667.518***	-14480.005***	-14506.841***	-11161.042***	-11535.069***	-11406.667***
	(813.330)	(942.634)	(927.317)	(1078.286)	(1040.761)	(969.328)
Log (Days from “Pending" to “Sold")	5996.536***	5819.319***	5948.616***	4390.895***	4484.136***	4563.159***
	(561.078)	(595.009)	(599.379)	(479.313)	(468.427)	(467.066)
Month Fixed Effects	No	Yes	Yes	No	Yes	Yes
Socio-Demographic Controls	No	No	Yes	No	No	Yes
Property Feature Controls	No	Yes	Yes	No	Yes	Yes
Observations	26,513	26,418	26,142	26,468	26,339	26,022
Panel B: %White Owners in Census Block Group						
	(1)	(2)	(3)	(4)	(5)	(6)
	%White Owners > Median			%White Owners ≤ Median		
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.292***	0.262***	0.219***	0.409***	0.301**	0.255**
	(0.051)	(0.063)	(0.071)	(0.122)	(0.129)	(0.127)
List Price	0.636***	0.647***	0.669***	0.535***	0.611***	0.636***
	(0.047)	(0.053)	(0.057)	(0.115)	(0.118)	(0.114)
Assessed Value	-0.024***	-0.028***	-0.028***	-0.020***	-0.033***	-0.036***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.008)	(0.009)
Log (Days on Market)	-15050.206***	-14853.645***	-14719.009***	-10570.085***	-11040.001***	-11104.400***
	(830.547)	(949.506)	(923.403)	(1194.037)	(1178.574)	(1100.818)
Log (Days from “Pending" to “Sold")	6218.804***	5988.047***	6066.440***	4070.516***	4264.311***	4396.335***
	(520.761)	(536.843)	(532.046)	(538.006)	(545.629)	(550.096)
Month Fixed Effects	No	Yes	Yes	No	Yes	Yes
Socio-Demographic Controls	No	No	Yes	No	No	Yes
Property Feature Controls	No	Yes	Yes	No	Yes	Yes
Observations	26,498	26,419	26,138	26,483	26,338	26,026
Panel C: Race of Individual Buyer						
	(1)	(2)	(3)	(4)	(5)	(6)
	White Buyers			Non-white Buyers		
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.354***	0.349***	0.299***	0.387*	0.419*	0.302
	(0.067)	(0.069)	(0.081)	(0.226)	(0.214)	(0.191)
List Price	0.590***	0.591***	0.623***	0.571***	0.541***	0.631***
	(0.062)	(0.061)	(0.070)	(0.212)	(0.199)	(0.175)
Assessed Value	-0.018***	-0.035***	-0.035***	-0.013	-0.025	-0.023
	(0.006)	(0.008)	(0.010)	(0.017)	(0.022)	(0.024)
Log (Days on Market)	-12343.447***	-11716.425***	-11842.400***	-10366.669***	-10322.607***	-12527.589***
	(933.093)	(920.590)	(975.997)	(3082.661)	(2871.303)	(2211.687)
Log (Days from “Pending" to “Sold")	5202.912***	5096.307***	5008.896***	2394.762*	2365.489*	3006.329***
	(548.518)	(498.374)	(505.707)	(1382.115)	(1232.515)	(1040.552)
Month Fixed Effects	No	Yes	Yes	No	Yes	Yes
Socio-Demographic Controls	No	No	Yes	No	No	Yes
Property Feature Controls	No	Yes	Yes	No	Yes	Yes
Observations	11,161	11,135	10,974	2,339	2,333	2,281

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. In Panel A, the regressions reported in Columns (1)-(3) use the purchases made in block groups where the share of white residents is greater than the median (0.858) for estimation while the regressions reported in Columns (4)-(6) use the purchases made in block groups where the share of white residents is less than the median. In Panel B, the regressions reported in Columns (1)-(3) use the purchases made in block groups where the share of white home owners is greater than the median (0.889) for estimation while the regressions reported in Columns (4)-(6) use the purchases made in block groups where the share of white home owners is less than the median. The regressions reported in Columns (1)-(3) use the purchases made by whites for estimation while the regressions reported in Columns (4)-(6) use the purchases made by non-whites. Buyers' races are identified from the names shown on public records.

Zestimate is more significant and greater in magnitude for less white neighborhoods. A similar subgroup analysis has been done based on the share of white homeowners in the neighborhoods (where the median is 0.885) and the results reported in Panel B in Table 1.9 show a similar pattern.[14] All of these findings show that people living in more diverse neighborhoods are relying more on the online summary statistic when making the important financial decision but this disparity driven by the racial makeup is much less significant than those differences among neighborhoods with various education or income level. Finally, I dig into the individual data obtained from the public land records where the buyers' race is identified from the public land records and check the racial difference on the individual level. I run regressions similar to Panel A of Table 1.9 but this time separately for white buyers and nonwhite buyers. The 2SLS results are reported in Panel C of Table 1.9. Consistent with previous results, nonwhites are more responsive to the changes in the Zestimate even though the difference is not significant.

1.6.2 Lingering Impact of Federal "Redlining": Home Value Gap between Whites and Nonwhites

I then investigate the lingering effects of "redlining", in other words, whether the minority neighborhoods are suffered from an unfair housing market. "Redlining" refers to the practice that the Federal Home Loan Bank Board (FHLBB)'s practices to create a map to indicate the level of security for real-estate investments in each neighborhood. The neighborhoods that were considered the riskiest for mortgages were outlined in red on this map and these neighborhoods are often minority neighborhoods. The existence of the map led to mortgage loan denials (Jackson, 1987) and a persistent decline in home values (Rutan & Glass, 2018) in these minority communities.

To quantify this effect, I run a set of OLS regressions where the independent variable of interest is the share of whites in the neighborhood. Two different measures are used here for the home value: the list price and the final sale price. Those variables are used to inspect the level of Redlining's lingering impacts on different aspects of the housing market. The

[14] The average share of white homeowners is 0.949 for the first group and 0.693 for the second group.

basic model for the prices is:

$$P_{itk} = \alpha^D + \beta^D Share\_Whites_k + \mu_t^D + \eta^D L_k + \zeta S_i^D + \epsilon_{ikt}^D,$$

where all the variables are the same as the ones described in Section 1.4.1 except the independent variable of interest is now $Share\_Whites_k$ and β^D shows the level of Redlining's lingering impact on the housing market.

Table 1.10: Lingering Impact of Federal "Redlining" on Housing Market

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	List Price		Sale Price			Δ
% White	100933.142***	101434.944***	94258.580***	95896.516***	101149.820***	-8374.646***
	(3151.928)	(4133.697)	(3029.549)	(3908.028)	(4687.799)	(1089.829)
Assessed Value		0.577***		0.454***	0.498***	-0.113***
		(0.034)		(0.026)	(0.031)	(0.012)
Log (Days on Market)					-8377.772***	-16609.433***
					(1105.969)	(546.676)
Log (Days from "Pending" to "Sold")					91.014	7160.211***
					(760.541)	(399.763)
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Socio-Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Property Feature Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	116,587	86396	116,587	87,254	52,585	52,585
R-squared	0.562	0.637	0.594	0.652	0.662	0.215
Adjusted R-squared	0.562	0.637	0.593	0.651	0.661	0.213

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is list price in Columns (1) and (2), sale price in Columns (3)-(5) and their difference (Δ = sale price- list price) in Column (6).

The results are reported in Column (1) of Table 1.10 for the list price and in Column (3) of the same table for the sale price. As we can see, both measures increase when the neighborhood is whiter: The average list price will increase by 964.8 dollars and the average sale price will increase by 900.1 dollars when the neighborhood is one percent whiter. The difference between the two gaps is significant. To test the robustness of the results, I add the assessed value into the model and rerun the regressions again. The results reported in Columns (2) and (4) show that the estimated racial biases are slightly greater and still very significant. A similar result for the sale price can be drawn from the robustness check where the number of days and the number of days from the pending sale to closing are also added to control the unobserved house/seller quality. Those findings are consistent with the findings in the previous studies (Aaronson *et al.* , 2017, Perry *et al.* , 2018). Another thing that is noticeable here is that the white-premium is higher for the list prices and it might have been moderated by some factors during the sale process, which leads to a lower premium in the final sale prices. To confirm this finding, I regress the logged sale-to-list ratios on the share of whites. The results reported in Column (6) suggest that the sale-to-list ratio is indeed significantly lower in those predominately white neighborhoods.

1.6.3 Racial Biases in Zestimate and its Moderation Effect on “Redlining"

One burning issue regarding algorithms is whether the outcomes they produced are biased (Cowgill & Tucker, 2020). So here I focus on investigating how biased the Zestimate is compared to the existing racial biases in the housing market – whether the Zestimate reflects this gap between white and nonwhites or the gap will attenuate through how the Zestimate is calculated. To study this question, I run regressions similar to those in Table 1.10 but use the log-transformed Zestimate-to-List ratio as the dependent variable.

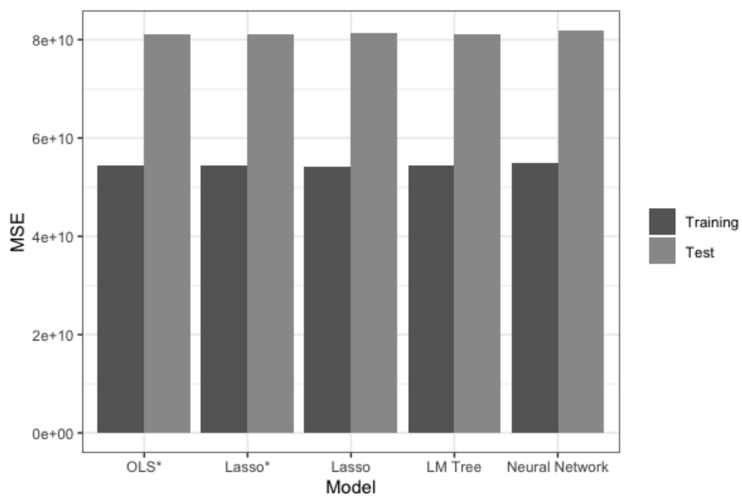
Table 1.11: Racial Biases in Zestimate

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log(Zestimate/List Price)			Log(Zestimate/Sale Price)		
% White	-0.034***	-0.033***	-0.020***	-0.052***	-0.052***	-0.015*
	(0.005)	(0.006)	(0.007)	(0.006)	(0.008)	(0.008)
Assessed Value		0.000**	0.000***		0.000***	0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
Log (Days on Market)			-0.037***			0.025***
			(0.002)			(0.002)
Log (Days from “Pending" to “Sold")			0.013***			-0.008***
			(0.001)			(0.002)
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
==Socio-Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Property Feature Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	116,423	86,306	52,869	116,423	87,164	52,533
R-squared	0.051	0.067	0.074	0.077	0.094	0.087
Adjusted R-squared	0.049	0.065	0.071	0.076	0.093	0.085

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is logged Zesimate-to-List ratio for Columns (1) -(3) and logged Zesimate-to-Sale ratio for Columns (4)-(6).

The OLS results are reported in Table 1.11. I compare the Zestimate with both the list price (Columns (1)-(3)) and the sale price (Columns (4)-(6)). The assessed home value, the time on the market, and the time between the pending sale and the final deal are controlled in various models. As we can see, the Zestimate home valuation predicted by algorithms doesn't fully reflect the white-premium in home owner's valuation. Based on the results in Column (3), on average, the logged Zestimate-to-list ratio is 0.02 lower for properties located in nonwhite neighborhoods (where $\%white = 0$) compared to similar properties sold in white neighborhoods (where $\%white = 1$). It indicates that even though the Zestimate is heavily influenced by the list price after the information becomes available [15], it doesn't fully incorporate the racial biases existing in the listing price. Similarly, the logged Zestimate-to-sale ratio is 0.01 lower for properties located in nonwhite neighborhoods (where $\%white = 0$) compared to similar properties sold in white neighborhoods (where $\%white = 1$). The difference is not as significant as the one observed for the logged Zestimate-to-list ratio. This finding is consistent with the previous conclusion drawn from Table 1.10 that the sale-to-list ratio is lower in those predominately white neighborhoods. It might be due to the attenuated racial biases in the Zestimate does affect the level of biases in the sale price, since buyers do use the Zestimate as a reference when making the final price decision.

Figure 1-6: Reverse Engineering: Model Comparison



Notes: In-sample and out-of-sample mean squared errors are reported here. 20% of the observations in each city are randomly selected into the holdout dataset for model evaluation.

To further investigate the racial biases in Zestimates, I made a few attempts at reverse-engineering the Zestimate by predicting the sale price using some of most popular prediction models. Most of the models are constructed with the property features used by Zillow.com

[15]See `https://www.inman.com/2018/08/08/whats-the-deal-with-zillow-changing-its-zestimates/`

and the city indicators.[16] Figure 1-6 presents the in-sample and out-of-sample goodness-of-fit for different models. First I fit a simple OLS linear regression model using the property features as predictors and report the in-sample and out-of-sample mean squared errors. Then variable selection and regularization are performed in a LASSO regression model with the same set of predictors. The LASSO model is further improved by adding the city indicators. Finally, I allow the heterogeneity in model parameters and fit a tree-based regression model with recursive partitioning. The data is stratified according to the city indicators and then separate regression models are fit to each stratum. As we can see from the figure, the goodness-of-fit measure doesn't improve significantly when the model becomes more complex. In addition, I train a neural network model which allows complex nonlinearities with the full set of predictors (property features and city indicators). It again doesn't show a better performance than the simple regression model. So the comparison here suggests that a simple regression model might be good enough to at least shed a light on how Zillow's prediction model works.

Table 1.12: Reverse Engineering: Racial Differences in Residuals

	(1)	(2)	(3)	(4)	(5)
Model	OLS	Lasso	Lasso	LM Tree	Neural Network
%White	9660.034**	9660.034**	10498.351***	10511.615***	9683.719**
	(3821.234)	(3821.234)	(3830.025)	(3820.765)	(3817.075)
Observations	53,176	53,176	53,176	53,176	53,176
R^2	0.023	0.024	0.031	0.025	0.034
Adjusted R^2	0.021	0.022	0.029	0.022	0.031

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is residuals.

I then regress the residuals ($P_{Sale} - \hat{P}_{Sale}$) from those models on the set of predictors including the list price, the assessed value, the time on the market, the time from pending to close, the month fixed effects, the property features, and the neighborhood socio-demographic controls including the white share of the population. The coefficients of the white share are reported in Table 1.12. Based on the results reported here, the gap between the predicted outcome based on the OLS model and the final sale price is $\$9,660$ smaller when the white share increases by 1. This difference significantly adjusts the existing racial biases in the

[16] See the Kaggle competition: `https://www.kaggle.com/c/zillow-prize-1`.

housing market by systematically underestimating the price markup on homes located in whiter neighborhoods.

Therefore thanks to the algorithmic prediction model it is calculated from, Zestimate is less in favor of those white-dominated neighborhoods than the actual list price. Combined with the fact that people who live in the diverse neighborhoods do follow the Zestimate, at least at the same level as their peers living in the white-dominated neighborhood, it is plausible that the Zestimate is one of the factors that moderate the lingering effect of “Redlining" in the home buying process.

1.7 Conclusion

In the previous analyses, I investigate how predictive algorithms change housing markets. Using data collected from Zillow.com and public records, I show that the Zestimate influences market participants’ decisions as a public source of market information. I also show that this estimate doesn’t fully reflect the racial biases in the housing market and thus it might have mitigated the home value gap between whites and nonwhites.

These results matter because sometimes policymakers and researchers might fear that the predictive algorithms may augment the inequality by reinforcing the privileged people’s advantages. However, the preliminary results here show that at least in our setting, the algorithm-powered home value estimates can actually mitigate the existing racial biases in the housing market by providing more neutral information. It can be generalized to other types of estimates which aim to provide summarized information to customers. Those statistics that summarize the unprejudiced market information may help customers understand the market better and make more objective decisions without putting more effects.

There are some limitations to the study. First, I do not know the browsing history of buyers and whether they did check the Zestimate, as well as the negotiation between buyers and sellers. Second, I do not have individual customer data on race and ethnicity for all the properties and instead focus on the neighborhood properties. Last, I have not considered the profitability of the information provider and whether they have incentives to provide less biased information.

1.8 Appendix

1.8.1 Pricing Game

This section consists of three parts. In the first part, I examine a strategic pricing game between two sellers and one buyer, and demonstrates the complementarity of the sellers' prices. In the second part, I incorporate sellers' information choices into the strategic pricing game, and use existing results in information economics to establish sellers' coordination motives in acquiring information about market demand. In the third part, I map the model into the housing markets and explain my empirical findings, that people rely more on the Zestimates when there are more private signals.

Strategic Pricing

Here I consider a two-stage game between two sellers and one buyer.[17] Each seller has a good for sale. Each seller's valuation for his good is normalized to 0. In stage 1, the two sellers simultaneously choose prices p_1 and p_2. In stage 2, the buyer observes p_1 and p_2, as well as his valuation for the two goods η_1 and η_2, drawn from some independent uniform distributions,[18] and then chooses which seller to buy from.

To highlight the key factors in the model, the buyer's choice is assumed to be binary: either she buys from seller 1 or she buys from seller 2. As a result, the buyer buys from seller 1 if

$$\eta_1 - p_1 \geq \eta_2 - p_2,$$

and vice versa. Seller i's utility is p_i if the buyer buys his good, and is 0 if the buyer buys from the other seller. Therefore, seller 1's utility can be written as the following function of p_1 and p_2:

$$U_1(p_1, p_2) \equiv p_1 \Pr_{\eta_1,\eta_2}(\eta_1 - p_1 \geq \eta_2 - p_2).$$

[17] It can be extended to model with two buyers and one seller and other market settings.

[18] The random variables η_1 and η_2 are interpreted as the difference in buyer's taste between the two products, which captures the differentiation across products. When the buyer's valuation is perfectly revealed to sellers, the two sellers will engage in a Bertrand competition.

and similarly,

$$U_2(p_1,p_2) \equiv p_2 \Pr_{\eta_1,\eta_2}(\eta_2 - p_2 \geq \eta_1 - p_1).$$

Following Bulow *et al.* (1985), I establish the strategic complementarity in sellers' price setting decisions by showing that $U_1(p_1,p_2)$ and $U_2(p_1,p_2)$ are both supermodular functions:

Proposition 1. *$U_1(p_1,p_2)$ and $U_2(p_1,p_2)$ are supermodular.*

Proof. Since U_1 and U_2 are symmetric, it is sufficient to show that U_1 is supermodular. For every $p_1^* > p_1'$ and $p_2^* > p_2'$, I show that

$$U_1(p_1^*,p_2^*) + U_1(p_1',p_2') - U_1(p_1^*,p_2') - U_1(p_1',p_2^*) > 0. \tag{1.1}$$

This is equivalent to:

$$p_1^* \int_{\eta_1} G_2(\eta_1 - p_1^* + p_2^*) - G_2(\eta_1 - p_1^* + p_2') dG_1(\eta_1) - p_1' \int_{\eta_1} G_2(\eta_1 - p_1' + p_2^*) - G_2(\eta_1 - p_1' + p_2') dG_1(\eta_1) \tag{1.2}$$

Given that $(\eta_1 - p_1^* + p_2^*) - (\eta_1 - p_1^* + p_2') = (\eta_1 - p_1' + p_2^*) - (\eta_1 - p_1' + p_2')$, and η_2 follows a uniform distribution,

$$G_2(\eta_1 - p_1^* + p_2^*) - G_2(\eta_1 - p_1^* + p_2') = G_2(\eta_1 - p_1' + p_2^*) - G_2(\eta_1 - p_1' + p_2') > 0.$$

Since $p_1^* > p_1^{'}$, we have (2) being strictly positive, which in turn implies that (1) is strictly positive. ∎

This result implies that the marginal benefit for seller 1 to increase his price increases with the price set by seller 2, in another word, sellers have incentives to coordinate when setting their prices. Such complementarities are well-known in monopolist competition models with sticky prices.

Strategic Pricing with Endogenous Information Acquisition

I expand the strategic price setting game in the first part by introducing an information acquisition stage before sellers choosing their prices. Suppose there are n informative signals $(s_1, s_2, ..., s_n)$ about the buyers' valuations $\overrightarrow{\eta} \equiv (\eta_1, \eta_2)$. I assume $s_1, s_2, ..., s_n$ are random variables that follow the same distribution and are independent conditional on $\overrightarrow{\eta}$.

The game proceeds in three stages. In stage 1, the two sellers simultaneously choose a subset of signals to observe, with $S_i \subset \{s_1, s_2, ..., s_n\}$ the set of signals observed by seller i. I assume that each seller faces a capacity constraint when acquiring information, in the sense that there exists $m \in \{1, 2, ..., n-1\}$ such that $|S_i| \leq m$ for every $i \in \{1, 2\}$. In stage 2, the two sellers simultaneously choose prices p_1 and p_2, after observing the realizations of the signals they choose to observe. Importantly, each seller cannot observe the other seller's informational choice, i.e., seller i cannot observe S_j. In stage 3, the buyer observes p_1, p_2, η_1, and η_2, and chooses between buying the item sold by seller 1 or from seller 2.

According to a well-known result in Hellwig & Veldkamp (2009) that establishes the strategic complementarity in sellers' informational choices, if players' actions in the price-setting stage are strategic complements, then players' informational choices are also strategic complements. In my setting, their result implies that when seller 2 observes signal s_i, the increase in seller 1's expected payoff by observing s_i strictly increases. It also suggests that seller i's pricing decision in the second stage becomes more responsive to signal s_i relative to the other signals he observe.

Application to Housing Market

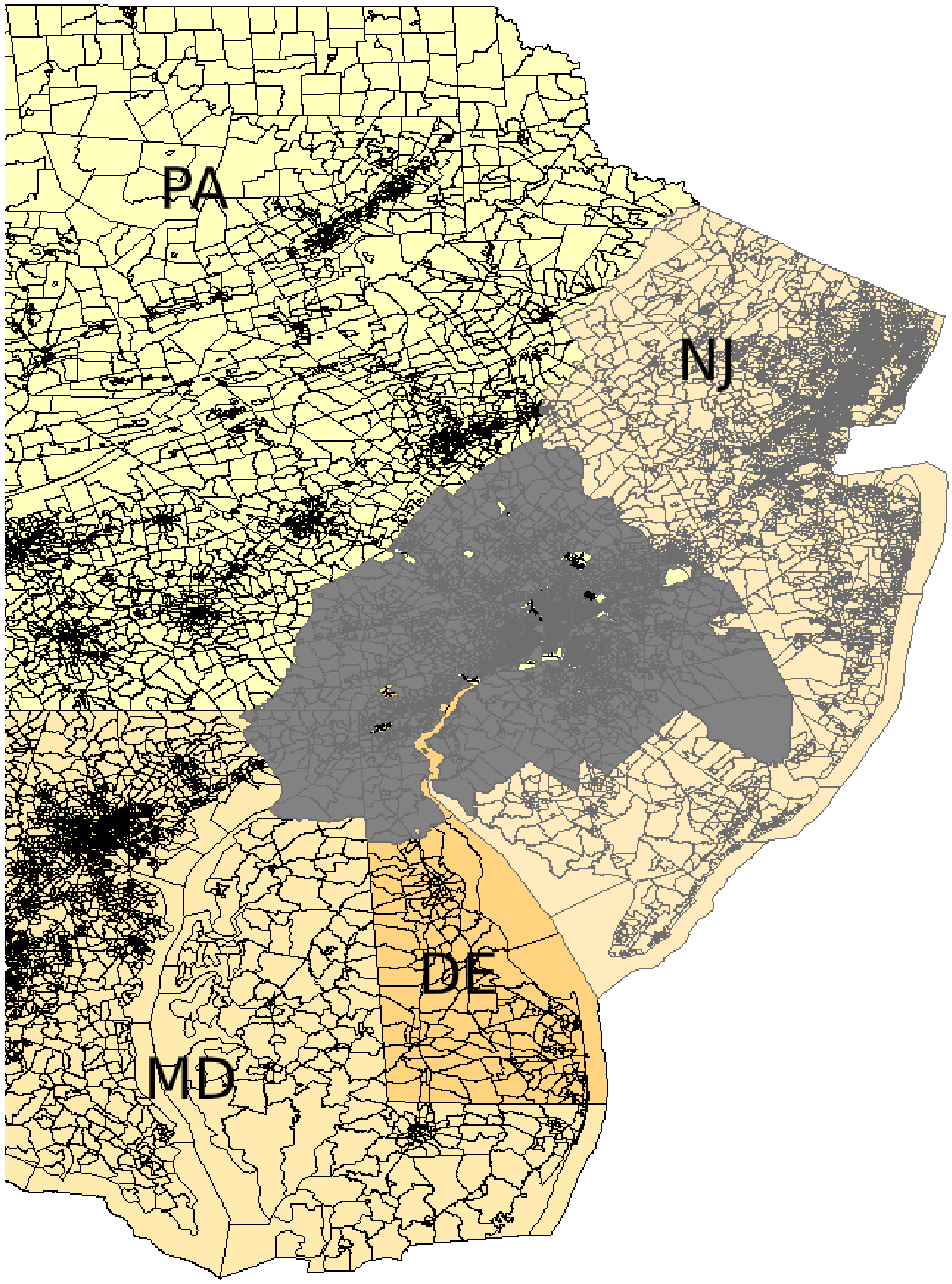
Consider a market with two sellers and one buyer. There are $n+1$ signals $\{s, s_1, s_2, ..., s_n\}$ available for the sellers to acquire, which are informative about $\overrightarrow{\eta}$ and are conditionally independent. s is the Zestimate (the public signal), and $s_1, s_2, ..., s_n$ are the prices (and conditions) of comparable sales (private signals). I assume that $s_1, s_2, ..., s_n$ are drawn from identical distributions.

Based on a modified three-stage game studied in the second part. In stage 1, each seller observes the public signal s for free, and choose to observe a subset of private signals

$\{s_1, s_2, ..., s_n\}$, while facing the constraint that the number of signals she can observe is no more than m. Stage 2 and stage 3 of the game remains the same as in the second part. Consider a symmetric equilibrium of the game in which each signal in the set $\{s_1, s_2, ..., s_n\}$ is observed by each seller with equal (ex ante) probability. Fixing each seller's capacity to process information m while increasing the number of available signals n, we know that for every $s_i \in \{s_1, s_2, ..., s_n\}$, the probability with which seller 1 observing s_i conditional on seller 2 observing s_i decreases. The conclusion in the second part then implies that as n increases, the seller's pricing decisions rely more on the public signal s compared to the other signals he can observe within the set $\{s_1, s_2, ..., s_n\}$.

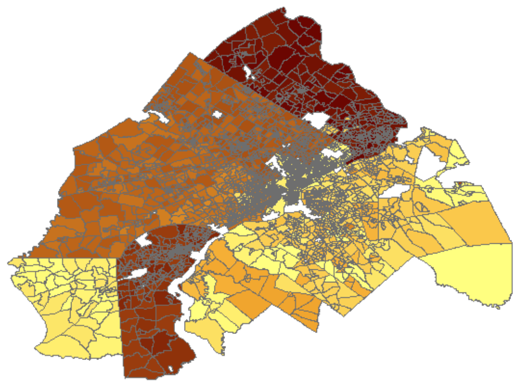
1.8.2 Figures

Figure 1-A1: Data Coverage

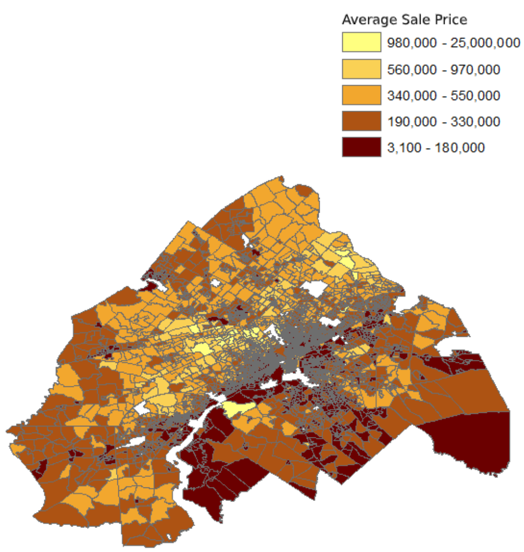


Notes: The census block groups included in the sample are shown in grey. It covers 4,108 census block groups in the Greater Philadelphia Area.

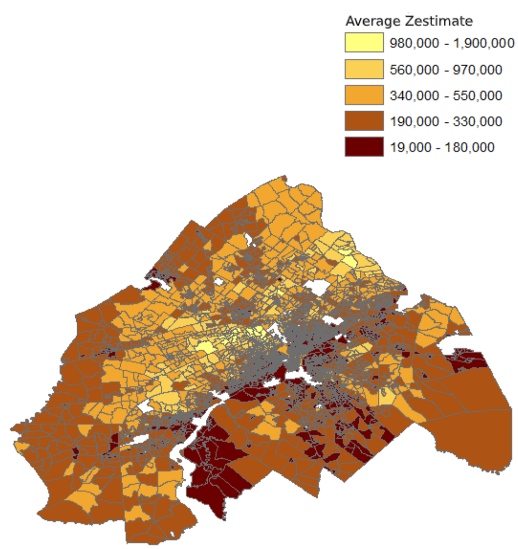
Figure 1-A2: Instrumental Variables: Exclusion Restriction



(a) IV: Months since Last Reassessment



(b) DV: Sale Price



(c) IDV: Zestimate

Notes: The average number of months since the last reassessment/revaluation are plotted for each census block in Panel A. The average sale price is plotted for each census block in Panel B and the average Zestimate is plotted for each census block in Panel C.

1.8.3 Tables

Table 1.A1: Summary of Covariates: Property Features

	Category	Covariates
S_i	Size	size
	Type	type
	Number of Bedrooms	number_of_bedrooms
	Number of Bathrooms	number_of_bathrooms
	Cooling	cooling_central, cooling_wall
	Heating	heating_electric, heating_gas, heating_wood, heating_air, heating_radiator, heating_baseboard
	Parking	parking_number, parking_carport, parking_detached, parking_attached
	Exterior Material	exterior_material_wood, exterior_material_vinyl, exterior_material_metal, exterior_material_brick, exterior_material_stone, exterior_material_cement, exterior_material_stucco
	Exterior Features	exterior_feature_deck, exterior_feature_porch, exterior_feature_garden, exterior_feature_lawn, exterior_feature_patio, exterior_feature_pool, exterior_feature_yard, exterior_feature_waterfront
	Views	view_park, view_mountain, view_water, view_territorial, view_city
	Water Sources	water_well, water_private
	Interior Flooring	interior_flooring_carpet, interior_flooring_hardwood
	Interior Heating	interior_heating_electric, interior_heating_gas, interior_heating_wood, interior_heating_air, interior_heating_radiator, interior_heating_baseboard
	Interior Appliances	interior_appliances_cleaning, interior_appliances_efficient, interior_appliances_stainless, interior_appliances_disposal, interior_appliances_efficiency, interior_appliances_hookups, interior_appliances_wall, interior_appliances_builtin, interior_appliances_island, interior_appliances_dishwasher, interior_appliances_washer

Table 1.A2: Summary of Covariates: Neighborhood Socio-demographic Characteristics

	Category	Covariates
L_k	Population	population
	Gender	male
	Age	age1, age2, age3, age4
	Race	white, black, asian, othersingle
	Mobility	samehouse, greatbostonarea, abroad
	Working Places	principalcity, workinarea, workincity, workoutcity
	Commute Methods	car, publictransportation,othermethods
	Time Leaving Home	before7, between7and9
	Commute Times	workers, time1, time2, time3, time4
	Children	householdwithchild, marriedparents, singlemothers
	Household Status	nonfamilyhousehold, familyhousehold_married, familyhousehold_female
	Household Size	size1, size2
	Education	highschooldiploma, somecollege, bachelor, graduateschool,averageyear
	Poverty	povertyratio_hou
	Household Income	medianincome
	Income Sources	withearning, withsalary, withselfemployment, withpublicassis
	Housing: Units	housingunit, occupiedhousingunits
	Housing: Rooms	tworooms, threerooms, fourrooms, fiverooms, sixrooms, sevenrooms, eightrooms, ninerooms
	Housing: Types	singlefamily, townhouse, mutiplefamily

Table 1.A3: The Effect of Zestimates on Final Sale Prices (Recent Sales)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Sales in Last Month						
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.955***	0.281	0.419	0.419	0.513*	0.598**
	(0.031)	(0.268)	(0.296)	(0.296)	(0.284)	(0.257)
List Price		0.670**	0.527*	0.527*	0.438	0.393*
		(0.267)	(0.298)	(0.298)	(0.277)	(0.232)
Assessed Value		-0.057**	-0.004	-0.004	-0.016	0.008
		(0.025)	(0.026)	(0.026)	(0.024)	(0.029)
Log (Days on Market)			-18268.811***	-18268.811***	-16708.126***	-13676.983***
			(3872.564)	(3872.564)	(3875.786)	(4187.314)
Log (Days from "Pending" to "Sold")			6705.032***	6705.032***	6833.305**	7489.551***
			(2487.337)	(2487.337)	(2747.382)	(2624.267)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	826	561	408	408	402	402
Panel B: Sales in Last Three Months						
	(1)	(2)	(3)	(4)	(5)	(6)
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	1.009***	0.685***	0.226**	0.226**	0.209*	0.208
	(0.014)	(0.123)	(0.104)	(0.103)	(0.112)	(0.155)
List Price		0.272**	0.712***	0.712***	0.719***	0.712***
		(0.118)	(0.098)	(0.098)	(0.101)	(0.135)
Assessed Value		-0.050***	-0.016*	-0.016*	-0.015*	-0.028**
		(0.009)	(0.009)	(0.009)	(0.009)	(0.013)
Log (Days on Market)			-17954.930***	-17896.563***	-17577.336***	-16699.422***
			(1715.162)	(1714.378)	(1712.232)	(2170.714)
Log (Days from "Pending" to "Sold")			7051.685***	7049.720***	6957.326***	6399.865***
			(1151.418)	(1151.035)	(1160.103)	(1271.241)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	10,238	6,483	4,225	4,225	4,188	4,173

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The regressions reported in Panel A use the purchases made in May 2019 for estimation while the regressions reported in Panel B use the purchases made from March to May 2019.

Table 1.A4: First Stages (Recent Sales)

Panel A: Full Sample						
	(1)	(2)	(3)	(4)	(5)	(6)
	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate
Months_Last_Update	228.281***	171.833***	171.089***	171.089***	178.617***	162.060**
	(75.280)	(36.251)	(58.787)	(58.787)	(64.663)	(63.281)
List Price		0.786***	0.798***	0.798***	0.762***	0.730***
		(0.041)	(0.068)	(0.068)	(0.078)	(0.087)
Assessed Value		0.246***	0.243**	0.243**	0.258**	0.202*
		(0.062)	(0.113)	(0.113)	(0.122)	(0.109)
Log (Days on Market)			-14261.522**	-14261.522**	-14640.705**	-14487.552***
			(6481.182)	(6481.182)	(5823.710)	(4841.351)
Log (Days from "Pending" to "Sold")			3399.035	3399.035	5605.201	2884.266
			(4379.311)	(4379.311)	(5119.440)	(4612.766)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	826	561	408	408	402	402
F statistic	15.04	72.90	48.68	48.68	44.99	32.11
R^2	0.018	0.971	0.975	0.975	0.978	0.983
Adjusted R^2	0.017	0.971	0.975	0.975	0.974	0.976
Panel B: Subsample without Missing Information						
	(1)	(2)	(3)	(4)	(5)	(6)
	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate
Months_Last_Update	221.339***	256.893***	102.498***	102.445***	92.046***	78.325***
	(36.837)	(96.481)	(17.705)	(17.722)	(15.770)	(13.981)
List Price		0.633***	0.835***	0.835***	0.816***	0.817***
		(0.126)	(0.022)	(0.022)	(0.023)	(0.024)
Assessed Value		0.430***	0.165***	0.165***	0.140***	0.102***
		(0.162)	(0.034)	(0.034)	(0.031)	(0.027)
Log (Days on Market)			-11947.337***	-11862.975***	-10970.462***	-10688.369***
			(1249.633)	(1251.189)	(1155.346)	(1198.952)
Log (Days from "Pending" to "Sold")			3138.087***	3137.065***	3257.768***	3247.184***
			(915.980)	(916.247)	(880.256)	(866.687)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
Socio-Demographic Controls	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	10,238	6,483	4,225	4,225	4,188	4,173
F statistic	372.74	1225.97	317.68	317.19	249.43	177.03
R^2	0.035	0.898	0.967	0.967	0.970	0.973
Adjusted R^2	0.035	0.898	0.967	0.967	0.970	0.972

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The regressions reported in Panel A use the purchases made in May 2019 for estimation while the regressions reported in Panel B use the purchases made from March to May 2019.

Table 1.A5: The Effect of Zestimates on Final Sale Prices (Alternative Specification)

Panel A: Full Sample						
	(1)	(2)	(3)	(4)	(5)	(6)
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	0.998***	0.507***	0.337***	0.335***	0.170**	0.094
	(0.008)	(0.064)	(0.048)	(0.048)	(0.075)	(0.083)
List Price		0.439***	0.599***	0.600***	0.715***	0.756***
		(0.060)	(0.045)	(0.045)	(0.061)	(0.066)
Assessed Value		-0.046***	-0.024***	-0.023***	0.017	0.005
		(0.004)	(0.003)	(0.003)	(0.018)	(0.017)
Log (Days on Market)			-12967.583***	-12778.470***	-13345.063***	-13641.824***
			(614.912)	(608.319)	(768.189)	(816.893)
Log (Days from "Pending" to "Sold")			5149.837***	5045.879***	5515.926***	5589.280***
			(364.152)	(359.637)	(426.878)	(435.927)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
City Fixed Effects	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	120,482	88,110	52,981	52,981	52,981	52,757
Panel B: Subsample without Missing Information						
	(1)	(2)	(3)	(4)	(5)	(6)
	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price	Sale Price
Zestimate_Sold	1.009***	0.391***	0.339***	0.337***	0.159**	0.094
	(0.004)	(0.049)	(0.050)	(0.050)	(0.077)	(0.083)
List Price		0.551***	0.597***	0.598***	0.725***	0.756***
		(0.046)	(0.047)	(0.047)	(0.063)	(0.066)
= Assessed Value		-0.037***	-0.024***	-0.024***	0.018	0.005
		(0.004)	(0.003)	(0.003)	(0.018)	(0.017)
Log (Days on Market)			-12924.107***	-12731.735***	-13411.288***	-13641.824***
			(631.588)	(624.879)	(782.676)	(816.893)
Log (Days from "Pending" to "Sold")			5146.118***	5042.462***	5552.183***	5589.280***
			(367.350)	(362.473)	(430.062)	(435.927)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
City Fixed Effects	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	52,757	52,757	52,757	52,757	52,757	52,757

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects.

Table 1.A6: First Stages (Alternative Specification)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate
Months_Last_Update	265.854***	163.572***	118.466***	118.341***	99.387***	87.254***
	(11.300)	(15.600)	(10.673)	(10.672)	(16.012)	(15.253)
List Price		0.748***	0.801***	0.800***	0.747***	0.733***
		(0.020)	(0.013)	(0.013)	(0.016)	(0.017)
Assessed Value		0.307***	0.223***	0.223***	0.285***	0.242***
		(0.031)	(0.022)	(0.022)	(0.026)	(0.026)
Log (Days on Market)			-10102.850***	-9828.774***	-8337.747***	-8217.160***
			(558.297)	(547.414)	(496.131)	(476.036)
Log (Days from "Pending" to "Sold")			3719.434***	3415.088***	3569.213***	3300.225***
			(370.919)	(370.504)	(354.101)	(343.639)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
City Fixed Effects	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	120,482	88,110	52,981	52,981	52,981	52,757
F statistic	3939.31	980.59	2397.47	2391.47	5088.59	1398.43
R^2	0.032	0.589	0.928	0.928	0.932	0.931
Adjusted R^2	0.032	0.589	0.928	0.928	0.931	0.931
Panel B: Subsample without Missing Information						
	(1)	(2)	(3)	(4)	(5)	(6)
	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate	Zestimate
Months_Last_Update	288.451***	116.164***	111.237***	111.105***	104.950***	87.254***
	(13.704)	(10.935)	(10.908)	(10.910)	(10.480)	(15.253)
List Price		0.804***	0.808***	0.807***	0.781***	0.733***
		(0.013)	(0.013)	(0.013)	(0.015)	(0.017)
Assessed Value		0.204***	0.205***	0.206***	0.190***	0.242***
		(0.022)	(0.022)	(0.022)	(0.021)	(0.026)
Log (Days on Market)			-10025.264***	-9774.361***	-9000.666***	-8217.160***
			(555.432)	(544.513)	(515.616)	(476.036)
Log (Days from "Pending" to "Sold")			3754.129***	3629.701***	3625.058***	3300.225***
			(366.551)	(365.620)	(354.691)	(343.639)
Month Fixed Effects	No	No	No	Yes	Yes	Yes
City Fixed Effects	No	No	No	No	Yes	Yes
Property Feature Controls	No	No	No	No	No	Yes
Observations	52,757	52,757	52,757	52,757	52,757	52,757
F statistic	2680.57	2293.42	2107.03	2100.93	1732.58	734.16
R^2	0.049	0.927	0.928	0.928	0.930	0.935
Adjusted R^2	0.049	0.927	0.928	0.928	0.930	0.934

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects.

Table 1.A7: Robustness Check: Excluding Irregular Transactions

Panel A: Interaction with Frequency of Similar Transactions Nearby ($t-6$)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Area	Block Group				City			
Measurement	per Capita		per Housing Unit		per Capita		per Housing Unit	
Zestimate_Sold	0.346***	0.241***	0.347***	0.239***	0.353***	0.238***	0.346***	0.238***
	(0.049)	(0.054)	(0.049)	(0.055)	(0.048)	(0.054)	(0.047)	(0.054)
# Transactions	-1283465.204***	-431214.423	-491146.326**	-144981.184	-928884.192**	-139023.185	-266809.429***	-59254.066
	(411187.553)	(381471.852)	(198683.449)	(161015.638)	(415417.446)	(178413.526)	(77349.750)	(54369.531)
Zestimate_Sold	4.234***	3.151**	1.776***	1.196**	2.441**	0.588	0.724***	0.251*
× # Transactions	(1.443)	(1.394)	(0.634)	(0.540)	(1.030)	(0.434)	(0.190)	(0.133)
List Price	0.584***	0.643***	0.581***	0.645***	0.579***	0.651***	0.586***	0.651***
	(0.045)	(0.047)	(0.045)	(0.047)	(0.046)	(0.048)	(0.044)	(0.048)
Assessed Value	-0.024***	-0.031***	-0.025***	-0.031***	-0.027***	-0.032***	-0.026***	-0.032***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Log (Days on Market)	-12846.030***	-12797.553***	-12839.669***	-12813.182***	-12627.432***	-12818.694***	-12743.050***	-12815.553***
	(622.769)	(608.088)	(619.970)	(607.690)	(644.643)	(612.469)	(618.371)	(609.689)
Log (Days from "Pending" to "Sold")	5120.181***	5223.773***	5131.294***	5234.423***	5088.570***	5241.073***	5117.284***	5241.056***
	(371.069)	(372.286)	(371.826)	(372.316)	(364.660)	(369.314)	(362.322)	(368.388)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	52,981	52,164	52,981	52,164	52,981	52,164	52,981	52,164

Panel B: Interaction with Frequency of Similar Transactions Nearby ($t-12$)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Area	Block Group				City			
Measurement	per Capita		per Housing Unit		per Capita		per Housing Unit	
Zestimate_Sold	0.346***	0.241***	0.347***	0.239***	0.353***	0.240***	0.347***	0.240***
	(0.048)	(0.054)	(0.048)	(0.055)	(0.048)	(0.054)	(0.047)	(0.054)
# Transactions	-794187.629***	-279127.294	-280325.184**	-83079.017	-497698.857**	-117310.936	-150053.503***	-46428.193
	(252223.437)	(229384.500)	(121488.689)	(100591.578)	(200713.777)	(100783.925)	(41717.756)	(31593.745)
Zestimate_Sold	2.689***	2.111**	1.053***	0.774**	1.322***	0.447*	0.415***	0.182**
× # Transactions	(0.868)	(0.834)	(0.382)	(0.332)	(0.490)	(0.240)	(0.102)	(0.076)
List Price	0.584***	0.643***	0.583***	0.645***	0.580***	0.649***	0.586***	0.649***
	(0.045)	(0.047)	(0.044)	(0.047)	(0.045)	(0.048)	(0.044)	(0.047)
Assessed Value	-0.024***	-0.031***	-0.024***	-0.030***	-0.026***	-0.032***	-0.025***	-0.032***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
Log (Days on Market)	-12873.928***	-12817.599***	-12871.119***	-12835.454***	-12668.480***	-12796.219***	-12751.256***	-12795.818***
	(621.320)	(607.848)	(620.390)	(608.736)	(629.564)	(609.994)	(615.125)	(608.208)
Log (Days from "Pending" to "Sold")	5129.426***	5211.745***	5140.891***	5224.837***	5107.555***	5232.039***	5126.767***	5232.833***
	(369.488)	(370.710)	(371.289)	(371.523)	(362.468)	(368.152)	(361.702)	(367.574)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	52,981	52,164	52,981	52,164	52,981	52,164	52,981	52,164

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The number of transactions (adjusted for the neighborhood size) is mean centered to allow easy interpretation of the main effects.

Table 1.A8: Robustness Check: Including Transactions in Other Segments

Panel A: Interaction with Average Deviation for Similar Transactions Nearby ($t-6$)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Area	Block Group				City			
Measurement	$Zestimate_t$		$Zestimate_{t-1}$		$Zestimate_t$		$Zestimate_{t-1}$	
Zestimate_Sold	0.327***	0.228***	0.327***	0.228***	0.335***	0.233***	0.335***	0.233***
	(0.048)	(0.055)	(0.048)	(0.055)	(0.048)	(0.054)	(0.048)	(0.054)
% Deviation	607.840	105.046	625.326	13.921	-3224.611**	-2328.539**	-3809.026**	-2818.507**
	(783.025)	(430.648)	(545.037)	(252.289)	(1321.064)	(997.843)	(1561.759)	(1181.979)
Zestimate_Sold	-0.006	0.001	-0.006*	0.001	0.005	0.012**	0.007	0.013**
× % Deviation	(0.004)	(0.002)	(0.003)	(0.001)	(0.007)	(0.005)	(0.007)	(0.005)
List Price	0.606***	0.660***	0.606***	0.660***	0.600***	0.657***	0.601***	0.657***
	(0.045)	(0.048)	(0.045)	(0.048)	(0.045)	(0.048)	(0.045)	(0.048)
Assessed Value	-0.024***	-0.031***	-0.024***	-0.031***	-0.023***	-0.032***	-0.023***	-0.032***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
Log (Days on Market)	-13193.512***	-13026.155***	-13158.554***	-12990.231***	-12973.777***	-12907.775***	-12937.282***	-12874.245***
	(610.837)	(616.487)	(610.593)	(616.226)	(619.673)	(608.580)	(618.863)	(608.071)
Log (Days from "Pending" to "Sold")	5253.598***	5316.434***	5241.672***	5303.866***	5147.748***	5250.566***	5135.670***	5239.306***
	(363.703)	(372.565)	(363.616)	(372.434)	(365.348)	(369.104)	(365.138)	(368.970)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	52,309	51,524	52,311	51,526	52,889	52,079	52,891	52,081

Panel B: Interaction with Average Deviation for Similar Transactions Nearby ($t-12$)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Area	Block Group				City			
Measurement	$Zestimate_t$		$Zestimate_{t-1}$		$Zestimate_t$		$Zestimate_{t-1}$	
Zestimate_Sold	0.327***	0.228***	0.327***	0.229***	0.329***	0.231***	0.329***	0.231***
	(0.048)	(0.055)	(0.048)	(0.055)	(0.048)	(0.054)	(0.048)	(0.054)
% Deviation	961.481	127.186	819.373	56.314	-4636.818**	-2581.407*	-5974.797**	-3684.205**
	(931.458)	(452.163)	(575.523)	(265.360)	(2055.943)	(1406.958)	(2645.534)	(1826.377)
Zestimate_Sold	-0.008*	0.000	-0.008**	0.000	0.001	0.008	0.006	0.012*
×% Deviation	(0.004)	(0.002)	(0.004)	(0.002)	(0.009)	(0.006)	(0.009)	(0.007)
List Price	0.606***	0.660***	0.606***	0.660***	0.605***	0.659***	0.605***	0.659***
	(0.045)	(0.048)	(0.045)	(0.048)	(0.045)	(0.048)	(0.045)	(0.048)
Assessed Value	-0.024***	-0.031***	-0.024***	-0.031***	-0.023***	-0.031***	-0.023***	-0.031***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
Log (Days on Market)	-13180.586***	-13018.715***	-13146.123***	-12982.399***	-13017.191***	-12931.334***	-12977.848***	-12898.397***
	(610.969)	(616.652)	(610.295)	(616.285)	(615.602)	(608.576)	(614.731)	(608.073)
Log (Days from "Pending" to "Sold")	5244.152***	5310.026***	5233.125***	5297.344***	5177.715***	5266.077***	5169.186***	5256.957***
	(363.594)	(372.422)	(363.334)	(372.255)	(364.037)	(369.471)	(363.827)	(369.377)
Month Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Socio-Demographic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Property Feature Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	52,351	51,561	52,353	51,563	52,901	52,090	52,903	52,092

Robust standard errors reported in parentheses are clustered at city×month level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is sale price. The time between listing and pending sale and the time between pending to closing are log-transformed to control their nonlinear effects. The accuracy of the Zestimate for a recent sale is calculated as $|Zestimate - Sale\_Price|/Zestimate$ and the average accuracy is mean centered to allow easy interpretation of the main effects.

Chapter 2

Challenges Facing Algorithm Decision Making: A Field Experiment on Repeated Marketing Campaigns

2.1 Introduction

The development of advanced machine learning algorithms has provided more tools in the process of customer relationship management (CRM). For example, unsupervised machine learning algorithms help marketers divide customers into groups and engage them with tailored content, and supervised machine learning algorithms enable marketers to predict values, such as purchase propensity and churn rate, based on pre-existing features and behavioral data, which in turn allows them to alter different aspects of the customer experience.

A quintessential application is recurring targeting promotions: Companies providing frequent services, such as Uber and Lyft, have moved quickly from simple geo-targeting to algorithm-driven targeting systems based on dynamic information, such as past behaviors and social networks, and utilize those dynamic targeting promotion campaigns to achieve personalized pricing and churn management.[1] And researchers have found that repetition of

This essay is based on the joint work with Yuting Zhu.

[1]See `https://www.theguardian.com/commentisfree/2018/apr/13/uber-lyft-prices-personalized-data`

marketing campaigns can make marketing campaigns more effective compared to displaying the content only once (Cacioppo & Petty, 1979, Batra & Ray, 1986, Anand & Sternthal, 1990). These firms usually send coupons to customers in a recurrent manner based on their behaviors in the previous periods and advanced algorithms are often involved in turning raw data into actionable recommendations.

Without observing previous marketing actions, most practitioners and methodologists have focused on prediction models that rest on the stationarity assumptions and tend to assume that consumers are unlikely to be able to learn the targeting rule and change their behaviors strategically. And algorithms are evaluated using their single-period performance. However, a massive theoretical literature suggests that firms should take the strategic behavior of forward-looking customers into account when making business decisions (for a review, see Fudenberg & Villas-Boas (2006)). The sophistication of customers also implies that the long-term performance of an algorithm might differ from its single-period performance if it is implemented recurrently and thus customers have chances to learn more about it.

Therefore, in this paper, we perform a field evaluation of this repeated-campaign strategy and compare both the single-period and long-term performance of targeting algorithms driven by two different linear models - an OLS model with only one explanatory variable, and a least absolute shrinkage and selection operator (Lasso) model with a larger set of predictor variables. We collaborate with a U.S. food delivery company and conduct a field experiment where targeted coupons are distributed by applying the same algorithms repeatedly. The results show that targeting rules that are driven by more complicated algorithms and utilize more information on customers' past behaviors perform better than the rules based on simpler algorithms. Moreover, we show that this difference only exists when the consumers have already been treated by the same policy a few times. We explore the purchase patterns and find that customers who are treated with simpler algorithms are more likely to switch between "purchase" and "not purchase" and not to make continuous purchases. It provides suggestive evidence that customers are able to learn the targeting rule through repeated campaigns, and they are sophisticated and forward-looking enough to game against the targeting rules.

These results should be important for practitioners. We suggest that companies would benefit from taking the strategic responses into account when designing algorithm-driven marketing strategies. It can be done by adopting more advanced machine learning techniques (such as adversarial learning, see Lowd & Meek (2005)), modelling the customer learning and strategic behaviors structurally, or simply changing the objective and focusing more on long-term goals if marketing actions are implemented in a recurrent manner.

2.2 Literature Review

Our paper is related to five distinct streams of literature. The first one is the literature that focuses on targeting across customers based on their past behaviors. Customers' past purchase patterns are usually used to predict their churn probabilities and then target incentives to those who are at risk to churn to induce them to stay (Lemmens & Croux, 2006, Neslin *et al.* , 2006, Schweidel *et al.* , 2011, Ascarza & Hardie, 2013, Godinho de Matos *et al.* , 2018; see Ascarza *et al.* , 2018, for a review). Theoretical works have also emphasized the important role of loyalty in coupon targeting (Shaffer & Zhang, 1995) and personalized promotion (Shaffer & Zhang, 2002). Moreover, purchase history has also been used in designing targeting strategies for new customers. For example, recommendation systems use the choice of a similar customer to recommend products to new customers (Ansari *et al.* , 2000, Moon & Russell, 2008) and to determine the face value of the customized coupon (Rossi *et al.* , 1996, Zhang & Krishnamurthi, 2004). Guided by this literature, we design our repeated targeting strategy by sending a coupon to the customers who are expected to make fewer purchases in the coming period. Our results contribute to the literature by evaluating the performance of the prediction-based churn management strategies in repeated campaigns, and it provides important managerial insights for marketers in the digital era.

The second stream is an emerging literature that discusses the application of advanced machine learning methods in targeting. Dubé & Misra (2017) use data generated from field experiments and apply machine learning algorithms to design third-degree price discrimination schemes based on customer features. Simester *et al.* (2020) compare the performance of model-driven methods and that of model-free methods in targeting and find that the model-driven methods perform best in general. We contribute to this literature by comparing not only the performance of different algorithms in designing one-shot targeting strategies but also extending the results to the repeated targeting campaigns and how they affect consumers' algorithm awareness and corresponding strategic behaviors.

Third, our paper is closely related to the literature on using dynamic optimization methods in repeated marketing campaigns. Gönül & Shi (1998) propose a structural dynamic programming model to design the optimal direct mail policy in a dynamic environment where customers understand the firm's mailing strategy and maximize long-term utility. Simester *et al.* (2006) solve the sequential catalog mailing policies using a dynamic optimization approach, which shows promise but under-performs for high-value customers. Zhang *et al.* (2014) find the optimal dynamic targeted pricing policy using a hierarchical Bayesian hidden Markov model in a B2B setting, and Hauser *et al.* (2009) apply a partially observable Markov decision process model to the website morphing problem. Instead of designing and evaluating dynamic policies, in this paper we consider the repetition of static algorithms. It is common in practice, especially for those firms whose engineering capabilities are limited.

Moreover, our findings contribute to the literature that studies consumer strategic behaviors. Fudenberg & Villas-Boas (2006) survey the theoretical literature on behavior-based price discrimination where the consumers are assumed to be strategic. Empirically, Zhang *et al.* (2018) provide some preliminary evidence for this assumption by discovering that customers tend to add more products into the shopping cart after receiving in-cart promotions. Misra & Nair (2011) show similar tendency in salespersons' decision-making process by analyzing how their incentives are changed by compensation ratcheting. This paper provides suggestive evidence that customer are forward-looking and play strategically against the rule when they are facing price discrimination based on past purchase behavior. Additionally, our results shed light on consumers' ability to learn the pricing policies from repeated observations.

Finally, our results also contribute to the recent debate on the relationship between big data and market competition. On the one hand, the legal literature suggests that big data can be a source of competitive advantages and monopoly power (Grunes & Stucke, 2015, Stucke & Grunes, 2016, Graef, 2017). Economic researchers also suggest that duopolists can benefit from competitive differential pricing allowed by big data (Belleflamme *et al.* , 2020) and data sharing (Choe *et al.* , 2020, Gu *et al.* , 2019). In addition, there are studies on how rich data affects firm performance (Bajari *et al.* , 2019, Shiller, 2020). For example, Shiller (2020) shows empirically that detailed web-browsing data will improve personalized prices and increase profits for those companies that have some pricing power. On the other hand, Lambrecht & Tucker (2017) use a strategic framework to evaluate the advantages brought by big data and suggest that big data can't contribute to a sustainable competitive advantage because it is neither rare nor exclusive. Moreover, Tucker (2019) show that digital data hardly ever contribute to antitrust concerns because it may have weakened network effects and the effects of switching costs. Our paper contributes to this debate by providing additional evidence that shows both the power of richer data and the ability of customers to play against the data-driven strategies.

2.3 Field Experiment

2.3.1 Setting

To implement the field experiment, we collaborate with a food delivery startup operating in 11 major U.S. cities (such as Atlanta, Boston, Chicago, New York City, and Seattle). One unique feature of their delivery service is that they not only offer the traditional delivery service for which dishes can be ordered à la carte but also allow users to order lunchboxes on weekdays for a small fixed delivery fee ($1, "lunch shuttle" service). Unlike the typical food delivery service that delivers food right to your door, those lunch boxes will be delivered to numerous pick-up points, and most of the pick-up locations are close to downtown office buildings, school buildings, or residential complexes. So drivers usually drop dozens of orders at the same locations, and it significantly reduces the cost of delivery for each single order. More than one dozen different dishes from different restaurants are available for ordering through the mobile App on every weekday. The menu and the selection of restaurants change every day but almost every dish appears at least once per week. Lunch shuttle orders can be placed from 11 am on the day prior to delivery to 11 am on the day of delivery[2], and the orders are delivered between 12 pm to 1 pm on every weekday.

In this study, we consider only lunch shuttle orders and users who have used the lunch shuttle service.[3] By focusing on the lunch shuttle service, this food delivery platform provides a perfect setting for the experiment: Demand is regular and easy to model because people have lunch once and only once every weekday, so the number of orders made in a given week can present most of the information on a customer's propensity to purchase and willingness to pay.

Table 2.1: Summary Statistics (Lunch Shuttle Orders)

	Number	Mean	Std. Dev.	Min	Max
Number of One-Time Users	1,268	26.6%			
Number of Multiple-Time Users	3,495	73.4%			
Number of Users	4,763	100.0%			
Number of Orders by One-Time Users	1,268	2.5%			
Number of Orders by Multiple-Time Users	50,030	97.5%			
Number of Orders	51,298	100%			
Average Number of Orders per One-Time Users		1	0	1	1
Average Number of Orders per Multiple-Time Users		14.3	21.3	2	260
Average Number of Orders per Users		10.8	19.1	1	260

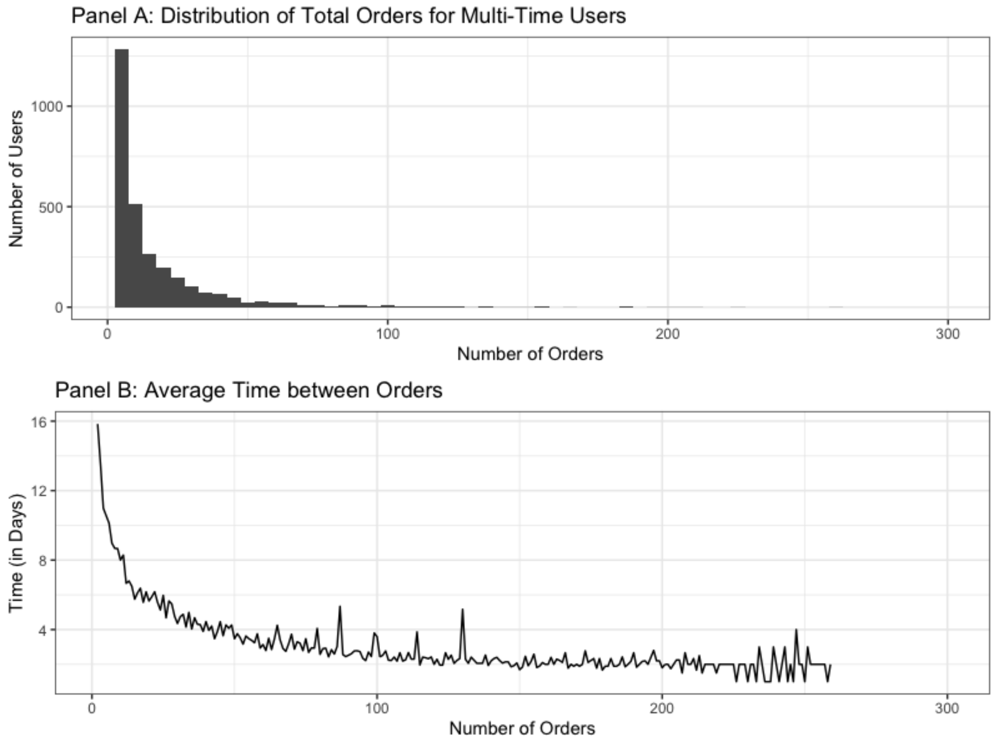
Table 2.1 presents several statistics which summarize users' purchase behavior for the lunch shuttle service. They are based on all the lunch shuttle orders placed between Jan 1, 2018 and Oct 31, 2018 in the Chicago area, which is the largest market for the company. 4,763 users made totally 51,298 orders during the 10-month period (about 1,300 orders per week). About 26.6% of the users only made one purchase during this time period (referred to as "one-time users" in Table 2.1), and the orders made by them make up about 2.5% of the total orders. Therefore, building long-term relationships with customers and increasing

[2] The cut-off time for lunch shuttle orders is either 10:30 am or 11 am in the Chicago area.

[3] Customers who use the lunch shuttle service and customers who use the traditional delivery service are basically two separated groups given the enormous differences in the nature of service.

customers' lifetime value will significantly boost the company's revenue.

Figure 2-1: Distribution of Total Orders and Average Time between Two Orders

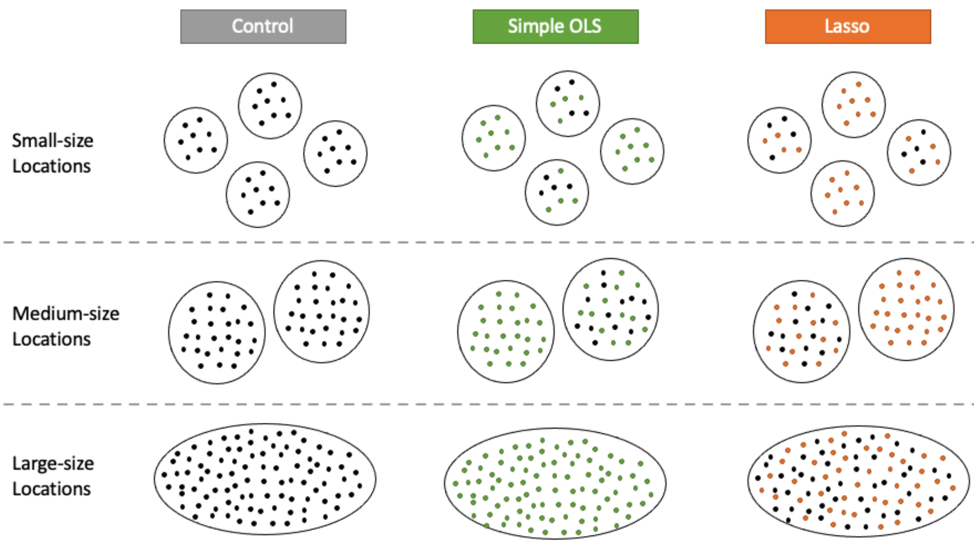


Notes: Panel A is based on the lunch shuttle orders placed by users who have placed more than one order during the 10-month period, and there is one observation per user per week. Panel B is based on the same set of users and shows the average time between the nth and (n+1)th orders over users.

For those users who have placed more than one order during the 10-month period (referred to as "multiple-time users" in Table 2.1), each of them placed on average 14.3 lunch shuttle orders during the time period (about 0.4 orders per week). The average expenditure per order is \$15.84 and there is no significant difference between orders placed by one-time users and orders placed by multiple-time users in the distribution of subtotals. Furthermore, we plot the distribution of the total number of orders for multiple-time users in Panel A of Figure 2-1. The distribution is right-skewed and there are only a few loyal users who ordered at least 50% of their lunches through the platform. In Panel B of the same figure, we plot the average time between orders over the number of orders have placed by the user. As we can see, the time elapsed between two orders drops significantly in the first few orders but becomes stable after that. Although the relationship is not causal and there are strong selection effects in those plots, those results encouraged the firm to strengthen customer relationships might be inviting customers who just “left" the platform for a few days to come back.

2.3.2 Experiment Design

Figure 2-2: Randomized Treatment Assignment With Stratified Cluster Sampling



Notes: Figure 2-2 illustrates the treatment assignment process. Users are grouped into clusters based on their pick-up locations. Then pick-up locations are divided into three strata (small-size, medium-size, and large-size) based on the number of users they have. Then simple random sampling is applied within each stratum, and one-third of the locations are selected into each treatment group. For those locations selected into the Simple OLS and Lasso groups, there is a 50% chance that all the users in the location will be treated by the corresponding treatment, otherwise, only half of the users will be treated. Those users who might receive coupons based on the two targeting rules are colored in green/orange in the figure.

We start our experiment on April 14, 2019, and the subjects recruited in our experiment are active lunch service users (who made at least one lunch shuttle order in the month prior to the experiment) in the Chicago area. We have one control group and two treatment groups – one for the targeting rule based on the simple OLS model and another for the targeting rule driven by the Lasso model. Each group includes about 200 users – There are 211 and 204 users in the two treatment groups, respectively, and 179 users in the control group. We use cluster sampling here due to the potential spillover effect caused by word-of-mouth learning.

There are 72 clusters and each of them consists of users picking up their orders at the same location. The clusters are then randomly selected into treatment groups with a stratification design to form the sample: We first divide the entire population of clusters into three strata by the size of those clusters (small-size, medium-size, and large-size),[4] and simple random sampling is then applied within each stratum. Furthermore, To generate variation in the intensity of word of mouth effects, we also vary the proportion of customers who are treated by the targeting rule across locations in the treatment groups. There are two arms: 50% of the users in the pickup location are in the treatment group or 100% of the users in the pickup location are in the treatment group. Each cluster in the treatment group has a 50% to be in the first arm. Figure 2-2 illustrates the randomized treatment assignment process.

Table 2.2: Randomization Check

	Control	Simple OLS	Lasso
Number of Orders	3.145	3.152	3.368
	(3.043)	(3.532)	(3.573)
Spending per Order	15.43	16.28	16.54
	(7.28)	(8.92)	(9.99)
Spending per User	45.48	51.39	54.78
	(42.25)	(73.90)	(64.00)
Observations	179	211	204

Notes: Standard deviations are reported in parentheses.

Given the complexity of our randomization process and the small number of users recruited in our experiment. We conduct a randomization check and report the results in Table 2.2. We calculate the average number of orders placed per user, the average amount of money spent per order, and the average amount of money spent per user during the one-month period before the experiment. Although users in the Lasso group placed more orders and spent slightly more on each order, none of those differences are significant.[5] Nevertheless, we still control the pre-experiment trends when analyzing the experiment results.

[4]Each of the three strata has 24 clusters. Small-size locations have less than 6 users, medium-size locations have 6-15 users, and large-size locations have more than 15 users.

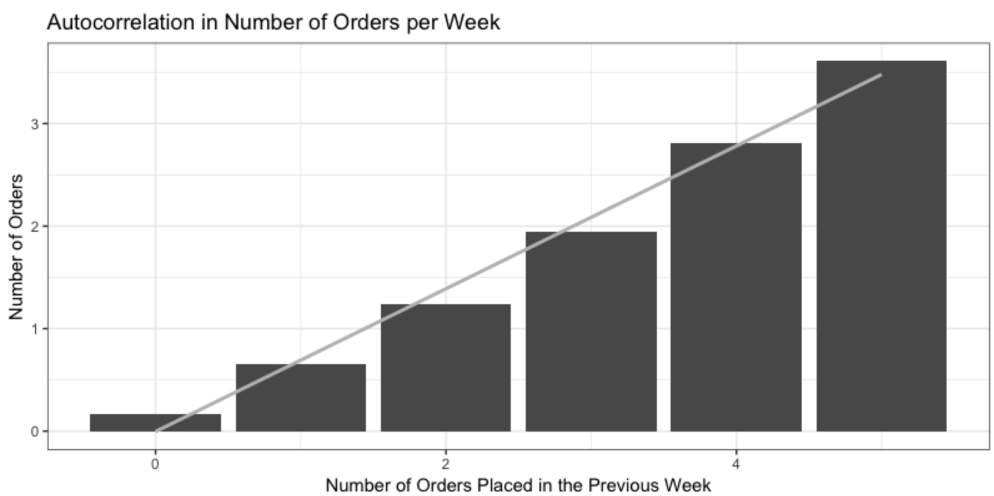
[5]The p-value of the two-sided t-test between the Lasso group and the other two groups is 0.46, the p-value of the two-sided t-test between the control group and the other two groups is 0.22, and the p-value of the two-sided t-test between the control group and the other two groups is 0.17.

On each Sunday, starting from April 14, 2019, users will receive a $5 coupon via email (see sample email message below) if they are selected by the corresponding targeting rule based on their purchase behaviors in the last week. The coupon expires in one week, and it can only be applied to lunch shuttle orders and is not transferable. The same campaign was run repeatedly for 6 weeks on the same group of users with the same targeting rule.

2.3.3 Targeting Rules

In this section, we discuss the targeting rules used in the experiment in detail. We have two different targeting rules which are designed based on two different prediction models – a simple OLS model and a Lasso model with more predictors. In those models, the number of purchases a user will make in week t is predicted based on the user's purchase behavior in week t-1. The two prediction models used in our study are:

Figure 2-3: Auto-correlation in Number of Orders



Notes: Figure 2-3 is based on the 45,267 observations between Jan 1, 2018 and Oct 31, 2018 for active users (who made at least one lunch shuttle order in the month prior to the observation) in the Chicago area. There is one observation per week per user. The average number of orders placed in a week is plotted over the the number of orders the users placed in the previous week.

- A simple OLS model with only one explanatory variable – the number of purchases made in the previous week. The number of purchases made in the previous week is selected here as the only variable here due to the strong auto-correlation in the number of orders. As shown in Figure 2-3, there is a strong linear correlation between the number of orders a user placed in week t and the number of orders she placed in week t-1.

- A least absolute shrinkage and selection operator (Lasso) model where more variables (including the number of orders placed in the previous week, the minimum/ average/ maximum expenditure in the previous week, the number of (unique) dishes/ categories/ restaurants per order in the previous week, the weekday and time of purchases, the customer tenure and race) are considered.

The reason why we choose to compare the simple OLS model with a Lasso model here is that both of them are linear models but differ in whether the predicted values (and the targeting policy) fully depend on one single variable – the number of purchases made in the previous week in this case.

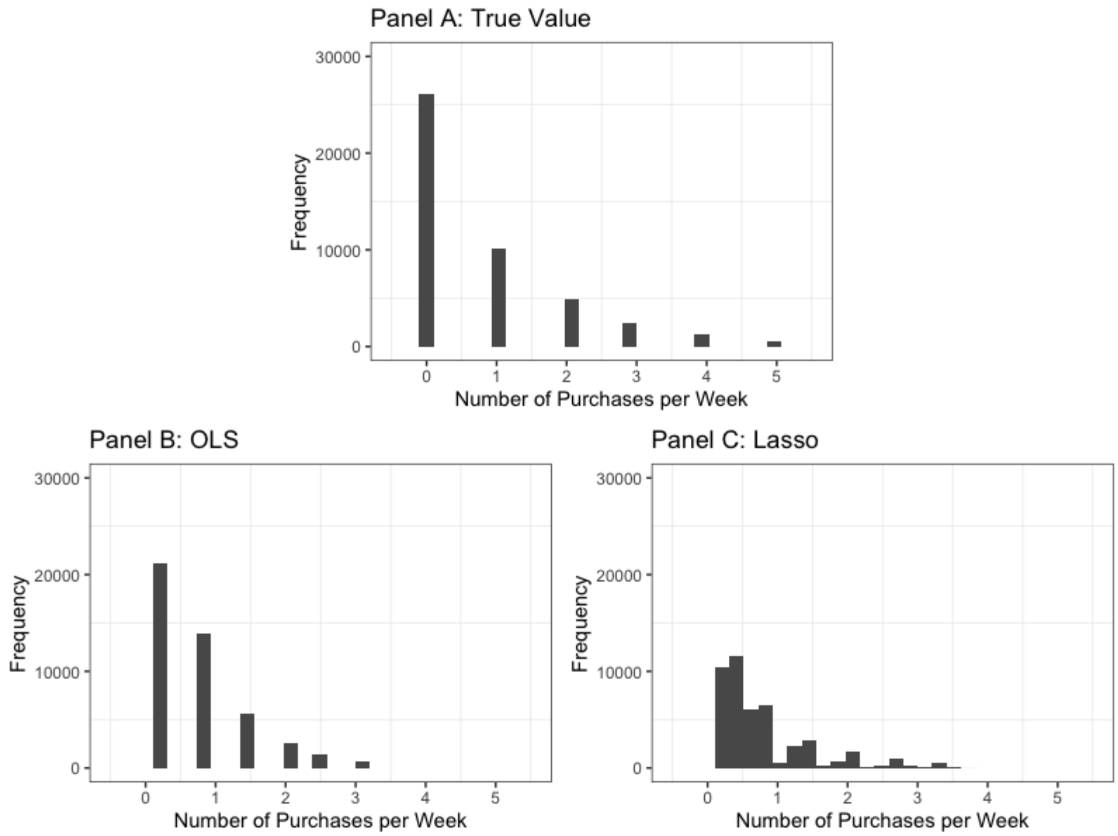
We train these two prediction models using the same training data set – 45,267 observations between Jan 1, 2018 and Oct 31, 2018 for active users (who made at least one lunch shuttle order in the month prior to the observation) in the Chicago area.[6] We use data collected on active users only to train the prediction models, in order to make sure that the training data set is consistent and comparable with the data facing in the targeting problem. In Panel A of Figure 2-4, we plot the distribution of total lunch shuttle orders per week per user. Panel B shows the distribution of the corresponding predicted values from the simple OLS model, and Panel C shows the results from the Lasso model where more variables are considered.[7]

Table 2.3: Model Comparison

	OLS	Lasso
MSE_Full	0.880	0.862
MSE_Training (0.7)	0.880	0.862
MSE_Test (0.3)	0.878	0.866

[6] There is one observation for each user each week.
[7] The Lasso model is fitted with cross-validation.

Figure 2-4: Predicted Values for Training Data Set (N=45,267)



Notes: Panel A is based on the 45,267 observations between Jan 1, 2018 and Oct 31, 2018 for active users (who made at least one lunch shuttle order in the month prior to the observation) in the Chicago area. There is one observation per week per user. Panel B and C are based on the 45,267 predicted values yielded the simple OLS model and the Lasso model, respectively.

Mean squared errors are reported in Table 2.3 and there is no significant difference between the two models in terms of that. To better compare the model performance, we randomly split the data set into two: a training set (includes 70% of the observations) and a test set (includes 30% of the observations). As we can see from Table 2.3, the Lasso model performs better on the holdout set than on the training set and the performance of the models becomes even closer.

Each Sunday, discount coupons are then sent out to users whose predicted number of lunch shuttle orders in the coming week is below a threshold ($\hat{n}$=0.3). This strategy is similar to the targeting policy used in customer re-engagement campaigns, where customers with high churn rates are targeted. So it should be fairly easy for users to learn the targeting rule generated by the simple OLS model– where whether a price discount will be available or not is fully determined by whether there was an order placed in the last week. With this group, we can easily identify whether consumers have algorithm awareness and are able to play strategically if the algorithm is easy to understand. The Lasso model can help us understand how sophisticated machine learning methods change this process by blocking the learning.

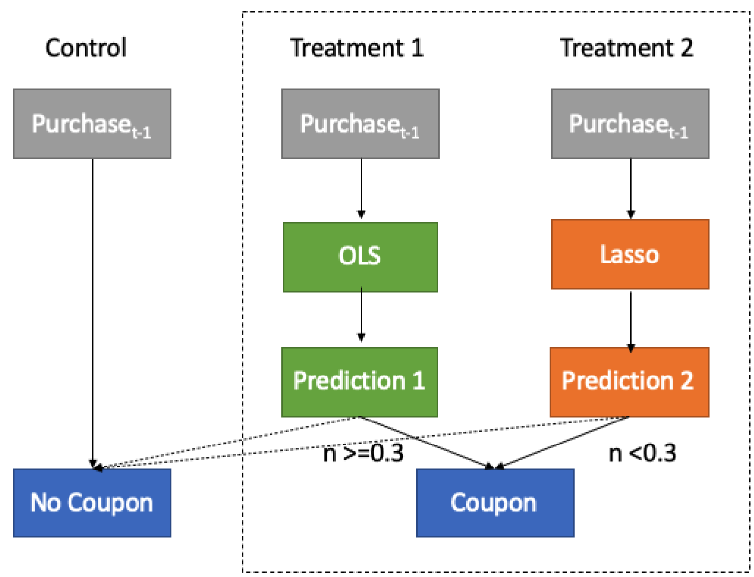


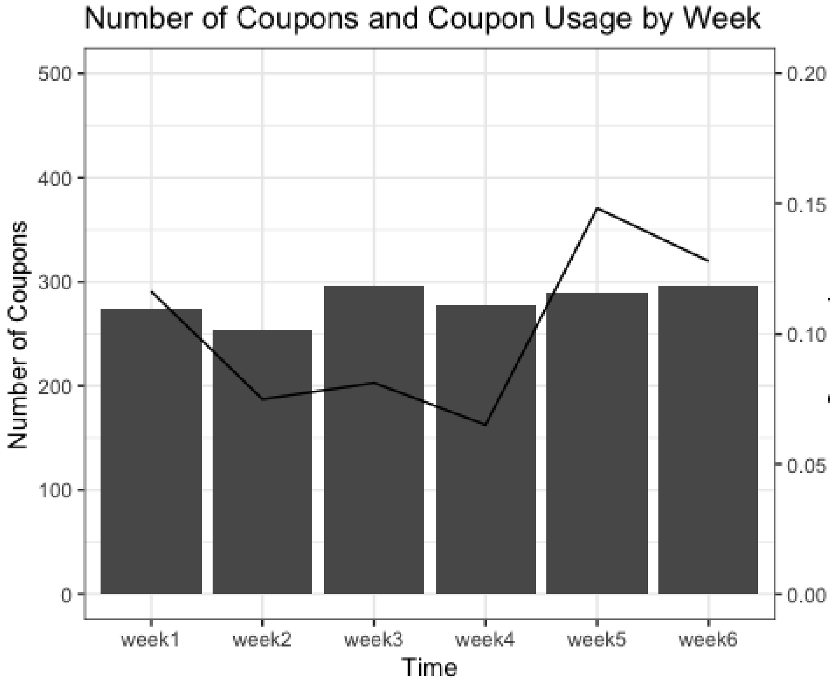
Figure 2-5: Experiment Design

Figure 2-5 summarizes the experiment design and targeting rules.

2.4 Main Effects

We run the experiment for six weeks (from April 14, 2019 to May 24l 2019) and record the purchase behavior for these 594 users over the six weeks, including when and what they purchased, and the coupon delivery and usage information. 1,689 coupons were sent out during the experiment. Approximately 11% of them were redeemed by users, it is higher than the average take-up rate for email coupon campaigns (this number is usually between 5%-7% for established online retailers). The coupon redemption rate varies across weeks and we can see a rising trend over the time in Figure 2-6. It is plausible that people are more

Figure 2-6: Coupon Delivery and Usage



Notes: Figure 2-6 is based on the 1,689 coupons sent out during the experiment. The vertical bars represent the number of coupons distributed via email in each week, and the markers on the line represent the coupon take-up rate in each week.

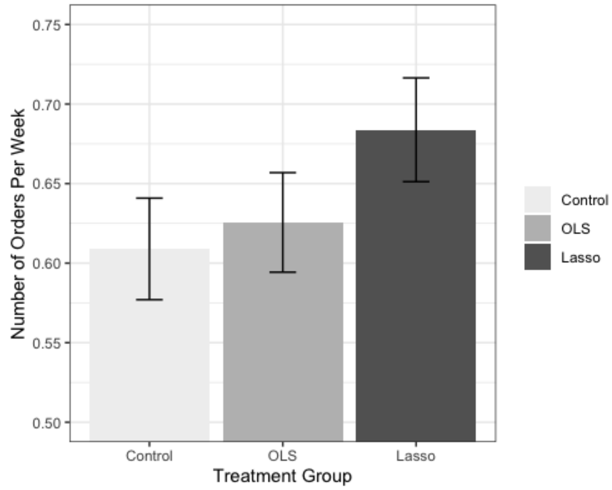
likely to start redeeming coupons after having received multiple ones. It suggests that, as we mentioned before, the repetition of actions can increase the effectiveness of marketing campaigns. This rate also varies across models – the Lasso group had a higher redemption rate (about 13%), compared to the 10% take-up rate for the simple OLS group.

2.4.1 Model-Free Evidence

We first use two different outcome measures to show the overall effects of these targeted coupon campaigns on customers: one is the average number of lunch shuttle orders placed per user per week, and another is the average (net) money expenditure on lunch shuttle orders per user per week.

Figure 2-7 shows the average number of lunch shuttle orders placed per user per week for control and two treatment groups. Without any interventions, these active users made on average 0.6 orders per week, which is consistent with the previous result from the training

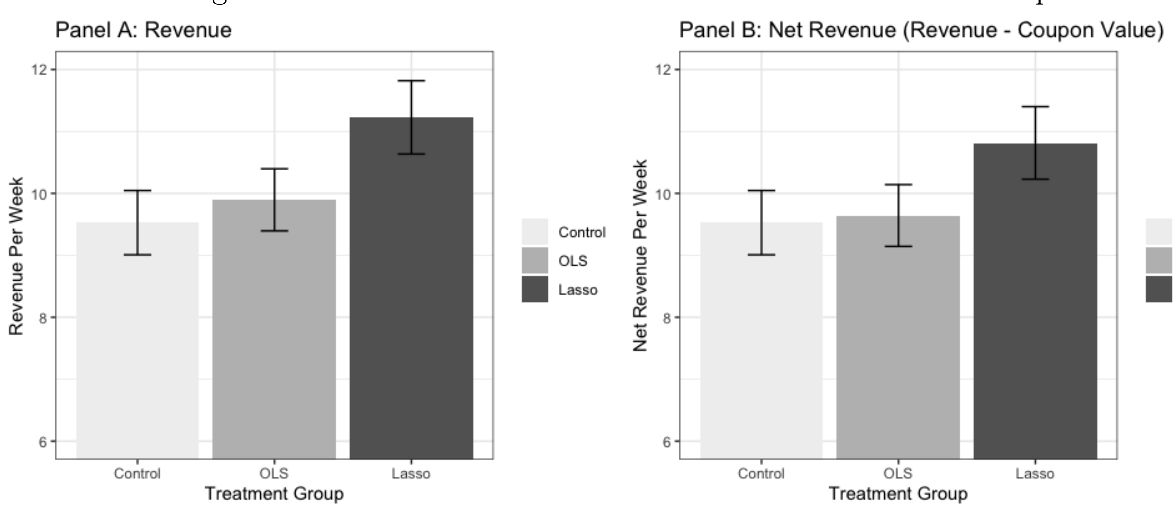
Figure 2-7: Number of Orders Across Treatment Groups



Notes: Figure 2-7 is based on 3,564 observations, and there is one observation per user per week.

data set. As we can see from the figure, there is a lift in the number of orders placed by users under both of the targeting rules, and this increase is larger and more significant for the group which is treated by the Lasso-driven rule. It suggests that even though the Lasso model slightly under-performs in predicting the churn rate, it actually increases the efficiency of this repeated customer retention campaign.

Figure 2-8: Revenue and Net Revenue Across Treatment Groups



Notes: Figure 2-8 is based on 3,564 observations, and there is one observation per user per week.

Panel A of Figure 2-8 displays the weekly expenditure results, and Panel B adjusts the weekly expenditure results by deducting the (applied) coupon value. They share the same pattern as the number of orders placed per user. An average user in the control group spends less than 10 dollars per week on the platform.[8] The coupons sent by the OLS-based targeting rule increase this number by about 0.5 dollars and this increase is negligible after the adjustment. The lift generated by the Lasso-based targeting policy is about three times larger than the one generated by the OLS-based campaign and is still significant even after deducting the coupon value from the revenue. It confirms the conclusion drawn from Figure 2-7. We will focus on analyzing the experiment data using the number of orders, rather than the revenue data, since the results are similar to each other.

Table 2.4: Summary Statistics by Location Type

	Commercial	Residential	School
Number of Orders	3.117	3.424	3.209
	(3.312)	(3.447)	(3.484)
Spending per Order	16.12	16.85	15.61
	(9.84)	(8.16)	(8.11)
Spending per User	50.58	54.07	48.73
	(72.28)	(55.63)	(54.17)
Coupon Take-up Rate	0.116	0.108	0.093
Observations	239	144	211

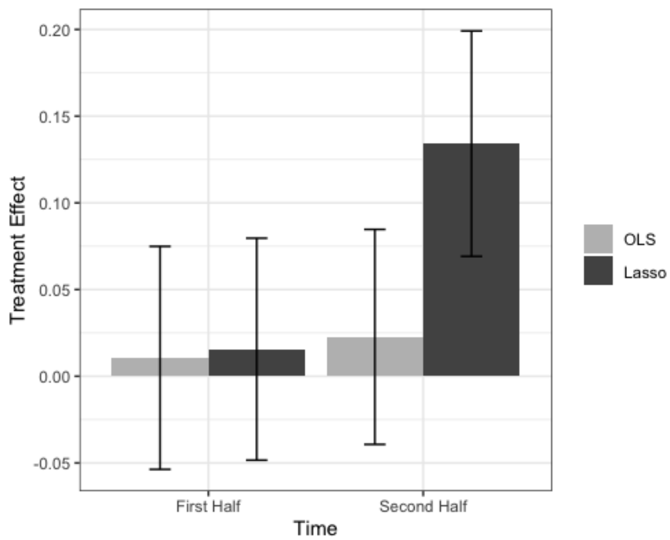
Notes: The first three rows are based on the orders made during the one-month period before the experiment. Standard deviations are reported in parentheses.

The effects can be heterogeneous across customers: Some customers may have a high response rate to the repeated coupon campaigns while others are not, because they are different in price sensitivity, outside options, and how familiar they are with the platform's promotion strategies. For example, users ordering from a commercial area are more responsive than others because they have more outside options and are more price-sensitive (see Table 2.4). We also report the pre-experiment purchase patterns by location type in Table 2.4. It seems that people made more orders and spent more money from home. Although

[8] Almost all bento boxes provided on the platform cost more than 10 dollars, so it makes sense that people spend on average 10/0.6=16.7 dollars for each order.

none of these differences are significant, we will control the location type in the regression analysis.

Figure 2-9: Short-Run and Long-Run Effects



Notes: Figure 2-9 is based on 3,564 observations, and there is one observation per user per week. Treatment effects are calculated as the average number of orders for the treatment group minus the average number of orders for the control group.

The treatment effects also change over the time. As shown in Figure 2-9, the treatment effects are larger in the second half (week 4-6) than those in the first half (week 3) for both groups. The potential explanations are the long-run effects of coupons (Anderson & Simester, 2004) and the advantages of repeated campaigns (Cacioppo & Petty, 1979). This difference is particularly large for the group of users where the coupons were generated using Lasso. It is the only group where the treatment effect is significantly larger than 0, where the coupons increase the average weekly number of purchases by almost 0.15. It shows the advantage of using a more advanced algorithm and including more variables into the model and might imply some level of learning and strategic behavior. We will provide more evidence and discuss it in detail in Section 2.5.

2.4.2 Reduced-form Results

In this section, we quantify the treatment effects using binomial logistic regression models. We choose the binomial logistic model because the number of orders a customer makes in a week follows a binomial distribution (N=5) if the purchase decisions are independent across days.

We model the number of purchases made in week t by user i as

$$N_{it} = \alpha + \beta_1 Simple\_OLS_i + \beta_2 Lasso_i + \gamma_t + \lambda S_i + \mu L_i + \eta N_{i0} + \zeta T_{it} + \theta C_i + \epsilon_{ikt},$$

where $Simple\_OLS_i$ and $Lasso_i$ are treatment indicators. $Simple\_OLS_i$ equals 1 if the user is assigned to the treatment group where the simple OLS-based targeting rule is used for coupon distribution and $Lasso_i$ equals 1 if the user is assigned to the treatment group where the targeting rule is driven by the Lasso model. β_1 and β_2 are coefficients of interest here and they show the estimated intent-to-treat effects of the treatments. γ_t controls the week fixed effects, and S_i is a vector of cluster (divided by stop size) indicators that control the cluster fixed effects. L_i is a vector of location indicators (commercial area, residential area, and on-campus pick-up points) that controls the fixed effects of different types of pick-up locations. N_{i0} is the pre-experiment order frequency of a user that takes users' different propensities to purchase into account. T_{it} is an indicator that equals 1 if the customer has been with the platform for more than 60 weeks at time t and controls the difference between loyal customers and new customers. Finally, C_i is an indicator which equals 1 if the customer is identified as Chinese from the last name – about 80% of the customers on the platform are Chinese and most of the restaurants on the platform are Asian restaurants.

The regression results are reported in Table 2.5, where the fixed effects and controls are gradually added into the model. As we can see, only the Lasso model increases the customer's propensity to purchase significantly – based on the results reported in Column (5), on average customers in the Lasso treatment group were $e^{0.109} - 1 = 11.5\%$ more likely to make a purchase compared to customers in the control group and $e^{0.109-(-0.007)} - 1 = 12.3\%$ more

Table 2.5: Intent-to-treat Effects

	(1) Logit	(2) Logit	(3) Logit	(4) Logit	(5) Logit
Simple_OLS	0.027	0.027	0.025	-0.021	-0.007
	(0.057)	(0.057)	(0.057)	(0.061)	(0.062)
Lasso	0.128^{**}	0.128^{**}	0.130^{**}	0.064	0.109^{*}
	(0.056)	(0.056)	(0.056)	(0.060)	(0.061)
Week Fixed Effects	No	Yes	Yes	Yes	Yes
Cluster Fixed Effects	No	No	Yes	Yes	Yes
Past Order Control	No	No	No	Yes	Yes
Tenure Fixed Effect	No	No	No	No	Yes
Location Fixed Effects	No	No	No	No	Yes
Chinese Fixed Effect	No	No	No	No	Yes
Observations	17,820	17,820	17,820	17,820	17820
AIC	13592.48	13591.23	13582.38	11840.65	11789.62

Notes: The population variable is days and a binomial logistic model is used. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

likely to make a purchase compared to customers in the simple-OLS treatment group. These results suggest that the Lasso-based targeting rule can increase the number of purchases and is more efficient than the rule driven by simpler algorithms.

To capture the trend of intent-to-treat effects over the time, we interact the treatment effect with time period indicators. Table 2.6 shows the results of the analysis where an indicator ($Second\_Half$) which equals 1 if the observation is from the second half (week 4-6) is adopted. Again, the fixed effects and controls are gradually added into the model for the results. The results again suggest that even though the effects are similar in the short term (none of the two algorithms generated a significant demand lift in the beginning), the Lasso prediction model beat the simple OLS model in terms of the long-term performance. Based on the results reported in Column (4), there is a declining trend in customers' purchase propensity: Compared to the first half, on average customers were $1 - e^{-0.215} = 19.3\%$ less likely to make a purchase in the second half of the time period. Meanwhile, the effect of the repeated coupon campaigns driven by the Lasso model increased by $e^{0.241} - 1 = 27.3\%$ in the last three weeks so that the purchase propensity of the users in the Lasso group actually

Table 2.6: Intent-to-treat Effects Over Time

	(1) Logit	(2) Logit	(3) Logit	(4) Logit
Simple_OLS	0.013	0.012	-0.031	-0.015
	(0.077)	(0.078)	(0.083)	(0.084)
Lasso	0.025	0.026	-0.052	-0.008
	(0.078)	(0.078)	(0.084)	(0.084)
Second_Half	-0.189**	-0.189**	-0.211**	-0.215**
	(0.084)	(0.084)	(0.088)	(0.088)
OLS $\times$ Second_Half	0.029	0.029	0.020	0.018
	(0.113)	(0.113)	(0.122)	(0.122)
Lasso $\times$ Second_Half	0.214*	0.214*	0.240**	0.241**
	(0.112)	(0.112)	(0.120)	(0.120)
Cluster Fixed Effects	No	Yes	Yes	Yes
Past Order Control	No	No	Yes	Yes
Tenure Fixed Effect	No	No	No	Yes
Location Fixed Effects	No	No	No	Yes
Chinese Fixed Effect	No	No	No	Yes
Observations	17,820	17,820	17,820	17,820
AIC	13588.87	13580.02	11838.73	11787.70

Notes: The population variable is days and a binomial logistic model is used. $Second\_Half$ equals 1 when the observation is from week 4-6. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

increased in the second half. It suggests that even though the complex targeting rule doesn't generate a better static targeting strategy, it may lead to consistent growth in the long run.

We run a robustness check where a vector of time indicators separates the treatment effects for the first two weeks, the middle two weeks, and the last two weeks. The results reported in Table 2.A1 show the same time trend as the previous ones: the baseline purchase propensity decreased over the time while there is an increasing trend in the performance of the Lasso-based targeting policy.

We have been focused on the intent-to-treat effects in the discussion above. However, the users in the Lasso group are more likely to receive a coupon (72.7% vs. 63.1%), and the difference between treatment groups can be driven by the uneven number of coupons sent by two targeting rules. To take it into account, in Table 2.7 we consider another set

Table 2.7: Average Treatment Effects on Treated

	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS
Coupon_OLS	0.023	0.023	0.021	0.016	0.029
	(0.072)	(0.072)	(0.072)	(0.059)	(0.060)
Coupon_Lasso	0.099	0.099	0.100	0.043	0.064
	(0.063)	(0.063)	(0.063)	(0.051)	(0.052)
Week Fixed Effects	No	Yes	Yes	Yes	Yes
Cluster Fixed Effects	No	No	Yes	Yes	Yes
Past Order Control	No	No	No	Yes	Yes
Tenure Fixed Effect	No	No	No	No	Yes
Location Fixed Effects	No	No	No	No	Yes
Chinese Fixed Effect	No	No	No	No	Yes
Observations	3,564	3,564	3,564	3,564	3,564

Notes: The population variable is weeks and the dependent variable is the number of orders placed within a week by a customer. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of estimators: average treatment effects on treated (ATTs). The ATTs are estimated using a set of two-stage least squares (2SLS) where the coupon usage is instrumented with the treatment assignment. Those estimators estimate the treatment effect of coupons per se (on compliers only). We replicate the results shown in Tables 2.5, 2.6, and 2.A1 with the average treatment effects on treated and report them in Tables 2.7, 2.A2, and 2.A3. Although most of the coefficients are not significant anymore (due to the change of population variable), all the previous results keep the same in terms of magnitude after taking the differences in compliance into account.

2.5 Mechanism: Customer Learning and Strategic Behavior

In this section, we explore a potential mechanism of why policies powered by more advanced algorithms have a better performance in repeated targeting campaigns, especially in the long run. We suggest that this phenomenon is driven by the fact that more advanced algorithms can prevent customers from learning the targeting policy and behaving strategically. Two pieces of suggestive evidence are provided.

2.5.1 Run-length Analysis: Purchase Pattern

We first look at the purchase patterns and investigate whether there are strategic behaviors. First of all, we cluster the purchase patterns by treatment group and find the representative patterns for each treatment group to simplify the question. The clustering method we use here is called dynamic time warping (DTW) distance for time series data (Sankoff & Kruskal, 1983, Berndt & Clifford, 1994). This method treats one time-series as one observation and the distance between two observations is measured using a mapping so that a specific distance measure between the coupled observations is minimized. It is able to find the similarity among observations that do not sync up perfectly.

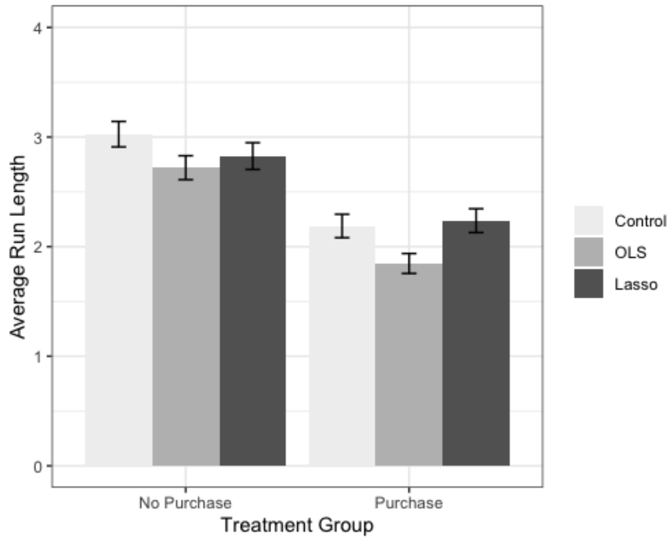
Table 2.8: Centroids from Time Series Clustering

(1) Simple_OLS		(2) Lasso	
Centroids	Observations	Centroids	Observations
(0, 0, 0, 0, 0, 0)	100	(0, 0, 0, 0, 0, 0)	81
(1, 0, 1, 0, 1, 0)	83	(1, 1, 0, 0, 1, 0)	54
(2, 2, 2, 3, 2, 2)	28	(1, 0, 1, 1, 2, 0)	69

Notes: The purchase and coupon delivery patterns (as a multivariate time series) are clustered into three clusters for each group. Only the representative purchase patterns are presented here.

We use a DTW-distance clustering method with global alignment kernels to divide the observations into three clusters for each treatment group. Table 2.8 presents the centroid and the number of observations for each cluster. As we can see from the table, there is a large group of users who almost never made a purchase during the six-week time period – 100 users in the simple OLS group and 81 users in the Lasso group. And there are also a few loyal customers who ordered every week and thus had never received a coupon in the simple OLS group. It is worth noting that more than one-third of people in the simple OLS group had a “perfect" strategic behavior. They rotated between “purchase" and “no purchase" so that they made and only made a purchase when there is a coupon. By contrast, it seems that users in the Lasso group were more likely to order consecutively and didn’t maximize the amount of discount they were able to receive.[9]

Figure 2-10: Average Run Length by Treatment Group



Notes: Figure 2-10 is based on 1,558 runs.

To further quantify this phenomenon using individual-level data, we test differences in the average length of run among groups. First, in order to better capture strategic behavior, we replace the number of orders placed in each week with a “purchase" indicator, which equals 1 if the user made one or more purchases in that week ("purchase") and equals 0 if the user didn't make any purchases in that week ("no purchase"). A run is defined as a succession of similar events preceded and followed by a different event. For example, we have (1, 0, 0, 1, 1, 0) as a sequence of observations. The first week is preceded and the last week is followed by a "no event". This sequence has four runs: first with a length of one, second and third with length two, fourth length one.

We calculate the run lengths for each customer's purchase pattern where only two values – “no purchase" (0) and “purchase" (1) – are considered. The average run length is plotted separately for “no purchase" and “purchase" by treatment group in Figure 2-10. The targeting policy driven by the simple OLS model not only reduces the average length of consecutive

[9]Although it is possible that users in the Lasso group would still receive a coupon even if they made purchases in the previous week, it is not the reason that drives the consecutive purchases. We take both purchase pattern and coupon delivery history (as a multivariate time series) into account in the clustering. Even users in the Lasso group didn't receive a coupon after purchases in those representative trends.

"no purchase" periods but also the average length of consecutive purchases. The repeated campaigns driven by the Lasso algorithm, on the other hand, only reduces the average length of consecutive "no purchase" periods but increases the chance of consecutive purchases. It suggests that even though the OLS model can successfully reactivate dormant customers and get them to come back, it is more likely to induce strategic behaviors and “train" some regular customers to increase their net consumption utility by avoiding consecutive purchases. More advanced algorithms like the Lasso model avoid this drawback by incorporating richer information to protect the targeting policy from being consumer learning.

Table 2.9: Treatment Effects on Run Lengths

	(1) Purchase	(2) Purchase	(3) No Purchase	(4) No Purchase
Simple_OLS	-0.347**	-0.232*	-0.312*	-0.334*
	(0.144)	(0.133)	(0.180)	(0.177)
Lasso	0.061	0.135	-0.219	-0.240
	(0.159)	(0.141)	(0.189)	(0.186)
Cluster Fixed Effects	No	Yes	No	Yes
Past Order Control	No	Yes	No	Yes
Tenure Fixed Effect	No	Yes	No	Yes
Location Fixed Effects	No	Yes	No	Yes
Chinese Fixed Effect	No	Yes	No	Yes
Observations	682	682	872	872
AIC	2517.80	2350.59	3661.74	3619.09

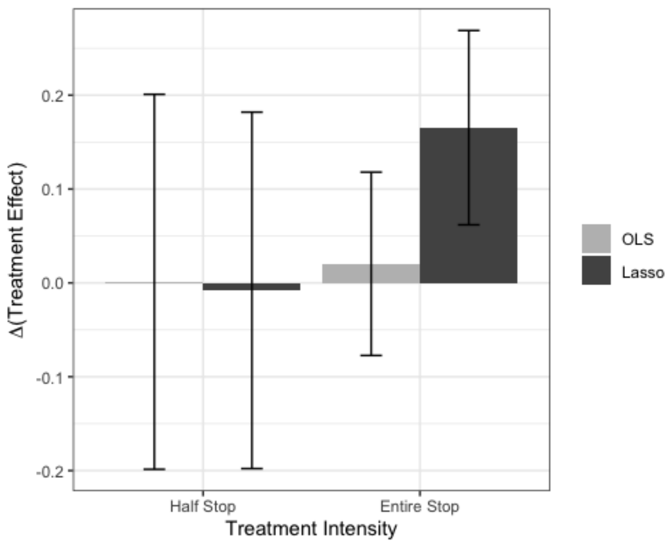
Notes: The population variable is runs and an OLS model is used. Robust standard errors clustered at user level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Simple regression models reported in Table 2.9 show the statistical significance of our results and confirm these observations with controlling user-level characteristics in the regressions.

2.5.2 Subgroup Analysis: Heterogeneous Learning Speed

To further investigate whether learning plays a role, we calculate the difference between the first half and second half separately for users in the stops where all users in the stop are in the treatment group (“Intensity =1") and users in the stops where only half of the users are in the treatment group (“Intensity=0.5’). The idea here is that the treatment intensity affects the speed of learning: For example, for two users picking up at the same location, if they are treated by the same rule and they talk to each other, the number of data points they observe will double, so both of them will learn more about the algorithm; however, if only one of them is treated by the rule and the other is in the control group, the latter one's observations will only confuse the former one and prevent her from learning the rule.

Figure 2-11: Short-Run and Long-Run Effects by Treatment Intensity



Notes: Figure 2-11 is based on 3,564 observations, and there is one observation per user per week. Δ(Treatment effect) is calculated as the treatment effect in the second half minus the treatment effect in the first half.

The increases in the second half from the first half are plotted in Figure 2-11 by treatment intensity. As we can see, the difference between the OLS group and the Lasso group is augmented when treatment is more intense.[10] This evidence supports the idea that the advantage of the Lasso algorithm is from reducing potential consumer learning and strategic behavior, since the treatment intensity affects people's ability to learn the rule.

Now we use regression models to re-test this mechanism. The results from a set of binomial logistic regressions are reported in Table 2.10. By comparing the regression results

[10]We note that the increase in the second half is larger for both of the treatment groups when the entire group is enrolled in the repeated coupon targeting campaign. A potential explanation here is spillover effects are amplified during the repetition. This is consistent with the theory of why repetition increases the efficiency of the marketing campaigns (Cacioppo & Petty, 1979).

Table 2.10: Intent-to-treat Effects Over Time: Subgroup Analysis

	(1) Intensity=1	(2) Intensity=0.5	(3) New Customers	(4) Established Customers
Simple_OLS	0.002	-0.050	0.113	-0.202*
	(0.092)	(0.131)	(0.124)	(0.120)
Lasso	-0.047	0.159	-0.066	0.031
	(0.092)	(0.127)	(0.126)	(0.114)
Second_Half	-0.215**	-0.218**	-0.125	-0.275**
	(0.088)	(0.089)	(0.138)	(0.115)
Simple_OLS × Second_Half	-0.054	0.191	-0.158	0.188
	(0.134)	(0.181)	(0.179)	(0.170)
Lasso × Second_Half	0.277**	0.147	0.281	0.171
	(0.130)	(0.182)	(0.180)	(0.163)
Cluster Fixed Effects	Yes	Yes	Yes	Yes
Past Order Control	Yes	Yes	Yes	Yes
Tenure Fixed Effect	Yes	Yes	No	No
Location Fixed Effects	Yes	Yes	Yes	Yes
Chinese Fixed Effect	Yes	Yes	Yes	Yes
Observations	14,700	8,490	9,390	8,430
AIC	9493.61	5721.39	5880.67	5898.01

Notes: The population variable is days and a binomial logistic model is used. Column (1) is based on the observations from the locations where the entire population is assigned to a treatment group and the observations from control group. Column (2) is based on the observations from the locations where half of the population is assigned to a treatment group and the observations from control group. Column (3) is based on the observations from the users have been on the platform for fewer than 60 weeks, and Column (4) is based on the observations from the users have been on the platform for more than 60 weeks. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, $p < 0.01$.

for the users in the stops where the entire population is in the treatment group (reported in Column (1)) with the results for the users in the stops where only half of the population is in the treatment group (reported in Column (2)), we show that the long-term advantage of the Lasso algorithm is positively correlated with the treatment intensity: This difference between the Lasso and OLS group is $e^{0.380} - 1 = 46.2\%$[11] when the treatment intensity equals 1, while this number decreases to $e^{-0.253} - 1 = -22.3\%$[12] when the treatment intensity drops to 0.5.

[11]It is calculated as the increase in the second half for the Lasso group minus the increase in the second half for the OLS group: $e^{0.277-(-0.047)-(-0.054-0.002)} - 1$

[12]Similarly to the previous result, it is calculated as $e^{0.147-0.159-(0.191-(-0.050))} - 1$

As discussed before, the learning speed is supposed to be faster when more people in the stop are receiving the same treatment, due to the increasing number of “accurate" information and decreasing number of “misleading" information shared within the group of users picking up their food at the same location (they are likely to be colleagues or classmates). Therefore, the differences between the two subgroups might indicate the possibility that the relative advantage of the Lasso algorithm comes from a more difficult consumer learning process.

To reinforce this argument, we also compare the results for new users who had been with the vendor for less than 60 weeks (reported in Column (3) of Table 2.10) with the results for established customers who had been with the vendor for more than 60 weeks (reported in Column (4) of Table 2.10). The results show that the long-term advantage of the Lasso algorithm is $e^{0.392} - 1 = 48.0\%$[13] for the new customers, larger than that for the established customers (-22.1%[14]). According to the principles of Bayesian learning, new customers with less experience and interactions with the platform are more likely to update their belief with the same number of observations and learn the correct rules, while more established customers tend to rely on their previous observations more. [15] Therefore, the fact that the advantages of more complicated algorithms are more prominent for new customers again supports our hypothesis and suggests that consumer learning plays a role in generating the observed differences.

We replicate the results shown in Table 2.10 with the average treatment effects on treated and report them in Table 2.A4, and the results keep the same direction.

2.6 Conclusion

In this study, we explore how different algorithms work in designing repeated marketing campaigns. By conducting a field experiment, we show that more advanced algorithms that utilize more information perform better than simpler algorithms in the long run, even though they may not yield better prediction results and out-perform the simpler ones in the short

[13]Similarly to the previous result, it is calculated as $e^{0.281-(-0.066)-(-0.158-0.113)} - 1$

[14]Similarly to the previous result, it is calculated as $e^{0.171-0.031-(0.188-(-0.202)} - 1$

[15]Customers occasionally received coupons if they were inactive for 15/30 days (only once in their lifetime) or if there were one-time special promotions.

run. We argue that this difference comes from the fact that advanced algorithms prevent customers from learning the algorithms and playing strategically against them. Supportive evidence is provided by exploring the heterogeneous effects and variation in the purchase patterns across subgroups.

These results are important to practitioners because they suggest that at least when algorithms are simple, consumers may have some level of algorithm awareness and are forward-looking enough to play strategically against the marketing strategies powered by algorithms. Marketers who are aware of it can benefit from taking the strategic responses into account. Those results can also be generalized to non-repeated marketing campaigns where consumers may still have a chance to learn and other types of targeting policies.

Although we conducted a field study, there are still limitations in this study. First, our data is limited. Due to restrictions in business operation, we only ran the repeated targeted-coupon campaign for six weeks. Moreover, the food delivery we collaborated with is a small startup with less than several dozen thousand users. These factors reduced the number of observations and the power of our analysis. Second, we adopted a repeated targeting strategy based on the repetition of static targeting rules, which are based on fairly bad targeting models: they are not trained using experiments and the threshold for sending a coupon is arbitrary. Modern tech companies might use more sophisticated targeting rules. Last, the evidence we provide is only suggestive, a detailed structural model can help further quantify the effect of consumer learning and strategic behavior. Notwithstanding these limitations, our study is a first step in understanding whether consumers have algorithm awareness and how they respond to algorithm-based marketing actions in the long run.

2.7 Appendix

2.7.1 Sample Email Message

Subject: $5 off your next lunch order, on us!

Content:

Enjoy $5 off your next lunch shuttle order!

Thank you for staying with us for another wonderful week! Looking forward to a new week? Here's a little pick-me-up to help you kick off the week. Simply enter D0F0AD0E (a personalized code) in the checkout page.

The coupon is already available for your account so just go ahead to use promo code for your next order.

But hurry - this coupon can be used once and expires on this Friday!

Love,

XXX Team

2.7.2 Tables

Table 2.A1: Intent-to-treat Effects Over Time (Alternative Independent Variables)

	(1)	(2)	(3)	(4)
	Logit	Logit	Logit	Logit
OLS	0.025	0.023	-0.019	-0.003
	(0.096)	(0.096)	(0.103)	(0.104)
Lasso	0.030	0.031	-0.048	-0.003
	(0.096)	(0.096)	(0.103)	(0.104)
Middle_Two	-0.015	-0.015	-0.017	-0.019
	(0.100)	(0.100)	(0.106)	(0.106)
OLS × Middle_Two	0.084	0.084	0.099	0.097
	(0.135)	(0.135)	(0.144)	(0.145)
Lasso × Middle_Two	0.053	0.053	0.062	0.061
	(0.136)	(0.136)	(0.146)	(0.146)
Last_Two	-0.187*	-0.187*	-0.209*	-0.214*
	(0.104)	(0.104)	(0.109)	(0.109)
OLS × Last_Two	-0.094	-0.094	-0.128	-0.132
	(0.142)	(0.142)	(0.152)	(0.152)
Lasso × Last_Two	0.250*	0.250*	0.283*	0.284*
	(0.138)	(0.138)	(0.148)	(0.148)
Cluster Fixed Effects	No	Yes	Yes	Yes
Past Order Control	No	No	Yes	Yes
Tenure Control	No	No	Yes	Yes
Chinese Fixed Effect	No	No	Yes	Yes
Observations	17,820	17,820	17,820	17,820
AIC	13584.98	13576.13	11832.83	11781.76

The population variable is days and a binomial logistic model is used. $Middle\_Two$ equals 1 when the observation is from week 3-4, and $Last\_Two$ equals 1 when the observation is from week 5-6. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A2: Average Treatment Effects on Treated Over Time

	(1)	(2)	(3)	(4)
	2SLS	2SLS	2SLS	2SLS
Coupon_OLS	0.024	0.022	0.016	0.031
	(0.134)	(0.134)	(0.110)	(0.111)
Coupon_Lasso	0.024	0.025	-0.034	-0.012
	(0.115)	(0.114)	(0.094)	(0.094)
Second_Half	-0.101	-0.101	-0.101*	-0.146**
	(0.067)	(0.067)	(0.055)	(0.070)
Coupon_OLS × Second_Half	0.020	0.021	0.021	0.020
	(0.144)	(0.144)	(0.118)	(0.118)
Coupon_Lasso × Second_Half	0.155	0.155	0.157	0.156
	(0.126)	(0.126)	(0.103)	(0.103)
Cluster Fixed Effects	No	Yes	Yes	Yes
Past Order Control	No	No	Yes	Yes
Tenure Fixed Effect	No	No	No	Yes
Location Fixed Effects	No	No	No	Yes
Chinese Fixed Effect	No	No	No	Yes
Observations	3,564	3,564	3,564	3,564

Notes: The population variable is weeks and the dependent variable is the number of orders placed within a week by a customer. $Second\_Half$ equals 1 when the observation is from week 4-6. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A3: Average Treatment Effects on Treated Over Time (Alternative Independent Variables)

	(1)	(2)	(3)	(4)
	2SLS	2SLS	2SLS	2SLS
Coupon_OLS	0.013	0.011	0.006	0.018
	(0.104)	(0.104)	(0.086)	(0.087)
Coupon_Lasso	0.020	0.021	-0.037	-0.016
	(0.091)	(0.091)	(0.074)	(0.075)
Middle_Two	-0.008	-0.008	-0.008	-0.009
	(0.082)	(0.082)	(0.067)	(0.067)
Coupon_OLS × Middle_Two	0.073	0.073	0.074	0.070
	(0.181)	(0.181)	(0.148)	(0.148)
Coupon_Lasso× Middle_Two	0.040	0.040	0.043	0.041
	(0.157)	(0.157)	(0.129)	(0.129)
Last_Two	-0.098	-0.098	-0.098	-0.100
	(0.082)	(0.082)	(0.067)	(0.067)
Coupon_OLS × Last_Two	-0.073	-0.073	-0.073	-0.075
	(0.179)	(0.179)	(0.147)	(0.147)
Coupon_Lasso × Last_Two	0.176	0.176	0.181	0.179
	(0.155)	(0.155)	(0.127)	(0.127)
Cluster Fixed Effects	No	Yes	Yes	Yes
Past Order Control	No	No	Yes	Yes
Tenure Fixed Effect	No	No	No	Yes
Location Fixed Effects	No	No	No	Yes
Chinese Fixed Effect	No	No	No	Yes
Observations	3,564	3,564	3,564	3,564

Notes: The population variable is weeks and the dependent variable is the number of orders placed within a week by a customer. $Middle\_Two$ equals 1 when the observation is from week 3-4, and $Last\_Two$ equals 1 when the observation is from week 5-6. Standard errors are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A4: Average Treatment Effects on Treated Over Time: Subgroup Analysis

	(1)	(2)	(3)	(4)
			New	Established
	Intensity=1	Intensity=0.5	Customers	Customers
Coupon_OLS	0.010	0.025	0.066	-0.085
	(0.091)	(0.131)	(0.115)	(0.132)
Coupon_Lasso	-0.042	0.087	-0.067	0.035
	(0.078)	(0.123)	(0.099)	(0.114)
Second_Half	-0.102*	-0.102*	-0.050	-0.144*
	(0.053)	(0.056)	(0.081)	(0.074)
Coupon_OLS × Second_Half	-0.019	0.159	-0.109	0.172
	(0.121)	(0.194)	(0.158)	(0.180)
Coupon_Lasso × Second_Half	0.177*	0.088	0.157	0.127
	(0.106)	(0.171)	(0.138)	(0.155)
Cluster Fixed Effects	Yes	Yes	Yes	Yes
Past Order Control	Yes	Yes	Yes	Yes
Tenure Fixed Effect	Yes	Yes	No	No
Location Fixed Effects	Yes	Yes	Yes	Yes
Chinese Fixed Effect	Yes	Yes	Yes	Yes
Observations	2,940	1,698	1,878	1,686

Notes: The population variable is weeks and the dependent variable is the number of orders placed within a week by a customer. Column (1) is based on the observations from the locations where the entire population is assigned to a treatment group and observations from the control group. Column (2) is based on the observations from the locations where half of the population is assigned to a treatment group and observations from the control group. Column (3) is based on the observations from the users have been on the platform for fewer than 60 weeks, and Column (4) is based on the observations from the users have been on the platform for more than 60 weeks. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 3

Does IT Lead to More Equal Treatment? An Empirical Study of the Effect of Smartphone Use on Customer Complaint Resolution

3.1 Introduction

All firms have to deal with angry customers. Anecdotal evidence suggests that vociferous customers attract more attention from firms and get their problems solved sooner, while customers who have equally serious complaints but are not good at advocating for themselves are usually ignored, and may eventually just leave without saying anything. A particular concern for organizations may be that customers' ability to advocate for themselves in a consumer complaint situation may also be related to underlying demographic factors, such as education.

In this study, we ask whether new communication technology mitigates unequal attention in resolution of customer complaints relative to the traditional phone call or letter. It is not clear *ex ante* whether new communication technologies should improve or worsen the lot of customers who are potentially less able to advocate for themselves. On the one hand, rich and educated customers are believed to be better at using technology, which would give them

This essay is based on the joint work with Catherine Tucker.

further advantages in complaint resolution. On the other hand, technologies may resolve the disadvantages facing less-educated communities and lead to fairer customer service by systematizing the communication.

We investigate this question using service performance data from customer complaint resolution in the public sector. We combine 364,189 Boston non-emergency public service operation records with demographic and socioeconomic census data. We find that complaints that originate in more highly educated census blocks, are more likely to be resolved quicker. However, we also find that the use of mobile information technologies improves the performance of customer service in the public sector and at least partially eliminates more educated customers' advantage in complaint resolution relative to people who submit complaints in neighborhoods with lower education levels, by providing a standardized communication tool. We present suggestive evidence that it is on occasions when these advanced digital tools are used to automate data and for more complex requests that apps are most effective at closing the gap between educated and less-educated customers.

An obvious concern about these findings is the endogeneity of mobile device use and how it might itself be related to education. To address this, we turn to an instrumental variables approach, where we use plausibly exogenous instruments which shift the ability of customers to submit complaints using mobile apps. The instrument we use captures the app use of city employees and the strength of the local cellphone signal. We present evidence that not only does this affect the ability to use the mobile application, but that also, due to the unusual topography and history of Boston, strength of local cellphone signal is not strongly correlated with the demographics of the local neighborhood. These instrumental variable results confirm our earlier findings.

We contribute to three distinct literatures. The first is the literature which explores the effects of complaint resolution. There is some evidence that customer complaint resolution is important for firm profitability. Satisfaction with how a complaint is resolved can have a positive effect on customer loyalty (Fornell & Wernerfelt, 1987, Andreassen, 1999, Tax *et al.* , 1998) and may evoke positive word-of-mouth behavior (Blodgett *et al.* , 1995).

However, there is far less empirical evidence regarding the process by which a firm can best resolve customer complaints. In general, research is either theoretical (Fornell & Wernerfelt, 1988), or based around qualitative frameworks – for example Davidow (2003) describes six dimensions of defensive marketing and summarizes studies about how each of them affects post-complaint responses. One exception is Homburg & Fürst (2005), which suggests that both having guidelines and a positive culture can help with consumer complaint resolution. We contribute to this literature by providing empirical evidence about the roles of technology in fighting potential inequality in the complaint resolution process.

Second, our study also contributes to a more general debate about the relationship between technology and inequality. The current literature focuses on labor supply, and mostly shows that using information technology in the workplace has been contributing to growing inequality because it complements the skills of the educated labor force (Acemoglu, 1998, 2002, Bresnahan *et al.* , 2002, Bartel *et al.* , 2007). This implies that more educated workers are likely to earn more due to the higher productivity (Black & Lynch, 2001, Bartel *et al.* , 2007, Bloom *et al.* , 2012) and the changed structure of firms (Bloom *et al.* , 2014) and industries (Tafti *et al.* , 2013), while many unskilled positions are replaced by new technologies. Our work differs from those papers by approaching the problem from the demand side and investigating the effect of IT for consumers. We show that, in contrast to the supply-side, information technology reduces inequality by providing a tool that substitutes the communication skills required for the resolution of complaints. This result builds on Morton *et al.* (2003), who find that the Internet has proved particularly beneficial to customers experiencing disadvantages in negotiating.

The final literature we contribute to is a literature studying how self-service technology affects the service performance and competitiveness of a business (Meuter *et al.* , 2000, Ray *et al.* , 2005, Jayachandran *et al.* , 2005, Dotzel *et al.* , 2013, Rust & Huang, 2014). Our work extends the literature by studying the use of those technologies in complaint handling and assessing the effect of self-service technologies on equality of treatment.

Our results are also important for managers who are interested in the emerging field of omni-channel marketing (Verhoef *et al.* , 2015). Much of the managerial excitement about omni-channel marketing has focused on how best to connect the customer experience of a firm's promotional marketing communications across mobile, other digital and offline channels. However, our paper also highlights the importance of an omni-channel approach for service resolution. In particular, it suggests that firms should consider directing portions of their customer-base towards channels which have the capacity to automate complaint submission in order to ensure effective resolution of service-issues. This supports recent movements by firms to expand their omni-channel messaging efforts towards apps for the purposes of improving consumer support. As a recent article about customer support platform Zendesk stated "[This] underscores the big impact that messaging apps are making in customer service. While phone and internet are massive points of contact, messaging apps is one of the most-requested features Zendesk's customers are requesting, 'because they want to be where their customers are."' [1] Other firms such as Hilton and Uber are also emphasizing increasing customer support services via their apps.[2] Our paper to our knowledge provides some of the first evidence on the efficacy of omni-channel approaches to providing customer support resolution.

3.2 Boston 311 Service

As the largest city in New England and the 23rd largest city in the United States, Boston has an estimated population of 667,137 distributed over an area of 89.6 square miles.[3] To provide better and more convenient public service, the city of Boston operates a multichannel system for non-emergency public service (311 constituent service) requests. All 311 service records since Jul 01, 2011 are available to the public on the website of the Boston city council.[4]

[1] See *Zendesk Acquires Smooch, Doubles down on Support via Messaging Apps like WhatsApp*: https://techcrunch.com/2019/05/22/zendesk-smooch/

[2] See *Hilton's HHonors App Redesign Unlocks Mobile Chat, Digital Key Expansion* (https://www.mobilemarketer.com/ex/mobilemarketer/cms/news/strategy/23419.html) and *Are You Providing A Frictionless Customer Experience?* (https://www.forbes.com/sites/shephyken/2019/06/09/are-you-providing-a-frictionless-customer-experience/\#1f8594fd4b8c)

[3] See *QuickFacts*: https://www.census.gov/quickfacts/table/PST045215/2507000.

[4] The data is made publicly available online on the city government's open data site as part of a commitment to increase transparency in government. However, sensitive personal information of request senders has been removed to protect individual privacy.

The dataset has been studied in other disciplines such as sociology and communications (Clark *et al.* , 2013, Buell *et al.* , 2017). However, the focus of this prior research has been on who adopts the 311 mobile application, rather than considering how the use of different technologies affects actual complaint resolution time.

A typical complaint is street cleaning, an abandoned shopping cart that needs to be removed, or snow clearing that has not been done thoroughly. Though we recognize that these are data from a governmental organization, we believe that the nature of the complaints, which are mainly focused on services that were inadequately provided (and the channels used to resolve them) are similar to a traditional commercial service-based organization.

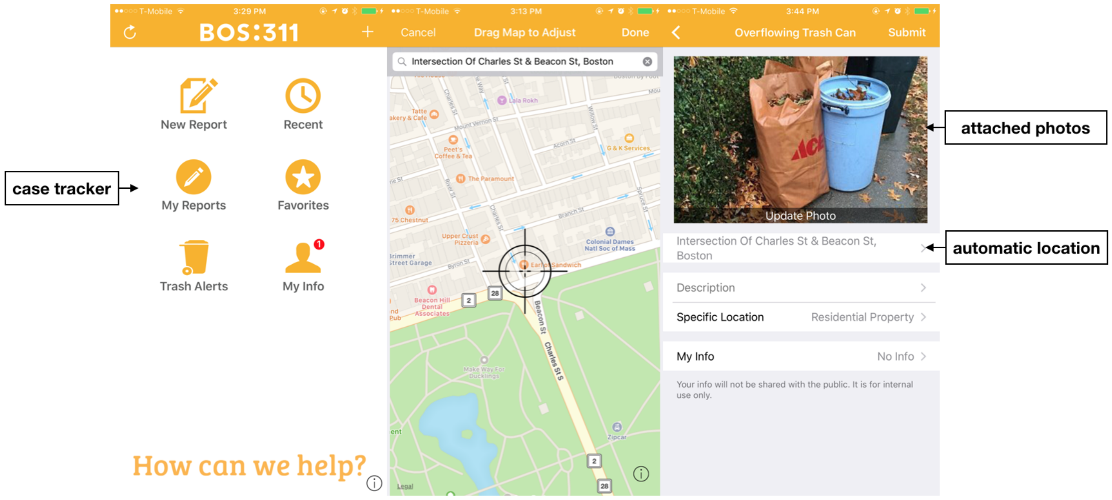
This complaint resolution service can be accessed in four ways:

- Phone Call
- Online Self-Service Website
- Mobile Application
- Social Media

Customers using a phone call access a 24-hour hotline (3-1-1 and previously 617-635-4500), where city workers take the call and log the service request into the computer system that routes requests and keeps records. This is the traditional means of submitting complaints. However, a unique feature of using the phone channel is that the interactive communication between representatives and reporters might make the accuracy of description depend on the oral communication skills of the person submitting the complaint and the extent to which the person taking down the complaint makes efforts to record the complaint completely.

The self-service website (`http://www.cityofboston.gov/311`) allows people to report non-emergency issues without help from a representative by filling in contact information, the location of the issue and a brief description of the request. After a successful form submission, requesters receive a tracking ID, with which they can check the status of their cases on the same website. A desktop computer is necessary for sending the form out.

Figure 3-1: BOS:311 App



Notes: The homepage of BOS:311 app (with navigation menu) is shown on the left. The page of case submission is shown on the right. A screenshot of the interactive map where people overwrite the auto-filled location by dragging a marker to the desired place is shown in the middle.

In contrast, the mobile application is not tied to a location. It was referred to as the 'Citizens Connect App' when it was first launched in 2009,[5] and now people can download it for free on iOS or Android as the BOS:311 App. Figure 3-1 shows a screenshot of the app. Three salient features make the app more technology- and data-intensive than the other channels for submitting complaints. First of all, high-quality photos can be taken and uploaded together with the description, which helps city employees obtain a better and quicker understanding of the complaint, especially where the description is not very clear. Second, the app can use GPS in mobile devices to locate the case and fill out the address automatically. This provides a more accurate address and avoids spelling mistakes in the address input. Third, people are able to track the status of their cases anytime and anywhere with their mobile devices, easing re-communication regarding the same case.

3.3 Data Description

We have two sources of data in this study: Public service operation records from the city of Boston and demographic and socioeconomic census data from the U.S. Census Bureau.

[5]We talked to a city employee and confirmed that the aim of this launch was to better serve citizens, and that the introduction was not to their knowledge influenced by any technological constraints.

Both of these are public datasets which are downloadable online.

3.3.1 311 Data

The public dataset on the website of Boston city council includes detailed information on each case opened after Jul 01, 2011: The open date/time, whether the case is still open or closed and the close date, reason, and result if it is closed, the completion time and the on-time status (the cut-off time for an "on-time" completion has been reported as target completion time for some but not all cases), the source of the case, the case type, the party responsible for the case, whether a photo is attached, the address, the latitude and the longitude of the case location, and the various districts or neighborhoods the case is within.

Since there was a major transition in the system – including changes in the website design, the name of the mobile application and the phone number, as well as some technological upgrades in the internal computer system – made on Aug 11, 2015, we use only records that were opened from Jul 01, 2011 to Aug 10, 2015. There are 603,694 cases during this period. 408,503 of them were generated by citizens, and the other 27.2% by city employees. To study the efficiency of complaint resolutions, we calculate the completion time for each record and discard 41,951 open cases. Most of those open cases are general complaints or complaints that can't be fixed with simple actions. We exclude those open cases because the overdue status here doesn't indicate that no proper attempt has been made punctually. We also exclude 35,082 internal cases logged by employees after the complaint was resolved, 15,346 duplicate cases, 15,022 invalid cases generated by errors, 14,654 cases for general comments with which no case location is associated, 9,405 cases closed administratively, and 558 cases with incomplete completion time.

Source and type of cases

For the remaining 464,683 cases in our dataset, the use of information technology is indicated by the source of case — citizens can submit the external request via the constituent call, the self-service website or Citizens Connect App, while city workers can report the internal case through the traditional system (employee generated) or City Worker App.[6] We don't study complaints submitted on Twitter because there are only 7 cases in our data.

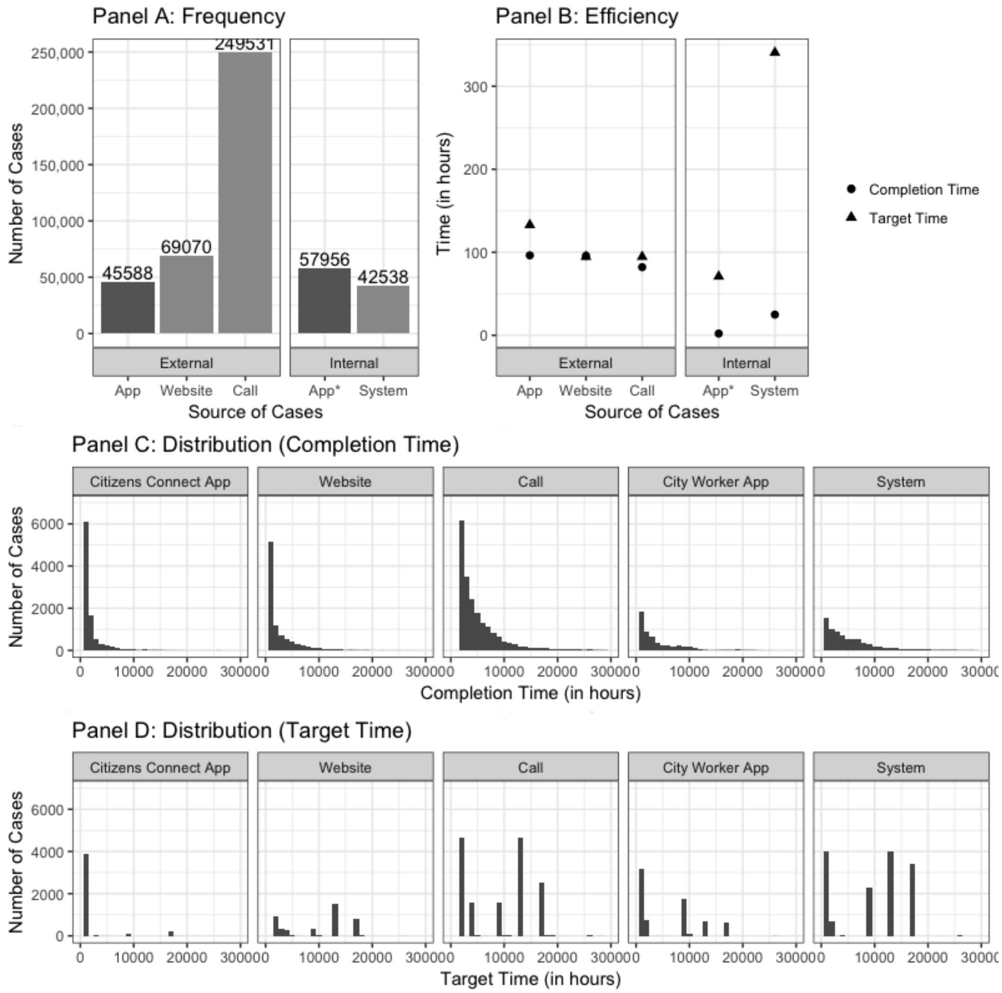
Panel A of Figure 3-2 reports the distribution of cases over those sources. It shows that city employees were generally more likely to use mobile devices to report a case than citizens. Panel B of the same figure shows the median of the actual completion time and the median of the target completion time. Cases generated internally were solved faster with an even longer target completion time. Though this of course does not condition for differences in case type, it still suggests that the internal cases and the external cases can be inherently different. Due to this reason we exclude all employee-generated cases in our initial analysis and focus on identifying the effect of using the Citizens Connect App and the associated mobile information technologies and distinguish the app from the self-service website or phone service. We return to employee-generated cases when we turn to identification.

Panel C of Figure 3-2 shows the histogram of actual completion times for different sources. The highly skewed distribution and the long tail on the right suggest that a survival model may be an appropriate way to capture the distribution of complaint time resolution. On the other hand, the distribution of target times is concentrated (Panel D), which suggests that this cut-off time for an "on-time" completion might be a preset number and case-by-case adjustments are rare. Therefore, we choose to focus on the actual completion time and use the "on-time" status calculated from target times as a robustness check when analyzing complaint resolution performance.

We summarize the type of cases in Table 3.A1. More than 80% of the cases belong to the top seven categories (reasons) - which have more than 20,000 cases: Sanitation, street cleaning, highway (and road) maintenance, street lights, recycling, signs & signals, and trees. We see that the average completion time varies greatly among different types in Figure 3-3: an issue about trees takes on average about 3,000 hours for city employees to resolve it while this number is only 100 hours for an average sanitation issue. To address this we use

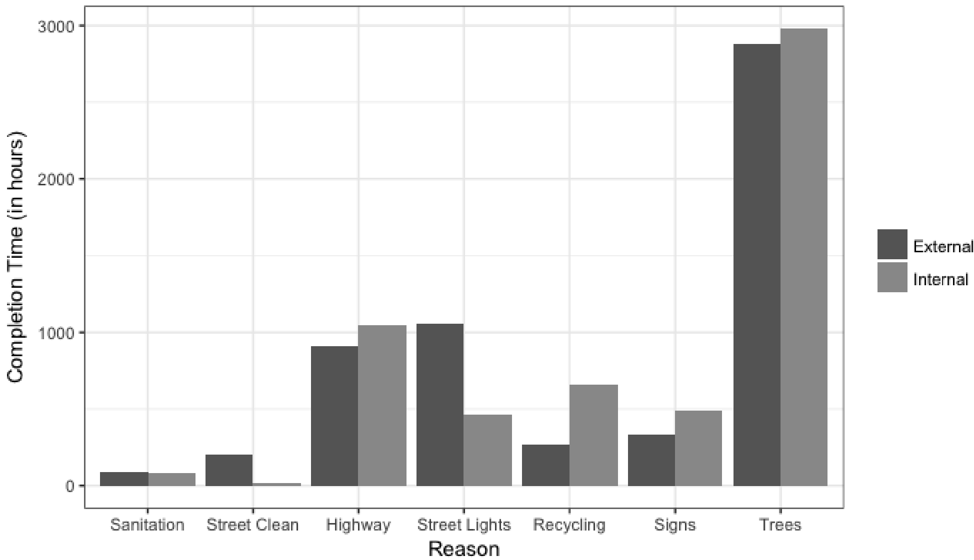
[6]There is also a Maximo Integration system, which serves as a central repository for all reports of issues concerning street lights, control boxes, fire houses that are automated. The lack of human intervention required means that we do not study complaints for this system.

Figure 3-2: Sources of Cases



Notes: Figure 3-2 is based on both external cases and internal cases. External cases were submitted by citizens through Citizen Connect App, the self-service website, or traditional phone calls. Internal cases were submitted by city employees through the traditional system or City Worker App. The number of cases for each source is plotted in Panel A. The average actual completion time and the average target completion time for each source are plotted in Panel B. The distribution of actual completion times is plotted in Panel C by source, and the distribution of target completion times is plotted in Panel D by source. There are in total 464,683 external and internal observations in Panel A, B, and C. There are in total 372,235 external and internal observations in Panel D due to missing values in target completion time. The target completion time is missing for 199 cases generated via Citizens Connect App, 20,210 cases generated via the self-service website, 59,153 cases generated via traditional phone calls, 1,468 cases generated via City Worker App, and 11,418 cases generated via the traditional internal system. The disproportionate reduction in missing records shows the advantage of advanced technology in automatic recording.

Figure 3-3: Completion Time by Type



Notes: Figure 3-3 is based on both external cases and internal cases. The average completion time of cases is plotted separately for external and internal cases in seven major case types where there are more than 20,000 cases. There are 94,962 external and 14,539 internal observations for sanitation, 69,434 external and 28,656 internal observations for street cleaning, 33,751 external and 31,301 internal observations for highway (and road) maintenance, 23,423 external and 3,475 internal observations for street lights, 21,633 external and 5,051 internal observations for recycling, 21,824 external and 3,734 internal observations for signs & signals, and 12,498 external and 9,565 internal observations for trees.

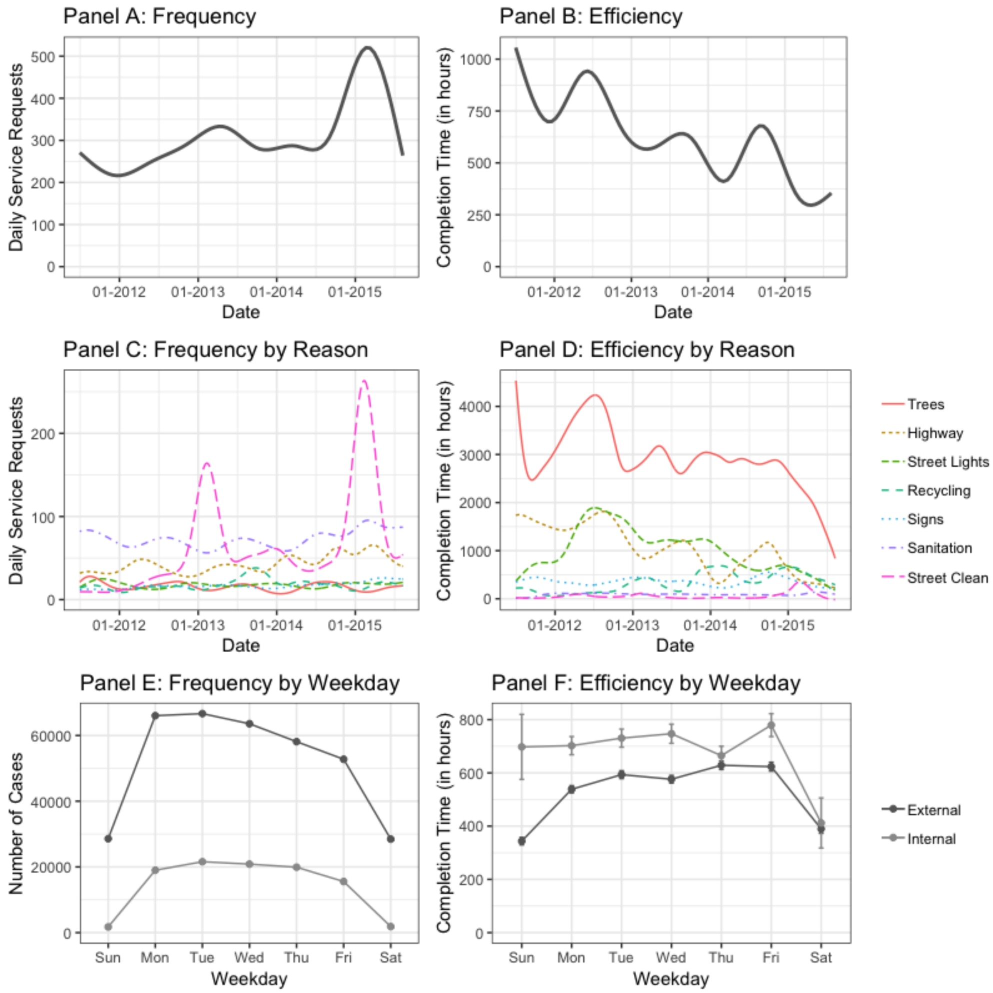
category (reason) stratification to shed light on the baseline efficiency of customer complaint resolution in our survival model.

3.3.2 Increasing efficiency, seasonality and weekly variations

Using a local polynomial regression, we display a smoothed trend over time of the number of cases opened in Panel A of Figure 3-4 and a trend over time of the average actual completion time in Panel B of the same figure. They suggest a strong seasonality in efficiency but not in frequency[7] – the peak of efficiency usually comes in January or February – while the average completion time is decreasing year by year. A further investigation in Panel D of Figure 3-4 shows this seasonality varies among different types of cases: Some types, like recycling and highway maintenance work, have obvious delays in extreme weather, while service

[7]There is a decrease on the right of Panel A because we have removed those incomplete cases which were opened late and had not been finished until our data collection date (Feb 09, 2016).

Figure 3-4: Trends in Completion Time and Number of Cases



Notes: Figure 3-4 is based on both external cases and internal cases. The number of daily service requests opened over time in Panel A has been smoothed using local polynomial regression, so does the average actual completion time in Panel B. Those smoothed trends are broken down by reason in Panel C and Panel D, respectively. There are in total 464,683 external and internal observations in Panel A and B. In Panel C and D, there are 109,501 observations for sanitation, 98,090 observations for street cleaning, 65,052 observations for highway (and road) maintenance, 26,898 observations for street lights, 26,684 observations for recycling, 25,558 observations for signs & signals, and 22,063 observations for trees. The total number of cases opened on each weekday have been plotted separately for external (submitted by citizens) and internal (submitted by city employees) cases in Panel E, so does the average actual completion time in Panel F. There are 364,189 external observations and 100,494 internal observations in both Panel E and F.

performance for other types, such as sanitation and signs & signals, is quite stable. Even though the composition of cases shown in Panel C of the same figure is more stable than the performance for all types except street cleaning work, the number of cases for most types still exhibits a seasonal pattern. For example, the number of street cleaning cases increases dramatically in winter, sanitation issues dominate the system in the summer, and the peak of tree-related cases always arrives in New England during the foliage season. These findings are in line with the previous argument that the baseline efficiency should be stratified by reason in our model.

We also plot the total number of cases and the average completion time by weekday in Figure 3-4. Panels E and F of Figure 3-4 implies variation across the day of the week the case was submitted in the waiting time for citizens, while this pattern is more subtle for the cases generated by employees. This is even though the number of cases submitted during the weekend is significantly lower than that on weekdays for both citizens and employees. For the requests sent by citizens, the weekly variation might be explained by a "stock effect," which is the difference in case completion time resulting from a varying number of open cases in the system when it was opened, and a "peer effect," which is the difference in case completion time resulting from a varying number of cases opened at the same time.

We control for these shifts over time by adding year, season and weekday fixed effects in our main model.

3.3.3 Census Data

We use 2010 census data for each block group, which is the smallest geographic unit used by the United States Census Bureau. There are 646 census block groups in the city of Boston and each of them has been labeled with a unique 12-digit ID number. Figure 3-A1 shows those block groups on the map. 10 out of those 646 census block groups are fully covered by lakes or parks and are reported to have no population living inside. The public service data also confirms that no requests have been sent from those block groups by citizens. We match each public service case with census block groups using the exact latitude and longitude of the case and approximate the reporter's characteristics using group-level socio-demographic variables about gender, age, race, language spoken, income and housing

Table 3.1: Summary of Demographic Variables

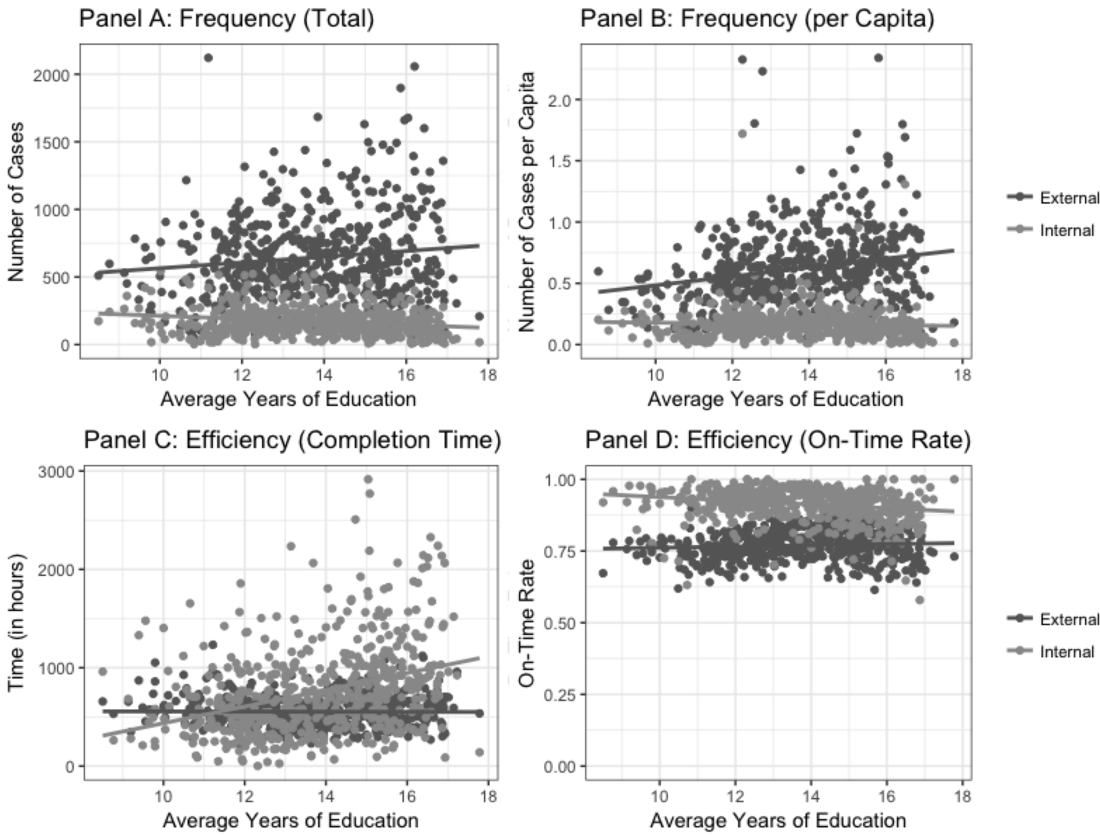
Variable	Mean	Std. Dev.	Min.	Max.
#Population	1160.191	526.878	13	3716
#Households	460.927	224.634	4	1424
Gender				
%Male	0.476	0.083	0.097	0.76
%Female	0.524	0.083	0.24	0.903
Age				
%< 18 years old	0.167	0.105	0	0.489
%18 - 29 years old	0.282	0.188	0	0.97
%30 - 44 years old	0.222	0.093	0	0.615
%45 - 59 years old	0.17	0.078	0	0.438
%≥ 60 years old	0.159	0.102	0	0.903
Race				
%White	0.537	0.317	0	1
%Black	0.257	0.297	0	1
%Asian	0.088	0.117	0	0.885
%Other Single Race	0.076	0.105	0	0.538
%Multiple Races	0.043	0.067	0	0.498
Education				
%Less than High school	0.146	0.13	0	0.583
%High School Diploma	0.225	0.137	0	0.844
%Some College	0.142	0.084	0	0.46
%Bachelor Degree	0.287	0.143	0	0.725
%Graduate School	0.199	0.166	0	0.781
Average Years of Education	13.763	1.857	8.505	17.781
Language				
%English Only	0.635	0.196	0	1
%Spanish	0.153	0.152	0	0.678
%Bilingual	0.248	0.13	0	0.697
%Limited English Speaking	0.117	0.132	0	1
Income				
Poverty Ratio for Households	0.217	0.168	0	1
Housing Status				
%Owner-occupied	0.358	0.24	0	1
%Renter-occupied	0.642	0.24	0	1
%Living in Same House 1 Year Ago	0.795	0.144	0.19	1
%Living in Greater Boston Area 1 Year Ago	0.142	0.098	0	0.788
%Living Abroad 1 Year Ago	0.017	0.031	0	0.188

Note: Observations = 545 and there is one observation for each block group.

status. There are 545 census block groups from which requests were sent, and a summary of those socio-demographic variables in these 545 census block groups is provided in Table 3.1. Each census block has 1,160 individuals and 460 households on average.

Panel A and B of Figure 3-5 shows an example of how the number of requests varies with those social-demographic characteristics. Both the total number of external requests

Figure 3-5: Completion Time and Number of Cases on Education



Notes: Figure 3-5 is based on both external cases and internal cases. The number of cases opened in a neighborhood is plotted over average years of education in the neighborhood in Panel A and the number of cases per capita is plotted over average years of education in Panel B. The average completion time in the neighborhood is plotted over average years of education in Panel C while the average on-time rate in the neighborhood is plotted in Panel D. All of them have been plotted separately for external (submitted by citizens) and internal (submitted by city employees) cases. There are 541 neighborhoods in those panels.

sent by citizens and the number of external requests per capita increase with the average years of education, while the number of internal requests sent by employees keeps almost constant across levels of education. Panel C of Figure 3-5 shows how complaint resolution performance varies with the average level of education. To take the difference in case type across internally vs externally generated complaints, we plot the on-time rate rather than completion time on the y-axis in Panel D. In both panels, the efficiency increases with the education level for external requests while the employee-generated cases show an opposite trend. Moreover, we see that the external cases are more standardized in terms of efficiency but have a large variation in the frequency.

3.4 Main Effect

3.4.1 Model

To identify the effect of the use of mobile app usage on the complaint resolution performance with controlling for all the factors mentioned in the previous section, we use a Cox proportional hazards model to analyze how long it takes for a complaint to be resolved. We focus on the distribution of completion times and model the time it takes for completion to occur. The validity of this survival analysis approach is supported by the model-free evidence regarding the distribution of completion time provided in the histogram depicted in Panel C of Figure 3-2.

Let

$$\lambda_i(t) = \lambda(t|CitizenConnectApp_i, WebSubmission_i, \omega, \mathbf{X}_j, \mathbf{Y}_k)$$

denote the hazard function[8] for a case i belongs to category $\omega(\omega \in \Omega)$ opened on day j and in census block group k at time t, where $CitizenConnectApp_i$ is an indicator which equals one if the case was submitted via the mobile app, and $WebSubmission_i$ is an indicator which equals one if the case was submitted on the self-service website. $\mathbf{X}_j$ is a vector of time-specific covariates, including year, season, and weekday indicators. And $\mathbf{Y}_k$ a vector of block-group-specific covariates, including gender, age, race, language, income, housing status, and education controls. Leaving the reason-specific baseline hazard function $\lambda_\omega(t) = \lambda(t|0, 0, \omega, \mathbf{0}, \mathbf{0})$ unspecific,[9] we models the log hazard ratio, which is the relative “risk" of the case closing at time t, as

[8] Hazard function assesses the instantaneous risk of demise at time t, conditional on survival to that time: $lim_{\Delta t \to 0} \frac{Pr[(t \leq T < t+\Delta t)|T \geq t]}{\Delta t}$ (Fox & Weisberg, 2010).

[9] Even though the Cox model is semi-parametric with unspecific baseline hazard and linear covariate terms, it can still be estimated by the method of partial likelihood (Cox, 1972).

$$\log(\frac{\lambda_i(t)}{\lambda_\omega(t)}) = \beta_0 + \beta_1 CitizenConnectApp_i + \beta_2 WebSubmission_i + \mathbf{X}'_j\lambda + \mathbf{Y}'_k\mu. \quad (1)$$

The parameters of interest are β_1 and μ, which capture the effect of using mobile information technologies and that of complainer's socio-demographic characteristics on the case completion time.

Furthermore, an extension of this model

$$\begin{aligned}\log(\frac{\lambda_i(t)}{\lambda_\omega(t)}) = {} & \beta_1 CitizenConnectApp_i + \beta_2 WebSubmission_i + \mathbf{X}'_j\lambda + \mathbf{Y}'_k\mu \\ & + CitizenConnectApp_i\tilde{\mathbf{Y}}'_k\tau_1 + WebSubmission_i\tilde{\mathbf{Y}}'_k\tau_2, \end{aligned} \quad (2)$$

where $\tilde{\mathbf{Y}}_k$ is a subset of $\mathbf{Y}_k$, incorporates the demographic characteristics of which the interactions with the channel used are studied.

To enhance interpretability of coefficients and reduce numerical instability caused by the multicollinearity in interaction models (Afshartous & Preston, 2011), we use mean-centered transformations for all socio-demographic variables ($\mathbf{Y}_k$) so that the baseline hazard function is $\lambda_\omega(t) = \lambda(t|0, 0, \omega, \mathbf{0}, \bar{\mathbf{Y}})$. β_1 captures the effect of using mobile information technologies for people in an average neighborhood, and τ_1 implies how smartphone app use changes the relative public service performance for different subpopulations. If the use of advanced technologies enhances complaint resolution for less educated people (i.e., τ_1 and μ have different signs), this suggests that app usage alleviates social inequality.

3.4.2 Initial Analysis

We initially focus on the 364,189 requests sent by citizens via phone calls, the self-service website or the Citizen Connect App. We pick “average years of education" as the variable of interest out of those socio-demographic variables $\mathbf{Y}_k$ since level of education is widely believed to affect a person's skills and inherent ability to advocate for themselves (Bresnahan *et al.* , 2002). This has also been widely documented in healthcare. Both

Table 3.2: Main Effects on the Completion Time of 311 Cases

	(1) Hazards	(2) Hazards	(3) Hazards	(4) Hazards	(5) Hazards	(6) Hazards	(7) Hazards
Average Year of Education	0.008***	0.017***	0.017***	0.023***	0.025***	0.028***	0.036***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Citizen Connect App					-0.048***	-0.039***	-0.042***
					(0.006)	(0.006)	(0.006)
Web Submission					-0.130***	-0.130***	-0.161***
					(0.005)	(0.005)	(0.005)
Average Years of Education × Citizen Connect App						-0.015***	-0.018***
						(0.003)	(0.003)
Average Years of Education × Web Submission						-0.001	-0.007**
						(0.003)	(0.003)
Reason Stratification	No	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes
Season Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes
Weekday Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes
Gender Controls	No	No	No	Yes	Yes	Yes	Yes
Age Controls	No	No	No	Yes	Yes	Yes	Yes
Race Controls	No	No	No	Yes	Yes	Yes	Yes
Language Controls	No	No	No	Yes	Yes	Yes	Yes
Income Controls	No	No	No	Yes	Yes	Yes	Yes
Housing Status Controls	No	No	No	Yes	Yes	Yes	Yes
Observations	364,189	364,189	364,189	364,189	364,189	364,189	310,736
AIC	8,598,758	6,842,089	6,835,968	6,835,666	6,834,835	6,834,811	5,968,633

Notes: The population variable is cases and a hazard model for the completion time is used. There are only 310, 736 observations in Column (7) because some minor categories are removed. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Zimmerman *et al.* (2015) and Berkman *et al.* (2011) suggest that more educated people received better healthcare service because education enhances their communication skills and ability to advocate for themselves. Willems *et al.* (2005) point out that higher patient educational levels are associated with better doctor-patient communication, which has a strong and positive influence on patients' satisfaction and compliance.

Table 3.2 reports the estimation results of the Cox hazard model. Column (1) in the table presents the result of a regression that includes only $AverageYearsOfEducation_k$ as the independent variable. The reason stratification, the time fixed effects X_j and other socio-demographic controls Y_k are added into the model incrementally in Columns (2)-(4). Column (5) adds the variables $CitizenConnectApp_i, WebSubmission_i$ to the model to take the effect of submission channels into account. Interactions between the channel and the level of education are added to the model in Column (6). Column (7) replicates the results in Column (6) but excludes cases in 36 smaller categories where there was no case has been submitted via the App. Categories included in Column (7) are starred in Table 3.A1.

The coefficient of $AverageYearsOfEducation_k$ is always significant and positive in Table 3.2. It implies that people who submit complaints in neighborhoods with lower education levels experience longer waiting times for their complaints to be resolved – particularly, based on Column (6), cases submitted via phone calls were $e^{0.028} - 1 = 2.8\%$ more likely to be completed at any time if the average level of education in the neighborhood increases by one year.[10] The coefficient of $CitizenConnectApp_i$ is negative in the table, and $WebSubmission_i$ also always has a negative effect on the service performance. These results suggest that the mobile app, on average, worsens the complaint resolution performance compared to channels where customers have real-time interaction with representatives such as phones (by $1 - e^{-0.039} = 3.8\%$ for an average neighborhood) but outperforms traditional technologies such as desktop computers (by $e^{-0.039+0.130} - 1 = 9.5\%$ for an average neighborhood). Meanwhile, app use leads to relatively better service for people who submit complaints in neighborhoods with lower average education levels. This mitigating effect is supported by the significantly negative coefficient of the interaction term between $AverageYearOfEducation_k$ and $CitizenConnectApp_i$ in Column (6) – the gap in the likelihood of case completion brought by a one-year increase in the average years of education in the neighborhood decreased by $1 - e^{-0.015} = 1.5\%$ if the app is used to submit the case, while it does not hold for the self-service website. Those numbers also imply that, the app expedited the complaint resolution compared to phones for all the citizens who live in a neighborhood where the average years of education are more than $-0.039/-0.015 = 2.6$ years below the city-wide average level.[11] All of those results hold for the robustness checks where the minor categories are removed (in Column (7)), which implies that the inclusion or exclusion of these smaller categories where the app is not used does not change our results.

[10] This number is $e^{0.028-0.015} - 1 = 1.3\%$ for cases submitted via the app and $e^{0.025-0.001} - 1 = 2.4\%$ for cases submitted via the website.

[11] In other words, it helps the neighborhoods where the average years of education are lower than 11.2 years.

A natural concern when interpreting these results is multicollinearity, given the fact that socio-demographic variables are usually correlated with each other. In Table 3.A2, we show the correlation matrix of all those socio-demographic variables at both group level (N=545, Panel A) and individual case level (N=364,189, Panel B). None of those correlation coefficients between $AverageYearsOfEducation_k$ and other socio-demographic controls Y_k has an absolute value higher than 0.7. Also, the variance inflation factor (VIF) of $AverageYears OfEducation_k$ in Column (3) is 4.1, below the threshold of 10, indicating that multicollinearity is not a likely threat to the parameter estimation (Cohen *et al.* , 2003).

Since this is both new, and somewhat complex, data which required us to make some judgment calls about how to structure it, we ran a battery of robustness checks to ensure that none of our judgment calls affected our results. First, we examine an alternative specification which uses logistic regression where the dependent variable equals one if the case was recorded as being solved "on time" on the same set of regressors to investigate robustness to functional form. The results in Table 3.A3 are consistent with the main results, which implies that our results will not be changed by taking the extra information in target completion times into account.

We also wanted to check that the way we specified our key explanatory variable for average education level in that census block did not affect our results. Table 3.A4 replicates the results in Column (6) of Table 3.2 for a series of alternative independent variables that measure the education level in different ways. We use the original independent variable "average years of education" in Column (1), the log of average years of education in Column (2), the percentage of people in neighborhoods with high school diploma and above in Column (3), the percentage of people with some college education (stay at least 2 years in college) and above in Column (4), the percentage of people with a bachelor's degrees and above in Column (5), and the percentage of people who attended graduate school in Column (6). The results in this table reinforce the previous conclusion – we see the same signs in all columns. The coefficient of the education level decreases from Column (3) to Column (5) while the absolute value of the interaction term increases from Column (4) to Column (6).[12] Those

[12] A series of Z-tests shows that the decrease in the coefficient of the education level is significant at the

trends suggest that the higher the degree is, the less important the skills gained during it are to the complaint resolution and the easier its effect is to be mitigated by digital technologies.

3.4.3 Block-by-block Analysis

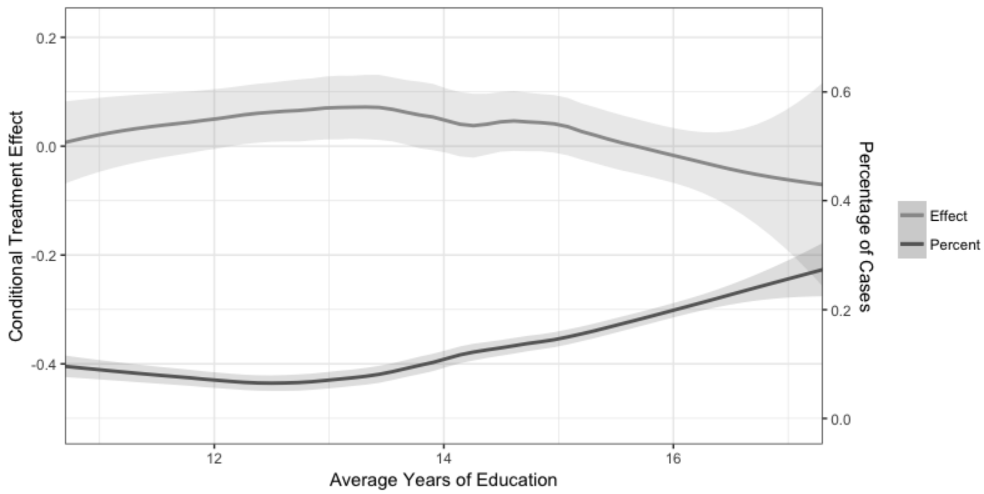
Adoption of information technology is not equally distributed across the population. This "digital divide" has been discussed for a long time. Evidence shows that more educated and high-income people are more likely to adopt Internet technologies (Chinn & Fairlie, 2007, Goldfarb & Prince, 2008).

One concern this raises is that there is likely to be uneven adoption of the app across census blocks which may affect the measurement of the treatment effect. The black line in Figure 3-6 shows that a higher proportion of cases are sent via the app in neighborhoods with higher education levels, even though the correlation between $AverageYearsOfEducation_k$ and $CitizenConnectApp_i$, is only 0.13. To deal with the potential issues arising from this dependency we use a potential outcomes approach to appropriately adjust our measured treatment effects.

Rubin (1977) suggests that if assignment to treatment group is made based only on the value of a covariate, then averaging conditional treatment effects over the distribution of those covariates will give a valid estimate of the treatment effect and any other sources of bias are ignorable. Therefore, if we believe that the treatment assignment (app adoption) is different among census block groups but uniform within each group, then the treatment effect calculated for each census block group should be valid and show the causality. This assumption might be realistic, as: (1) The census block group is small in size – on average there are only 1160 people in a block group; (2) socio-demographic characteristics that affect the app adoption most significantly, such as education and income, vary dramatically between block groups but only slightly within block groups; (3) socio-demographic characteristics that vary within block groups, like age and gender, matter less for advanced technologies' adoption than for the adoption of traditional technology (Chinn & Fairlie, 2007). We will use instrumental variables to address the endogeneity if this assumption does not hold.

$\alpha = 0.1$ level from Column (4) to Column (5) and the increase in the absolute value of the interaction term

Figure 3-6: Treatment Assignment and Conditional Treatment Effects



Notes: Figure 3-6 is based on external cases only. The percentage of cases reported via Citizens Connect App is plotted over the average years of education in the neighborhood in Panel A. The conditional treatment effect of the app use in each census block, calculated using a Cox proportional hazards model with only the case type stratification and the date controls X_j, are plotted separately over average years of education in Panel B. There are 541 observations in both panels. Local linear regressions with a 50% smoothing span are used here to smoothen the trends.

Therefore, we calculate the conditional treatment effect for each census block using a Cox proportional hazards model with only the case stratification and the date controls X_j. The effect of the Citizen Connect App use conditional on average years of education is represented by the grey line in Figure 3-6. The results are consistent with our main results: use of the app doesn't improves the complaint resolution performance for everyone, and use of the app helps less educated people (people submitting complaints in neighborhoods with lower education levels). Surprisingly, we notice that the mitigating effect of the app is not monotonic in Figure 3-6 and the app amplifies the disadvantages of people with fewer than about 13 years of education. One interpretation is that using the app competently does require a baseline level of skills.

3.5 Mechanism: Standardized Communication

We then turn to investigate the mechanism behind our key result which is that the use of the app appears to mitigate the influence of education on complaint resolution time.

is significant at the $\alpha = 0.1$ level from Column (5) to Column (6).

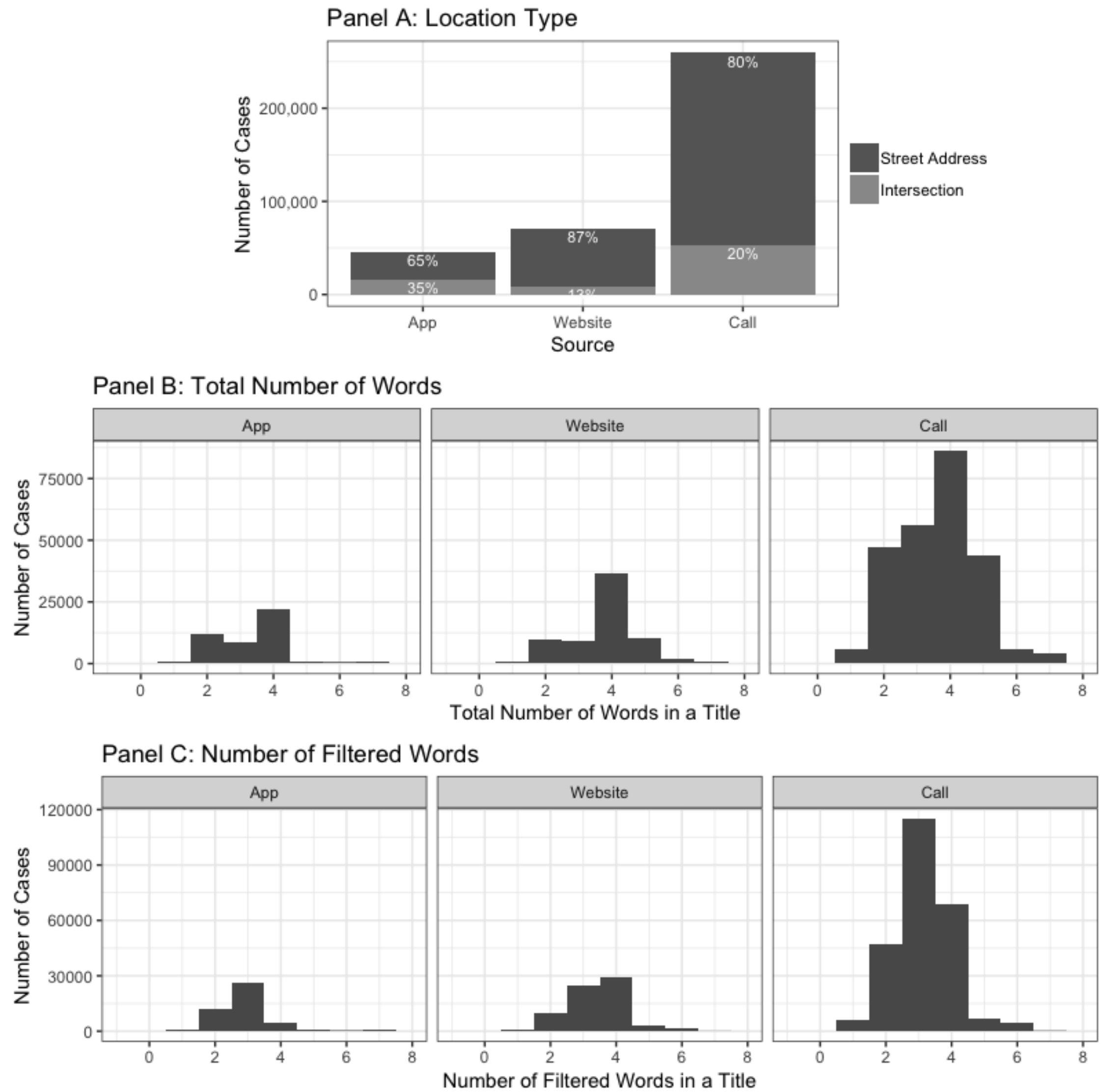
3.5.1 Standardization of Case Locating

One potential way the app may be able to mitigate the influence of education on complaint resolution time, is simply by making it easier for the city worker to find the issue geographically. Unlike the cellphone and website channels, the app has the feature that it makes it easier to pinpoint where precisely the issue is and thus substitutes for the required communication skills.

There are two types of locations in the record – the cases can be either submitted with a street address, such as "100 Second St." or filled as an intersection, such as "the intersection of Second St and Park St." The number of cases where the location was reported as a street address and the number of cases at an intersection are plotted by source in Panel A of Figure 3-7. The mobile app automatically fills the location with either a street address or an intersection depends on the real-time geo-information captured by smartphones. This input can be overwritten only by dragging a marker to the desired place on an interactive map. Since the location can never be filled manually in the app, the app has the largest proportion (35%) of cases reported with an intersection. Complaints submitted by the website by contrast tend to have street addresses, because its autocomplete function is based on Google maps which does not use intersections – only 13% of the cases submitted via the website was reported at an intersection. Complaints that are submitted by phone are more likely to be reported with an intersection, perhaps reflecting the reporting preferences of service providers themselves. Indeed our interviews suggested that intersections are considered more useful, as they to not rely on having to identify a street number in a town where street numbers are often not visible and may refer to buildings that span city blocks and corners making locations imprecise. It seems possible that complaints that involve these street addresses generated by the app may need less context and explanation to pinpoint the location of the complaint than other channels. Therefore such cases are where the ability of the app to use GPS to pinpoint a precise location may be very helpful for people who are less able to communicate a complex location.

We investigate these hypotheses by interacting the main effects with an indicator which

Figure 3-7: Length and Informativeness of Titles by Source



Notes: Figure 3-7 is based on external cases only. The distribution of address types is plotted by source in Panel A. The histogram of total numbers of words in a title and that of numbers of filtered words in a title are plotted by source in Panel B and Panel C, respectively. There are 364,189 external observations.

equals one if the case was submitted with a street address. First, based on the results reported in Column (2) of Table 3.3,[13] we find that street addresses do slow down the complaint resolution process: cases submitted with street addresses were $1 - e^{-0.198} = 18.0\%$ less likely to be completed at any time compared to the cases submitted at intersections after reason

[13]The results about the effects of website submission have been cut from Table 3.3 due to limited space. They are reported in a full version of this table in the appendix (Table 3.A5).

stratification. Second, supported by the positive coefficient of the interaction term between $AverageYearOfEducation_k$ and the street address indicator, the advantage of educated people is significantly amplified by the street address use. It supports the hypothesis that the skills of communicating a complex location play a more crucial role in reporting these cases. Meanwhile, as expected the app provides better customer service performance when the location is complex: the mobile app, on average, expedited the complaint resolution of a case reported with a street address in an average neighborhood by $e^{0.185-0.025} - 1 = 17.4\%$ compared to a similar case submitted via a phone call. More importantly, the effect of app use on the gap between educated and less educated consumers is more significant when the communication about the location is complicated. This moderating effect is even larger than the amplification effect of the street address use on the advantage of educated people.

3.5.2 Standardization of Case Description

As a corollary to this evidence that the automation of data input is most beneficial when it is difficult to describe location, it is possible too that the app may substitute for communication for complex cases more generally. One hypothesis is that case complexity may be related to how much information it conveys. It may also be that, as case complexity increases, the case becomes more difficult to describe and needs more communication. Due to the limitations of the available data, we create proxies for case complexity from case titles. Following the standard process of neuro-linguistic programming (NLP) here, we first count the total number of words in each title and then filter out the insignificant words (like stop words) to count the number of meaningful words (Jurafsky & Martin, 2009).

The distribution of the total number of words in a case title is plotted in Panel B of Figure 3-7 and that of the number of filtered words is plotted in Panel C of the same figure. On average there are 3.6 words and 3.1 filtered words in a title. Titles are slightly more informative in the website channel while the representatives tended to write down a title with fewer words during the phone call, but the correlation coefficient between the total number of words and the app use is only about -0.08. The gap becomes even smaller and the distributions look more similar when the case complexity is measured by the number of filtered words, which have a narrower definition.

Table 3.3: Mechanism

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Effect	Citizen Connect App				Photo Attachment		
Moderating Factor	None	Address	#Total	#Filtered	None	#Total	#Filtered
Average Year of Education	0.028***	0.026***	0.028***	0.029***	0.025***	0.025***	0.026***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Citizen Connect App	-0.039***	-0.025***	-0.060***	-0.063***			
	(0.006)	(0.007)	(0.006)	(0.007)			
Average Years of Education	-0.015***	-0.014***	-0.013***	-0.015***			
× Citizen Connect App	(0.003)	(0.003)	(0.003)	(0.003)			
Photo					-0.040***	-0.061***	-0.058***
					(0.009)	(0.009)	(0.010)
Average Years of Education					-0.006	-0.004	-0.007
× Photo					(0.005)	(0.005)	(0.005)
Moderating Factor		-0.198***	-0.147***	-0.034***		-0.143***	-0.060***
		(0.006)	(0.003)	(0.003)		(0.002)	(0.003)
Moderating Factor		0.023***	0.012***	0.013***		0.011***	0.009***
× Average Year of Education		(0.003)	(0.001)	(0.001)		(0.001)	(0.001)
Moderating Factor		0.185***	0.038***	-0.093***			
× Citizen Connect App		(0.012)	(0.005)	(0.006)			
Moderating Factor		-0.026***	-0.011***	-0.013***			
× Avg. Education × App		(0.006)	(0.002)	(0.003)			
Moderating Factor						0.064***	-0.030***
× Photo						(0.009)	(0.011)
Moderating Factor						-0.009**	-0.014**
× Avg. Education × Photo						(0.005)	(0.006)
Interactions with Web Use	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reason Stratification	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weekday Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Language Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing Status Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	364,189	364,189	364,149	364,149	364,189	364,149	364,149

Notes: The population variable is cases and a hazard model for the completion time is used. The presence of missing values leads to a smaller sample size in Columns (3), (4), (6), and (7). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Columns (3) and (4) of Table 3.3, we interact the main effects with the total number of words and the number of filtered words, respectively. Consistent with our hypotheses, we find that cases where the title contains more information were resolved less efficiently: cases were $1-e^{-0.147}=13.7\%$ less likely to be completed at any time if there was one more word in the title and this number was $1-e^{-0.034}=3.3\%$ for one more “meaningful" word in the title. And the lack of communication skills tends to widen this gap. Although the significantly positive coefficient of the interaction term between $AverageYearOfEducation_k$ and the title informativeness implies that educated people enjoyed more advantages and privileges coming from better communication skills when the title conveyed more information, this effect is mitigated by the app use: the mitigating effect of app use on the advantage of educated people was intensified by more than 50% if there was one more word in the case title for an average case. This analysis provides supportive evidence for the hypothesis that the app substitutes for communication skills needed for describing complex cases.

Another hypothesis is that the ability of the app to convey visual information in the form of photos also substitutes for the need for communication skills. To investigate this we examine the effect of submitting cases included photos (photos are only conveyable through the app). 35.4% cases submitted with the app included photos. Therefore, we replace the app-use indicator $CitizenConnectApp_i$ with $Photo_i$, a binary variable indicating whether the case submission included photos.[14] As we can see from Columns (5) -(7) of Table 3.3, all the results mentioned before remain the same after changing the independent variable, which implies that the app improves the efficiency of complaint resolution by allowing those who live in less educated areas to upload the photos that can substitute accurate case descriptions.

To summarize, we find that the use of mobile information technology appears to reduce inequality by providing a standardized communication tool that substitutes for the need to communicate and have this communication appropriately recorded in complicated cases. Similar results are found after replacing the average level of education by the percentage of Spanish speakers in the census block (see Table 3.A6), which reinforces our conclusion about the diminished role of communication skills when mobile information technologies are used during request submission.

[14] A photo can be attached only when the case is submitted via the app, i.e., $CitizenConnectApp_i \leq Photo_i$ always holds.

3.6 Endogeneity of App Adoption: Instrumental Variable Approach

We have addressed the issues arising from the uneven adoption of information technology across census blocks using the potential outcome approach in Section 3.4.3. However, an even more serious concern is that even within a census block there may be differences in the people who use the app and the people who do not. To address this, we turn to instrumental variables.

Given uneven app adoption within census block groups, there may be concerns over endogeneity of adoption within a census block. First, the adoption decision and the complaint resolution efficiency can be affected by individuals' wealth and education level, which is unobservable to researchers. Second, people who adopt an app are more likely to be the ones who had a better experience with the app (Buell *et al.* , 2017). Third, selection issues may also arise from the possibility that people who adopt the app might be the ones that really care about public service and are aware of the app launch. Those people are more likely to report non-emergencies which won't be resolved immediately.

3.6.1 Geographic Variation in Cell Tower Proximity

To address these potential endogeneity issues, we instrument the use of mobile devices with a proxy of cellular signal strength. There is reason to think that the quality of cellular signal will affect usage of the mobile app – people will not be able to use the app when the signal is very weak, even if they wished to do so. The instrument we use is the geographic distance to the closest cell tower of the complaint. The idea is that if cell phone towers are reasonably randomly distributed across Boston, then this will affect mobile app usage but will not directly affect complaint resolution time.

Based on the records found on antennasearch.com, we identify 84 registered cell towers located in the city of Boston and its surroundings. The average distance to the closest cell tower is 1.14 kilometers. Even though channel choices were affected by other factors such as the availability of in-home WiFi networks and case-by-case considerations, this instrument did significantly shift citizens' app uses and provides a valid estimate of a local average treatment effect (LATE) of the app use. Test statistics suggest that the first stage of the instrument is strong and that the employee submission channel is a good predictor of citizens' behavior – as shown in Table 3.A7, the probability of submitting a case via the mobile app increases more than 1.2% when the case is located one kilometer closer to a cell tower.

Another key question is whether this instrument meets the exclusion restriction which requires that the location of the cell phone tower be unrelated to any underlying geographic features which might also explain complaint resolution time. In an old city such as Boston, cell towers are attached to existing tall buildings like church towers. The number of tall buildings is further restricted by building codes, which mean that the presence (or absence) of tall buildings is related to the social geography of Boston many decades prior to the present. The pattern of settlement by different socio-economic groups in Boston has changed over time, meaning that the presence of historic "tall" buildings is not related to individual characteristics, especially when considering only the relative wealth or education level within census blocks. It is not even strongly related to neighborhood wealth. The difference between the distribution of the education level (Panel A of Figure 3-A2) and that of distance to the closest tower (Panel B of Figure 3-A2) provides supportive evidence for this.

3.6.2 Instrumental Variables Estimation

In order to apply the instrument variables method here, we first obtain a discrete-time approximation for the proportional-hazards model by running a logistic regression on a set of pseudo-observations generated as follows. Let $\mathbf{t} = (t_1, t_2, ...)$ be a series of cut points which divide the timeline into small intervals: Each day in the first week, each week of the remaining weeks in the first month, and each month of the remaining months in the first year serves as a time interval, while the rest of time constitutes the remainder.[15] Suppose case i is resolved at time t where $t \in [t_{n-1}, t_n)$, we generate one resolution indicator d_{im} (m=i,...,n) for each time interval from $[0, t_1)$ to $[t_{n-1}, t_n)$. This variable takes the value one for the interval $[t_{n-1}, t_n)$ and zero otherwise. To each of these indicators we associate a copy of the covariate vector

[15] An approximation with fewer small intervals is used here to reduce the number of observations. We conducted various tests and proved that neither increasing the number of intervals nor changing the cut points significantly improves the precision of our results.

($CitizenConnectaApp_i$, $WebSubmission_i$, X_j, Y_k), a group of indicators that identify the time interval, and a group of reason indicators and their interactions with the time interval indicators (for reason stratification). Treating the d_{in} as independent Bernoulli observations with probability given by the individual hazard rate λ_{in} is equivalent to fitting the Cox proportional hazard model in equation (1) (Holford, 1980, Laird & Olivier, 1981). We fit the new model using a logit regression specification and replicate the results from Columns (1)-(4) of Table 3.3 with this discrete-time approximation, while marginal effects at means, rather than the original coefficients, are reported in Panel A of Table 3.4. The results are very close to what we get from the Cox model in terms of significance and relative magnitude.

We report the 2SLS estimates, where mobile app usage for a similarly located employee-generated case is used as the instrument in Panel B of Table 3.4. The first stages for Column (1) are reported in Table 3.A7. The Wald statistic is 539.83, which suggests a strong first stage. Comparing the marginal effects at means from the original logit estimates reported in Panel A, the measured treatment effects are similar in sign but are larger in magnitude for most of the mitigating effects – now the mitigating effect of app use on case complexity will increase by 1.6% if there is a one-year decrease in the education level of the neighborhood. However, the coefficient of the interaction term between $AverageYearOfEducation_k$ and $CitizenConnectApp_i$ is not significant on average. One possible explanation for this is that the effect of signal strength on the use of the mobile app is heterogeneous among people. For example, people whose app-use behavior can be switched by the mobile signal constraint are less likely to have other access to digital technologies, like WiFi at home or workspace. With little exposure to digital technologies, they are more likely to experience difficulties in using the app so that the app can not substitute the important communication skills correlated with education. So we might be overconfident in the effect of the app use itself and the mitigating effect of it has been underestimated in the model with endogenous app adoption. This issue is addressed in the next subsection by allowing the first stage coefficients to vary by neighborhood.

Table 3.4: Instrumental Variables Estimation

Panel A: Logit Approximation of Cox Model

Moderating Factor	(1) None	(2) Address	(3) #Total	(4) #Filtered
Average Year of Education	0.007***	0.006***	0.007***	0.007***
	(0.000)	(0.000)	(0.001)	(0.001)
Citizen Connect App	-0.016***	-0.011***	-0.016***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
Average Years of Education	-0.002***	-0.002**	-0.002**	-0.002***
× Citizen Connect App	(0.000)	(0.001)	(0.001)	(0.001)
Moderating Factor		-0.054***	-0.035***	-0.002***
		(0.002)	(0.001)	(0.001)
Moderating Factor		0.006***	0.002***	0.002**
× Average Year of Education		(0.001)	(0.000)	(0.000)
Moderating Factor		0.050***	0.023***	-0.018***
× Citizen Connect App		(0.003)	(0.001)	(0.002)
Moderating Factor		-0.008***	-0.003***	-0.003***
× Avg. Education × App		(0.002)	(0.001)	(0.001)

Panel B: Instrumental Variables Estimation (2SLS)

Moderating Factor	(1) None	(2) Address	(3) #Total	(4) #Filtered
Average Year of Education	0.008***	0.008***	0.008***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Citizen Connect App	-0.142***	-0.152***	-0.164***	-0.142***
	(0.050)	(0.055)	(0.050)	(0.050)
Average Years of Education	0.002	-0.000	0.003	0.002
× Citizen Connect App	(0.010)	(0.012)	(0.010)	(0.010)
Moderating Factor		-0.108***	-0.051***	-0.011***
		(0.008)	(0.002)	(0.003)
Moderating Factor		0.001	0.003***	-0.001
× Average Year of Education		(0.003)	(0.001)	(0.003)
Moderating Factor		0.222***	0.108***	0.017
× Citizen Connect App		(0.031)	(0.011)	(0.013)
Moderating Factor		-0.008	-0.019***	0.007
× Avg. Education × App		(0.016)	(0.005)	(0.005)

Panel C: 2SLS where First Stage Coefficients Vary by Neighborhood

Moderating Factor	(1) None	(2) Address	(3) #Total	(4) #Filtered
Average Year of Education	0.012***	0.011***	0.011***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)
Citizen Connect App	0.224***	0.235***	0.202***	0.234***
	(0.014)	(0.015)	(0.014)	(0.014)
Average Years of Education	-0.049***	-0.053***	-0.046***	-0.049***
× Citizen Connect App	(0.006)	(0.005)	(0.005)	(0.005)
Moderating Factor		-0.090***	-0.047***	-0.005*
		(0.007)	(0.002)	(0.003)
Moderating Factor		0.009***	0.001	-0.002*
× Average Year of Education		(0.003)	(0.001)	(0.001)
Moderating Factor		0.244***	0.106***	0.029**
× Citizen Connect App		(0.026)	(0.011)	(0.012)
Moderating Factor		-0.064***	-0.012***	0.006
× Avg. Education × App		(0.012)	(0.005)	(0.005)
Interactions with Web Use	Yes	Yes	Yes	Yes
Reason Stratification	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Season Fixed Effects	Yes	Yes	Yes	Yes
Weekday Fixed Effects	Yes	Yes	Yes	Yes
Gender Controls	Yes	Yes	Yes	Yes
Age Controls	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes
Language Controls	Yes	Yes	Yes	Yes
Income Controls	Yes	Yes	Yes	Yes
Housing Status Controls	Yes	Yes	Yes	Yes
Observations	1,941,987	1,941,987	1,941,578	1,941,578

Notes: The population variable is time intervals and the dependent variable is whether the case was resolved during the time interval. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6.3 First-Stage Heterogeneity

The instrumental variable estimation described above assumes that the effect of cell tower proximity on app adoption is the same across the population. However, it is possible that some people are more likely to overcome the mobile signal constraint by installing in-home WiFi or switching to a provider which has better network coverage.

As a robustness check, we allow the first stage coefficients to vary by neighborhood, which is equivalent to having a set of interacted instruments where $MinDistance_i$ is interacted with neighborhood dummies (Abadie *et al.* , 2019). To tackle the over-identification caused by the large number of instruments (and controls), we use high-dimensional machine-learning methods to select which instruments and which controls to include (Chernozhukov *et al.* , 2015). Lasso estimators are used twice here: first, to select the controls by estimating a Lasso regression of d_{in} on all the controls; second, to select the instruments and controls by by estimating a Lasso regression of $CitizenConnectaApp_i$ on all the instruments and controls. The final choice of control variable is the union of the controls selected in the two Lasso regressions.

The 2SLS estimates are reported in Panel C of Table 3.4. The first stages for Column (1) are reported in Table 3.A7. Although the Wald statistic does not suggest a stronger first stage, multiple interacted instruments show significant effects on app use. The results shown in Column (1) are consistent with our hypothesis that the first-stage heterogeneity leads to the confusion between the effect of app use on the service performance and its mitigating effect which reduces inequality in the resolution of customer complaints. Now the coefficient of $CitizenConnectApp_i$ becomes positive – the mobile app, on average, improves the complaint resolution performance by 22.4% for those people whose app-use behavior would be shifted by the mobile signal strength. It is different from not only the instrumental variable estimate reported in Panel B but also the marginal effect from the logit estimation shown in Panel A. It suggests that the app might be selected to report some difficult non-emergency complaints. The coefficient of the interaction term between $AverageYearOfEducation_k$ and $CitizenConnectApp_i$ is -0.049, which indicates a strong mitigating effect of the use of mobile app on social inequality in the complaint resolution process. The results shown in Columns (2)-(4) reinforce our conclusion about the underlying mechanism behind this mitigating effect.

3.7 Conclusion

In this study, we investigate the extent to which technology can help reduce inequality in the resolution of customer complaints. Using extensive data from non-emergency complaints issued about public services, we show that for older technologies such as phone calls, complaint resolution tends to be slower for people living in neighborhoods with less educated people. We show that this inequality is mitigated by the use of newer technologies such as mobile apps which help consumers more accurately describe and locate their complaint. Since there are endogeneity concerns surrounding the adoption of new mobile technologies, we confirm this finding using instrumental variables and exogenous variation in mobile app adoption which can be explained by differences in cell tower proximity.

These results matter because usually policy makers and firms might fear that promoting new technologies would lead them to be less inclusive. However, our results show that in our setting mobile communication technologies can actually mitigate potential inequality in the treatment of customers by standardizing communication. It can be generalized to other types of new technology which ease personal interactions by boosting standardized communication. A great example is the menu-based in-app customer complaint resolution system, which fully automates the communication in the app by analyzing keywords and does not require any human interaction.[16]

There are of course limitations to our study. First, we do not have individual customer data on levels of education and instead infer education from the surrounding census block. Second, since this is a public service setting, we are not able to relate our findings to individual customer profitability. Last, we do not know how the availability of a mobile app changed the likelihood of a complaint being reported. Notwithstanding these limitations, we

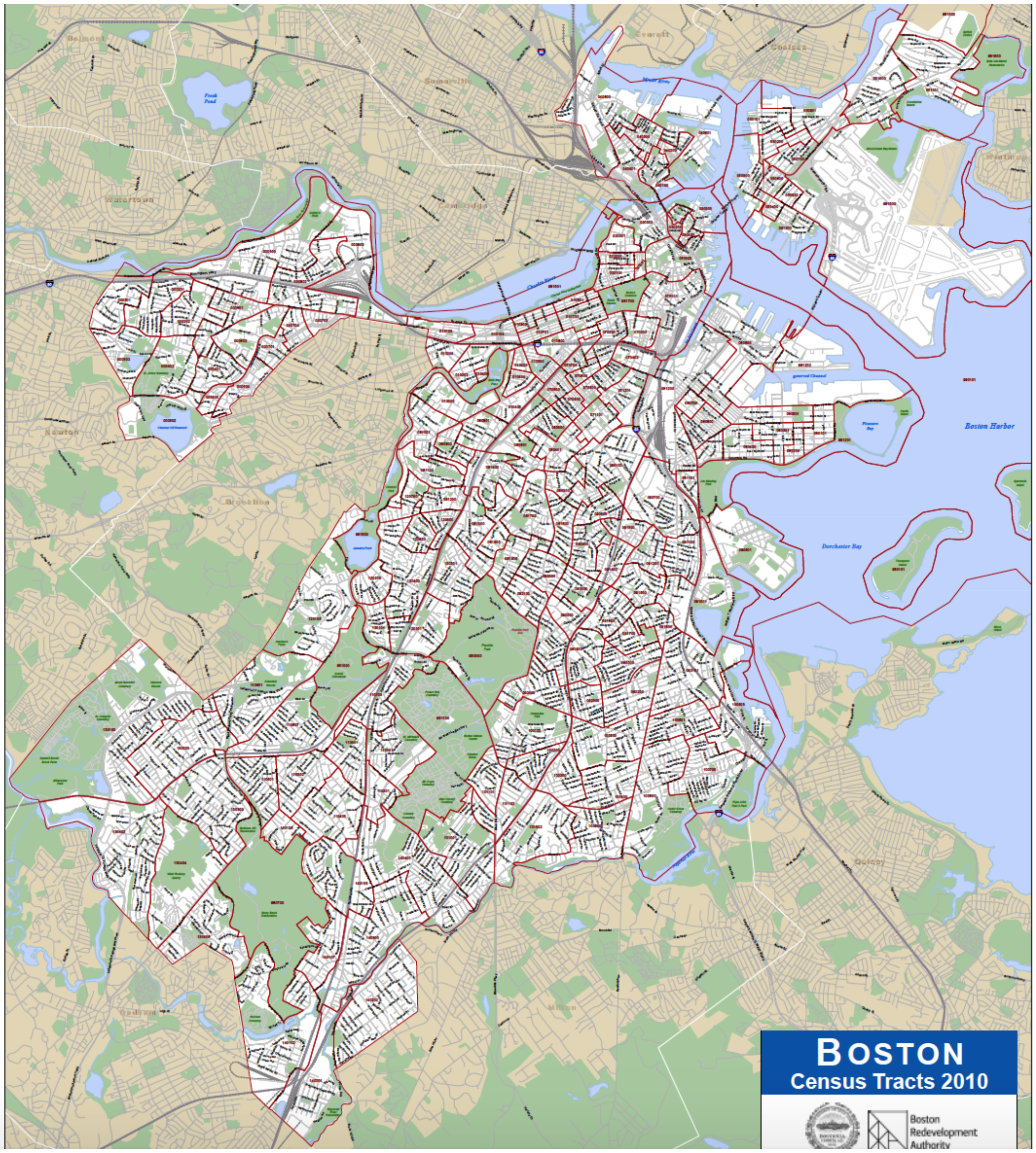
[16] See *Uber's Customer Support is about to Get a Lot Better*: `http://www.businessinsider.com/uber-beefs-up-customer-support-in-app-2016-3`.

feel our study is a useful first step in understanding how technology can alter inequality in the complaint resolution process.

3.8 Appendix

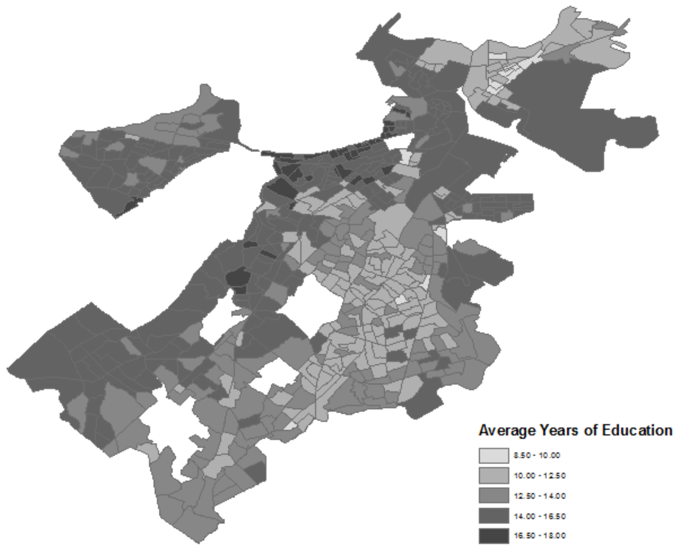
3.8.1 Figures

Figure 3-A1: Boston Census Block Groups Boundary 2010

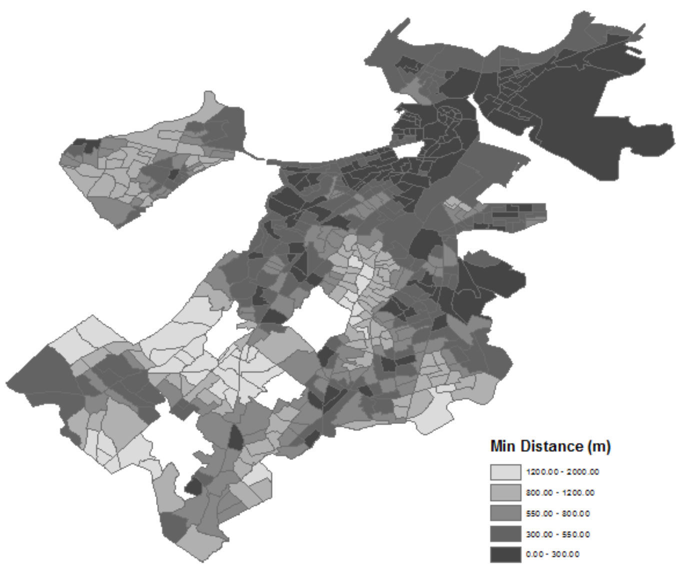


Note: The map shows the boundaries for 646 census block groups in the city of Boston. From Boston Redevelopment Authority: `http://www.bostonredevelopmentauthority.org/research-maps/maps-and-gis/census-and-demographic-maps`

Figure 3-A2: Instrumental Variables: Exclusion Restriction



(a) Average Years of Education



(b) IV: Employee's Choice

Notes: The average years of education are plotted for each census block in Panel A. The average value of the indicator whether a city worker used a mobile to submit a complaint in the same location for an external case is plotted for each census block in Panel B.

3.8.2 Tables

Table 3.A1: Types of Cases

Case Type	Freq.	Percent	Cum. Percent
Sanitation*	109,501	23.56	23.56
Street Cleaning*	98,090	21.11	44.67
Highway Maintenance*	65,052	14.00	58.67
Street Lights*	26,898	5.79	64.46
Recycling	26,684	5.74	70.20
Signs & Signals*	25,558	5.50	75.70
Trees*	22,063	4.75	80.45
Housing*	16,769	3.61	84.06
Graffiti*	14,515	3.12	87.18
Building	13,814	2.97	90.16
Enforcement & Abandoned Vehicles*	12,202	2.63	92.78
Environmental Services*	10,285	2.21	95.00
Administrative & General Requests	5,944	1.28	96.28
Notification	3,655	0.79	97.06
Park Maintenance* & Safety	2,464	0.53	97.59
Health	2,215	0.48	98.07
Catch Basin	1,450	0.31	98.38
Employee & General Comments	1,388	0.30	98.68
Traffic Management & Engineering	1,296	0.28	98.97
Operations	1,203	0.26	99.22
Sidewalk Cover / Manhole	724	0.16	99.37
Abandoned Bicycle*	664	0.14	99.52
Fire Hydrant	591	0.13	99.64
General Request*	320	0.07	99.71
Code Enforcement*	259	0.06	99.77
Weights and Measures	237	0.05	99.82
Water Issues	168	0.04	99.85
Needle Program*	150	0.03	99.89
Programs	92	0.02	99.91
Animal Issues	91	0.02	99.93
Pothole	77	0.02	99.94
Bridge Maintenance	54	0.01	99.95
Billing	42	0.01	99.96
Boston Bikes	41	0.01	99.97
Parking Complaints	38	0.01	99.98
Fire Department	23	0.00	99.99
Valet	18	0.00	99.99
Office of The Parking Clerk	8	0.00	99.99
Air Pollution Control	7	0.00	99.99
Volunteer & Corporate Groups	7	0.00	99.99
Noise Disturbance	6	0.00	100.00
Administrative	5	0.00	100.00
Disability	4	0.00	100.00
Cemetery	3	0.00	100.00
Generic Noise Disturbance	3	0.00	100.00
Investigations and Enforcement	2	0.00	100.00
Call Center Intake	1	0.00	100.00
Consumer Affairs Issues	1	0.00	100.00
Metrolist	1	0.00	100.00
Total	464,683	100.00	100.00

Table 3.A2: Correlation Matrix

Panel A: Census Block Level

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 Average Years of Education	1.00																
2 %Male	-0.03	1.00															
3 %< 18 years old	-0.57	-0.17	1.00														
4 %18 - 29 years old	0.37	0.03	-0.63	1.00													
5 %30 - 44 years old	0.14	0.18	0.08	-0.39	1.00												
6 %45 - 59 years old	-0.24	0.07	0.25	-0.61	-0.03	1.00											
7 %White	0.68	0.07	-0.58	0.27	0.14	-0.20	1.00										
8 %Black	-0.52	-0.11	0.54	-0.32	-0.14	0.25	-0.87	1.00									
9 %Asian	0.07	0.01	-0.23	0.25	-0.11	-0.16	-0.00	-0.33	1.00								
10 %Other Single Race	-0.44	-0.03	0.40	-0.15	-0.03	0.03	-0.50	0.21	-0.12	1.00							
11 %English Only	0.61	-0.05	-0.28	-0.01	0.12	0.08	0.50	-0.17	-0.27	-0.50	1.00						
12 %Spanish	-0.60	0.05	0.43	-0.17	0.03	0.05	-0.41	0.22	-0.23	0.57	-0.68	1.00					
13 %Limited English Speaking	-0.57	0.02	0.15	-0.01	-0.11	-0.09	-0.32	0.03	0.31	0.30	-0.75	0.48	1.00				
14 Poverty Ratio for Households	-0.36	-0.15	0.17	0.30	-0.40	-0.26	-0.44	0.27	0.22	0.28	-0.48	0.32	0.45	1.00			
15 %Owner-occupied	0.28	0.06	0.04	-0.44	0.21	0.41	0.30	-0.12	-0.26	-0.23	0.47	-0.34	-0.44	-0.70	1.00		
16 %Living in Greater Boston Area 1 Year Ago	0.18	0.04	-0.36	0.49	-0.09	-0.37	0.15	-0.17	0.09	-0.07	-0.01	-0.07	0.04	0.14	-0.31	1.00	
17 %Living Abroad 1 Year Ago	0.17	-0.01	-0.27	0.44	-0.18	-0.29	0.08	-0.17	0.27	-0.04	-0.14	-0.06	0.11	0.18	-0.23	0.20	1.00

Panel B: Individual Case Level

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 Average Years of Education	1.00																
2 %Male	-0.00	1.00															
3 %< 18 years old	-0.53	-0.27	1.00														
4 %18 - 29 years old	0.30	0.13	-0.61	1.00													
5 %30 - 44 years old	0.22	0.14	-0.01	-0.34	1.00												
6 %45 - 59 years old	-0.21	0.02	0.23	-0.59	-0.08	1.00											
7 %White	0.70	0.10	-0.56	0.23	0.23	-0.19	1.00										
8 %Black	-0.54	-0.12	0.53	-0.27	-0.23	0.23	-0.88	1.00									
9 %Asian	0.03	-0.02	-0.20	0.19	-0.09	-0.10	-0.03	-0.28	1.00								
10 %Other Single Race	-0.45	-0.02	0.36	-0.12	-0.07	0.04	-0.53	0.26	-0.12	1.00							
11 %English Only	0.66	-0.03	-0.29	0.02	0.15	0.04	0.55	-0.24	-0.27	-0.51	1.00						
12 %Spanish	-0.60	0.02	0.40	-0.15	-0.02	0.05	-0.44	0.25	-0.21	0.55	-0.69	1.00					
13 %Limited English Speaking -0.61	0.02	0.15	-0.02	-0.13	-0.06	-0.37	0.08	0.32	0.33	-0.77	0.50	1.00					
14 Poverty Ratio for Households -0.41	-0.12	0.16	0.29	-0.42	-0.22	-0.48	0.34	0.19	0.29	-0.49	0.33	0.47	1.00				
15 %Owner-occupied 0.30	-0.03	0.10	-0.47	0.18	0.40	0.30	-0.16	-0.21	-0.21	0.44	-0.32	-0.43	-0.66	1.00			
16 %Living in Greater Boston Area 1 Year Ago	0.14	0.22	-0.39	0.51	-0.00	-0.33	0.13	-0.13	0.03	-0.03	-0.02	-0.05	0.01	0.14	-0.35	1.00	
17 %Living Abroad 1 Year Ago	0.12	0.08	-0.21	0.38	-0.13	-0.25	0.08	-0.16	0.18	-0.00	-0.14	-0.01	0.10	0.15	-0.20	0.20	1.00

Notes: For Panel A, observations=545 and there is one observation for each block group. For Panel B, observations=364,189 and there is one observation for each case.

Table 3.A3: Main Effects (Alternative Model – Logit Regression)

	(1) On-Time	(2) On-Time	(3) On-Time	(4) On-Time	(5) On-Time	(6) On-Time	(7) On-Time
Average Years of Education	0.002	0.005**	0.006**	0.005	0.007	0.020***	0.020***
	(0.002)	(0.003)	(0.003)	(0.005)	(0.005)	(0.006)	(0.006)
Citizen Connect App					-0.040***	-0.022	-0.009
					(0.014)	(0.015)	(0.015)
Web Submission					-0.405***	-0.401***	-0.442***
					(0.014)	(0.014)	(0.015)
Average Years of Education						-0.042***	-0.037***
× Citizen Connect App						(0.007)	(0.007)
Average Years of Education						-0.027***	-0.029***
× Web Submission						(0.008)	(0.009)
Reason Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes
Season Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes
Weekday Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes
Gender Controls	No	No	No	Yes	Yes	Yes	Yes
Age Controls	No	No	No	Yes	Yes	Yes	Yes
Race Controls	No	No	No	Yes	Yes	No	No
Language Controls	No	No	No	Yes	Yes	Yes	Yes
Income Controls	No	No	No	Yes	Yes	Yes	Yes
Housing Status Controls	No	No	No	Yes	Yes	Yes	Yes
Observations	364,189	364,189	364,189	364,189	364,189	364,189	310,736
AIC	397,523	296,664	277,617	277,394	276,550	276,514	240,078

Notes: The population variable is cases and a logit model for the "on-time" indicator is used. There are only 310, 736 observations in Column (7) because some minor categories are removed. Standard errors are in parentheses. sym* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A4: Main Effects (Alternative Independent Variables)

	(1) Hazards	(2) Hazards	(3) Hazards	(4) Hazards	(5) Hazards	(6) Hazards
Citizen Connect App	-0.039***	-0.040***	-0.042***	-0.039***	-0.033***	-0.036***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)
Web Submission	-0.130***	-0.130***	-0.128***	-0.130***	-0.131***	-0.130***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Average Years of Education	0.028***					
	(0.002)					
Average Years of Education × Citizen Connect App	-0.015*** (0.003)					
Average Years of Education × Web Submission	-0.001 (0.003)					
ln(Average Years of Education)		0.356***				
		(0.027)				
ln(Average Years of Education) × Citizen Connect App		-0.181*** (0.038)				
ln(Average Years of Education) × Web Submission		-0.010 (0.035)				
%High School and Above			0.226***			
			(0.025)			
%High School and Above × Citizen Connect App			-0.129*** (0.043)			
%High School and Above × Web Submission			-0.035 (0.039)			
%Some College and Above				0.213***		
				(0.016)		
%Some College and Above × Citizen Connect App				-0.118*** (0.024)		
%Some College and Above × Web Submission				-0.008 (0.021)		
%Bachelor and Above					0.189***	
					(0.014)	
%Bachelor and Above × Citizen Connect App					-0.151*** (0.020)	
%Bachelor and Above × Web Submission					0.005 (0.017)	
%Graduate School						0.215***
						(0.018)
%Graduate School × Citizen Connect App						-0.220*** (0.032)
%Graduate School × Web Submission						-0.049* (0.028)
Reason Stratification	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Season Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Weekday Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes
Language Controls	Yes	Yes	Yes	Yes	Yes	Yes
Income Controls	Yes	Yes	Yes	Yes	Yes	Yes
Housing Status Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	364,189	364,189	364,189	364,189	364,189	364,189
AIC	6,834,811	6,834,829	6,834,932	6,834,823	6,834,801	6,834,850

Notes: The population variable is cases and a hazard model for the completion time is used. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A5: Mechanism (Full Table)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Effect	Citizen Connect App				Photo Attachment		
Moderating Factor	None	Address	#Total	#Filtered	None	#Total	#Filtered
Average Year of Education	0.028***	0.026***	0.028***	0.029***	0.025***	0.025***	0.026***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Citizen Connect App	-0.039***	-0.025***	-0.060***	-0.063***			
	(0.006)	(0.007)	(0.006)	(0.007)			
Web Submission	-0.142***	-0.098***	-0.148***	-0.118***	-0.126***	-0.140***	-0.114***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Average Years of Education	-0.015***	-0.014***	-0.013***	-0.015***			
× Citizen Connect App	(0.003)	(0.003)	(0.003)	(0.003)			
Average Years of Education	-0.001	-0.007***	-0.002	-0.006**	0.001	0.000	-0.003
× Web Submission	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Photo					-0.040***	-0.061***	-0.058***
					(0.009)	(0.009)	(0.010)
Average Years of Education					-0.006	-0.004	-0.007
× Photo					(0.005)	(0.005)	(0.005)
Moderating Factor		-0.198***	-0.147***	-0.034***		-0.143***	-0.060***
		(0.006)	(0.003)	(0.003)		(0.002)	(0.003)
Moderating Factor		0.023***	0.012***	0.013***		0.011***	0.009***
× Average Year of Education		(0.003)	(0.001)	(0.001)		(0.001)	(0.001)
Moderating Factor		0.185***	0.038***	-0.093***			
× Citizen Connect App		(0.012)	(0.005)	(0.006)			
Moderating Factor		0.247***	-0.022***	-0.043***		-0.025***	-0.027***
× Web Submission		(0.014)	(0.005)	(0.005)		(0.004)	(0.005)
Moderating Factor		-0.026***	-0.011***	-0.013***			
× Avg. Education × App		(0.006)	(0.002)	(0.003)			
Moderating Factor		0.006	0.000	0.003		0.001	0.007**
× Avg. Education × Web		(0.008)	(0.003)	(0.003)		(0.003)	(0.003)
Moderating Factor						0.064***	-0.030***
× Photo						(0.009)	(0.011)
Moderating Factor						-0.009**	-0.014**
× Avg. Education × Photo						(0.005)	(0.006)
Reason Stratification	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weekday Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Language Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing Status Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations Observations	364,189	364,189	364,149	364,149	364,189	364,149	364,149

Notes: The population variable is cases and a hazard model for the completion time is used. The presence of missing values leads to a smaller sample size in Columns (3), (4), (6), and (7). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table 3.A6: Mechanism (Language Proficiency)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Effect	Citizen Connect App				Photo Attachment		
Moderating Factor	None	Address	#Total	#Filtered	None	#Total	#Filtered
% Spanish	0.026	0.032	0.029	0.025	0.045**	0.043**	0.043**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Citizen Connect App	-0.044***	-0.030***	-0.063***	-0.068***			
	(0.006)	(0.007)	(0.006)	(0.006)			
Web Submission	-0.130***	-0.141***	-0.147***	-0.119***	-0.125***	-0.138***	-0.115***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
% Spanish	0.185***	0.189***	0.163***	0.188***			
× Citizen Connect App	(0.038)	(0.040)	(0.038)	(0.038)			
% Spanish	-0.012	-0.016	-0.010	-0.013	-0.036	-0.031	-0.036
× Web Submission	(0.033)	(0.034)	(0.033)	(0.034)	(0.033)	(0.033)	(0.034)
Photo					-0.042***	-0.060***	-0.059***
					(0.009)	(0.009)	(0.009)
% Spanish					0.111*	0.104*	0.131**
× Photo					(0.060)	(0.061)	(0.062)
Moderating Factor		-0.197***	-0.148***	-0.034***		-0.143***	-0.059***
		(0.006)	(0.003)	(0.003)		(0.002)	(0.003)
Moderating Factor		-0.187***	-0.025**	-0.040**		-0.024**	-0.027*
× % Spanish		(0.037)	(0.011)	(0.016)		(0.011)	(0.014)
Moderating Factor		0.184***	0.040***	-0.093***			
× Citizen Connect App		(0.012)	(0.005)	(0.006)			
Moderating Factor		0.253***	-0.017***	-0.036***		-0.021***	-0.022***
× Web Submission		(0.014)	(0.004)	(0.005)		(0.004)	(0.005)
Moderating Factor		0.267***	0.038	0.053			
× % Spanish × App		(0.081)	(0.031)	(0.036)			
Moderating Factor		-0.006	-0.011	0.046		-0.010	0.031
× % Spanish × Web		(0.101)	(0.033)	(0.040)		(0.033)	(0.040)
Moderating Factor						0.066***	-0.033***
× Photo						(0.009)	(0.010)
Moderating Factor						0.064	0.102
× % Spanish × Photo						(0.061)	(0.071)
Reason Stratification	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weekday Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Language Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Housing Status Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	364,189	364,189	364,149	364,149	364,189	364,149	364,149

Notes: The population variable is cases and a hazard model for the completion time is used. The presence of missing values leads to a smaller sample size in Columns (3), (4), (6), and (7). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table 3.A7: Instrumental Variables Estimation (First Stages)

	(1)	(2)	(3)	(4)
	App	Interaction	App (Lasso)	Interaction (Lasso)
Min Distance (in km)	-0.012***	-0.022***	-0.004***	-0.010***
	(0.000)	(0.001)	(0.000)	(0.001)
Average Year of Education	-0.005***	-0.046***	-0.003***	-0.042***
× Min Distance (in km)	(0.000)	(0.000)	(0.000)	(0.000)
Interactions with Web Use	Yes	Yes	Yes	Yes
Reason Stratification	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Season Fixed Effects	Yes	Yes	Yes	Yes
Weekday Fixed Effects	Yes	Yes	Yes	Yes
Gender Fixed Effects	Yes	Yes	Yes	Yes
Age Fixed Effects	Yes	Yes	Yes	Yes
Race Fixed Effects	Yes	Yes	Yes	Yes
Language Fixed Effects	Yes	Yes	Yes	Yes
Income Fixed Effects	Yes	Yes	Yes	Yes
Housing Status Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,941,987	1,941,987	1,941,987	1,941,987
Wald Statistic	539.83		444.58	

Notes: The population variable is time intervals and the dependent variable is whether the case was resolved during the time interval. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Bibliography

Aaronson, Daniel, Hartley, Daniel A, & Mazumder, Bhashkar. 2017. *The Effects of the 1930s HOLC "Redlining" Maps*. FRB of Chicago Working Paper No. WP-2017-12.

Abadie, Alberto, Gu, Jiaying, & Shen, Shu. 2019. *Instrumental Variable Estimation with First Stage Heterogeneity*. Working paper.

Acemoglu, Daron. 1998. Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly Journal of Economics*, **113**(4), 1055–1089.

Acemoglu, Daron. 2002. Technical change, inequality, and the labor market. *Journal of Economic Literature*, **40**(1), 7–72.

Afshartous, David, & Preston, Richard A. 2011. Key results of interaction models with centering. *Journal of Statistics Education*, **19**(3), 1–24.

Ali, Muhammad, Sapiezynski, Piotr, Bogen, Miranda, Korolova, Aleksandra, Mislove, Alan, & Rieke, Aaron. 2019. *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*. arXiv preprint arXiv:1904.02095.

Anand, Punam, & Sternthal, Brian. 1990. Ease of message processing as a moderator of repetition effects in advertising. *Journal of Marketing Research*, **27**(3), 345–353.

Anderson, Eric T., & Simester, Duncan I. 2004. Long-run effects of promotion depth on new versus established customers: Three field studies. *Marketing Science*, **23**(1), 4–20.

Andreassen, Tor Wallin. 1999. What drives customer loyalty with complaint resolution? *Journal of Service Research*, **1**(4), 324–332.

Angwin, Julia, Larson, Jeff, Mattu, Surya, & Kirchner, Lauren. 2016. Machine bias. *ProPublica*, **23**, 2016.

Ansari, Asim, Essegaier, Skander, & Kohli, Rajeev. 2000. Internet recommendation systems. *Journal of Marketing Research*, **37**(3), 363–375.

Ascarza, Eva, & Hardie, Bruce GS. 2013. A joint model of usage and churn in contractual settings. *Marketing Science*, **32**(4), 570–590.

Ascarza, Eva, Neslin, Scott A, Netzer, Oded, Anderson, Zachery, Fader, Peter S, Gupta, Sunil, Hardie, Bruce GS, Lemmens, Aurélie, Libai, Barak, Neal, David, *et al.* . 2018. In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, **5**(1), 65–81.

Bajari, Patrick, Chernozhukov, Victor, Hortaçsu, Ali, & Suzuki, Junichi. 2019. The impact of big data on firm performance: An empirical investigation. *Pages 33–37 of: AEA Papers and Proceedings*, vol. 109.

Bartel, Ann, Ichniowski, Casey, & Shaw, Kathryn. 2007. How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills. *Quarterly Journal of Economics*, **122**(4), 1721–1758.

Batra, Rajeev, & Ray, Michael L. 1986. Situational effects of advertising repetition: The moderating influence of motivation, ability, and opportunity to respond. *Journal of Consumer Research*, **12**(4), 432–445.

Bayer, Patrick, Casey, Marcus, Ferreira, Fernando, & McMillan, Robert. 2017. Racial and ethnic price differentials in the housing market. *Journal of Urban Economics*, **102**, 91–105.

Belleflamme, Paul, Lam, Wing Man Wynne, & Vergote, Wouter. 2020. Competitive imperfect price discrimination and market power. *Marketing Science*, **39**(5), 996–1015.

Berkman, Nancy D, Sheridan, Stacey L, Donahue, Katrina E, Halpern, David J, & Crotty, Karen. 2011. Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, **155**(2), 97–107.

Berndt, Donald J, & Clifford, James. 1994. Using dynamic time warping to find patterns in time series. *Pages 359–370 of: KDD Workshop*, vol. 10. Seattle, WA, USA.

Berry, Brian JL. 1976. Ghetto expansion and single-family housing prices: Chicago, 1968–1972. *Journal of Urban Economics*, **3**(4), 397–423.

Black, Sandra E. 1999. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, **114**(2), 577–599.

Black, Sandra E., & Lynch, Lisa M. 2001. How to Compete: The Impact of Workplace Practices and Information Technology on Productivity. *Review of Economics and Statistics*, **83**(3), 434–445.

Blodgett, Jeffrey G, Wakefield, Kirk L, & Barnes, James H. 1995. The effects of customer service on consumer complaining behavior. *Journal of Services Marketing*, **9**(4), 31–42.

Bloom, Nicholas, Sadun, Raffaella, & Van Reenen, John. 2012. Americans Do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review*, **102**(1), 167–201.

Bloom, Nicholas, Garicano, Luis, Sadun, Raffaella, & Reenen, John Van. 2014. The Distinct Effects of Information Technology and Communication Technology on Firm Organization. *Management Science*, **60**(12), 2859–2885.

Bresnahan, Timothy F., Brynjolfsson, Erik, & Hitt, Lorin M. 2002. Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *Quarterly Journal of Economics*, **117**(1), 339–376.

Brown, Jeffrey R, & Goolsbee, Austan. 2002. Does the Internet make markets more competitive? Evidence from the life insurance industry. *Journal of Political Economy*, **110**(3), 481–507.

Buell, Ryan W, Porter, Ethan, & Norton, Michael I. 2017. *Surfacing the Submerged State: Operational Transparency Increases Trust in and Engagement with Government*. Working paper.

Bughin, Jacques, Seong, Jeongmin, Manyika, James, Chui, Michael, & Joshi, Raoul. 2018. *Notes From the AI Frontier: Modeling the Impact of AI on the World Economy*. McKinsey Global Institute.

Bulow, Jeremy I, Geanakoplos, John D, & Klemperer, Paul D. 1985. Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy*, **93**(3), 488–511.

Cacioppo, John T, & Petty, Richard E. 1979. Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of Personality and Social Psychology*, **37**(1), 97.

Chambers, Daniel N. 1992. The racial housing price differential and racially transitional neighborhoods. *Journal of Urban Economics*, **32**(2), 214–232.

Chernozhukov, Victor, Hansen, Christian, & Spindler, Martin. 2015. Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review*, **105**(5), 486–90.

Chinn, Menzie D., & Fairlie, Robert W. 2007. The determinants of the global digital divide: A cross-country analysis of computer and internet penetration. *Oxford Economic Papers*, **59**(1), 16–44.

Choe, Chongwoo, Matsushima, Noriaki, & Tremblay, Mark J. 2020. *Behavior-Based Personalized Pricing: When Firms Can Share Customer Information*. Working Paper.

Clark, Benjamin Y, Brudney, Jeffrey L, & Jang, Sung-Gheel. 2013. Coproduction of government services and the new information technology: Investigating the distributional biases. *Public Administration Review*, **73**(5), 687–701.

Claxton, John D, Fry, Joseph N, & Portis, Bernard. 1974. A taxonomy of prepurchase information gathering patterns. *Journal of Consumer Research*, **1**(3), 35–42.

Cohen, Jacob, C. Cohen, P, G. West, Stephen, & S. Aiken, Leona. 2003. *Applied Multiple Regression/Correlation Analysis For The Behavioral Sciences*. Routledge.

Cowgill, Bo, & Tucker, Catherine E. 2020. Algorithmic fairness and economics. *Journal of Economic Perspectives*.

Cox, David R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2), 87–22.

Cui, Ruomeng, Li, Jun, & Zhang, Dennis J. 2020. Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Science*, **66**(3), 1071–1094.

Davidow, Moshe. 2003. Organizational responses to customer complaints: What works and what doesn't. *Journal of Service Research*, **5**(3), 225–250.

Dietvorst, Berkeley J, Simmons, Joseph P, & Massey, Cade. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, **144**(1), 114.

Dotzel, Thomas, Shankar, Venkatesh, & Berry, Leonard L. 2013. Service innovativeness and firm value. *Journal of Marketing Research*, **50**(2), 259–276.

Dubé, Jean-Pierre, & Misra, Sanjog. 2017. *Scalable Price Targeting*. NBER Working Paper.

Edelman, Benjamin, Luca, Michael, & Svirsky, Dan. 2017. Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, **9**(2), 1–22.

Ellison, Glenn, & Ellison, Sara Fisher. 2009. Search, obfuscation, and price elasticities on the Internet. *Econometrica*, **77**(2), 427–452.

Fornell, Claes, & Wernerfelt, Birger. 1987. Defensive marketing strategy by customer complaint management: A theoretical analysis. *Journal of Marketing Research*, **24**(4), 337–346.

Fornell, Claes, & Wernerfelt, Birger. 1988. A model for customer complaint management. *Marketing Science*, **7**(3), 287–298.

Fox, John, & Weisberg, Sanford. 2010. *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA.

Fudenberg, Drew, & Villas-Boas, J Miguel. 2006. Behavior-based price discrimination and customer recognition. *Handbook on Economics and Information Systems*, **1**, 377–436.

Godinho de Matos, Miguel, Pedro, Ferreira, & Rodrigo, Belo. 2018. Target the ego or target the group: Evidence from a randomized experiment in proactive churn management. *Marketing Science*, **37**(5), 793–811.

Goldfarb, Avi, & Prince, Jeff. 2008. Internet adoption and usage patterns are different: Implications for the digital divide. *Information Economics and Policy*, **20**(1), 2 – 15.

Gönül, Füsun, & Shi, Meng Ze. 1998. Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models. *Management Science*, **44**(9), 1249–1262.

Graef, Inge. 2017. Algorithms and fairness: What role for competition law in targeting price discrimination towards end consumers? *Columbia Journal of European Law*, **24**, 541.

Grunes, Allen P, & Stucke, Maurice E. 2015. No mistake about it: The important role of antitrust in the era of big data. *Antitrust Source*.

Gu, Yiquan, Madio, Leonardo, & Reggiani, Carlo. 2019. *Exclusive Data, Price Manipulation and Market Leadership*. CESifo Working Paper.

Hauser, John R, Urban, Glen L, & Weinberg, Bruce D. 1993. How consumers allocate their time when searching for information. *Journal of Marketing Research*, **30**(4), 452–466.

Hauser, John R, Urban, Glen L, Liberali, Guilherme, & Braun, Michael. 2009. Website morphing. *Marketing Science*, **28**(2), 202–223.

Hellwig, Christian, & Veldkamp, Laura. 2009. Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*, **76**(1), 223–251.

Hendel, Igal, Nevo, Aviv, & Ortalo-Magné, François. 2009. The relative performance of real estate marketing platforms: MLS versus FSBOMadison. com. *American Economic Review*, **99**(5), 1878–98.

Holford, Theodore R. 1980. The analysis of rates and of survivorship using log-linear models. *Biometrics*, **36**(2), 299–305.

Homburg, Christian, & Fürst, Andreas. 2005. How organizational complaint handling drives customer loyalty: An analysis of the mechanistic and the organic approach. *Journal of Marketing*, **69**(3), 95–114.

Ihlanfeldt, Keith, & Mayock, Tom. 2009. Price discrimination in the housing market. *Journal of Urban Economics*, **66**(2), 125–140.

Jackson, Kenneth T. 1987. *Crabgrass Frontier: The Suburbanization of the United States*. Oxford University Press.

Jayachandran, Satish, Sharma, Subhash, Kaufman, Peter, & Raman, Pushkala. 2005. The role of relational information processes and technology use in customer relationship management. *Journal of Marketing*, **69**(4), 177–192.

Jensen, Robert. 2007. The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *Quarterly Journal of Economics*, **122**(3), 879–924.

Jurafsky, Dan, & Martin, James H. 2009. *Speech and Language Processing : An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Kanamori, T, & Shimodaira, H. 2009. Geometry of covariate shift with applications to active learning. *Dataset Shift in Machine Learning*, 87–105.

Kiel, Geoffrey C, & Layton, Roger A. 1981. Dimensions of consumer information seeking behavior. *Journal of Marketing Research*, **18**(2), 233–239.

Kiel, Katherine A, & Zabel, Jeffrey E. 1996. House price differentials in US cities: Household and neighborhood racial effects. *Journal of Housing Economics*, **5**(2), 143–165.

King, A Thomas, & Mieszkowski, Peter. 1973. Racial discrimination, segregation, and the price of housing. *Journal of Political Economy*, **81**(3), 590–606.

Kuruzovich, Jason, Viswanathan, Siva, & Agarwal, Ritu. 2010. Seller search and market outcomes in online auctions. *Management Science*, **56**(10), 1702–1717.

Laird, Nan, & Olivier, Donald. 1981. Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, **76**(374), 231–240.

Lambrecht, Anja, & Tucker, Catherine. 2019. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science*, **65**(7), 2966–2981.

Lambrecht, Anja, & Tucker, Catherine E. 2017. Can big data protect a firm from competition? *CPI Antitrust Chronicle*, **1**(1), 1–8.

Lemmens, Aurélie, & Croux, Christophe. 2006. Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, **43**(2), 276–286.

Logg, Jennifer M, Minson, Julia A, & Moore, Don A. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, **151**, 90–103.

Lowd, Daniel, & Meek, Christopher. 2005. Adversarial learning. *Pages 641–647 of: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining.*

Meuter, Matthew L, Ostrom, Amy L, Roundtree, Robert I, & Bitner, Mary Jo. 2000. Self-service technologies: Understanding customer satisfaction with technology-based service encounters. *Journal of Marketing*, **64**(3), 50–64.

Misra, Sanjog, & Nair, Harikesh S. 2011. A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics*, **9**(3), 211–257.

Moon, Sangkil, & Russell, Gary J. 2008. Predicting product purchase from inferred customer similarity: An autologistic model approach. *Management Science*, **54**(1), 71–82.

Morton, Fiona Scott, Zettelmeyer, Florian, & Silva-Risso, Jorge. 2003. Consumer information and discrimination: Does the Internet affect the pricing of new cars to women and minorities? *Quantitative Marketing and Economics*, **1**(1), 65–92.

Myers, Caitlin Knowles. 2004. Discrimination and neighborhood effects: Understanding racial differentials in US housing prices. *Journal of Urban economics*, **56**(2), 279–302.

Neslin, Scott A, Gupta, Sunil, Kamakura, Wagner, Lu, Junxiang, & Mason, Charlotte H. 2006. Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, **43**(2), 204–211.

Newman, Joseph W, & Staelin, Richard. 1972. Prepurchase information seeking for new cars and major household appliances. *Journal of Marketing Research*, **9**(3), 249–257.

Obermeyer, Ziad, Powers, Brian, Vogeli, Christine, & Mullainathan, Sendhil. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, **366**(6464), 447–453.

Payne, J, Bettman, JR, & Johnson, EJ. 1991. Consumer decision making. *Handbook of Consumer Behaviour*, 50–84.

Perry, Andre, Rothwell, Jonathan, & Harshbarger, David. 2018. The devaluation of assets in black neighborhoods: The case of residential property. *Metropolitan Policy Program at Brookings*.

Ratchford, Brian T, Lee, Myung-Soo, & Talukdar, Debabrata. 2003. The impact of the Internet on information search for automobiles. *Journal of Marketing Research*, **40**(2), 193–209.

Ray, Gautam, Muhanna, Waleed A., & Barney, Jay B. 2005. Information Technology and the Performance of the Customer Service Process: A Resource-Based Analysis. *MIS Quarterly*, **29**(4), 625–652.

Rossi, Peter E, McCulloch, Robert E, & Allenby, Greg M. 1996. The value of purchase history data in target marketing. *Marketing Science*, **15**(4), 321–340.

Rubin, Donald B. 1977. Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational and Behavioral Statistics*, **2**(1), 1–26.

Rust, Roland T, & Huang, Ming-Hui. 2014. The service revolution and the transformation of marketing science. *Marketing Science*, **33**(2), 206–221.

Rutan, Devin Q, & Glass, Michael R. 2018. The lingering effects of neighborhood appraisal: Evaluating redlining's legacy in Pittsburgh. *Professional Geographer*, **70**(3), 339–349.

Sankoff, David, & Kruskal, Joseph. 1983. *Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison*. Addison Wesley.

Schaninger, Charles M, & Sciglimpaglia, Donald. 1981. The influence of cognitive personality traits and demographics on consumer information acquisition. *Journal of Consumer Research*, **8**(2), 208–216.

Schweidel, David A, Bradlow, Eric T, & Fader, Peter S. 2011. Portfolio dynamics for customers of a multiservice provider. *Management Science*, **57**(3), 471–486.

Seagraves, Philip, & Gallimore, Paul. 2013. The gender gap in real estate sales: Negotiation skill or agent selection? *Real Estate Economics*, **41**(3), 600–631.

Shaffer, Greg, & Zhang, Z John. 1995. Competitive coupon targeting. *Marketing Science*, **14**(4), 395–416.

Shaffer, Greg, & Zhang, Z John. 2002. Competitive one-to-one promotions. *Management Science*, **48**(9), 1143–1160.

Shiller, Benjamin Reed. 2020. Approximating purchase propensities and reservation prices from broad consumer tracking. *International Economic Review*, **61**(2), 847–870.

Simester, Duncan, Timoshenko, Artem, & Zoumpoulis, Spyros I. 2020. Targeting prospective customers: Robustness of machine-learning methods to typical data challenges. *Management Science*, **66**(6), 2495–2522.

Simester, Duncan I, Sun, Peng, & Tsitsiklis, John N. 2006. Dynamic catalog mailing policies. *Management science*, **52**(5), 683–696.

Stucke, Maurice E, & Grunes, Allen P. 2016. *Big Data and Competition Policy*. Oxford University Press.

Tafti, Ali, Mithas, Sunil, & Krishnan, M. S. 2013. The Effect of Information Technology–Enabled Flexibility on Formation and Market Value of Alliances. *Management Science*, **59**(1), 207–225.

Tax, Stephen S., Brown, Stephen W., & Chandrashekaran, Murali. 1998. Customer evaluations of service complaint experiences: Implications for relationship marketing. *Journal of Marketing*, **62**(2), 60–76.

Tucker, Catherine. 2019. Digital data, platforms and the usual [antitrust] suspects: Network effects, switching costs, essential facility. *Review of Industrial Organization*, **54**(4), 683–694.

Tucker, Catherine, Zhang, Juanjuan, & Zhu, Ting. 2013. Days on market and home sales. *RAND Journal of Economics*, **44**(2), 337–360.

Verhoef, Peter C, Kannan, Pallassana K, & Inman, J Jeffrey. 2015. From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. *Journal of Retailing*, **91**(2), 174–181.

Westbrook, Robert A, & Fornell, Claes. 1979. Patterns of information source usage among durable goods buyers. *Journal of Marketing Research*, **16**(3), 303–312.

Willems, Sara, De Maesschalck, Stéphanie, Deveugele, Myriam, Derese, Anselme, & De Maeseneer, Jan. 2005. Socio-economic status of the patient and doctor–patient communication: Does it make a difference? *Patient Education and Counseling*, **56**(2), 139–146.

Zettelmeyer, Florian, Morton, Fiona Scott, & Silva-Risso, Jorge. 2006. How the Internet lowers prices: Evidence from matched survey and automobile transaction data. *Journal of Marketing Research*, **43**(2), 168–181.

Zhang, Dennis J, Dai, Hengchen, Dong, Lingxiu, Qi, Fangfang, Zhang, Nannan, Liu, Xiaofei, Liu, Zhongyi, & Yang, Jiang. 2018. How do price promotions affect customer behavior on retailing platforms? Evidence from a large randomized experiment on Alibaba. *Production and Operations Management*, **27**(12), 2343–2345.

Zhang, Jie, & Krishnamurthi, Lakshman. 2004. Customizing promotions in online stores. *Marketing Science*, **23**(4), 561–578.

Zhang, Jonathan Z, Netzer, Oded, & Ansari, Asim. 2014. Dynamic targeted pricing in B2B relationships. *Marketing Science*, **33**(3), 317–337.

Zimmerman, Emily B, Woolf, Steven H, & Haley, Amber. 2015. *Understanding the Relationship Between Education and Health: A Review of the Evidence and an Examination of Community Perspectives*. Agency for Healthcare Research and Quality and Office of Behavioral and Social Sciences Research, National Institutes of Health.