Who Should be Subsidized for Electric Vehicles? Demand Estimation and Policy Design under Network Effects

Jiarui Liu*

September 26, 2022

Click here to download the latest draft

Abstract

I quantify the heterogeneous network effects and price elasticities in consumers' Electric Vehicles (EV) purchases. I consider both direct network effects from social influences and indirect effects from charging stations. Given the estimated effects, I design targeted pricing policies for the firms. To examine the equilibrium effects of counterfactual policies under social influences, I model consumers' decisions jointly as equilibrium outcomes. Multiple equilibria might arise, posing challenges to estimation and identification. I show that if the average social influence effect is within my derived bounds, then unique equilibrium is guaranteed even under counterfactual policies. I investigate whether the data patterns suggest unique equilibrium; if so, estimation requires searching for parameters within the derived bounds. Another challenge comes from the endogenous charging stations and prices. I construct instruments for the endogenous variables and show identification of each effect. Using zip code level vehicles and charging stations data in Texas over six years, I find positive heterogeneous social influence effects; moreover, ignoring social influence effects under-estimates the mean price elasticities by 11%. The socio-economically disadvantaged group is less affected by social influences and more sensitive to prices. I design a profit-maximizing targeted pricing policy which charges this group \$15k less than the others on average. Under the recommended policy, the average firm's annual sales increase by 35.7% and annual profits increase by 2.7%. The recommended policy improves not only firms' profits but also distributional equity among consumers.

1 Introduction

The transportation sector emits the most greenhouse gas of all economic sectors in the United States: 1,632 million metric tons in 2020 (EPA, 2022). In addition, 66% of the United States petroleum consumption goes to the transportation sector, amounting to 12 million barrels per

*University of Chicago; jiarui@uchicago.edu. I am grateful to my advisors Pradeep K. Chintagunta, Ali Hortaçsu and Alexander Torgovitsky for their invaluable guidance and support. Thanks go to Giovanni Compiani, Jing Li and participants at various seminars for helpful comments. All errors are my own.

day in 2020 (EIA, 2022). One potentially effective solution to the environmental and energy challenges is Electric Vehicles (EV). Thus, governments implement many policies to incentivize EV purchases. For example, the 2022 Inflation Reduction Act offers $7,500 in tax credits to EV owners with annual income less than $150,000. By subsidizing the lower-income consumers, such policy also improves distributional equity among consumers.

In addition to the governments' efforts, there are increasing interests from firms to design policies that are equitable and sustainable. According to the 2021 Fortune/Deloitte CEO Survey, 94% of CEOs indicated that Diversity, Equity and Inclusion (DEI) was a personal strategic priority, and 90% agreed that their organization aspired to be a leader on the topic[1]. More than 90% of S&P 500 companies now publish Environmental, Social, and Governance (ESG) reports in some form, as do approximately 70% of Russell 1000 companies[2]. Therefore, an important question arises: can firms design profitable pricing policies that are also equitable in the EV market?

In order to design targeted pricing policies in the EV market, we should account for heterogeneous direct and indirect network effects. The direct network effects arise from social influences. The choices of neighbors directly affect consumers' utilities through social norms[3]. The choices of neighbors indirectly affect consumers' purchase decisions through neighborhood charging stations, which constitutes the indirect network effects. Most of the literature on EV focuses on the indirect effects from charging stations, yet the direct effects from social influences have not been well-studied.

In this paper, I quantify the heterogeneous network effects and price elasticities across consumer demographic groups. Given the estimated effects, I design optimal targeted pricing policies for the firms taking into account the heterogeneous network effects and price elasticities. I address three methodological and empirical challenges.

The first challenge is modeling social influences under a new policy. Under a new pricing

[1] https://www2.deloitte.com/us/en/insights/topics/marketing-and-sales-operations/global-marketing-trends/2022/diversity-and-inclusion-in-marketing.html

[2] https://www.mckinsey.com/capabilities/sustainability/our-insights/does-esg-really-matter-and-why

[3] See Section 3.1 for more detailed discussions about social influences effects in EV purchases.

policy, neighbors' decisions change. Therefore, we need a structural model with social influences in order to model neighbors' decisions endogenously. I build a discrete choice model with heterogeneous social influences which models all consumers' decisions jointly as equilibrium outcomes.

Identification and estimation of such model pose another challenge: potential multiple equilibria. Multiple equilibria happens when one set of parameters and market characteristics can be consistent with multiple equilibrium market shares. Intuitively, suppose social influence effects are so strong that consumers' purchase decisions only depend on the neighborhood market shares, as an extreme example. The consumers can all be satisfied with either very high or very low market shares in equilibrium. With potential multiple equilibria, usual demand estimation strategies no longer work because researchers do not know which equilibrium is selected by consumers.

To address the concerns of potential multiple equilibria, I first investigate whether the data patterns are consistent with unique equilibrium. Formally, we can use hypothesis testing to evaluate whether the conditional choice probabilities are homogeneous across neighborhoods (De Paula and Tang, 2012; Otsu et al., 2016; Bugni et al., 2020). Informally, we can compare the market shares of neighborhoods with similar characteristics. If there is unique equilibrium in the data, then these market shares should be similar; if the observed market shares in the data are very different, then multiple equilibria likely exist. In my data, I find patterns that are consistent with unique equilibrium.

However, even if the data exhibits unique equilibrium, there can still be multiple equilibria under a new counterfactual policy. The number of equilibria depends on the values of both the parameters and the market characteristics. Under a new pricing policy, the number of equilibria can change from unique to multiple, then the researchers still do not know which equilibrium will be selected by the consumers.

Therefore, I provide theoretical conditions under which unique equilibrium is guaranteed given *any* market characteristics; this guarantees unique equilibrium even under a new counterfactual policy. In particular, I show that if the average social influence effect is within my derived bounds, then unique equilibrium is guaranteed. The derived bounds measure the variance of consumers' idiosyncratic taste shocks. The intuition is: if the social influence effects completely dominate consumers' private heterogeneity, then consumers can be equally satisfied with both very high and very low EV shares. To summarize, since my data patterns suggest unique equilibrium, I only search for candidate parameters within the derived bounds in the estimation.

The third challenge is the identification of the effects of endogenous variables, because the endogenous variables can be correlated with unobserved demand factors. The endogenous variables include: car prices, neighborhood market shares (direct network effects), and neighborhood charging stations (indirect network effects). I show that the effects of the endogenous variables can be separately identified by their respective instruments. For price instruments, I use exogenous characteristics of competing products, also known as the BLP instruments (Berry et al., 1995). For market share instruments, I use one normalized exogenous car characteristics - miles per dollar[4]. Since the coefficient of miles per dollar is normalized, its exogenous variations identify the direct network effects from neighborhood market shares. For charging stations instruments, I propose a novel set of instruments which leverage the exogenous location assignment of new charging stations.

The data consists of vehicle prices and characteristics, zip code level vehicle sales, and charging stations information in Texas from 2015 to 2020. I find that the social influence effect is significantly positive for the baseline consumer with average neighborhood income and no college degree. For example, if the neighborhood market share of Chevrolet Volt increases by 1 percentage point, the baseline consumer's utility increase from the associated social influence effects is equivalent to a decrease of $66 in price. Moreover, socio-economically disadvantaged consumers with lower income or lower education are less affected by social influences; they are also more sensitive to prices. Lastly, if we do not account for social influences, then the mean

[4]This is similar argument as with essential instruments in Berry and Haile (2014).

absolute price elasticities are under-estimated by 11%.

Given the estimated heterogeneous network effects and price elasticities, I design targeted pricing policies by consumer demographics. I segment consumers into a disadvantaged and an advantaged group based on their income and education. Given the segmentation, I find targeted prices that maximize firms' profits subject to the constraint that the targeted prices for all consumers do not exceed their observed uniform prices in the data. I find that on average, firms should charge $41,800 for the advantaged group and $19,150 for the disadvantaged group. The average firm's annual sales increase by 35.68% and annual profits increase by 2.71%. Under the recommended policy, both firms' profits and equity among consumers improve. Thus, profitable firm policies can also be equitable, even when there are heterogeneous network effects.

The rest of the paper proceeds as the follows. Section 2 discusses relevant literature. Section 3 describes the empirical setting and the data. Section 4 presents the empirical model and explains how the multiple equilibria concern is addressed. Section 5 shows the GMM estimator and its identification. Section 6 presents the estimation results, and Section 7 describes the counterfactual policy design of optimal targeted group pricing. Section 8 concludes.

2 Relevant Literature

This paper is closely related to a growing literature on electric vehicles and alternative fuel vehicles. Much attention has been devoted to studying the indirect network effects (Narayanan and Nair, 2013; Shriver, 2015; Li et al., 2017; Remmy et al., 2022), and evaluating the effectiveness of various policies promoting EV adoption, including tax credit incentives and subsidies (Springel, 2016; He et al., 2022), unifying incompatible charging standards (Li, 2019), zero emissions vehicle regulation (Sinyashin, 2021). I contribute to this literature by studying the direct network effects - social influence - in consumers' decisions to purchase EVs.

Secondly, this paper is related to the literature on social interactions and peer effects[5]. Peer

[5]For a detailed review of social interactions and peer effects, see Blume et al. (2011), De Paula (2017), Epple and Romano (2011), Chyn and Katz (2021).

effects from both peers' installed base (Bollinger and Gillingham, 2012; Narayanan and Nair, 2013; Bollinger et al., 2021) and contemporaneous peers' choices (Nair et al., 2010; De Giorgi et al., 2010) have been well-studied. I supplement this literature by modeling peer's choices endogenously in order to perform counterfactual policy analysis.

The seminal paper of Manski (1993) lays out three major challenges in estimating social effects: the endogenous effects, contextual effects, and correlated effects. Since then, more progress has been made in linear models (Manski, 1993; Bramoullé et al., 2009) and binary models (Brock and Durlauf, 2001; Bhattacharya et al., 2021) with homogeneous coefficients. This paper studies multinomial choice models with random coefficient. Brock and Durlauf (2002, 2006) and Bayer and Timmins (2007) study multinomial choice models, but abstract away from multiple equilibria concerns. Allende (2021) studies social interactions in school choice with a multinomial choice and random coefficient model. She focuses more on the contextual effects of peers (peers' characteristics), whereas this paper focuses more on the endogenous effects (neighbors' choices). This paper also explicitly addresses with the multiple equilibria concern and provides theoretical conditions under which unique equilibrium is guaranteed for *any* market characteristics. Guerra and Mohnen (2020) studies occupational choice in Victorian London with a multinomial choice model using data on individual level outcomes and information on network structure. They use a model with homogeneous preferences and social interactions effects, whereas this paper incorporates heterogeneous preferences and network effects.

Thirdly, this paper is related to the literature on the estimation of discrete games with incomplete information. As opposed to games with complete information (Bresnahan and Reiss, 1991; Ciliberto and Tamer, 2009), this paper estimates games with incomplete information. Ellickson and Misra (2011) provides a clear summary of various estimation methods. I contribute to this literature by proposing an estimation method that allows for counterfactual policy design.

This paper also builds on the literature of demand estimations (Berry et al., 1995; Berry and

Haile, 2014, 2021; Gandhi and Nevo, 2021; Compiani, 2022; Tebaldi et al., 2021). I contribute to this literature by introducing social influence effects, and establishes identification conditions for the heterogeneous demand elasticities and social influence effects. Hartmann (2010) estimates demand with social influence effects under complete information. This assumption fits his empirical context because the two agents (golfers) are likely to be friends.

Another relevant literature is indirect network effects. Nair et al. (2004) studies the indirect network effects of personal digital assistants. Dubé et al. (2010) studies the indirect network effects of video game consoles. This paper is also related to the diffusion of new products in a network (Heutel and Muehlegger, 2015; Kumar and Sudhir, 2021; Zhang et al., 2021; Conley and Udry, 2010).

3 The Empirical Setting and Data

There are three types of electric vehicles[6]: (1) battery EV (BEV), (2) plug-in hybrids (PHEV), (3) hybrids (HEV). BEVs, also known as all-electric vehicles, use battery packs to store the electrical energy that powers the motor. The batteries are charged by plugging the vehicle in to an electric power source. They are zero-emission vehicles because they produce no tailpipe emissions. PHEVs use batteries to power an electric motor, as well as another fuel, such as gasoline or diesel, to power an internal combustion engine or other propulsion source. They can be refueled by electric power sources or petroleum. HEVs are powered by an internal combustion engine in combination with electric motors that use energy generated from regenerative braking. HEVs are refueled with petroleum instead of external electric power sources[7].

BEVs and PHEVs can be recharged by plugging into an ordinary electrical outlet (level 1), or faster outlets (level 2 and 3). Level 2 outlets can be attached to the outlet typically dedicated to laundry dryers and electric ovens in residential homes, and some employers and owners of shopping malls, restaurants, and hotels have installed Level 2 charging stations as

[6]See https://afdc.energy.gov/vehicles/electric.html for more details about different types of electric vehicles.

[7]See Archsmith et al. (2021); Rapson and Muehlegger (2021); Gillingham (2021); Gillingham et al. (2021) for more detailed discussions on the EV market.

an amenity to their employees and customers. Level 3 is the fastest outlets, and are mostly public stations as they require more power and fixed costs for installation. 82% of consumers' charging events happen at homes, and only 18% are conducted elsewhere (Smart and Schey, 2012).

3.1 Social Influences in EV purchases

One important factor in determining household's adoption of EVs is social influence. Zhuge and Shao (2019) finds that social influence accounts for 9.7% of the importance weights in a survey conducted in China. The other important factors are vehicle price (32.3%), vehicle usage (28.1%), environmental awareness (9.6%), purchase restrictions (12.4%), and traffic restrictions (7.8%). Another survey in UK shows that social influence explains 12.6% of the variance in a factor composition analysis of hybrid vehicle adoption (Ozaki and Sevastyanova, 2011).

Social influence affects vehicle choices through visibility and social norms (Pettifor et al., 2017). Vehicles are highly visible physical products used primarily in public environments, and thus seeing more EVs in the neighborhood can affect consumers' decisions to purchase EVs. McShane et al. (2012) finds that consumers purchase more new cars when they have seen others around them do so recently. Jansson et al. (2017) uses the Sweden national registry data and finds that neighbors have a stronger influence than family or co-workers in the adoption of alternative fuel vehicles. Similar results of neighborhood effects in (hybrid) vehicle purchases can be found in Mau et al. (2008); Zhu and Liu (2013); Grinblatt et al. (2008); Shemesh et al. (2022); Yang and Allenby (2003). Consumers' EV purchases can be influenced by neighbors' purchases through social norm, since vehicle choices can be guided by beliefs as to the social acceptability of owning a particular vehicle. Jansson et al. (2010) find that social norms are a stronger predictor of alternative fuel vehicles adoption in Sweden than education, income and current ownership. Schuitema et al. (2013) find that people who believe that a green image fits with their self-image are more likely to have positive perceptions of EVs.

Moreover, There are ample evidence that social influences affect consumers' energy related behavior. Examples include adoption of solar photovoltaic panels (Gillingham and Bollinger, 2021; Bollinger et al., 2021; Bollinger and Gillingham, 2012), water consumption (Bollinger et al., 2020; Burkhardt et al., 2021), residential energy audits (Gillingham and Tsvetanov, 2018) [8]. Notably, Bollinger and Gillingham (2012) shows evidence of social influences and peer effects in the adoption of solar panels by showing spatial clustering of installations.

We can see similar spatial clustering in the EV market shares in my data. I will describe the data source in Section 3.2. In Figure 1, I plot cumulative sales per 10,000 households as of first quarter of 2021 for all zip codes in Texas. The darker the green color, the more household-weighted sales. The blue dots are all charging stations of Texas. We can see a clear clustering pattern of both the sales and charging stations. The clustering pattern is suggestive evidence of social influence effects. Other possible explanations of the clustering pattern are indirect network effects and unobserved neighborhood characteristics. I will model all these channels in Section 4, and show that these channels are separately identified in Section 5.

3.2 Data

I combine several data sources to study the empirical question at hand. First data source is from the Texas Department of Motor Vehicles (DMV). I observe the VIN number, transaction prices, and mileage information of all vehicles registered at the Texas DMV from 2015 to 2020. More importantly, the data contains the zip code of where the registrant resides[9]. This allows me to study network effects at a granular zip code level. Because the state of Texas requires all residents to register their vehicles in Texas, this data covers the vehicle ownership of all Texan residents.[10]

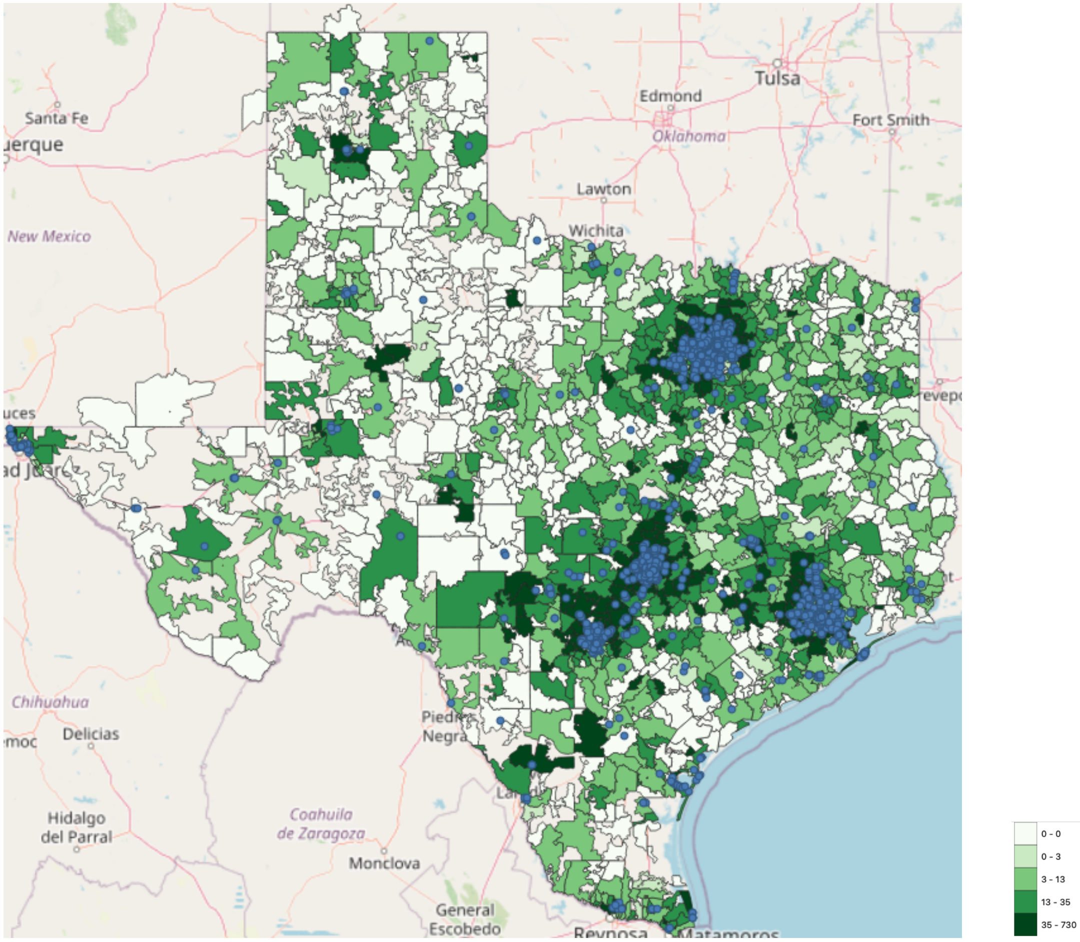
However, the information about the make, model and fuel type of the vehicles in this data set has many missing entries and is sometimes unreliable. Thus, I supplement it with the

[8] See Wolske et al. (2020) for a detailed review.

[9] If the registrant moves, I have both the old and new zip codes.

[10] Note that this data does not contain household level covariates: the vehicle registration does not contain any household demographics information (except for their zip code), and the vehicle characteristics are the same for all households. Thus, we need an estimator that uses market level data, which will be presented in Section 5.1.

Figure 1: Spatial distribution of cumulative EV sales and public charging stations in Texas



Notes: This figure plots all public charging stations (blue dots) and cumulative sales per 10,000 households as of first quarter of 2021 for all zip codes in Texas (green blocks).

second data source: VinAudit.com, Inc., a leading vehicle data and software solutions provider for the U.S. automotive market. More reliable and detailed make and model information is obtained by decoding the VIN number.

I then merge these by car model with the third data source, which is the United States Environmental Protection Agency fuel economy[11]. This data contains vehicle characteristics including MPG and MPG equivalent (MPGe) for both local city streets and highway, battery range, electric motor power, vehicle size, etc.

[11] https://www.fueleconomy.gov/feg/download.shtml

The fourth data source contains charging stations information from the Department of Energy Alternative Fuels Data Center (AFDC)[12]. It includes information of all charging stations in the U.S.: the exact location (in longitude and latitude), the opening date, type of facility that the charger is located in, etc. This allows me to pinpoint all charging stations at high precision, and will be useful for constructing the instruments for charging stations, as will be discussed later in Section 5.3.

A shortcoming of the AFDC data set is that the opening dates of some charging stations are missing. So I supplement it with the fifth data source, PlugShare. PlugShare is an online platform that allows users to find charging stations, check-in online with charging stations and leave reviews. PlugShare is a large and active community of EV drivers with 2 million registered users[13]. I approximate the year in which a station opens with the first check-in date of any EV drivers on PlugShare.

I then collect all federal, state and local incentives from the Department of Energy Alternative Fuels Data Center. In order to construct the MPG and MPGe that are relevant for each zip code, I obtain the local city streets and highway information for each zip code from the Texas Department of Transportation's Roadway Inventory. In order to make MPG and MPGe comparable, I divide them by the annual electricity or gas prices from Texas to construct miles per dollar. The electricity and gas prices are from the U.S. Energy Information Administration.

I then collect additional observed zip code characteristics. The demographics information of each zip code is from American Community Survey. The urban versus rural population in each zip code is from Decennial Census. Whether the zip code is Frontier and Remote Area (FAR) comes from the U.S. Department of Agriculture; FAR describes territory characterized by some combination of low population size and high geographic remoteness.

After merging the eleven data sources from above, I keep only new vehicles and non-obscure make and model[14]. Table 1 presents the summary statistics of all types of EVs in the data. To

[12] https://afdc.energy.gov/data_download

[13] https://www.evgo.com/press-release/plugshare-platform-reaches-2-million-registered-users-worldwide/

[14] I define all models with 6 year cumulative sales less than 1000 as obscure models. I define new vehicles as vehicles with mileage less than 100 miles.

deal with observed zero market shares, I follow Li (2019) and use an empirical bayes estimator. Figure 6 in Appendix A plots the estimated posterior market shares against the observed market shares.

Table 1: Summary statistics

	mean	sd	min	max
BEV dummy	0.19	0.39	0	1
PHEV dummy	0.28	0.45	0	1
Conventional HEV dummy	0.53	0.50	0	1
price ($1000)	47.49	26.73	21.76	149.59
battery range for BEV (miles)	230.98	50.98	132.80	321.73
battery range for PHEV (miles)	19.35	13.85	4.68	44.23
electric motor power for BEV (kw-hrs)	159.36	40.42	110	239
electric motor power for PHEV (kw-hrs)	68.51	32.67	16	135
$>$ midsize dummy	0.35	0.48	0	1
$<$ midsize dummy	0.09	0.29	0	1
urban dummy	0.30	0.45	0	1
FAR dummy	0.11	0.31	0	1
miles per dollar	21.67	7.54	8.98	46.25
market share for each inside product	0.003	0.009	$5e$-12	0.628
number of years	6			
number of zip codes	1,502			
number of households per zip code	5,000			
number of inside products	25			
number of markets	100,084			

Notes This table presents summary statistics of the data from 2015 to 2020. A market is defined as a zip code/product/year combination.

4 The Empirical Framework

4.1 Decisions of each household

To model each household's decision to purchase a new car, I use a static random coefficient discrete choice model with heterogeneous social influences. We observe households from $g = 1, ..., G$ neighborhoods (zip codes) and $t = 1, ..., T$ time periods (years), where each household i belongs to one neighborhood $g(i, t)$ at time t. Each household i chooses one car that maximizes its utility, among available products $j = 0, 1, ..., J_t$. Product $j = 0$ is the outside product of a regular gas car. Products $j = 1, ..., J_t$ are inside products, which include all types of electric vehicles: BEV, PHEV, and HEV. Each inside product is defined as a car brand and EV type. Household i's utility v_{ijt} of purchasing $j > 0$ at time t is:

$$\begin{aligned} v_{ijt} = {} & x'_{g(i,t)jt}\,\beta^x_{it} - \beta^p_{it}\,p_{jt} + c'_{g(i,t)jt}\,\beta^c_{it} + \beta^n_{it}\,n_{g(i,t)t} + \sum_{\tilde{t}=2}^{T} \beta^t_{\tilde{t}}\,\mathbb{1}_{\{\tilde{t}=t\}} \\ & + \gamma_{ijt}\,s^e_{g(i,t)jt} + \xi_{g(i,t)jt} + \epsilon_{ijt} \end{aligned} \tag{1}$$

The utility for the outside product $j = 0$ is normalized as $v_{i0t} = \epsilon_{i0t}$. For an inside product j, each household's utility depends on the log of car's price p_{jt}[15] and observed characteristics $x_{g(i,t)jt}$. The latter includes miles per dollar, battery range, fuel type, vehicle size, and electric motor power. The miles per dollar for each product is computed by dividing the MPG (or MPG equivalent) by the annual gas (or electricity) prices in Texas. Since each car's MPG (or MPG equivalent) is different for highway and local city streets, I compute an average MPG (or MPG equivalent) weighted by each neighborhood's composition of highway and local city streets. Thus, miles per dollar varies across g, j, t.

To capture the indirect network effects β^c_{it} from neighborhood charging stations, I model $c_{g(i,t)jt}$ as the number of charging stations that are potentially useful for the daily activities of households in neighborhood g at time t. According to the National Household Travel Survey[16], an average driver drives 29 miles a day; an analysis by the Idaho National Laboratory finds that an EV driver on average drives 28.8-30.3 miles per day (Smart and Schey, 2012). Thus, I define charging stations within 30 miles of the focal neighborhood g as the potentially useful stations[17]. In order to allow for differential charging station effects for different EV types and neighborhoods, $c_{g(i,t)jt}$ is the (weighted by distance) number of charging stations within 30 miles

[15] This is equivalent to a Cobb-Douglas utility function in expenditures on other products and characteristics of the product purchased. We can think of v as the log-utility. See Berry et al. (1995) for a detailed discussion.

[16] https://www.bts.gov/statistical-products/surveys/national-household-travel-survey-daily-travel-quick-facts

[17] An average neighborhood (zip code) has a radius of ∼5 miles.

of g, interacted with EV type dummies and whether neighborhood g is FAR[18].

The direct network effects γ_{ijt} come from the social influences of neighbors' decisions to purchase vehicles. Since each zip code in Texas has 5000 households on average, households do not necessarily observe the decisions of all neighbors in the zip code. Thus, I model that households form expectations of the market shares of all products in their neighborhood $s^e_{g(i,t)jt}$, and make decisions based on their expectations. Households' information structure to form such expectations will be discussed later in section 4.2.

Other observed components of the household's utility include observed neighborhood characteristics $n_{g(i,t)t}$ and time dummies $\{\mathbb{1}_{\{\tilde{t}=t\}}\}_{\tilde{t}=2,...,T}$. The time dummies intend to account for common time shocks. For the observed neighborhood characteristics, I include whether the neighborhood is a mostly urban area.

The unobserved components of the household's utility include unobserved neighborhood-car demand factors $\xi_{g(i)jt}$ and the household's idiosyncratic taste shock ϵ_{ijt}. Examples of $\xi_{g(i)jt}$ are: the neighborhood's level of environmental awareness, the car culture, the historical backgrounds, the neighborhood's loyalties and perceptions of certain car brands, etc. Therefore, $\xi_{g(i)jt}$ can be correlated with prices p_{jt} and neighborhood charging stations $c_{g(i,t)jt}$: both p_{jt} and $c_{g(i,t)jt}$ are endogenous. Identification of the effects of these endogenous variables requires instruments, which will be discussed in section 5.3.

Households' taste shock ϵ_{ijt} is assumed to be independently and identically distributed as Type-1 extreme value distribution with location parameter 0 and scale parameter α for all i, j, t, conditional on market characteristics. Market characteristics include the unobserved demand factors ξ_{gt} and the observed market characteristics $\chi_{gt} = \left(x'_{gt}, p_t, c_{gt}, n_{gt}, \mathbb{1}_{\{\tilde{t}=2\}}, ..., \mathbb{1}_{\{\tilde{t}=T\}}\right)$. To simplify notations, I use subscript gt to denote the matrix containing the values for all inside products: for example, $x_{gt} = (x_{g1t}, ..., x_{gJ_tt})$. Note that the taste shocks are assumed to be uncorrelated with market characteristics χ_{gt} and ξ_{gt}, but they can be correlated with household's expectations of the market shares s^e_{gt}.

[18]Frontier and Remote Area (FAR) describes territory characterized by some combination of low population size and high geographic remoteness, according to U.S. Department of Agriculture.

Households' heterogeneous parameters follow a distribution F, which is known up to some parameter θ to be estimated. In particular, denote $\beta_{it} = (\beta^x_{it}, \beta^p_{it}, \beta^c_{it}, \beta^n_{it}, \beta^t_2, ..., \beta^t_T)$. The distribution is $F(\beta_{it}, \gamma_{i1t}, ..., \gamma_{iJ_tt}|\theta)$, which will be specified in Section 4.4.

4.2 Equilibrium household decisions

We first describe the information structure of the households. At each t, households in neighborhood g know the observed market characteristics χ_{gt} and the unobserved demand factors ξ_{gt}. They also know the distribution of the parameters $(\beta_{it}, \gamma_{ijt})$ and taste shocks ϵ_{ijt}. Households know their own taste shocks, but do not know their neighbors'. Since a neighborhood (zip code) has around 5000 households on average in my data, it is unlikely that households know all their neighbors' idiosyncratic taste shocks.

Given these information, households form expectations about the market shares of all products in their own neighborhood at a given time. I assume rational expectations in equilibrium. Thus, the expected equilibrium market shares s^e_{gjt} equal the realized equilibrium market shares s_{gjt} for all g, j, t. The rational expectation assumption is common in the incomplete information games literature (Brock and Durlauf, 2001).

For each g and t, the equilibrium market shares solve the following system of equations:

$$s_{gjt} = \int \frac{\exp((\chi'_{gjt}\beta_{it} + \gamma_{ijt}s_{gjt} + \xi_{gjt})/\alpha)}{1 + \sum_{k=1}^{J_t}\exp((\chi'_{gkt}\beta_{it} + \gamma_{ikt}s_{gkt} + \xi_{gkt})/\alpha)} dF(\beta_{it}, \gamma_{i1t}, ..., \gamma_{iJ_tt}|\theta) \quad \forall j = 1, ..., J_t \quad (2)$$

The market share of the outside product is $s_{g0t} = 1 - \sum_{j=1}^{J_t} s_{gjt}$. This is a system of J_t nonlinear multivariate functions with respect to the market share vector s_{gt}. These functions have complicated forms and involve integrals, so there are no closed form solutions for the market share vector s_{gt}, given market characteristics (χ_{gt}, ξ_{gt}). The solutions can only be computed numerically using computational algorithms such as successive approximations (i.e., fixed-point iteration).

4.3 Potential multiple equilibria

Most importantly, the system of equations (2) can have multiple solutions of the market share vector s_g, given market characteristics (χ_{gt}, ξ_{gt}) and parameters (θ, α). This is the well-known multiple equilibria problem. When there are multiple equilibria, the demand functions are no longer well-defined functions. A well-defined demand function should map market characteristics to a *unique* vector of market shares given the parameters. In other words, a demand function is well-defined only if it can be written in the following form:

$$s_{gjt} = \sigma_j \left(\chi_{gt}, \xi_{gt} | \theta, \alpha\right) \quad \forall j = 1, ..., J_t \tag{3}$$

where $\sigma_j(.)$ is a well-defined function[19]. Equation (2) is not of this form because market shares appear on both sides of the equation. If we set $\gamma_{ijt} = 0$, i.e. shutting down the direct network effect, then market shares no longer appear on the right hand side of equation (2), and thus we are guaranteed a well-defined demand function. Therefore, demand functions are well-defined if and only if equations (2) admit a unique solution. If $\gamma_{ijt} \neq 0$, well-defined demand functions are no longer guaranteed. This is because one set of market characteristics and parameters can be mapped to *multiple* values of equilibrium market shares.

For illustration, consider a simple example with 2 inside products, homogeneous coefficients and α fixed at 1. Figure 2 plots the market share of product 1 against the characteristics of product 1 for various values of γ, holding the characteristics of all other products fixed. The characteristics of good j is $\chi'_{gjt}\beta + \xi_{gjt}$. When $\gamma = 0$, the demand function of product 1 is well-defined, as shown in Figure 2a. The market share of product 1 is a strictly increasing function of it's characteristics[20]. But when γ increases to 3 as in Figure 2b, the demand function of product 1 is no longer well-defined. Each value of the characteristics of product 1 can map to multiple values of its market share, depending on the value of the characteristics. Thus, the

[19]Recall that a function is well-defined only if it maps each input to a *unique* output.

[20]Note that the market shares of all inside products are also strictly increasing functions of their characteristics. Due to space limitations, their plots are not shown.

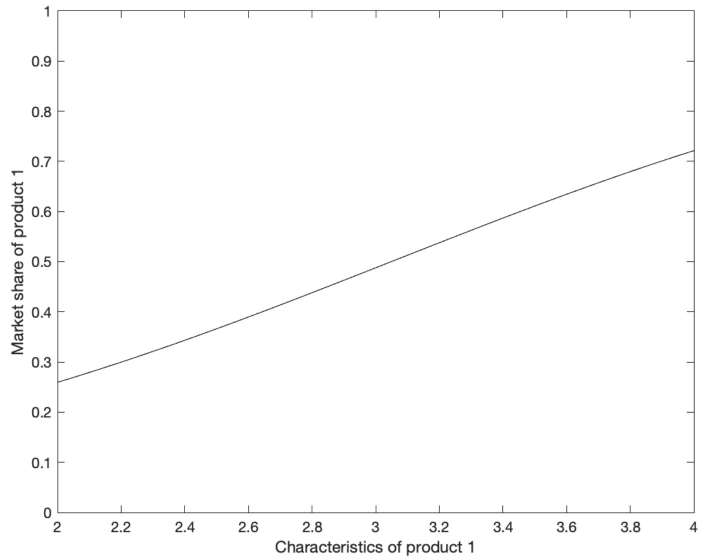
number of equilibria depends on not only the parameters but also the characteristics.

Multiple equilibria poses challenges to the identification and the estimation of the model. Researchers do not observe which equilibrium is selected by the consumers. Thus, the model is not estimable with a full-information approach, e.g. likelihood. The state-of-the-art demand estimation method (BLP) also no longer applies because it requires well-defined demand functions. For example, in Berry and Haile (2014), all analysis is built upon the premise that there exists a system of well-defined demand functions (equation (1) in their paper, which resembles equation (3) in this paper).

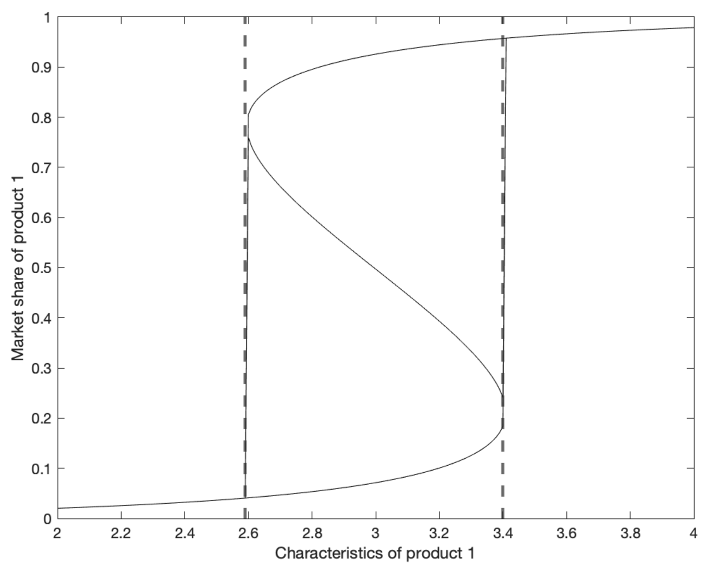
Existing solutions to the potential multiple equilibria issue are unsuitable for the setting of this paper. One existing solution is to assume an equilibrium selection rule. This assumption likely holds in contexts where the consumers can coordinate with each other (Hartmann, 2010). In my setting, each neighborhood has about 5000 households, so coordination among all households to achieve a certain equilibrium is unlikely. Other solutions include using data patterns that are robust to multiple equilibria or partial identification. These methods are suitable for contexts with a small number of consumers and choices. It is unclear how these methods can be extended to study many consumers and many choices, as in the setting of this paper.

Other existing solutions include assuming one equilibrium is selected, or estimating equilibrium selection probabilities. The former assumes that the potential multiple equilibria issue does not need to be accounted for in the estimation. The latter faces computational challenges. The number of equilibria depends the values of both the parameters and the market characteristics, as shown in Figure 3. Thus, for each candidate value of the parameters and characteristics, we need to search for all equilibria using computational algorithms such as successive approximations. Because we do not know the number of equilbria ex-ante and the number of equilibria can change when either the parameters or the characteristics change, solving for all equilibria can be computationally very difficult if not impossible. In particular, no computational methods can guarantee finding all the equilibria of a game, unless the equilibrium equations form a system of polynomial equations (Su, 2014).

Figure 2: Illustration of unique vs. multiple equilibria



(a) $\gamma = 0$, unique equilibrium,
demand function is well-defined



(b) $\gamma = 3$, multiple equilibria,
demand function is not well-defined

Notes: This figure plots market share of product 1 against characteristics of product 1, holding characteristics of all other products fixed for various values of γ. Characteristics of good j is $\chi'_{gjt}\beta + \xi_{gjt}$. α is fixed at 1. Number of inside products is 2, and coefficients are homogeneous.

The more important issue with the solutions above is counterfactuals. Under a new counterfactual policy, e.g. a new pricing policy, market characteristics change, so the number of equilibria can change as well. If so, then the estimated selection probabilities will also change under the new policy. Moreover, even if we have assumed or tested that the data exhibits unique equilibrium, multiple equilibria can still occur under a new counterfactual policy. If so, the researchers still do not know which equilibrium will be the selected under the new counterfactual policy. Since no computational methods can guarantee finding all the equilibria of any game numerically, the researchers can not reliably check whether the equilibrium is unique under the new policy using numerical methods. Therefore, in the next Section 4.4, I provide theoretical conditions under which unique equilibrium is guaranteed given *any* market characteristics; this guarantees unique equilibrium even under new counterfactual policies.

4.4 Unique equilibrium conditions

I now present three theorems that characterize the conditions under which unique equilibrium is guaranteed given *any* market characteristics. Theorem 1 shows results assuming: all households in the same neighborhood-year have the same parameters, but the parameters can vary by products. Theorem 2 allows households in the same neighborhood-year to have different parameters, but the social influence parameters are the same for all products. Theorem 3 allows parameters to vary by both households and products, after specifying an assumption on the distribution of the parameters.

Theorem 1. *For any g, t, j and $J_t \geq 2$, let $\beta_{it} = \beta_{gt}$ and $\gamma_{ijt} = \gamma_{gjt}$ for all $i \in g$. If $|\gamma_{ijt}| < 2\alpha$ for all i, j, t, then there exists a vector of unique equilibrium market shares s_{gt} given any market characteristics (χ_{gt}, ξ_{gt}). Thus, demand functions for all products are well-defined.*

Proof. Proof is in Appendix C.1. □

Theorem 2. *For any g, t and $J_t \geq 2$, let $\gamma_{ijt} = \gamma_{it}$ for all $j = 1, ..., J_t$. If $\mathbb{E}_{i \in g}\left[|\gamma_{ijt}|\right] < 2\alpha$, then there exists a vector of unique equilibrium market shares s_{gt} given any market characteristics*

(χ_{gt}, ξ_{gt}). *Thus, demand functions for all products are well-defined.*

Proof. Proof is in Appendix C.2. $\square$

I model households' heterogeneous parameters as functions of their demographics D_{it}, which includes income, education, and daily commuting time to work. I let

$$\begin{pmatrix} \beta_{it} \\ \gamma_{i1t} \\ \vdots \\ \gamma_{iJ_t t} \end{pmatrix} = \begin{pmatrix} \bar{\beta} \\ \bar{\gamma}_1 \\ \vdots \\ \bar{\gamma}_{J_t} \end{pmatrix} + \begin{pmatrix} \Pi_\beta \\ \Pi_\gamma^1 \\ \vdots \\ \Pi_\gamma^{J_t} \end{pmatrix} D_{it} \tag{4}$$

The parameter to be estimated is $\theta = \left(\bar{\beta}, \bar{\gamma}_1, ..., \bar{\gamma}_{J_t}, \Pi_\beta, \Pi_\gamma^1, ..., \Pi_\gamma^{J_t}\right)$.

Theorem 3. *For any g, t and $J_t \geq 2$, let $\Pi_\gamma^j = \Pi_\gamma$ for all $j = 1, ..., J_t$ and $\bar{\gamma}_t^* = \max_{j=1,...,J_t}\{|\bar{\gamma}_j|\}$. If $\bar{\gamma}_t^* + \mathbb{E}_{i \in g}\left[|\Pi_\gamma D_{it}|\right] < 2\alpha$, then there exists a vector of unique equilibrium market shares s_{gt} given any market characteristics (χ_{gt}, ξ_{gt}). Thus, demand functions for all products are well-defined.*

Proof. Proof is in Appendix C.3. $\square$

The theorems above imply that when the average social influence effect is small enough compared to the variance of households' idiosyncratic taste shocks, the equilibrium market shares exist and are unique. Intuitively, if the social influence effect is too large and dominates households' heterogeneous private shocks, then multiple equilibria can occur because there is not enough pre-determined private incentives for some households to choose one product over another. Households can be equally satisfied with either very high market shares or very low market shares. Similar intuition holds for the case of binary choice models with homogeneous coefficients (Brock and Durlauf, 2001).

To further illustrate how the number of equilibria depends on the value of the social influence effect γ, I consider the same simple example as with Figure 2. In Figure 3, I plot the market share of product 1 against different values of γ, holding characteristics of all products fixed. The number of equilibria changes sharply from 1 to 3 as γ passes the bound 2α. Since α is fixed at 1 in this example, the bound is 2.

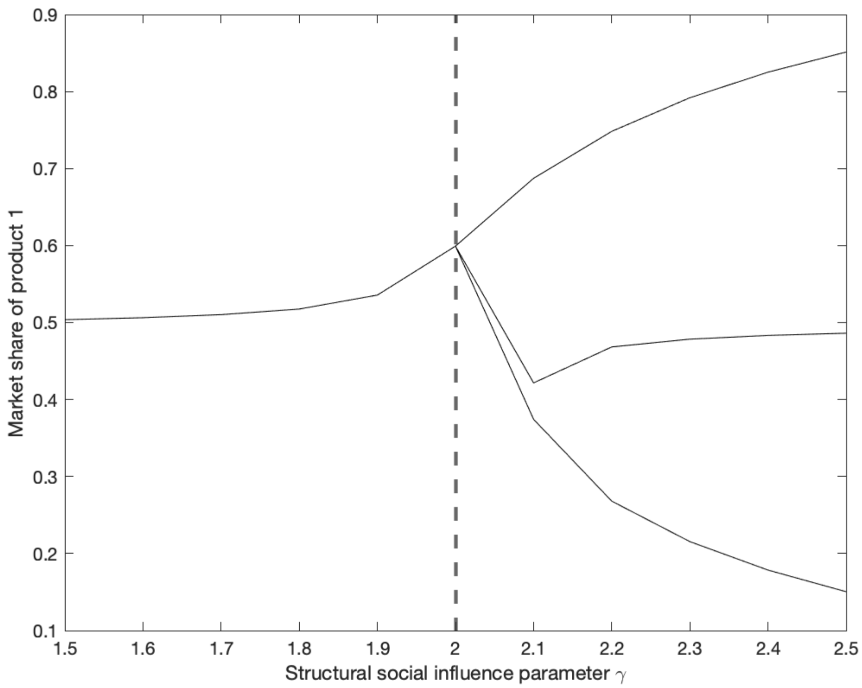
Deriving the theoretical conditions for unique equilibrium is quite challenging. This requires characterizing the behavior of a system of nonlinear multivariate functions; these functions have complicated forms and involve integrals. In binary choice models with homogeneous coefficients, these functions reduce to one singlevariate function without integral. For example in Brock and Durlauf (2001), this function becomes a *tanh* function. Moreover, we need to derive bounds that are implementable in the estimation. Simply showing that some bounds exist is not sufficient for estimation.

How exactly are these theorems useful in the estimation? If we know that the data exhibits unique equilibrium, then we just need to search for the candidate parameters within the bounds 2α in the estimation. The estimator will be described in Section 5.1. At the estimated parameters, unique equilibrium is therefore guaranteed even under a new counterfactual policy. The remaining question is: does the data exhibit unique equilibrium?

4.5 Do the data patterns suggest unique equilibrium?

To investigate whether the data patterns are consistent with unique equilibrium, I lay out both formal and informal methods in this section. Recent progress in homogeneity tests proposes various formal statistical tests for unique equilibrium. Otsu et al. (2016) form a chi squared test statistics based on nonparameteric estimates of conditional choice probabilities. The test compares directly the set of conditional choice probabilities estimated from the sample pooling all groups with those estimated from each group separately. Under the null hypothesis of unique equilibrium, the two sets of conditional choice probabilities should be the same. This test is useful when the data has a large number of groups. In the situation where the data has finite number of groups, we can use the homogeneity tests proposed by Bugni et al. (2020). Their test resembles a randomization test. Under the null hypothesis that the conditional choice

Figure 3: Number of equilibria for different values of social influence effects



Notes: This figure plots the market share of product 1 against the social influence effect γ, holding characteristics of all products fixed. α is fixed at 1. Number of inside products is 2, and coefficients are homogeneous.

probabilities are homogeneous across groups, permuting the state variables and the resulting choices (subject to some restrictions) should not change the likelihood of the data. Their test is implemented by a MCMC algorithm.

If, in addition to market level data, data on consumer level characteristics and outcomes are also available, we can use the test in De Paula and Tang (2012). Under the null hypothesis of unique equilibrium, the consumers' choices should be uncorrelated with each other conditional on characteristics, because their private values are independent. Their test uses stepwise multiple testing procedure to infer whether each consumer adopts multiple strategies conditional on characteristics.

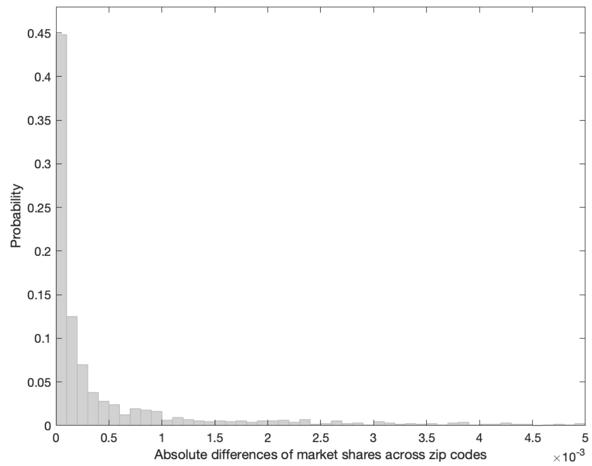
In the situations where the formal tests can not be applied, I propose an informal method to investigate whether the data patterns are consistent with unique equilibrium. I compare market shares across neighborhoods with similar market characteristics. If there is unique equilibrium in the data, then these market shares should be similar to each other. If we observe very different market shares in the data, then this is evidence that multiple equilibria likely exist.

In my data, I find data patterns that are consistent with unique equilibrium using the informal method[21]. Figure 4a plots the probability density of the absolute differences of market shares across zip codes with similar characteristics in 2020. The absolute differences are very close to zero with high probability. This suggests that zip codes with similar characteristics have similar market shares, and thus the data is likely to be generated from unique equilibrium. In comparison, the absolute differences of market shares across zip codes with different characteristics are distributed further away from zero in Figure 4b. I consider the following characteristics: distribution of consumer demographics including average household income, share of college graduates, share of long commuters, street composition, urban, FAR, and number of charging stations. The data patterns for other years are in Appendix Figure 7.

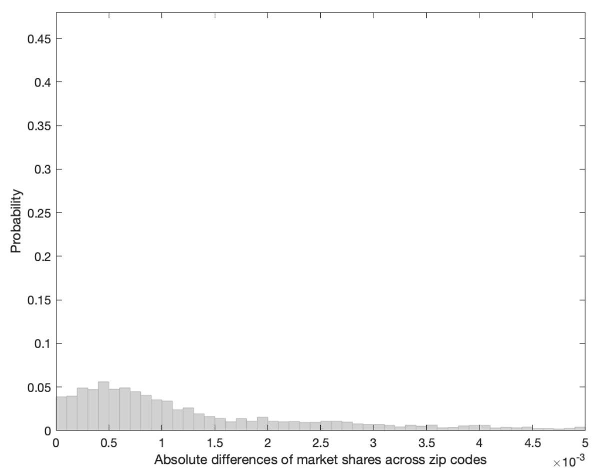
Given that the patterns in the data suggest unique equilibrium, we now impose the parameter constraints specified in Theorem 2. Therefore, we have a system of well-defined demand

[21] The results for the formal tests are still in progress.

Figure 4: Absolute differences of market shares across zip codes



(a) zip codes with similar characteristics



(b) zip codes with different characteristics

Notes: This figure plots the probability density of the absolute differences of market shares across zip codes with similar characteristics vs. across zip codes with different characteristics in 2020. The characteristics include: distribution of consumer demographics, street composition, urban, FAR, number of charging stations.

functions $\sigma = (\sigma_1, ..., \sigma_{J_t})$, where each σ_j is defined in equation (3). Since the scale parameter of the taste shocks is not identified and is generally normalized to 1, we let $\alpha = 1$ from now on.

5 Estimator and Identification

5.1 The GMM estimator

The GMM estimator minimizes moment conditions that involve the product of instrumental variables Z and the unobserved demand factors ξ. The instruments Z should be uncorrelated with unobserved demand factors ξ. More detailed requirements on the instruments will be discussed in Section 5.2; Section 5.3 will describe how the instruments are constructed. Z is a $N \times K$ matrix, where N is the total number of observations (each observation is an inside product $\times$ zip code $\times$ year) and K is the dimension of valid instruments. $\xi(\theta)$ is a $N \times 1$ vector of unobserved demand factors computed from inverting the system of demand equations (3). Formally, the GMM estimator is:

$$\hat{\theta} = \arg\min_{\theta} \ \xi(\theta)' Z W Z' \xi(\theta), \quad \text{s.t. } \hat{\mathbb{E}}_{i \in g} \left[|\gamma_{ijt}| \right] < 2 \quad \forall g, t, j > 0 \tag{5}$$

W is the optimal weighting matrix of a GMM estimator, which is a consistent estimate of the inverse of $\mathbb{E}\left[Z'\xi\xi'Z\right]$. For any g, t, j, $\hat{\mathbb{E}}_{i \in g}\left[|\gamma_{ijt}|\right]$ is the sample analog of $\mathbb{E}_{i \in g}\left[|\gamma_{ijt}|\right]$. Since we have normalized α to be 1, imposing the constraints above guarantees unique equilibrium and well-defined demand functions. Note that if the parameters are specified as in Theorem 3, then we just need to replace the constraints in the GMM estimator with the sample analog of the parameter constraints in Theorem 3.

5.2 Identification

Identification of the proposed GMM estimator requires: (1) the demand functions are invertible, i.e. we can solve for unique ξ from inverting demand equations (3), and (2) there are sufficient instruments for the endogenous variables. This section will discuss each of these requirements.

Let $x^{(1)}_{gjt}$ be a scalar component of the observed car characteristics. I choose miles per dollar because it has variations across g, j, t as discussed in Section 4.1. Let the rest of the observed market characteristics be $\tilde{\chi}_{gjt}$, so $\chi_{gjt} = (x^{(1)}_{gjt}, \tilde{\chi}_{gjt})$. Since the unobservable demand factors have no natural location or scale, we must normalize them for unique representation of preferences. Following Berry and Haile (2014), we normalize the scale by setting the coefficient of $x^{(1)}_{gjt}$ as $\beta^{x^{(1)}}_{it} = 1$,[22] and the location by setting $\mathbb{E}\left[\xi_{gjt}\right] = 0$. Define a linear index:

$$\delta_{gjt} = x^{(1)}_{gjt} + \xi_{gjt} \quad \forall g, j, t \tag{6}$$

Now we can rewrite the demand functions as:

$$s_{gjt} = \sigma_j(\delta_{gt}, \tilde{\chi}_{gt}|\theta) \quad \forall j = 1, ..., J_t \tag{7}$$

In order to prove that the demand functions above are invertible with respect to δ, I first show an important feature of these demand functions: the Jacobian matrices can be derived analytically, and they are non-singular if we impose the parameter constraints in Theorem 2.[23] The formal result is in Lemma 1, and the derivation of the Jacobian matrices are in Appendix C.4. Given this feature of the Jacobian matrices, I then show that demand functions are invertible if all products are weak substitutes after imposing the parameter constraints in Theorem 2. Now, let me formally introduce the results.

Lemma 1 (Non-singularity of the Jacobian Matrix). *If the parameter constraints in Theorem 2 are satisfied, then the demand functions $\sigma(\delta_{gt}, \tilde{\chi}_{gt}|\theta)$ are differentiable with respect to δ_{gt} and the associated Jacobian matrix is non-singular.*

Proof. Proof is in Appendix C.5. □

Definition (Weak Substitutes). Goods $(0, 1, ..., J_t)$ are weak substitutes if $\sigma_k(\delta_{gt}, \tilde{\chi}_{gt}|\theta)$ is

[22] $\beta^{x^{(1)}}_{it}$ can be normalized to a constant instead of 1, but this constant is not identified.
[23] The results can be extended to the parameter constraints in Theorem 3.

non-increasing in δ_{gjt}, for all $j > 0, k \neq j$, given any $\tilde{\chi}_{gt}$.

Theorem 4 (Invertibility of Demand Functions)**.** *Let the parameter constraints in Theorem 2 hold. Consider any $\tilde{\chi}_{gt}$ and any market shares $s_{gt} > 0$ such that $\sum_{j=1}^{J_t} s_{gjt} < 1$. If all products are weak substitutes, then there is at most one vector δ such that $s_{gjt} = \sigma_j(\delta, \tilde{\chi}_{gt}|\theta)$ for all $j > 0$. Thus, δ is uniquely determined and all demand functions are invertible.*

Proof. Proof is in Appendix C.6. □

Given the results of Theorem 4, we can write

$$\delta_{gjt} = \sigma_j^{-1}(s_{gt}, \tilde{\chi}_{gt}|\theta) \quad \forall g, t, j > 0 \tag{8}$$

$$\text{or} \quad x_{gjt}^{(1)} = \sigma_j^{-1}(s_{gt}, \tilde{\chi}_{gt}|\theta) - \xi_{gjt} \quad \forall g, t, j > 0 \tag{9}$$

$$\text{or} \quad x_{gjt}^{(1)} = \sigma_j^{-1}(s_{gt}, \tilde{\chi}_{gt}^{(\text{exg})}, \tilde{\chi}_{gt}^{(\text{end})}|\theta) - \xi_{gjt} \quad \forall g, t, j > 0 \tag{10}$$

In equation (10), $\tilde{\chi}_{gt}^{(\text{exg})}$ denotes the exogenous variables in $\tilde{\chi}_{gt}$, including car characteristics except for miles per dollar, observed neighborhood characteristics, and time dummies. $\tilde{\chi}_{gt}^{(\text{end})}$ denotes the endogenous variables in $\tilde{\chi}_{gt}$, including the log of car prices p_t and the number of charging stations c_{gt}. Since equation (10) takes the same form as in Berry and Haile (2014), I invoke their identification arguments: if we have instruments z_{gt} for the endogenous p_t and c_{gt} that satisfy exclusion restrictions and completeness[24], then each of the inverse demand functions σ_j^{-1} is identified. Consequently, the demand functions and the unobserved demand factors ξ_{gjt} are also identified.

We do not need additional instruments for the endogenous market share vector s_{gt} because $x_{gt}^{(1)}$ serves as an essential instrument for the endogenous market shares. See Berry and Haile (2014, 2016) for detailed discussions for the validity of the essential instrument. Intuitively, since the coefficient of $x_{gt}^{(1)}$ is normalized, the exogenous variations in $x_{gt}^{(1)}$ identify the social influence effects. Given the identified social influence effects, the instruments for the endogenous

[24]The exclusion restriction is $\mathbb{E}\left[\xi_{gjt}|z_{gt}, x_{gt}^{(1)}, \tilde{\chi}_{gt}^{(\text{exg})}\right] = 0$ a.s.; the completeness restriction is that for all functions $B(s_{gt}, \tilde{\chi}_{gt})$ with finite expectations, if $\mathbb{E}[B(s_{gt}, \tilde{\chi}_{gt})|z_{gt}, x_{gt}^{(1)}, \tilde{\chi}_{gt}^{(\text{exg})}] = 0$ a.s., then $B(s_{gt}, \tilde{\chi}_{gt}) = 0$ a.s..

charging stations identify the indirect network effects; the instruments for the endogenous prices identify the price effects.

5.3 Instruments

As discussed in Section 5.2, identification requires constructing instruments for the endogenous prices and number of charging stations. For the price instruments, I use exogenous characteristics of competing products, also known as the BLP instruments (Berry and Haile, 2021; Gandhi and Nevo, 2021). More specifically, I use two sets of price instruments: (1) functions of the difference of the exogenous characteristics of the focal product and the competing products, which measures product differentiation[25], (2) functions of the number of competing products with similar characteristics, which measures local competition[26]. The exclusion restriction holds as long as the exogenous characteristics are mean independent of the local unobserved demand shocks. Because firms are setting the car characteristics at aggregate level, they are unlikely to be correlated with local demand shocks.

Before describing the instruments for neighborhood public charging stations, let us briefly review how we model households' utility. In equation (1), c_g is the number of charging stations that are potentially useful for the daily activities of households in zip code g. According to studies from National Household Travel Survey and Idaho National Laboratory, I model that stations within 30 miles are within households' daily activity range. In Figure 5, households in the focal zip code g (the small circle with 5 mile radius in the center) care about the number of stations within 30 miles (the big circle with solid line). c_g is endogenous because it can be correlated with the unobservables of the focal zip code g. A valid instrument for c_g should satisfy exogeneity and relevance. Exogeneity requires that the instrument is uncorrelated with the unobservables of the focal zip code g. Relevance requires that the instrument can shift c_g, and thus the instrument needs variations at the zip code level.

[25]The functions are: sum of squared differences, which captures a continuous measure of product isolation, and interactions of differences between characteristics dimensions, which captures the covariance between two dimensions of differentiation.

[26]The functions are: the number of products within a certain bandwidth of the focal product, and the interactions of the numbers between characteristics dimensions.

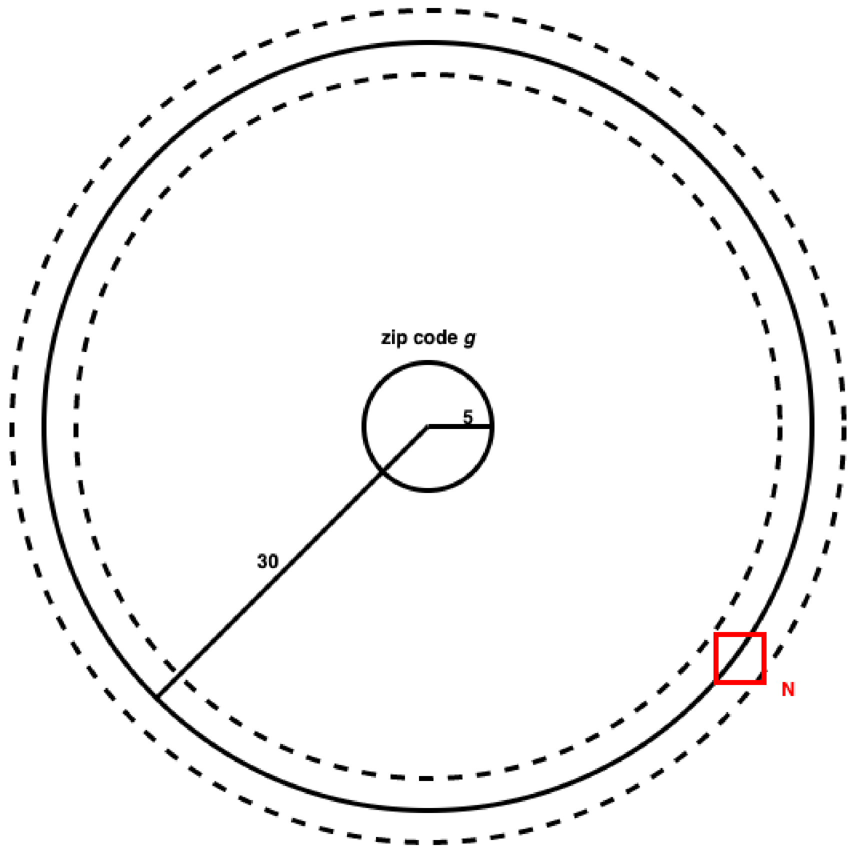
Instruments that have been proposed in the literature are unsuitable for my setting. One type of instruments involves using the number of grocery stores and supermarkets (Li et al., 2017). These instruments have zip code level variations, but are unlikely to satisfy exogeneity. This is because the number of grocery stores and supermarkets can be correlated with unobserved neighborhood characteristics. Another type of instruments use government subsidies for charging stations (Li, 2019). These instruments are likely to be exogenous but usually lack zip code level variations.

I propose a novel instrument which leverages the exogenous location assignment of new charging stations. Among all new stations built within a small bandwidth along the boundary of the 30-mile circle (within the dotted lines), whether the stations are built inside or outside the boundary is uncorrelated with the unobservables at g. For example, if a new stations is built at N in Figure 5, it's location is determined by the unobservables of many zip codes within 30 miles of N, not just the unobservables of the focal zip code g. Moreover, since the average radius of a zip code in Texas is about 5 miles, the focal zip code g is relatively far away from the 30-mile boundary. The location assignment of new stations along the boundary is unlikely to be correlated with the unobservable of one particular zip code that is far away.

Therefore, a valid instrument can be: the number of stations inside the 30-mile circle and within the small bandwidth along the boundary. It satisfies exogeneity because it is uncorrelated with the unobservables of g, as described above. It satisfies relevance because it can shift the number of stations within the 30-mile circle with zip code level variations.

There might be concerns about common time shocks or regional shocks, e.g. regional investment in charging stations. In this case, we can add time and regional fixed effects. Alternatively, we can take the difference of the number of stations inside the boundary within the bandwidth and the number of stations outside the boundary within the bandwidth. This method can difference out common shocks if we assume that the common shocks affect equally the stations inside and those outside within the bandwidth. In the estimation, I use the differenced-out instrument. First stage results are in Appendix Table 6.

Figure 5: Graphical illustration of the instruments for charging stations



Notes: This graph illustrates the instruments for the charging stations for a focal zip code g. The focal zip code in the center has 5 mile radius. Households in g care about the number of stations within 30 miles, which is the big circle with solid line. The dotted lines represent a small bandwidth along the boundary of the 30-mile circle. N is an example location of a new station along the boundary.

6 Estimation Results

The estimation results are in Table 2. The first column reports the households' baseline parameters $(\bar{\beta}, \bar{\gamma}_1, ..., \bar{\gamma}_{J_t})$; the last three columns report how households' parameters deviate from the baseline due to their demographics (Π), as described in equation (4). The demographics include: indicator for college degree, standardized log income, and indicator for daily commute time longer than one hour. For the current set of results, I let $\bar{\gamma}$ vary by the brands' continent of origin[27] and $\Pi_\gamma^j = \Pi_\gamma$ for all $j = 1, ..., J_t$.

[27] In my data, all North American brands are from the US. The results for when baseline social influence effects vary by brands are in progress.

The baseline parameters describe the effects for a baseline household with average income of own neighborhood, no college degree and daily commute time less than one hour. The social influence effect for the baseline household is significantly positive. In particular, if the neighborhood market share of an Asian brand (e.g. Toyota Prius) increases by 1 percentage point, then the utility increase from social influence effects is 0.532 percentage point for the baseline household. This increase in utility is equivalent to a decrease of $55 in price[28]. The effect for a US brand is larger: a 1 percentage point increase in the market share of a US brand (e.g. Chevrolet Bolt) increases the baseline household's utility by 0.646 percentage point. This increase in utility is equivalent to a decrease of $66 in price.

Compared to the baseline household, the social influence effects are significantly stronger for households with college degree, or higher income, or longer daily commute time. For example, if the neighborhood market share of an Asian brand increases by 1 percentage point, then the utility increase for households with college degree is 0.779, which is equivalent to a price decrease of $84. For a US brand, the equivalent price decrease is $96. For households whose log income is one standard deviation above the average, their utility increase is equivalent to a price decrease of $175 for an Asian brand, and $191 for a US brand.

Compared to the baseline household, the estimated absolute price elasticity is smaller for households with college degree, or higher income, or longer daily commute time; they are less price sensitive. The consumer groups who are less price sensitive are also the groups who are more affected by social influences. This finding has important implications for designing the optimal targeted group pricing policy in Section 7.

To estimate the indirect network effects from the public charging stations within households' daily activity range (30 miles), I allow for differential effects due to both the EV types and the neighborhoods' characteristics. In neighborhoods with low population size and high geographic remoteness (FAR areas), the baseline household values more public charging stations in their daily activity range. Compared to the baseline household, households with higher education or

[28]To convert the utility increase into dollar value, I divide the utility increase by the estimated price elasticity, and evaluate the percentage increase in prices at the average price.

higher income value public charging stations less. One possible explanation is: these households likely have home charging installations or charging facilities at work, thus are less reliant on public charging stations. Long commuters also value less public charging stations less, compared to the baseline household. This is intuitive because long commuters can charge at work, which is very likely to be outside of the 30-mile radius.

For neighborhoods that are not FAR, the population is denser, so there is potential concern for congestion (Chen et al., 2017) and electrical overload (Muratori, 2018). Muratori (2018) shows that uncoordinated charging can significantly increase peak demand for electricity, and strains the electricity distribution infrastructure, even if the market share is low. More public charging stations in the focal neighborhood decrease the baseline household's utility from purchasing a battery EV (BEV). Compared to the baseline household, households with higher income are not as negatively affected by more neighborhood public charging stations. These households likely have access to more stable electricity supply systems and are less affected by high electricity bills. Long commuters are also not as negatively impacted, since they are able to charge at work. They work far away from the focal neighborhood, thus are less affected by the electricity supply and congestion in the focal neighborhood.

When it comes to households' utility for purchasing Plug-in hybrid EV (PHEV), the effect of neighborhood public charging stations is generally insignificant. Households have more flexibility: they can refuel in gas stations when there is congestion or electrical overload. PHEVs are also easier to charge at home: it takes 5-6 hours to fully charge a PHEV with a common residential AC outlet[29].

Longer battery range increases the baseline households' utility for purchasing BEVs, whereas it decreases the utility for purchasing PHEVs. In fact, cars with longer battery range have much heavier batteries; larger weights lead to less efficiency and more safety problems. Anderson and Auffhammer (2014) show that being hit by a vehicle that is 1000 pounds heavier generates a 40-50% increase in fatality risk and by itself generates a societal cost equivalent to a $0.97 per gallon

[29] https://www.transportation.gov/rural/ev/toolkit/ev-basics/charging-speeds

gas tax. Long range batteries also scale the problems of resource extraction, manufacturing emissions and battery recycling. The information above is easily accessible on search engines[30], so potential buyers are likely aware of the problems of long battery range. Since PHEVs can be refueled by gas, they are less constrained by low battery range. Compared to the baseline household, the long commuters care less about the battery range of BEVs. Even a BEV with low battery range is more than sufficient for long daily commutes; moreover, the longer the daily commute time is, the longer the driver is exposed to safety problems. But long commuters could be constrained by PHEVs with low battery range, so they prefer PHEVs with longer battery range compared to the baseline household.

To understand the importance of social influence effects, I estimate the model without accounting for social influence effects. The results are in Table 3. If we ignore social influence effects, the baseline household's absolute price elasticity is under-estimated by 11%. The bias is larger for households with college degree (13%), households whose log income that are one standard deviation above average (16%), and longer commuters (13%). The baseline charging stations effects are also under-estimated without accounting for social influence effects.

7 Optimal Targeted Pricing

Given the estimated demand elasticities and social influence effects, I design the firm's optimal targeted pricing policies by consumer demographics. In particular, I segment consumers into two groups: group D consists of socio-economically disadvantaged consumers without college degree and below average income of their neighborhood; group A consists of the rest of the consumers who are advantaged. Other thresholds for segmenting the consumers will be implemented in later versions of the paper. For example, group D consumers can satisfy: (1) income below average of their neighborhood, or (2) income below a certain threshold, e.g. the income cap proposed by the 2022 Inflation Reduction Act.

[30]For example, one google search shows articles like this: https://www.cnet.com/roadshow/news/how-much-range-you-really-need-in-an-electric-car/

Table 2: Estimates of the full model with social influences

	Baseline	Deviations by demographic variables		
		college	log(income)	long commute
social influence effect, Asian brands	0.532***	0.247***	0.758***	0.401***
	(0.003)	(0.004)	(0.009)	(0.002)
social influence effect, US brands	0.646***	0.247***	0.758***	0.401***
	(0.003)	(0.004)	(0.009)	(0.002)
social influence effect, European brands	0.021***	0.247***	0.758***	0.401***
	(0.000)	(0.004)	(0.009)	(0.002)
$-\log(\text{price})$	4.530***	-0.203***	-1.107***	-1.100***
	(0.134)	(0.041)	(0.067)	(0.049)
# charging stations $\leq$ 30 miles x FAR	3.634**	-0.424***	-2.635***	-0.643***
	(1.502)	(0.001)	(0.005)	(0.002)
# charging stations $\leq$ 30 miles x BEV	-0.126***	0.008	0.068***	0.066***
	(0.040)	(0.013)	(0.017)	(0.022)
# charging stations $\leq$ 30 miles x PHEV	-0.069	-0.003	0.035	0.029
	(0.119)	(0.033)	(0.046)	(0.041)
battery range (miles) x BEV	0.081***	-0.009***	-0.043***	-0.041***
	(0.005)	(0.002)	(0.002)	(0.006)
battery range (miles) x PHEV	-1.199***	0.060	0.421***	0.441***
	(0.096)	(0.045)	(0.044)	(0.030)
BEV dummy	-10.024***			
	(0.519)			
PHEV dummy	-2.869***			
	(0.304)			
$>$ midsize dummy	-0.217***			
	(0.039)			
$<$ midsize dummy	0.399***			
	(0.068)			
electric motor power x BEV	0.062***			
	(0.003)			
electric motor power x PHEV	0.012***			
	(0.001)			
urban dummy	0.551**			
	(0.249)			
Observations	100,084			

Notes This table shows the estimates of the full model with social influences. A unit of observation is a product, zip code, year. Estimation includes year fixed effects. Bootstrapped standard errors are in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3: Estimates of the model without social influences

	Baseline	Deviations by demographic variables		
		college	log(income)	long commute
$-$log(price)	4.016***	-0.237**	-1.139***	-0.701***
	(0.163)	(0.098)	(0.084)	(0.117)
# charging stations x FAR	1.237	0.141***	0.787***	0.239***
	(1.450)	(0.009)	(0.014)	(0.008)
# charging stations x BEV	-0.194***	0.016	0.100***	0.061
	(0.042)	(0.027)	(0.016)	(0.038)
# charging stations x PHEV	-0.143	0.002	0.068	0.037
	(0.146)	(0.072)	(0.059)	(0.071)
battery range x BEV	0.081***	-0.010**	-0.048***	-0.029***
	(0.006)	(0.004)	(0.003)	(0.005)
battery range x PHEV	-0.938***	-0.047	0.353***	0.347***
	(0.262)	(0.053)	(0.108)	(0.099)
BEV dummy	-7.037***			
	(0.440)			
PHEV dummy	-2.738***			
	(0.272)			
> midsize dummy	-0.049			
	(0.038)			
< midsize dummy	0.336***			
	(0.058)			
electric motor power x BEV	0.053***			
	(0.002)			
electric motor power x PHEV	0.010***			
	(0.001)			
urban dummy	0.493*			
	(0.287)			
Observations	100,084			

Notes This table shows the estimates of the model without social influences. A unit of observation is a product, zip code, year. Estimation includes year fixed effects. Bootstrapped standard errors are in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

To compute the optimal targeted prices, I first estimate firms' marginal costs from their first order conditions of the profit functions, taking into account the ownership structures of the firms[31]. I assume that the uniform prices observed in the data are the optimal uniform prices chosen by the firms. The estimated marginal costs and markups are in Table 4. I find that markups are on average \$33,430, and the average marginal costs are \$13,350.

I define the firm's objective as profit-maximizing but subject to the constraint that the targeted prices for all households do not exceed their observed uniform prices in the data. Thus, the firm is maximizing profits while making sure the consumers are not worse-off under the targeted pricing policy. There are other ways to define the firm's objective, e.g. profit-maximizing only, or a combination of profit and consumer equity; these will also be implemented in future versions of the paper. I compute the optimal targeted prices for each firm, given the observed uniform prices charged by the other firms. For future versions of the paper, I will compute the optimal targeted prices when all firms are setting prices simultaneously.

The optimal group targeted pricing results are in Table 5. The optimal prices are on average \$19,150 for group D who are socio-economically disadvantaged, and \$41,800 for group A who are more advantaged. The price differences between the two groups can be viewed as price subsidies to the disadvantaged group D. Under these optimal prices, firms' annual sales in Texas increase by 35.7% on average; firms' annual profits in Texas increase by 2.7% (\$4.42 millions) on average.

Price subsidies to the disadvantaged group D increase firms' sales and profits. Intuitively, since the disadvantaged households are more price sensitive, their purchases increase significantly due to the price subsidies. By increasing the purchases of the disadvantaged households in the neighborhood, we also increase the purchases of the advantaged households in the neighborhood because the latter are more affected by social influences. Thus, the effect of the price subsidies propagates to the non-subsidized group. However, price subsidies to the advantaged group A do not have the same effect. The purchases of the advantaged group do not increase

[31] See Nevo (2001) for a detailed discussion of the estimation of firms' marginal costs.

as much, since they are less price sensitive; the effect of the subsidies also do not propagate to the non-subsidized group as much, since the non-subsidized group D is less affected by social influences.

Under the recommended pricing policy both firms' profits and equity among consumers improve, even when there are heterogeneous network effects. The private incentives of profits and the public incentives of equity align. Thus, some common concerns of targeted pricing are alleviated, such as fairness and antitrust concerns (Kahneman et al., 1986; OECD, 2018; Zhang and Misra, 2022).

Table 4: Estimated markups and marginal costs

	mean	10th prctile	median	90th prctile
prices (\$1000)	46.77	24.34	35.19	82.06
markups (\$1000)	33.43	11.23	28.07	63.91
marginal costs (\$1000)	13.35	2.92	11.32	27.02

Notes This table shows the summary statistics of the estimated markups and marginal costs of all EV manufacturers from 2015 to 2020. They are estimated from firms' first order conditions of the profit functions, given the estimated demand elasticities from the full model with social influence in Table 2.

Table 5: Counterfactual outcomes with optimal targeted pricing

	mean	10th prctile	median	90th prctile
Optimal targeted prices ($thousands)				
uniform prices	46.77	24.34	35.19	82.06
optimal targeted prices				
for group A (p_A)	41.80	22.75	34.32	78.12
for group D (p_D)	19.15	6.27	16.39	33.39
differences across groups ($p_A - p_D$)	22.66	5.33	15.80	46.51
Demand response (thousands)				
annual sales with uniform pricing	2.07	0.23	0.77	6.60
annual sales with optimal targeted pricing	2.82	0.24	1.02	6.83
annual sales increase	0.75	0.00	0.07	1.15
annual sales increase (%)	35.68	0.09	4.93	123.30
Profits ($millions)				
annual profits with uniform pricing	55.08	7.18	27.07	145.50
annual profits with optimal targeted pricing	59.51	7.23	27.12	150.41
annual profits increase	4.42	0.00	0.06	3.92
annual profits increase (%)	2.71	0.01	0.23	10.83

Notes This table presents the optimal targeted prices for two groups, as well as the resulting changes in firms' annual sales and profits in Texas. Group D consists of households without college degree and below average income of their zip code; the rest of the households are in group A. The optimal targeted prices for all households do not exceed their observed uniform prices in the data.

8 Conclusion

The empirical contributions are twofold. First, I quantify heterogeneous social influence effects in EV purchases. Second, given the estimated effects, I design firms' targeted pricing policies that are both profitable and equitable. In terms of the methodological contributions, I first build a heterogeneous coefficient discrete choice model with social influence effects that allows for counterfactual policy design. Second, I develop a demand estimation strategy that addresses the concerns of potential multiple equilibria. Third, I show how to separately identify the endogenous direct (social influences) and indirect (charging stations) network effects using instruments. Lastly, the proposed method can be used for demand estimation and policy design in other contexts with social influences, e.g. energy, environment, and social media.

There are several avenues for future work. First, we can evaluate the equilibrium effects of sustainability policies. Some sustainability policies can have externalities such as social influence effects. It is important to design such policies taking into account the externalities. Second, we can extend the current framework to account for the dynamic effects of both direct and indirect network effects. With the dynamic framework, we can study the long-term effects of sustainability policies.

References

Allende, C. (2021). Competition under social interactions and the design of education policies.

Anderson, M. L. and M. Auffhammer (2014). Pounds that kill: The external costs of vehicle weight. *Review of Economic Studies 81*(2), 535–571.

Archsmith, J. E., E. Muehlegger, and D. S. Rapson (2021). Future paths of electric vehicle adoption in the united states: Predictable determinants, obstacles and opportunities.

Bayer, P. and C. Timmins (2007). Estimating equilibrium models of sorting across locations. *The Economic Journal 117*(518), 353–374.

Berry, S., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica 81*(5), 2087–2111.

Berry, S. and P. Haile (2016). Identification in differentiated products markets. *Annual Review of Economics 8*(1), 27–52.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Berry, S. T. and P. A. Haile (2021). Foundations of demand estimation. In *Handbook of Industrial Organization*.

Bhattacharya, D., P. Dupas, and S. Kanaya (2021). Demand and welfare analysis in discrete choice models with social interactions.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2011). Identification of social interactions. In *Handbook of social economics*, Volume 1, pp. 853–964. Elsevier.

Bollinger, B., J. Burkhardt, and K. T. Gillingham (2020). Peer effects in residential water conservation: Evidence from migration. *American Economic Journal: Economic Policy 12*(3), 107–33.

Bollinger, B. and K. Gillingham (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science 31*(6), 900–912.

Bollinger, B., K. Gillingham, A. J. Kirkpatrick, and S. Sexton (2021). Visibility and peer influence in durable good adoption. *Available at SSRN 3409420*.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of econometrics 150*(1), 41–55.

Bresnahan, T. F. and P. C. Reiss (1991). Empirical models of discrete games. *Journal of Econometrics 48*(1-2), 57–81.

Brock, W. and S. N. Durlauf (2006). Multinomial choice with social interactions. *The economy as an evolving complex system III*.

Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Brock, W. A. and S. N. Durlauf (2002). A multinomial-choice model of neighborhood effects. *American Economic Review 92*(2), 298–303.

Bugni, F. A., J. Bunting, and T. Ura (2020). Testing homogeneity in dynamic discrete games in finite samples. *Working paper*.

Burkhardt, J., N. W. Chan, B. Bollinger, and K. T. Gillingham (2021). Conformity and conservation: Evidence from home landscaping and water conservation. *American Journal of Agricultural Economics*.

Chen, H., H. Zhang, Z. Hu, Y. Liang, H. Luo, and Y. Wang (2017). Plug-in electric vehicle charging congestion analysis using taxi travel data in the central area of beijing. *arXiv preprint arXiv:1712.07300*.

Chyn, E. and L. F. Katz (2021). Neighborhoods matter: Assessing the evidence for place effects. *Journal of Economic Perspectives 35*(4), 197–222.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

Compiani, G. (2022). Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics 13*(2), 545–591.

Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in ghana. *American economic review 100*(1), 35–69.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics 2*(2), 241–75.

De Paula, A. (2017). Econometrics of network models. In *Advances in Economics and Econometrics: Theory and Applications: Eleventh World Congress*, Volume 1, pp. 268–323. Cambridge University Press Cambridge.

De Paula, A. and X. Tang (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica 80*(1), 143–172.

Dubé, J.-P. H., G. J. Hitsch, and P. K. Chintagunta (2010). Tipping and concentration in markets with indirect network effects. *Marketing Science 29*(2), 216–249.

EIA (2022). Montly energy review. *U.S. Energy Information Administration*.

Ellickson, P. B. and S. Misra (2011). Structural workshop paperestimating discrete games. *Marketing Science 30*(6), 997–1010.

EPA (2022). Inventory of u.s. greenhouse gas emissions and sinks: 1990-2020. *U.S. Environmental Protection Agency, EPA 430-R-22-003*.

Epple, D. and R. E. Romano (2011). Peer effects in education: A survey of the theory and evidence. In *Handbook of social economics*, Volume 1, pp. 1053–1163. Elsevier.

Gandhi, A. and A. Nevo (2021). Empirical models of demand and supply in differentiated products industries.

Gillingham, K. (2021). Designing fuel-economy standards in light of electric vehicles.

Gillingham, K., M. Ovaere, and S. M. Weber (2021). Carbon policy and the emissions implications of electric vehicles.

Gillingham, K. and T. Tsvetanov (2018). Nudging energy efficiency audits: Evidence from a field experiment. *Journal of Environmental Economics and Management 90*, 303–316.

Gillingham, K. T. and B. Bollinger (2021). Social learning and solar photovoltaic adoption. *Management Science*.

Grinblatt, M., M. Keloharju, and S. Ikäheimo (2008). Social influence and consumption: Evidence from the automobile purchases of neighbors. *The review of Economics and Statistics 90*(4), 735–753.

Guerra, J.-A. and M. Mohnen (2020). Multinomial Choice with Social Interactions: Occupations in Victorian London. *The Review of Economics and Statistics*, 1–44.

Hartmann, W. R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing science 29*(4), 585–601.

He, C., O. C. Ozturk, C. Gu, and P. K. Chintagunta (2022). Consumer tax credits for evs: Some quasi-experimental evidence on consumer demand, product substitution, and carbon emissions. *accepted at Management Science*.

Heutel, G. and E. Muehlegger (2015). Consumer learning and hybrid vehicle adoption. *Environmental and resource economics 62*(1), 125–161.

Jansson, J., A. Marell, and A. Nordlund (2010). Green consumer behavior: determinants of curtailment and eco-innovation adoption. *Journal of consumer marketing*.

Jansson, J., T. Pettersson, A. Mannberg, R. Brännlund, and U. Lindgren (2017). Adoption of alternative fuel vehicles: Influence from neighbors, family and coworkers. *Transportation Research Part D: Transport and Environment 54*, 61–73.

Kahneman, D., J. L. Knetsch, and R. Thaler (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American economic review*, 728–741.

Kumar, V. and K. Sudhir (2021). Can friends seed more buzz and adoption?

Li, J. (2019). Compatibility and investment in the us electric vehicle market.

Li, S., L. Tong, J. Xing, and Y. Zhou (2017). The market for electric vehicles: indirect network effects and policy design. *Journal of the Association of Environmental and Resource Economists 4*(1), 89–133.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Mau, P., J. Eyzaguirre, M. Jaccard, C. Collins-Dodd, and K. Tiedemann (2008). The neighbor effect: Simulating dynamics in consumer preferences for new vehicle technologies. *Ecological Economics 68*(1-2), 504–516.

McShane, B. B., E. T. Bradlow, and J. Berger (2012). Visual influence and social groups. *Journal of Marketing Research 49*(6), 854–871.

Muratori, M. (2018). Impact of uncoordinated plug-in electric vehicle charging on residential power demand. *Nature Energy 3*(3), 193–201.

Nair, H., P. Chintagunta, and J.-P. Dubé (2004). Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics 2*(1), 23–58.

Nair, H. S., P. Manchanda, and T. Bhatia (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research 47*(5), 883–895.

Narayanan, S. and H. S. Nair (2013). Estimating causal installed-base effects: A bias-correction approach. *Journal of Marketing Research 50*(1), 70–94.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

OECD (2018). Personalised pricing in the digital era - oecd.

Otsu, T., M. Pesendorfer, and Y. Takahashi (2016). Pooling data across markets in dynamic markov games. *Quantitative Economics 7*(2), 523–559.

Ozaki, R. and K. Sevastyanova (2011). Going hybrid: An analysis of consumer purchase motivations. *Energy policy 39*(5), 2217–2227.

Pettifor, H., C. Wilson, J. Axsen, W. Abrahamse, and J. Anable (2017). Social influence in the global diffusion of alternative fuel vehicles–a meta-analysis. *Journal of Transport Geography 62*, 247–261.

Rapson, D. S. and E. Muehlegger (2021). The economics of electric vehicles.

Remmy, K. et al. (2022). Adjustable product attributes, indirect network effects, and subsidy design: The case of electric vehicles.

Schuitema, G., J. Anable, S. Skippon, and N. Kinnear (2013). The role of instrumental, hedonic and symbolic attributes in the intention to adopt electric vehicles. *Transportation Research Part A: Policy and Practice 48*, 39–49.

Shemesh, J., F. Zapatero, and Y. Zenou (2022). Heterogenous peer effects: How community connectivity affects car purchases.

Shriver, S. K. (2015). Network effects in alternative fuel adoption: Empirical analysis of the market for ethanol. *Marketing Science 34*(1), 78–97.

Sinyashin, A. (2021). Optimal policies for differentiated green products: Characteristics and usage of electric vehicles.

Smart, J. and S. Schey (2012). Battery electric vehicle driving and charging behavior observed early in the ev project. *SAE International Journal of Alternative Powertrains 1*(1), 27–33.

Springel, K. (2016). Network externality and subsidy structure in two-sided markets: Evidence from electric vehicle incentives.

Su, C.-L. (2014). Estimating discrete-choice games of incomplete information: Simple static examples. *Quantitative Marketing and Economics 12*(2), 167–207.

Tebaldi, P., A. Torgovitsky, and H. Yang (2021). Nonparametric estimates of demand in the california health insurance exchange.

Wolske, K. S., K. T. Gillingham, and P. W. Schultz (2020). Peer influence on household energy behaviours. *Nature Energy 5*(3), 202–212.

Yang, S. and G. M. Allenby (2003). Modeling interdependent consumer preferences. *Journal of Marketing Research 40*(3), 282–294.

Zhang, W., P. K. Chintagunta, and M. U. Kalwani (2021). Social media, influencers, and adoption of an eco-friendly product: Field experiment evidence from rural china. *Journal of Marketing 85*(3), 10–27.

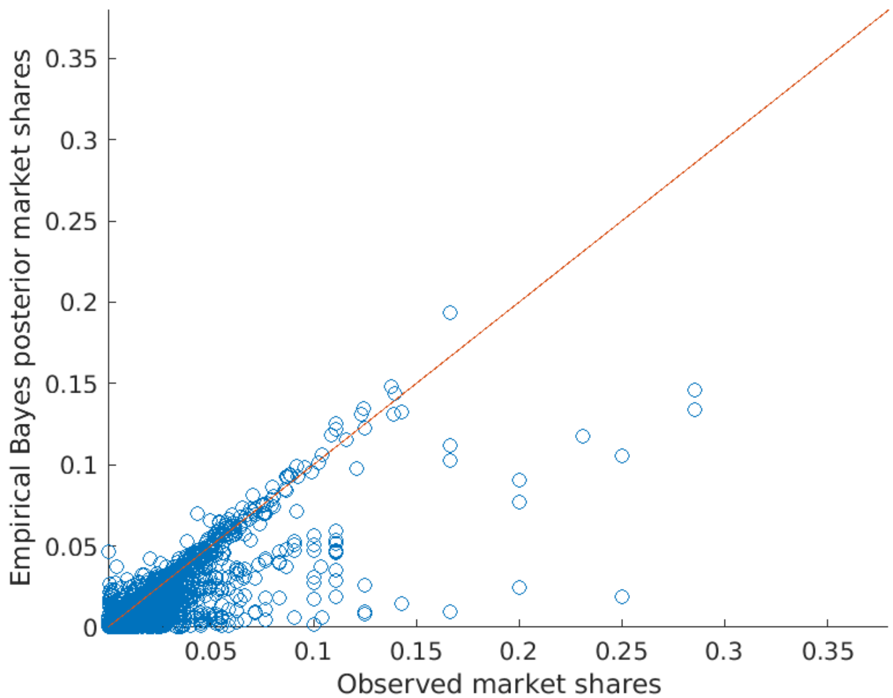
Zhang, W. W. and S. Misra (2022). Coarse personalization. *arXiv preprint arXiv:2204.05793*.

Zhu, X. and C. Liu (2013). Investigating the neighborhood effect on hybrid vehicle adoption. *Transportation research record 2385*(1), 37–44.

Zhuge, C. and C. Shao (2019). Investigating the factors influencing the uptake of electric vehicles in beijing, china: Statistical and spatial perspectives. *Journal of cleaner production 213*, 199–216.

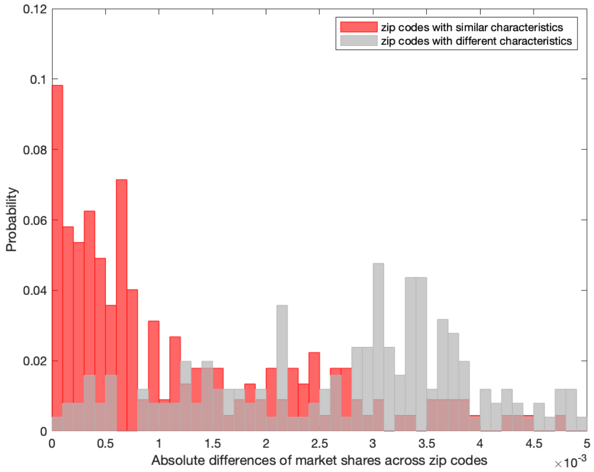
A Additional Figures

Figure 6: Empirical bayes posterior market shares vs. observed market shares

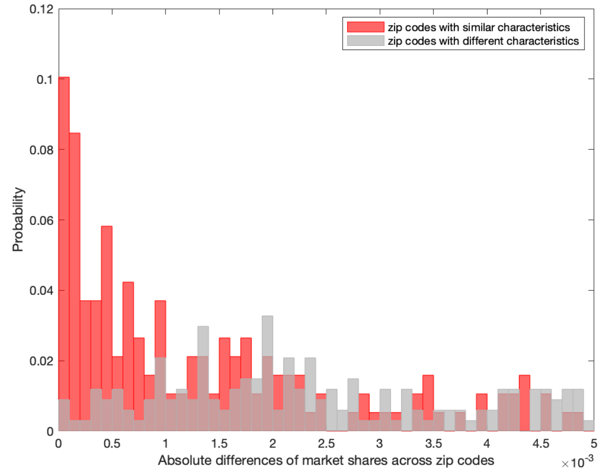


Notes: This figure plots the estimated empirical bayes posterior market shares against observed market shares for 2020.

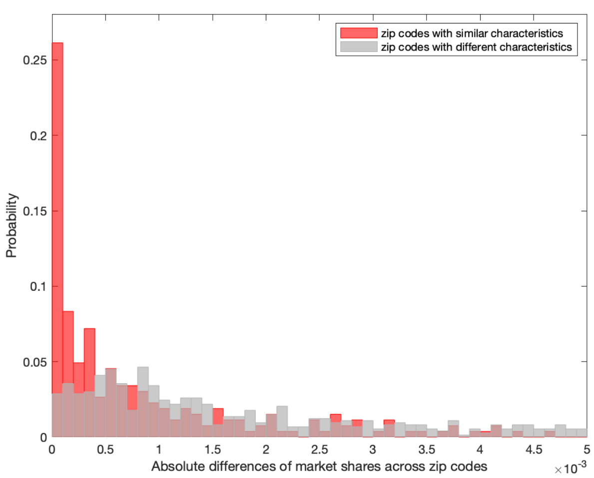
Figure 7: Absolute differences of market shares across zip codes for all years



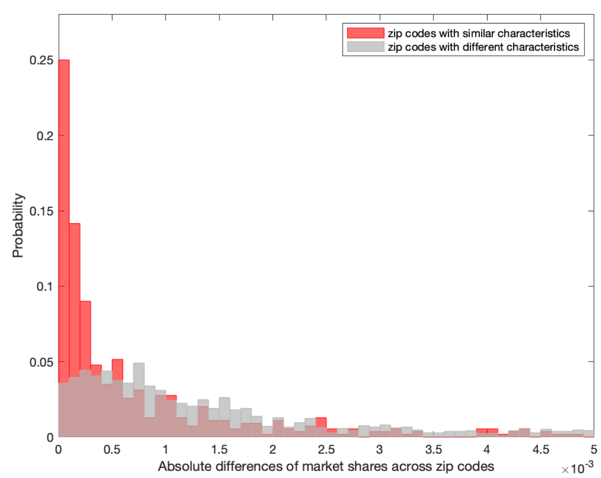
(a) 2015



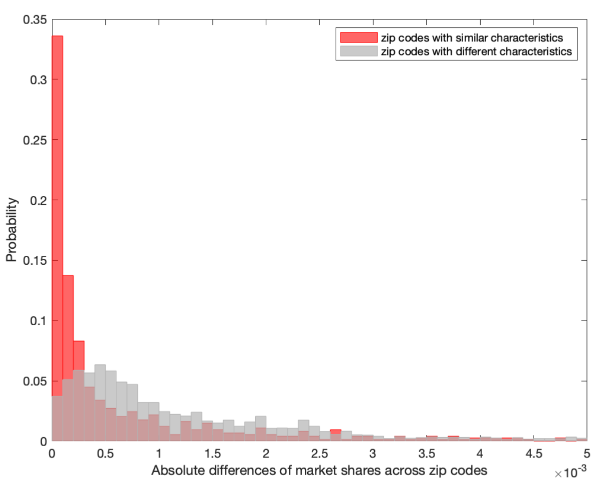
(b) 2016



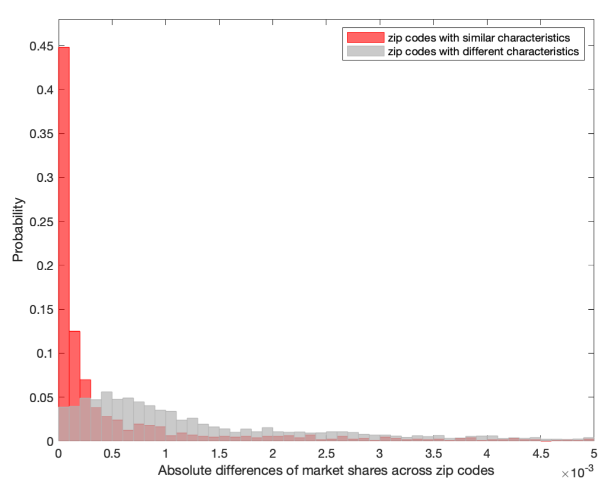
(c) 2017



(d) 2018



(e) 2019



(f) 2020

B Additional Tables

Table 6: First stage results of the instruments for charging stations

	# stations x BEV	# stations x PHEV	# stations x FAR
Intercept	121.329^{***}	-531.650^{***}	-0.334^{***}
	(9.785)	(11.734)	(0.114)
(inside-outside) x BEV	0.522^{***}	-0.038	0.000
	(0.041)	(0.049)	(0.000)
(inside-outside) x PHEV	-0.039	0.520^{***}	0.000
	(0.034)	(0.041)	(0.000)
(inside-outside) x FAR	0.660	0.797	-0.255^{***}
	(0.460)	(0.551)	(0.005)
mpd	-0.804^{***}	-1.715^{***}	0.000
	(0.048)	(0.058)	(0.001)
BEV dummy	153.596^{***}	453.413^{***}	-0.191
	(13.518)	(16.212)	(0.158)
PHEV dummy	118.016^{***}	-81.040^{***}	0.324^{***}
	(5.221)	(6.261)	(0.061)
> midsize dummy	-5.104^{***}	-12.700^{***}	-0.029^{***}
	(0.859)	(1.030)	(0.010)
< midsize dummy	-56.830^{***}	378.763^{***}	-0.273^{**}
	(11.614)	(13.928)	(0.136)
battery range x BEV	1.148^{***}	-1.287^{***}	0.002^{***}
	(0.041)	(0.049)	(0.000)
battery range x PHEV	-2.768^{***}	5.395^{***}	0.013^{***}
	(0.114)	(0.136)	(0.001)
electric motor power x BEV	-2.011^{***}	-0.162^{**}	-0.001
	(0.056)	(0.067)	(0.001)
electric motor power x PHEV	-0.149^{***}	-0.893^{***}	-0.004^{***}
	(0.038)	(0.046)	(0.000)
urban dummy	12.006^{***}	20.201^{***}	-0.043^{***}
	(0.189)	(0.226)	(0.002)
R^2	0.414	0.410	0.050
Adj. R^2	0.414	0.410	0.050
Num. obs.	100,084	100,084	100,084
F statistic	1157.849	1141.214	86.819

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes This table presents first stage results for the instrument for charging stations. The instruments are # stations inside the boundary within bandwidth − # stations outside the boundary within bandwidth. The regression includes all price instruments as well (43 in total), but their estimates are not presented due to space limitations.

C Proofs

C.1 Proof of Theorem 1

Proof. For any g and t, given a set of parameter values $(\alpha, \beta, \gamma_1, ..., \gamma_J)$ and characteristics (ξ_{gt}, χ_{gt}), let us denote $h_{gjt} = (\chi'_{gjt}\beta + \xi_{gjt})/\alpha$ and $\tilde{\gamma}_j = \gamma_j/\alpha$ for notational simplicity. From now on, suppress g and t subscript.

Now let us further define a function $\tilde{\sigma} : \mathcal{S} \to \mathcal{S}$ where $\mathcal{S} = \{(s_1, ..., s_J) : 0 \leq s_j \leq 1 \quad \forall j = 1, ..., J \text{ and } \sum_{j=1}^J s_j \leq 1\} \subset \mathcal{R}^J$. Each component j of the function $\tilde{\sigma}$ is

$$\tilde{\sigma}_j(s|\tilde{\gamma}, h) = \frac{\exp(h_j + \tilde{\gamma}_j s_j)}{1 + \sum_{k=1}^J \exp(h_k + \tilde{\gamma}_k s_k)} \quad \forall j > 0$$

The goal is to show that if $|\tilde{\gamma}_j| < 2$ for all $j = 1, ..., J$, then $\tilde{\sigma}$ has a unique fixed point $s^* \in \mathcal{S}$, i.e. $s^* = \tilde{\sigma}(s^*|\tilde{\gamma}, h)$ given any values of h. Banach fixed point theorem implies that $\tilde{\sigma}$ has a unique fixed point as long as it is a contraction mapping. So it suffices to show that there exists $q \in [0, 1)$ such that $\|\tilde{\sigma}(\bar{s}) - \tilde{\sigma}(\underline{s})\|_\infty \leq q\|\bar{s} - \underline{s}\|_\infty$ for all $\bar{s}, \underline{s} \in \mathcal{S}$.

For any j, we know that $\tilde{\sigma}_j$ is continuous on $\mathcal{S}$ and differentiable on $\mathcal{S}^o$, where $\mathcal{S}^o = \{(s_1, ..., s_J) : 0 < s_j < 1 \quad \forall j = 1, ..., J \text{ and } \sum_{j=1}^J s_j < 1\} \subset \mathcal{R}^J$. We can then apply mean value theorem. Thus, there exists $c \in (0, 1)$ such that $\tilde{\sigma}_j(\bar{s}) - \tilde{\sigma}_j(\underline{s}) = \nabla\tilde{\sigma}_j(c\bar{s} + (1-c)\underline{s}) \cdot (\bar{s} - \underline{s})$ for all $\bar{s}, \underline{s} \in \mathcal{S}$. We now have for any $j = 1, ..., J$

$$\begin{aligned}
|\tilde{\sigma}_j(\bar{s}) - \tilde{\sigma}_j(\underline{s})| &= |\nabla\tilde{\sigma}_j(c\bar{s} + (1-c)\underline{s}) \cdot (\bar{s} - \underline{s})| \\
&= \left|\sum_{k=1}^J \frac{\partial\tilde{\sigma}_j(\tilde{s})}{\partial\tilde{s}_k}|_{\tilde{s}=c\bar{s}+(1-c)\underline{s}} \; (\bar{s}_k - \underline{s}_k)\right| \\
&\leq \sum_{k=1}^J \left|\frac{\partial\tilde{\sigma}_j(\tilde{s})}{\partial\tilde{s}_k}|_{\tilde{s}=c\bar{s}+(1-c)\underline{s}} \; (\bar{s}_k - \underline{s}_k)\right| \quad \text{by triangle inequality} \\
&\leq \left[\sum_{k=1}^J \left|\frac{\partial\tilde{\sigma}_j(\tilde{s})}{\partial\tilde{s}_k}|_{\tilde{s}=c\bar{s}+(1-c)\underline{s}}\right|^p\right]^{\frac{1}{p}} \left[\sum_{k=1}^J |\bar{s}_k - \underline{s}_k|^m\right]^{\frac{1}{m}} \quad \text{by Hölder's inequality} \\
&= \left[\sum_{k=1}^J \left|\frac{\partial\tilde{\sigma}_j(\tilde{s})}{\partial\tilde{s}_k}|_{\tilde{s}=c\bar{s}+(1-c)\underline{s}}\right|\right] \|\bar{s} - \underline{s}\|_\infty \quad \text{by letting } p = 1 \text{ and } m = \infty
\end{aligned}$$

Given the equations above, we realize that it is enough to show

$$\sum_{k=1}^J \left|\frac{\partial\tilde{\sigma}_j(s)}{\partial s_k}\right| < 1 \quad \forall s \in \mathcal{S}^o, \quad \forall j = 1, ..., J \tag{11}$$

This is because we can let $q = \sum_{k=1}^J \left|\frac{\partial\tilde{\sigma}_j(\tilde{s})}{\partial\tilde{s}_k}|_{\tilde{s}=c\bar{s}+(1-c)\underline{s}}\right|$ for any $\bar{s}, \underline{s} \in \mathcal{S}$. Then we have $|\tilde{\sigma}_j(\bar{s}) - \tilde{\sigma}_j(\underline{s})| \leq q\|\bar{s} - \underline{s}\|_\infty$ for all j, so $\|\tilde{\sigma}(\bar{s}) - \tilde{\sigma}(\underline{s})\|_\infty = \max_j\{|\tilde{\sigma}_j(\bar{s}) - \tilde{\sigma}_j(\underline{s})|\} \leq q\|\bar{s} - \underline{s}\|_\infty$.

Now we show that equation (11) holds. To do that, let us first compute the derivatives:

$$\frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} = \frac{-\tilde{\gamma}_k \exp(h_j + \tilde{\gamma}_j s_j)\exp(h_k + \tilde{\gamma}_k s_k)}{[1+\sum_{k=1}^J \exp(h_k + \tilde{\gamma}_k s_k)]^2} = -\tilde{\gamma}_k s_j s_k, \quad k \neq j$$
$$\frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} = \frac{\tilde{\gamma}_j \exp(h_j + \tilde{\gamma}_j s_j)[1+\sum_{k=1}^J \exp(h_k + \tilde{\gamma}_k s_k)] - \tilde{\gamma}_j \exp(h_j + \tilde{\gamma}_j s_j)^2}{[1+\sum_{k=1}^J \exp(h_k + \tilde{\gamma}_k s_k)]^2} = \tilde{\gamma}_j s_j (1 - s_j)$$

Therefore, we have for all j

$$\begin{aligned}
\sum_{k=1}^J \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| &= |\tilde{\gamma}_j| s_j (1-s_j) + s_j \big(\sum_{k \neq j} |\tilde{\gamma}_k| s_k \big) \\
&= s_j \left[|\tilde{\gamma}_j|(1-s_j) + \sum_{k \neq j} |\tilde{\gamma}_k| s_k \right] \\
&< 2 s_j \left[(1-s_j) + \sum_{k \neq j} s_k \right] \quad \text{since } |\tilde{\gamma}_j| < 2 \quad \forall j \\
&= 2 s_j (1 - s_j + 1 - s_j - s_0) \quad \text{since } \sum_{k=0}^J s_k = 1 \\
&= 2 s_j (2 - 2 s_j - s_0) \\
&= -4 \left(s_j - \frac{2 - s_0}{4} \right)^2 + 4 \left(\frac{2 - s_0}{4} \right)^2 \\
&\leq 4 \left(\frac{2 - s_0}{4} \right)^2 \\
&= (1 - \frac{1}{2} s_0)^2 < 1 \quad \text{because } 0 < s_0 < 1
\end{aligned}$$

□

C.2 Proof of Theorem 2

Proof. Suppress the subscript g and t. Define a function $\tilde{\sigma} : \mathcal{S} \rightarrow \mathcal{S}$ where $\mathcal{S} = \{(s_1, ..., s_J) : 0 \leq s_j \leq 1 \quad \forall j = 1, ..., J \text{ and } \sum_{j=1}^J s_j \leq 1\} \subset \mathcal{R}^J$. Each component j of the function $\tilde{\sigma}$ is

$$\tilde{\sigma}_j(s|\theta, \alpha, \xi, \chi) = \int \frac{\exp((\chi_j' \beta_i + \gamma_i s_j + \xi_j)/\alpha)}{1 + \sum_{k=1}^J \exp((\chi_k' \beta_i + \gamma_i s_k + \xi_k)/\alpha)} dF(\beta_i, \gamma_i|\theta) \tag{12}$$

Similar to the argument in the proof of Theorem 1 in Appendix C.1, it suffices to show

$$\sum_{k=1}^J \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| < 1 \quad \forall s \in \mathcal{S}^o, \quad \forall j = 1, ..., J \tag{13}$$

Let $\tilde{\gamma}_i = \frac{\gamma_i}{\alpha}$. The derivatives are

$$
\begin{aligned}
\frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} &= \int -\tilde{\gamma}_i r_{ij} r_{ik} \, dF(\beta_i, \gamma_i|\theta), \quad k \neq j \\
\frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} &= \int \tilde{\gamma}_i r_{ij}(1 - r_{ij}) \, dF(\beta_i, \gamma_i|\theta) \\
\text{where } r_{ij} &= \frac{\exp((\chi_j \beta_i + \gamma_i s_j + \xi_j)/\alpha)}{1 + \sum_{k=1}^{J} \exp((\chi_k \beta_i + \gamma_i s_k + \xi_k)/\alpha)}
\end{aligned}
$$

Furthermore, let us define $r_{i0} = \frac{1}{1+\sum_{k=1}^{J} \exp((\chi_k \beta_i + \gamma_i s_k + \xi_k)/\alpha)}$. So $\sum_{k=0}^{J} r_{ik} = 1$. We have for all j

$$
\begin{aligned}
\sum_{k=1}^{J} \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| &\leq \int |\tilde{\gamma}_i| r_{ij} \left[(1 - r_{ij}) + \sum_{k \neq j} r_{ik} \right] \, dF(\beta_i, \gamma_i|\theta) \\
&= \int |\tilde{\gamma}_i| r_{ij} (2 - 2r_{ij} - r_{i0}) \, dF(\beta_i, \gamma_i|\theta) \\
&\leq \int -2|\tilde{\gamma}_i| \left(r_{ij} - \frac{2 - r_{i0}}{4} \right)^2 + 2|\tilde{\gamma}_i| \left(\frac{2 - r_{i0}}{4} \right)^2 \, dF(\beta_i, \gamma_i|\theta) \\
&\leq \int 2|\tilde{\gamma}_i| \left(\frac{2 - r_{i0}}{4} \right)^2 \, dF(\beta_i, \gamma_i|\theta)) \\
&= \frac{1}{2} \int |\tilde{\gamma}_i| (1 - \frac{1}{2} r_{i0})^2 \, dF(\beta_i, \gamma_i|\theta) \\
&\leq \frac{1}{2} \int |\tilde{\gamma}_i| \, dF(\beta_i, \gamma_i|\theta) \quad \text{because } 0 \leq r_{i0} \leq 1 \text{ and } (1 - \frac{1}{2} r_{i0})^2 \leq 1 \\
&= \frac{1}{2\alpha} \mathbb{E}\left[|\gamma_i| \right] \\
&< 1 \quad \text{because } \mathbb{E}\left[|\gamma_i| \right] < 2\alpha
\end{aligned}
$$

□

C.3 Proof of Theorem 3

Proof. Suppress the subscript g and t. Define a function $\tilde{\sigma} : \mathcal{S} \to \mathcal{S}$ where $\mathcal{S} = \{(s_1, ..., s_J) : 0 \leq s_j \leq 1 \quad \forall j = 1, ..., J \text{ and } \sum_{j=1}^{J} s_j \leq 1\} \subset \mathcal{R}^J$. Each component j of the function $\tilde{\sigma}$ is

$$
\tilde{\sigma}_j(s|\theta, \alpha, \xi, \chi) = \int \frac{\exp((\chi_j' \beta_i + \gamma_{ij} s_j + \xi_j)/\alpha)}{1 + \sum_{k=1}^{J} \exp((\chi_k' \beta_i + \gamma_{ik} s_k + \xi_k)/\alpha)} dF(\beta_i, \gamma_i|\theta) \tag{14}
$$

Similar to the argument in the proof of Theorem 2 in Appendix C.2, it suffices to show

$$
\sum_{k=1}^{J} \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| < 1 \quad \forall s \in \mathcal{S}^o, \quad \forall j = 1, ..., J \tag{15}
$$

Let $\tilde{\gamma}_{ij} = \frac{\gamma_{ij}}{\alpha}$. The derivatives are

$$
\begin{aligned}
\frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} &= \int -\tilde{\gamma}_{ik} r_{ij} r_{ik} \; dF(\beta_i, \gamma_i|\theta), \quad k \neq j \\
\frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} &= \int \tilde{\gamma}_{ij} r_{ij}(1 - r_{ij}) \; dF(\beta_i, \gamma_i|\theta) \\
\text{where } r_{ij} &= \frac{\exp((\chi_j \, \beta_i + \gamma_{ij} s_j + \xi_j)/\alpha)}{1 + \sum_{k=1}^{J} \exp((\chi_k \, \beta_i + \gamma_{ik} s_k + \xi_k)/\alpha)}
\end{aligned}
$$

Take the absolute value of the equations above, we have

$$
\begin{aligned}
\left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| &\leq \int |\tilde{\gamma}_{ik}| r_{ij} r_{ik} \; dF(\beta_i, \gamma_i|\theta), \quad k \neq j \\
&= \int |\bar{\gamma}_k/\alpha + \Pi_\gamma D_i/\alpha| \; r_{ij} r_{ik} \; dF(D_i|\theta) \\
&\leq \int (|\bar{\gamma}_k/\alpha| + |\Pi_\gamma D_i/\alpha|) \; r_{ij} r_{ik} \; dF(D_i|\theta) \\
&\leq \int (\bar{\gamma}^*/\alpha + |\Pi_\gamma D_i/\alpha|) \; r_{ij} r_{ik} \; dF(D_i|\theta) \\
\left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right| &\leq \int |\tilde{\gamma}_{ij}| r_{ij}(1 - r_{ij}) \; dF(\beta_i, \gamma_i|\theta) \\
&\leq \int (\bar{\gamma}^*/\alpha + |\Pi_\gamma D_i/\alpha|) \; r_{ij}(1 - r_{ij}) \; dF(D_i|\theta)
\end{aligned}
$$

Summing up the absolute values, we get

$$
\begin{aligned}
\sum_{k=1}^{J} \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| &\leq \int (\bar{\gamma}^*/\alpha + |\Pi_\gamma D_i/\alpha|) \; r_{ij} \left[(1 - r_{ij}) + \sum_{k \neq j} r_{ik} \right] \; dF(D_i|\theta) \\
&\leq \int (\bar{\gamma}^*/\alpha + |\Pi_\gamma D_i/\alpha|) \; r_{ij}(2 - 2r_{ij} - r_{i0}) \; dF(D_i|\theta) \\
&\leq \frac{1}{2} \int \bar{\gamma}^*/\alpha + |\Pi_\gamma D_i/\alpha| \; dF(D_i|\theta) \\
&= \frac{1}{2\alpha} \int \bar{\gamma}^* + |\Pi_\gamma D_i| \; dF(D_i|\theta) \\
&= \frac{1}{2\alpha} \big(\bar{\gamma}^* + \mathbb{E}\left[|\Pi_\gamma D_i| \right] \big) \\
&< 1 \quad \text{because } \bar{\gamma}^* + \mathbb{E}\left[|\Pi_\gamma D_i| \right] < 2\alpha
\end{aligned}
$$

$\square$

C.4 Derivation of the Jacobian Matrices

Because we know on the reduced space, we can write market shares as well-defined functions of market characteristics. In particular, we have

$$s_{gj} = \sigma_j \left(\xi_g, \chi_g | \theta, \alpha\right)$$

Thus, we can derive the following

$$\sigma_j \left(\xi_g, \chi_g | \theta, \alpha\right) = \int \frac{\exp((\chi_j \, \beta_i + \gamma_i \sigma_j(\xi_g, \chi_g) + \xi_j)/\alpha)}{1 + \sum_{k=1}^{J} \exp((\chi_k \, \beta_i + \gamma_i \sigma_k(\xi_g, \chi_g) + \xi_k)/\alpha)} dF(\beta_i, \gamma_i | \theta)$$

By law of total differentiation, we get Jac the Jacobian matrix with respect to ξ satisfies the following:

$$Jac = A + BJac \tag{16}$$

$$\Rightarrow (I - B)Jac = A \tag{17}$$

where A is the direct effect of ξ, and B is the social influence weights. More specifically,

$$\begin{aligned}
A_{jj} &= \int r_{ij}(1 - r_{ij}) \; dF(\beta_i, \gamma_i | \theta) \\
A_{jk} &= -\int r_{ij} r_{ik} \; dF(\beta_i, \gamma_i | \theta) \\
B_{jj} &= \int \tilde{\gamma}_i r_{ij}(1 - r_{ij}) \; dF(\beta_i, \gamma_i | \theta) \\
B_{jk} &= -\int \tilde{\gamma}_i r_{ij} r_{ik} \; dF(\beta_i, \gamma_i | \theta) \\
\text{where } r_{ij} &= \frac{\exp((\chi_j \, \beta_i + \gamma_i s_j + \xi_j)/\alpha)}{1 + \sum_{k=1}^{J} \exp((\chi_k \, \beta_i + \gamma_i s_k + \xi_k)/\alpha)} \text{ and } \tilde{\gamma}_i = \frac{\gamma_i}{\alpha}
\end{aligned}$$

I show that $I - B$ is invertible in Appendix C.5, so the Jacobian matrix can be computed from

$$Jac = (I - B)^{-1} A \tag{18}$$

Because we have defined $\delta_{gj} = x_{gj}^{(1)} + \xi_{gj}$, the Jacobian matrix with respect to δ is the same as the one with respect to ξ.

C.5 Proof of Lemma 1

Proof. As discussed in Appendix C.4, the Jacobian matrix satisfies equation (17). In order to show that the Jacobian matrix is invertible, I first show that $I - B$ is invertible.

Recall from Appendix C.2 that we have already shown $\sum_{k=1}^{J} \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| < 1$ for all $s \in \mathcal{S}^o$ and for all $j = 1, ..., J$. Thus, we have

$$\sum_{k \neq j} \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_k} \right| < 1 - \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right| \leq \left| 1 - \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right| \tag{19}$$

The second inequality holds because (1) if $\frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} < 0$, then $1 - \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right| < 1 < \left| 1 - \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right|$; (2) if $0 \leq \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \leq 1$, then $1 - \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right| = 1 - \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} = \left| 1 - \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right|$ since $1 - \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} > 0$; (3) if $\frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} > 1$, then $1 - \left| \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right| < 0 < \left| 1 - \frac{\partial \tilde{\sigma}_j(s)}{\partial s_j} \right|$.

Equation (19) implies the following:

$$\sum_{k \neq j} \left| - \int \tilde{\gamma}_i r_{ij} r_{ik} \, dF(\beta_i, \gamma_i | \theta) \right| < \left| 1 - \int \tilde{\gamma}_i r_{ij} (1 - r_{ij}) \, dF(\beta_i, \gamma_i | \theta) \right| \quad \forall j \tag{20}$$

The LHS of the equation above is also $\sum_{k \neq j} |(I - B)_{jk}|$ and the RHS is also $|(I - B)_{jj}|$. Therefore, the matrix $(I - B)$ is strictly diagonally dominant. By the Levy Desplanques theorem, the matrix $(I - B)$ is non-singular.

□

C.6 Proof of Theorem 4

Proof. I have shown in Lemma 1 that the Jacobian matrix with respect to δ is nonsingular on the reduced parameter space. If all goods are also weak substitutes, then the system of demand functions σ is inverse isotone on its domain. The arugment follows from Theorems 1 and 2 in Berry et al. (2013). Since inverse isotone functions are injective, the demand functions are invertible.

□