UNIVERSITAT DE BARCELONA

Prof. Dr. Joan-Ramon Borrell
Professor Titular d'Universitat

Departament d'Econometria, Estadística i Economia Aplicada

Secció de Polítiques Públiques

Facultat d'Economia i Empresa
John M. Keynes, 1-11
E-08034 Barcelona

http://www.ub.edu/gim/membre/joan-ramon-borrell/

Tel. +34 93 403 97 22
jrborrell@ub.edu

June 14th, 2024

Student Condition of the Author

To Whom It May Concern,

I am writing to confirm that Ragıp Kaan Erdemli is a 4th-year Ph.D. student in Economics at the University of Barcelona, currently under my supervision.

Do not hesitate to contact me if you need further details.

Sincerely yours,

BORRELL ARQUE, JOAN RAMON (AUTENTICACIÓN)

Firmado digitalmente por BORRELL ARQUE, JOAN RAMON (AUTENTICACIÓN)
Fecha: 2024.06.14 15:55:46 +02'00'

Prof. Dr. Joan-Ramon Borrell
Tenured Associate Professor in Economics -UB School of Economics
University of Barcelona
BA Barcelona (1993), MA Warwick (1996), PhD Barcelona (1999), Post-Doc Harvard (2001)

Do transparency policies work as expected? Evidence from the retail gasoline market

(Work in Progress)

Ragıp Kaan Erdemli*

June 2024

Abstract

Market transparency policies aim to increase the proportion of informed consumers and increase competition among firms. However, these policies may also facilitate coordination. This paper examines how a price transparency policy influences spatial competition among firms. Specifically, I use data from the fuel retail market in the Sydney Metropolitan Area, Australia, where a transparency policy regarding fuel prices was implemented. To understand the policy's impact on profit margins, I develop an empirical model of demand and supply. A key feature of this model is that it accommodates both fully informed consumers and those with limited information. At the beginning of each period, consumers are informed only about a subset of petrol stations close to their daily commuting path. Each period, consumers decide whether to obtain information about all petrol stations in the market through online transparency tools. Access to these online transparency tools is costly, so consumers who decide not to use them make purchasing decisions considering only already known close stations. Additionally, by incorporating multi-market contact into the model, I account for heterogeneous coordination among firms. My findings indicate that there is coordination among firms and as there are more informed consumers, profit margins decline.

Keywords: market transparency, multi-market contact, retail gasoline prices

JEL Classification: D22, D43, D83, L13, L50, Q41

*rkerdemli@ub.edu; University of Barcelona, Department of Econometrics, Statistics & Applied Economics

1 Introduction

The regulation of market transparency is an important issue that interests both economists and policymakers. As a general definition, market transparency encompasses all informational aspects of a market. More precisely, it implies how well-informed the participating agents are about prices, quality, and other factors of products in the market. Market transparency policies aim to increase the proportion of informed consumers who have perfect information about prices. In this way, these policies are expected to intensify competition.

One of the main concerns regarding these policies is that improved market transparency may lead to tacit collusion. Although transparency policies primarily target consumers, they also provide firms with more information about their competitors. As market transparency increases, detecting deviations from collusion becomes easier for firms, which may help sustain collusive behavior. On the other hand, increasing market transparency makes price undercutting more profitable for competing firms and renders it more challenging for firms to uphold cooperative agreements. The results of transparency policies will depend on the strengths of these opposing forces (Møllgaard & Overgaard 2001, Schultz 2005).

Empirical studies on market transparency provide conflicting results. Albek et al. (1997), focusing on the Danish ready-mixed concrete market and using post-transparency data, find that the prices of ready-mixed concrete rose significantly after firms were required to disclose price information. Bryne & de Roos (2019) examine 15 years of post-transparency data from the fuel retailing market. Using the price data, they demonstrate how tacit collusion was formed after the implementation of the price transparency policy in Western Australia. Luco (2019) examines the impact of the price transparency policy on competition in the Chilean retail gasoline industry. His findings suggest that the policy reduced the intensity of competition on average. Furthermore, he finds that the policy had distributional effects. More precisely, margins increased the least and even decreased in high-income areas, while they increased the most in low-income areas.

On the other hand, there are empirical studies that provide evidence in favor of transparency policies. Montag & Winter (2020) examine the transparency policy in Germany's fuel retail market.

Using data on fuel prices from stations near the French-German border, they employ a difference-in-differences approach to show that the transparency policy resulted in decreased prices. Ater & Rigbi (2017) examine mandatory online disclosure for supermarket prices in Israel. Using prices before and after the transparency policy, they demonstrate that supermarket prices declined.

In this study, I examine a price transparency policy in the fuel retailing market in New South Wales (NSW), Australia. Since the majority of the branded petrol stations were already informed about prices through a private entity, Informed Sources, the policy primarily changes price information available to the consumers. This allows me to identify the effect of consumer-side information changes on margins separately from the effect of coordination among the stations. Given this context, the study aims to address how consumer-side transparency influences competition in a market and how existing coordination among firms affects the effectiveness of transparency policies on market competition and profit margins.

The NSW fuel transparency policy operates through the website and web application of the local government. Initially, only a small percentage of people used these information channels, but over time, the user base gradually increased. To examine how consumer-side transparency influences profit margins, I extend the demand estimation model developed by Berry, Levinsohn, and Pakes (1995) to accommodate both informed consumers and those with limited information. Unlike traditional models assuming a Bertrand-Nash pricing game among firms, I relax this assumption by introducing multi-market considerations on the supply side. This extension allows for heterogeneous coordination among firm pairs based on the number of markets in which they compete against each other, facilitating the construction of counterfactual scenarios with varying levels of market coordination.

This paper makes two main contributions to the existing literature. First, the available data in my study exclusively pertains to the post-policy period, which poses a challenge due to the absence of a reliable control group. In this context, employing reduced-form techniques for causal inference may not be viable. To address this empirical challenge, I introduce a novel model that enables me to assess the impact of price transparency policies on competition among gas stations, utilizing only post-transparency data without the need for a control group. Specifically, I integrate the available data on online transparency tool users into a demand estimation model, construct cost functions for firms, and jointly estimate the model's parameters. Once the model parameters are estimated, I construct counterfactual scenarios with varying percentages of informed consumers.

Second, findings in the previous literature on transparency policies are puzzling regarding overall effect of them on profit margins. In some markets it could be the case that cost of being informed is relatively low for firms compared to consumers and firms get informed immediately while it takes some time for consumers to get informed. In an extreme case all firms get informed immediately and start coordinating perfectly. The unique context allows me to construct counterfactuals which mimic this extreme case with varying informed consumer shares. Utilizing such counterfactual scenarios, this study is expected to provide an explanation for null results and provide further empirical evidence on effectiveness of transparency policies when firms are already coordinating.

2 Background and Data

In this section, first I describe the fuel retailing market, and fuel transparency policy. Later, I explain the data sources. and last subsection shows descriptive statistics.

2.1 Market description

In New South Wales (NSW) retail gasoline market, there are five 5 major players. Namely major brands are Caltex, BP, 7-Eleven, Metro Fuel and Coles Express[1]. Among major players only BP and Caltex are vertically integrated. Coles Express and 7-Eleven are supplied by Viva Energy (Shell) and Mobil respectively. Both Coles Express and 7-Eleven are supermarket chains and in general their station convenience stores are larger. Caltex had a partnership with supermarket chain Woolworths and its stations' convenience stores vary in size depending on the location. In Sydney Metropolitan Area between 2017 and 2019 major brands had a share around 60%. In this period their total station share is increase slightly due to establishment of several new stations and brand changes in existing stations.

[1] Caltex and Woolworths, a supermarket chain, are partners and from the data it is difficult to differentiate which station is owned by which. Therefore for the analysis I assume that all Caltex and Woolworths stations belong to same firm.

In August 2016 NSW Government introduced fuel transparency scheme that covers all sites in the state. With this policy reporting prices has became mandatory for petrol stations. If the reported price and site price do not match stations have to pay a penalty. Therefore, all stations have strong incentive to report prices correctly. Although petrol stations are expected to comply as soon as possible, it took several months for all stations to adjust to the new policy. For that reason, in my analysis I focus on the 2017-2019 period. The reported fuel prices shared with public via a free online service called FuelCheck which has a website and a mobile app. FuelCheck shows current fuel prices in all stations in NSW.

It is important to stress that, by 2014, before the transparency policy majority of the branded petrol stations, namely BP, Coles Express, Caltex, Woolworths and 7-Eleven, were already informed about prices (Bryne et al 2023). Most notable channel they used was Informed Sources, a company that provided information on rival prices as a service. In August 2014 Australian Competition and Consumer Commission (ACCC) instituted proceedings against Informed Sources. At the end of 2015 case resolved and as a part of resolution through its subsidiary MotorMouth, Informed Sources started to provide pricing information for consumers at daily level. following the proceedings Informed Sources started to provide price information for public as well through its subsidiary MotorMouth.

2.2 Data sources

Price and Registry Data

The NSW Government's FairTrade unit is responsible for maintaining the NSW FuelCheck mobile app and its associated website. This unit provides price report records for fuel stations in New South Wales (NSW) and information about the registrants of these stations, who are the individuals or entities that own them. The registrant data provides insights into station ownership, including location, station name, and date of registration.

While petrol price records are available from August 2016 onward, it's worth noting that in the initial months following the policy's implementation, some stations did not promptly adjust and begin reporting their prices. Additionally, the COVID-19 pandemic introduced significant disruptions to the factors governing supply and demand, making it challenging to separate and analyze these changes. Due to these complexities, this study focuses on the period from 2017 to 2019.

According to the transparency act, petrol stations are responsible for reporting any changes in fuel prices to NSW FairTrade. Leveraging the price report records, I have constructed daily price datasets for each station, covering the period from 2017 to 2019. These datasets encompass daily prices for all fuel types [2].

The NSW FuelCheck data also includes address information. Figure 4 visually represents the distribution of stations within Greater Sydney as of January 2017, while Table 1 provides a summary of station statistics for January 2017 and December 2019. As indicated in the table, during this timeframe of interest, the number of stations in the region increased. In early 2017, there were 906 stations, which rose to 938 by the end of 2019. Although major brands saw a slight increase in their market share over time, the overall market structure remained relatively stable.

Petrol Station Characteristics

The characteristics of petrol stations, including their size, the number of pumps, and the presence of additional amenities such as car wash and repair shops, play a significant role in influencing consumers' purchasing decisions. Consequently, I conducted manual data collection on these station attributes using Google Maps' street view feature. In most cases, Google has archived multiple street views of each station over time, enabling me to monitor changes in station characteristics. Through street views, I gathered data on the number of service islands, pump counts, and the presence of car wash and repair shop facilities.

The primary reason for collecting data on both the number of islands and the number of pumps is their distinct informational value. In my dataset, there are instances where stations possess the same number of pumps but differ in the number of islands. Typically, older service islands

[2] Sometimes stations stop reporting a certain fuel type for a long time. While constructing daily price data, I assumed that if a station is not reporting price for a fuel type more than a month then that fuel type is not available after 30 days of last price reporting.

accommodate fewer pumps, and including service islands as a variable provides supplementary insights into the quality of petrol stations.

I can track changes in station characteristics by comparing street views captured on different dates. However, street views, in general, do not provide precise dates for station upgrades or downgrades. In many cases, they offer a timeframe during which the change occurred, and occasionally, I can pinpoint the exact month of maintenance based on street views. I have observed that substantial maintenance activities usually entail station closures for at least a few months. Utilizing price reporting data, I can identify the months during which a station was temporarily out of service for updates. However, there are instances where stations undergo changes without requiring closure, such as the addition of a new car wash or the replacement of some service islands with newer ones while the station remains operational. In such cases, price data is not informative, and I make an assumption by selecting the midpoint of the change period as an estimate.

It is crucial to emphasize that in many cases, it remains unclear whether repair shops and car wash facilities are owned by the nearby petrol station. While it is evident in some cases, particularly when fully automated car washes are present, that car washes are directly owned by the station, making similar conclusions about repair shops is challenging. Repair shops may be operated by other entities under contractual agreements with the station brand. The available data does not provide insights into ownership arrangements. Due to the ambiguity surrounding ownership, in the model, these characteristics will be considered solely from the demand side.

Location of Terminals and Terminal Gate Prices

I gathered data on terminal locations and the monthly average terminal gate prices for gasoline from the Oil Price Information Service (OPIS). This dataset also provides insights into which upstream brands are responsible for supplying fuel from specific terminals. Notably, there are cases where different upstream brands source fuel from the same terminal but employ varying pricing strategies.

In my analysis, I've made certain assumptions to address these scenarios. Vertically integrated brands are assumed to procure fuel from their own upstream facilities, while brands with established upstream contracts are assumed to exclusively purchase fuel from their designated upstream supplier. For instance, 7-Eleven and Coles Express source their fuel from Mobil and Shell (Viva Energy), respectively. Additionally, there are downstream brands and independent stations that lack a designated supplier. In such cases, I assume that these entities exclusively obtain their fuel from independent upstream brands, specifically Freedom and Park Fuels.

With these assumptions in place, I proceed to calculate the shortest driving route from each station to its supplier terminal. It's important to note that a few upstream brands provide fuel from multiple terminals. In these instances, I select the nearest terminal when determining the distance between a station and its supplier. The calculated distance to the supplier terminal, along with the corresponding terminal gate prices, serves as instrumental variables in my analysis. These instruments are instrumental in identifying the price coefficient on the demand side.

Market Shares

In this study, I have utilized market share data sourced from the Oil Price Information Service (OPIS), spanning a twelve-month period from 2017 to 2019. This dataset, collected on a monthly basis, is derived from a subset of transactions occurring at petrol stations. OPIS obtains this subset of transactional data through its partnership with Fleet Card [3]. This dataset provides essential insights, including information on how frequently customers visit each station and the type of fuel they purchase, distinguishing between gasoline and diesel. Subsequently, the determination of market shares for each station relies on this transactional dataset.

To ascertain the reliability and representativeness of the market shares derived from Fleet Card transactions, OPIS conducted a comprehensive market examination. The findings of this examination revealed that Fleet card users exhibit purchasing patterns similar to those of the average

[3] There are a few stations/chains that do not accept fleet cards. Within Greater Sydney, there are no chains that do not accept fleet cards but few stations. Following Houde 2012 ,for those stations, I imputed the missing shares via linear regression methods. The explanatory variables include average market shares of branded stations (within 3km and 6km), average market shares of unbranded stations (within 3km and 6km), brand of station, number of pumps and service islands, presence of car wash, presence of repair shop, variety of gasoline (ULP,P95,P98 and E10) served in the station, monthly gasoline price, number of close rivals (1km, 3km and 10km), average size (in pumps and service islands) of rival stations in 6km, ratio of car washes and repair shops in rivals within 6km, population density of its grid (1km), population density of adjacent grids, monthly average terminal gate prices for the station and employment within 3km,6km and 9km.

consumer when it comes to fuel procurement. Notably, the global business development team at OPIS informed me of their prior research in the United States' fuel retailing market. In this research, they conducted a comparative analysis, aligning demand data obtained from actual station sales with market share data constructed through Fleet card usage. They found a strong correlation between the market share data derived from Fleet cards and the actual market shares calculated from station sales data.

Within the context of my analysis, it is essential to establish the outside good share. To achieve this, I turned to data collected from the New South Wales Household Travel Survey, an annual endeavor aimed at gathering data concerning personal travel behaviors and transportation modes. This survey provides valuable insights into the percentage of the metropolitan area's population that utilizes various travel methods, including active transport (such as cycling and walking), public transportation (including buses and trains), and a combination of these options. In the framework of this study, the outside good share is defined as the percentage of the metropolitan population opting for travel methods other than using private vehicles.

Website - App Usage and User Distribution

The NSW Fair Trading unit also maintains records of monthly website visits to the FuelCheck platform and compiles data on monthly downloads of the FuelCheck mobile application for both Android and iOS devices. I use this information to figure out what percentage of consumers are aware of fuel prices. NSW Fair Trading has also provided geolocation of price viewing request of subsample of FuelCheck. Using this information I deduce what portion of users are within Greater Sydney and how they are distributed.

In the process of constructing this monthly share of informed consumers, I make certain assumptions. Firstly, I assume that each visit to the website corresponds to a different individual, and I consider consumers who have downloaded the NSW FuelCheck app as recurring users. With this assumption in mind, I aggregate the monthly app downloads with the monthly website visits to determine the total number of users. Using sample user geolocation data I deduce what share resides in Greater Sydney. Then using within Greater Sydney geolocation of sample users I determine how total number of informed consumer are distributed among markets. Subsequently, I divide this figure by the total number of adults in corresponding market to find share of consumers in that market..

Commuting Patterns

The Australian Bureau of Statistics (ABS) provides population and housing data at the Statistical Area (SA)1 level, with SA1s representing the smallest units as defined in the Australian Statistical Geography Standard. On average, SA1s encompass a population of approximately 400 individuals. The 2016 Australian Census data offers comprehensive information on daily commuting patterns, including origin-destination data of citizens. This origin-destination data is made publicly available at the SA1 to SA2[4]. Moreover, data pertaining to weekly income, which serves as a proxy for capturing variations in individual preferences, is accessible at the SA1 to SA2. I combine this information with commuting patterns data.

To enhance the granularity of destination areas for commuters, I leverage employment levels within the spatial units known as travel zones (TZs). These TZs are defined by Transport Performance and Analytics and provide detailed geographic segmentation. Although TZs are not classified as statistical areas, they often closely align in terms of size with SA1s. To define smaller destination areas for commuters, I utilize employment data within TZs situated within a specific SA2. First, I compute the cumulative employment levels within a given SA2. Then, based on the employment distribution across TZs within that SA2, I assign probabilities to each TZ, reflecting the likelihood of each being the destination for commuters. The center of these travel zones is designated as the focal point.

2.3 Overview of the Data

In this section, I provide descriptive statistics and examine markups over time by only using price data and terminal gate prices. During the 2017-2019 period there has been some changes in the market structure and station characteristics. However, as Table 1 illustrates overall these were

[4]SA2s are the second smallest statistical area and they contain multiple SA1S.

not major changes. In Sydney Metropolitan Area (SMA) during this period, there were some new station constructions but only few stations permanently closed. With the new stations and change in the ownership of some stations major brands increased their station share slightly.

	January 2017	**December 2019**
Number of pumps	27.15	29.07
	(13.269)	(13.1)
Car wash	0.160	0.165
	(0.367)	(0.371)
Repair shop	0.347	0.336
	(0.476)	(0.472)
Rivals within 1km	1.486	1.601
	(1.284)	(1.352)
Rivals within 3km	10.56	10.87
	(7.094)	(7.188)
Major brands' share	0.627	0.646
Station number	910	938

Table 1: Summary Statistics on Stations
Standard deviations in parentheses

I have chosen to focus my primary research efforts on unleaded petrol (gasoline) for several well-defined reasons. First and foremost, in contrast to many other countries, the majority of private vehicles in Australia are equipped with gasoline engines, with diesel vehicles constituting a minority portion[5]. Diesel fuel is primarily used by heavy vehicles, and it is highly likely that establishments with heavy vehicle fleets have already established long-term contracts with specific petrol stations or brands. Consequently, the introduction of transparency policies is expected to have limited impact on the selection of petrol stations for these establishments. Furthermore, due to constraints in the available data, it is challenging to determine the market share attributed to private diesel vehicles for each individual petrol station. If diesel is included in the analysis unobservable long-term contracts between petrol stations and the establishments will be contaminating the findings. Additionally, the absence of available data regarding job commuting patterns poses a significant limitation. This constraint renders it unfeasible to identify which petrol stations would serve as optimal refueling choices for heavy vehicles based on commuting data.

[5]During the period of interest diesel vehicle share was less than 7%.

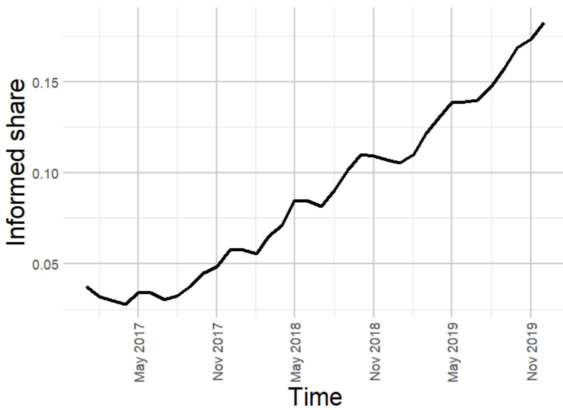

Figure 1: Informed consumer share over time

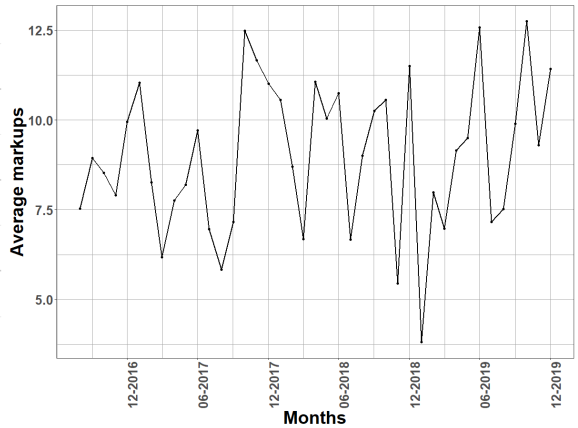

Figure 2: Average markups (in cents) over time

Figure 1 shows how informed consumer share in the metropolitan area changes over time. In the beginning of the policy a small portion of the population utilize the FuelCheck app and website and till the final months of 2017 informed consumer share remained below 10%. However, starting with 2018 informed share begun to rise more steadily and in December 2019 the share reached to 18%. Figure 2 illustrates average markups over time in the metropolitan area. Average markups are calculated as the difference between average gasoline prices and average terminal gate prices [6]. The striking feature is that there is no clear pattern of decline in markups over time. These descriptive findings suggest that transparency policy is not pushing markups down as expected. To examine how consumer side transparency affects markups in detail I develop a model, estimate parameter values and finally construct counterfactuals.

3 Market Definition

Total number of petrol station are too many to be in the choice set of any consumer. For that reason, I divide metropolitan area to largest statistical areas (SA4s) and assume that they are isolated markets. Under this assumption I have 15 isolated markets.

[6]It is important to note that since petrol stations are buying fuel in large amounts they are likely to pay less that reported terminal gate prices per litre. Therefore, average markups are likely to be larger than the calculated markups and these markup lines are mostly parallel to each other over time

4 Economic Model

4.1 Indirect Utility and Demand Model

Each commuter has an origin (residence location) and a destination (work location). I assume that commuters choose between driving to work and other means of transportation which is interpreted as outside option. Decision to drive also involves choosing a petrol station to visit. If n is the number of stations in the market, for informed consumer there are n+1 options including outside option.

Formally, indirect utility of consumer i's from buying fuel from store j is given by

$$\mathrm{U}_{i,j}(p_j, X_j, \xi_j|\theta) = \begin{cases} X_j\beta + \xi_j + \alpha p_j + \lambda_0 D((o_i, d_i), s_j) + \lambda_1 dist_i + \epsilon_{i,j} & \text{if } j \neq 0 \\ \epsilon_{i,0} & \text{otherwise} \end{cases}$$

where X_j is vector of observed characteristics of station j[7], p_j is the posted prices, $D((o_i, d_i), s_j)$[8] is the minutes deviation from daily commuting pattern of consumer i to visit station j, $dist_i$ is distance to work and ξ_j is the unobserved station attributes that affects all consumers in the same way. θ denotes the all of the parameters and θ_2 will be used to represent nonlinear parameters. I assume that $\epsilon_{i,j}$ is an i.i.d random utility shock with type-1 extreme value distribution.

I extend the model developed by Berry, Levinsohn and Pakes (BLP) (1995) to allow two types of commuters. BLP assumes that all consumers are perfectly informed about prices. The modified model allows consumers with perfect information and consumers with limited information. The percentage of informed consumers is determined in two steps. First, using subsample of users whose search location recorded for each month I find what percentage of users are within the metropolitan area. Then I multiply this share with total number of users and finally using within

[7]Observed station characteristics consist of number of service islands, number of pumps, presence of car wash and repair shop.

[8]Let o_i be location of residence for consumer i, d_i be work place of the consumer and finally let s_j be the location of station j. Define $r(x, y)$ as optimal driving pattern between location x and location y in minutes. Finally $D((o_i, d_i), s_j)$ is defined as follows:

$$D((o_i, d_i), s_j) = r(o_i, s_j) + r(s_j, d_i) - r(o_i, d_i) \tag{1}$$

metropolitan area distribution of subsample of users with location information I calculate number of informed consumers in each SA4. Finally, by dividing calculated informed share for each SA4 to its adult population I find share of informed consumers in each SA4. I assume that uninformed commuters are aware of a subset of stations in the market and the subset consists of stations that are in the 1st quartile in term of required deviation for consumer i. Meanwhile, informed consumers are aware of all petrol stations in their market.

Given the assumption on the distribution of $\epsilon_{i,j}$, commuters' conditional choice probability for buying fuel at station j has multinomial logit form with heterogeneous coefficients

$$\mathrm{P}_j((o_i,d_i)|\delta,p) = \begin{cases} P_j((o_i,d_i)|\delta,p) = \frac{exp(\delta_j+\mu_{i,j})}{1+\sum_{k\in J} exp(\delta_k+\mu_{i,k})} & \text{informed} \\ \\ P_j((o_i,d_i)|\delta,p) = \frac{exp(\delta_j+\mu_{i,j})}{1+\sum_{k\in J_i^{Q1}} exp(\delta_k+\mu_{i,k})} & \text{uninformed} \end{cases}$$

where J and J_i^{Q1} indicates all stations and stations within 1st decile of deviation needed for the commuter i, $\delta_j = \alpha p_j + X_j\beta + \xi_j$ and $\mu_{i,j} = \lambda_0 D((o_i,d_i),s_j) + \lambda_1 dist_i$. δ_j contains observed and unobserved characteristics that influence utility homogeneously for all consumers. $\mu_{i,j}$ represents heterogeneous tastes of consumers which is essential to capture substitution patterns.

From now on conditional choice probabilities of uninformed consumers and informed consumers will be denoted with P_j^{UN} and P_j^I respectively. Assuming there is no pair of stations that are equally desired, market share of each station is defined as follows:

$$S_j(p) = \sum_o \sum_d \int \big(P_j^I(y_i,(o_i,d_i)|\delta,p)u_i + P_j^{UN}(y_i,(o_i,d_i)|\delta,p)(1-u_i)\big)dF(y|o,d)T_{o,d} \qquad (2)$$

where u_i is consumer specific probability of being informed and $T_{o,d}$ denotes the distribution of drivers across origin/destination.

4.2 Consumers' Decision to be Informed

In each period, consumers consider whether to use transparency tools(app/website). This decision is modeled in a way that resembles sequential search models. At the beginning of each period consumers are only aware of petrol stations within 1st decile of deviation needed. If consumers decide be informed they observe all the petrol stations in the market. In other words, at the beginning of each period, consumers commit from which set of products to search on. Using transparency tools has a cost and only consumers whose expected benefit exceed the cost gets informed. Expected benefit is defined as value of information (VOI). Assuming that consumers are aware of the distribution of prices, VOI is difference between expected consumer surplus from becoming informed (E_i^{in}) and staying uninformed.(E_i^{un}).

$$
\begin{aligned}
E_i^{in} &= \frac{1}{-(\alpha)} ln\Bigg(\int_P \left(\sum_{l \in J \setminus J_i^{D1}} (exp(U_{il}))\right) dF(P) + \left(\sum_{k \in J_i^{D1}} (exp(U_{ik}))\right)\Bigg) \\
E_i^{un} &= \frac{1}{-(\alpha)} \left(ln \sum_{k \in J_i^{D1}} (exp(U_{ik}))\right) \\
VOI_i &= E_i^{in} - E_i^{un}
\end{aligned}
$$

The right side in inequality (3) is the cost of being informed. Cost of being informed is consumer specific and x_{imt} is a vector of cost variables including log income, log time, log market size and a constant. I assume that v_{imt} is distributed i.i.d type 1 extreme value and probability of being informed for consumer i in market m at time t can be represented as in equation (4).

$$VOI_i > \frac{\Gamma x_{imt} + v_{imt}}{-\rho} \tag{3}$$

$$u_{imt} = \frac{\rho VOI_i + \Gamma x_{imt}}{1 + \rho VOI_i + \Gamma x_{imt}} \tag{4}$$

4.3 Firm Behaviour

The supply side is included for several reasons. First, to examine the impact of consumer side transparency, I need to calculate margins. First order conditions allows me to calculate margins. Once I construct counterfactual cases with varying informed commuter shares, I can compare calculated margins to understand effectiveness of the policy. Secondly, I intend to allow heterogeneous coordination among market players by introducing multi-market contact. Finally, joint estimation of demand and supply helps obtaining more precise estimates.

4.3.1 Price Competition with Coordination

I denote set of players as F and I assume that they are non-cooperative Bertrand-Nash competitors. I assume that implies branded stations are making the pricing decision on brand level not in station level. I denote set of petrol stations at market m by J^m and subset of petrol stations owned by brand f is denoted by J_f^m. Similar to demand side, cost function has an observed and an unobserved component. For each station j, I denote observed characteristics of cost by w_j and unobserved characteristics of it by ω_j. Observed cost characteristics consist of number of pumps, number of fuel islands, terminal gate prices and weekly wage.

Marginal cost of station j is expressed as

$$mc_j = w_j \gamma + \omega_j \tag{5}$$

where w_j denotes observed cost variables, ω denotes unobserved cost term and the vector of parameters are represented by γ. Following Ciliberto & Williams, profit function of brand $f \in F$ at market m is expressed as

$$\begin{aligned} \Pi_{j \in J_f^m} = & \sum_{j \in J_f^m} (p_j^m - mc_j^m) S_j^m(p, X, \xi | \theta) M_m - C_j^m \\ & + \sum_{f' \in (F \setminus f)} \sum_{k \in J_{f'}^m} f_{mmc}(MMC_{ff'}; \psi_{mmc})(p_k^m - mc_k^m) S_k^m M_m \end{aligned} \tag{6}$$

where

$$f_{mmc}(MMC_{ff'};\psi_{mmc}) = \frac{exp(\psi_{mmc} + Y2\psi_{mmc2} + Y3\psi_{mmc3})}{1 + exp(\psi_{mmc} + Y2\psi_{mmc2} + Y3\psi_{mmc3})} MMC_{ff'} \tag{7}$$

where M_m, C_j^m and $MMC_{ff'}$ indicate the size of the market m, fixed cost and the percentage of markets where brand f and f' contact respectively. Each brand is assumed to choose prices to maximize their profits given their stations' characteristics, prices of rivals and rivals' station characteristics. Second line in equation (6) is referred as multi-market consideration. Depending on in what portion of markets firms are facing each other $MMC_{ff'}$ takes a value between 0 and 1 and first component of equation 7 determines the magnitude of multi-market consideration between f-f' pair. If the term f_{mmc} is 0, brands facing each other in different market do not coordinate and if the term is 1 then brands in contact perfectly collude. I allow coordination among brands to change over years. As equation 9 shows there are 3 new parameters and $Y2$, $Y3$ are index functions that returns 1 if it is 2nd and 3rd year respectively. Although in reality firms are unlikely to take into account profits of each other firms, this modelling function serves as a proxy to capture the possibility of coordination among pairs of firms facing each other in multiple-markets.

Given the profit function, for any station $j \in J_f^m$ first order condition is

$$\begin{aligned} & S_j(p, X, \xi|\theta) + \sum_{k \in J_f^m} (p_k - mc_k) \frac{\partial S_k(p, X, \xi|\theta)}{\partial p_j} \\ & + \sum_{f' \in (F \setminus f)} \sum_{l \in J_{f'}^m} f_{mmc}(MMC_{ff'};\psi_{mmc})(p_l^m - mc_l^m) \frac{\partial S_l(p, X, \xi|\theta)}{\partial p_j} = 0 \end{aligned} \tag{8}$$

As an example, below I illustrate first condition order condition in vector form for a market with three different branded petrol stations. Let Λ_O be ownership matrix and Λ_M be multi-market consideration matrix:

$$\Lambda_O = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Lambda_M = \begin{bmatrix} 0 & f_{mmc}(MMC_{12}) & f_{mmc}(MMC_{13}) \\ f_{mmc}(MMC_{21}) & 0 & f_{mmc}(MMC_{23}) \\ f_{mmc}(MMC_{31}) & f_{mmc}(MMC_{32}) & 0 \end{bmatrix} \tag{9}$$

Next redefine Δ as:

$$\Delta = -(\Lambda_O + \Lambda_M) \odot \begin{bmatrix} \partial S_1/\partial p_1 & \partial S_1/\partial p_2 & \partial S_1/\partial p_3 \\ \partial S_2/\partial p_1 & \partial S_2/\partial p_2 & \partial S_2/\partial p_3 \\ \partial S_3/\partial p_1 & \partial S_3/\partial p_2 & \partial S_3/\partial p_1 \end{bmatrix} \tag{10}$$

where $\odot$ is Hadamard product. Now first order conditions can be represented as

$$p = \Delta(p, X, \xi|\theta)^{-1} S(p, X, \xi|\theta) + mc \tag{11}$$

5 Identification and Estimation Methodology

5.1 Identification

In this section I discuss the main assumptions to identify the key parameters. I focus the identification of price coefficient α and multi-market contact coefficients ψ_{mmc}, ψ_{mmc2} and ψ_{mmc3}.

In order to resolve the endogeneity between prices and unobserved demand and cost shocks, I employ instrumental variables. My main instruments are cost-shifters (brand specific terminal gate prices and weekly wages). Following BLP (1995) and Houde (2012) I also include instruments related to distance and station characteristics, namely inverse distance weighted rival station characteristics and number of rival stations within 5min distance. These instruments are correlated with prices and they are assumed to be orthogonal to unobserved demand and cost characteristics.

Distinguishing whether the source of high prices lies in unobserved cost shocks or some degree of coordination among firms poses an empirical challenge. To separate markups from unobserved cost shocks, I use additional instruments. Instrumental variables should exhibit correlations with the contact variables while remaining orthogonal to unobserved cost shocks. Following Ciliberto and Williams (2014) and Koh (2021), I introduce three add instrumental variables:presence ratio of the firm in other markets, the fraction of rival supermarket major brands[9] in the same market, and the fraction of low-cost brands[10] in the same market.

[9] Coles Express and 7-Eleven
[10] Metro Fuel

I argue that these variables are uncorrelated with unobserved cost shocks, thus establishing their validity as instruments. There are two primary reasons supporting this assertion. Firstly, station entry and exit decisions are typically long-term in nature, primarily related to fixed costs rather than marginal cost shocks. Secondly, in my analysis, I incorporate fixed effects for brand, market, and season. This inclusion effectively removes a significant portion of the persistence from unobservable demand and cost shocks. Any residual demand and cost shocks that remain are unlikely to be driver of these decisions.

5.2 The GMM Estimator

The set of parameters are given by $\theta = \{\alpha, \beta, \lambda_0, \lambda_1, \psi_{mmc}, \psi_{mmc1}, \psi_{mmc2}\}$. In estimation of parameters of the model I employ the nonlinear GMM estimator developed by Berry (1994) and BLP (1995). Assuming that observed data are the equilibrium outcomes, I estimate demand and supply parameters jointly. Following Nevo (2001), I incorporate fixed-effects for demand and supply moments. For supply side I include brand and market fixed effects and for demand side I include season, brand and market fixed effects.

$$g = \begin{bmatrix} \frac{1}{N}\sum_{m\in M}\sum_{j=1}^{J^m} Z_{d,j}\xi_j(\delta,\theta_1) \\ \frac{1}{N}\sum_{m\in M}\sum_{j=1}^{J^m} Z_{s,j}\omega_j(\delta,\gamma) \\ \frac{1}{MT}\sum_{mt=1}^{MT} Z_{mt}(U^o_{mt} - U^p_{mt}) \end{bmatrix} \tag{12}$$

where $Z_{\xi,jm}$,$Z_{\omega,jm}$ and Z_{mt} are instruments for demand and supply and decision to be informed respectively. The decision to informed moment is similar to micromoments of Petrin (2002)[11] Given sample moments, I use GMM to find the parameter values and the objective function is $g'A^{-1}g$, where A is a weighting matrix.

5.3 Simulation

The distribution of consumer demographics is empirical. Therefore, there is no closed-form solution for predicted market shares. To find an approximate solution, I simulate a sample of

[11] Instruments for the decision to be informed are log time, log income, second, third quantile of rents in the market and employment ratio in the market.

working individuals. The censored data from the 2016 Census provides me with the distribution of commuters and income from SA1 to SA2. Based on the distribution of commuters, I sample 1000 individuals from each market. Given a sample of consumers with origin and destination, for each one, I sample an income value from the income distribution of the corresponding SA1-SA2 pair. In order to define smaller and more realistic destinations for sampled commuters, I make use of employment data at the travel zone (TZ) level. In general, TZs have a size similar to that of SA1s. For each TZ within the same SA2, I assign a probability of being the destination for commuters based on observed employment levels. More specifically, I first obtain total employment in each TZ from ABS and then find the total employment for each SA2. Then, I assign a probability to each TZ within a certain SA2 based on its employment share in that SA2. Using these probabilities, I assign a TZ to each commuter, and their destination is defined as the center of the corresponding TZ. Finally, I use the Open Source Routing Machine (OSRM), which provides the optimal driving path given the starting point, visit point, and destination point. OSRM allows me to define the minimum deviation distance and time from the optimal origin-destination path for each station.

Estimation Algorithm

Below I explain the estimation procedure in steps:

- Draw a sample as explained above based on distribution of individuals' commuting patterns and income
- Make an initial guess for δ and nonlinear parameters (θ_2).
- Given values for δ and θ_2, approximate integral for market share of stations (equation 2) for each market/time period using simulated draws.
- Use a contraction mapping below (for each time period/market separately) until convergence is achieved. To accelerate the process, following Nash et al. (2012), employ the squared polynomial extrapolation method (SQUAREM).

$$\delta_m^{h+1} = \delta_m^h + (ln(S_m) - ln(s_m)) \tag{13}$$

where S_t is the observed market share and s_t is the estimated market share. This step implies that previous step where market shares are estimated will be iteratively repeated for new δ values until convergence is achieved.

- After convergence, use new $\hat{\delta}$ to recover markups $\hat{\eta}_m = \Delta_m(p_m, X_m, \xi|\theta)^{-1} s_m(p_m, X_m, \xi|\theta)$.
- Define unobserved residual demand and supply shocks:

$$\hat{\xi}_{jm}(\theta_2) = \hat{\delta}_{jm}(\theta_2) - X_{jm}\beta \tag{14}$$

$$\hat{\omega}_{jmt}(\theta_2) = (p_{j,m} - \hat{\eta}_{j,m}) - w_{j,m}\gamma \tag{15}$$

- Construct moments (equation 12), define the GMM objective function $g(\theta_2)'Ag(\theta_2)$ and find the linear parameter values θ_1 that minimizes the objective function given θ_2. Finally, find value of objective function
- GMM objective value is a function of nonlinear parameters. Using a nonlinear optimization algorithm search over θ_2 values to minimizes the GMM objective function. The optimization algorithm repeats all steps except first one until the minimizing set of parameters are found.

6 Results

6.1 Structural Estimation Results

Table 2 shows parameter estimates for the model with multi-market contact and standard errors are provided in parentheses. First I discuss demand side coefficients and then multi-market coefficients. As expected price coefficient is negative and significant. Distance to work has a positive coefficient and this suggests that driving is more comfortable mode of travel for longer distances. Final taste parameter is deviation time required to visit a petrol station. Although the sign of the coefficient is negative, it is insignificant. Deterministic choice set restriction could be the reason that this parameter is not well identified. Interestingly, coefficient for service islands is negative. This could be explained by the fact that newer petrol stations and recently renovated petrol stations

Table 2: Parameter Estimates

Demand Variables		Supply Variables	
Price	-1.96*** (0.15)	Base conduct	-1.18*** (0.17)
Distance to work	0.095*** (0.014)	2nd year	-0.15 (0.09)
Deviation	-0.03* (0.016)	3rd year	-0.27** (0.12)
Pumps	0.015 (0.04)	Pumps	0.10* (0.06)
Islands	-0.037*** (0.01)	Islands	0.08*** (0.01)
Repairshop	-0.15*** (0.027)	Average TGP	0.89*** (0.07)
Carwash	0.02 (0.03)	Wage	0.0012** (0.005)
Constant	-0.96** (0.39)	Constant	-0.21*** (0.02)
User Variables		Mean own price elasticity	-1.73
Constant	-2.98*** (0.84)		
VOI	7.1*** (1.31)		
Log time	-0.05** (0.02)		
Log income	-0.42*** (0.1)		
Log market size	0.1** (0.04)		

Brand and region fixed effects included. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$

have few service islands with more pumps. On the other hand, older petrol stations, in general, have more service islands with fewer pumps. Therefore this variable serves as a proxy for novelty of stations. Coefficient for repair shop is also negative and significant. Most of the repair shops are present in independent petrol stations and small branded chains. Similar to service island variable, this variable could be capturing lower service quality in these petrol stations.

When we look estimated coefficients for multi-market, we see that the base variable is significant and negative. This suggest that there in each observation year there is some level of coordination among brands that face each other in different markets. This finding is in the line with findings of Bryne & de Roos (2019). Coefficients for second and third year is negative but only the coefficient for the 3rd year is significant. This suggest that level of coordination declines in the 3rd year.

6.2 Counterfactual Simulation Results

At the beginning of 2017, on average, informed consumer share was around 4% and by December 2019 informed share reached to approximately 18%. In order to examine effect of consumer-side transparency, using January 2017 data I construct a counterfactual scenario where informed share was 18% and compare profit margins with the actual profit margins. In January 2017, average margins were around 68.2 cents. In the counterfactual scenario, margins are approximately 63.4 cents. This represents a decline in margins of about 7.02%. Following Ryan (2012), to construct confidence intervals I used bootstrapping by sampling random subsamples without replacement. The calculated 95% confidence interval is [-8.51, -2.1], suggesting that the decline in margins is statistically significant.

7 Discussion

In this section, I discuss next steps for this study and limitations of the current model. I start by discussing the tests for goodness of fit and other additional analysis to measure the sensitivity of the results.

Goodness-of-Fit Tests

In order to examine robustness of the model I will perform several goodness to fit tests. First, I will employ the chi-square test that is proposed by Andrews (1988). I will partition the region in which response variables lie into disjoint cells and then I will calculate a quadratic form based on the difference between the observed number of outcomes in each cell and the conditionally expected number in each cell given the observed covariates. If the parametric model is correct, then the quadratic form converges in distribution to a chi-square random variable as the sample size increases.

Second, I will compare my model with the BLP model where all consumers are informed. More specifically, I will estimate parameters for both models and compare difference between estimated markets shares and observed market shares. This simple test will provide information on relative fit of the models.

Finally, Andrews, Gentzkow, and Shapiro (2017) proposes of the relationship between parameter estimates and the moments of the data they depend on. Compared to linear regression analysis structural empirical methods are much less transparent and it may be difficult to understand which features of the data drive results. This measure increases the transparency of structural methods by allowing readers to predict how the reported results would change when the identification assumptions are slightly perturbed. This measure will be reported to alleviate concerns regarding transparency of the methodology.

Appendix

Maps

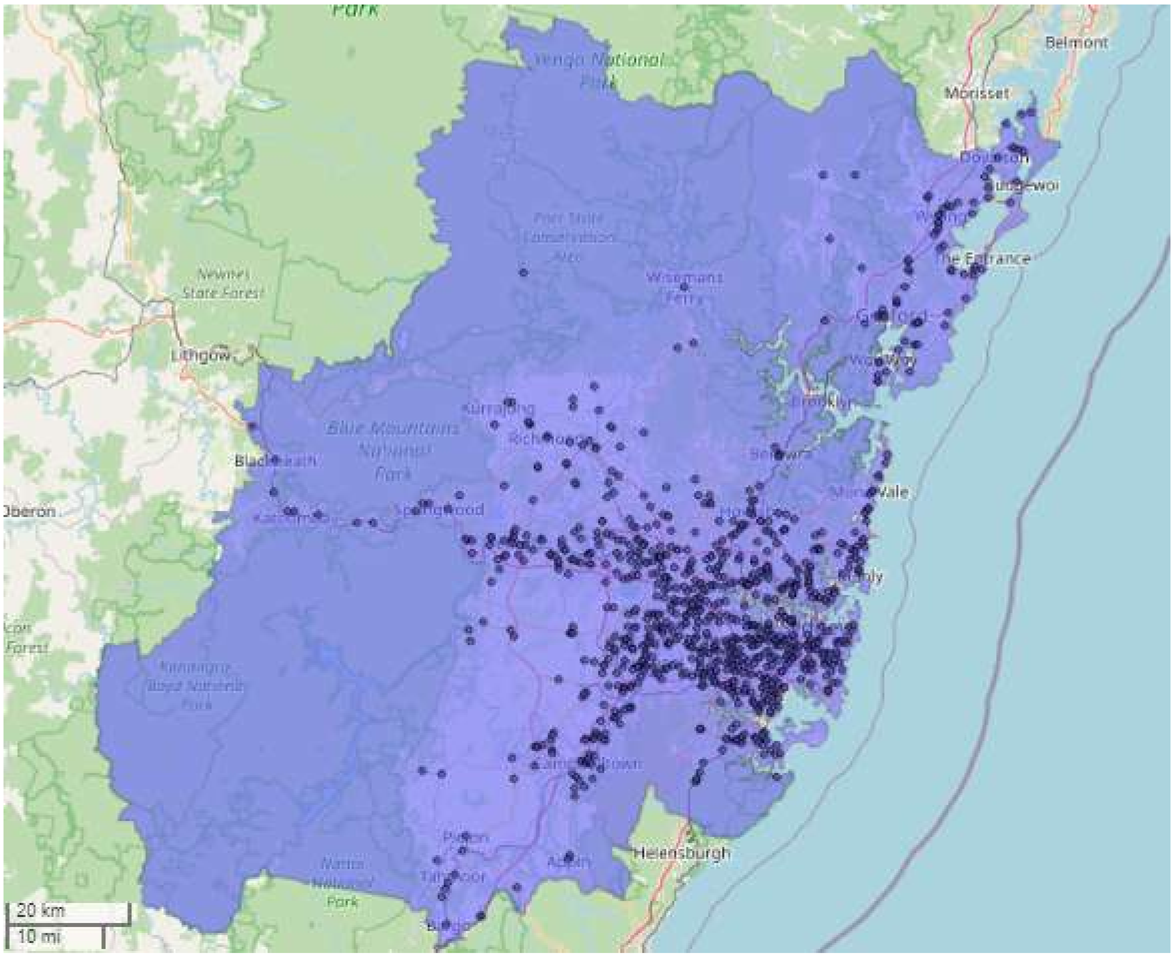

Figure 3: Distribution of petrol station in Greater Sydney

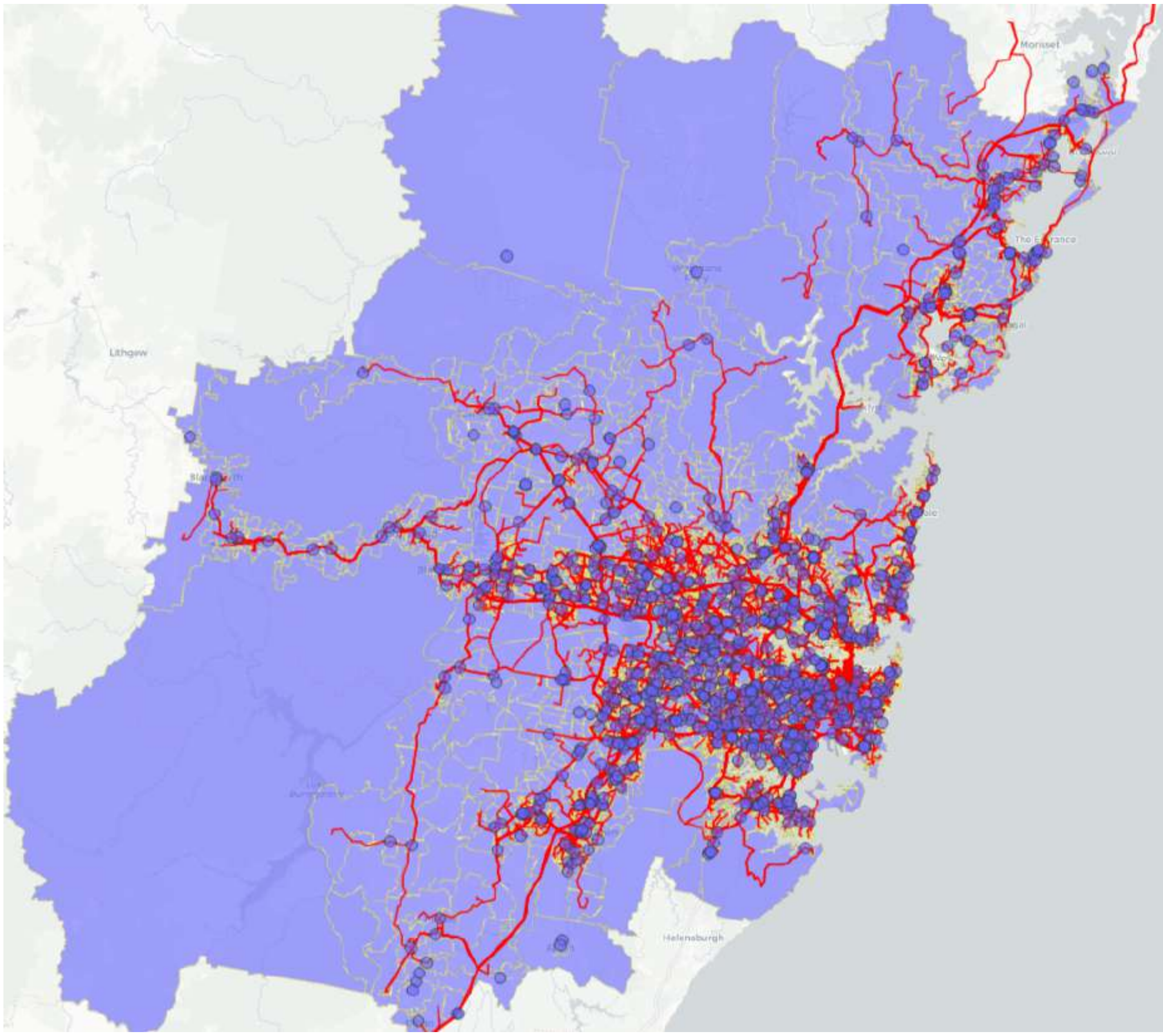

Figure 4: Simulated commuting patterns

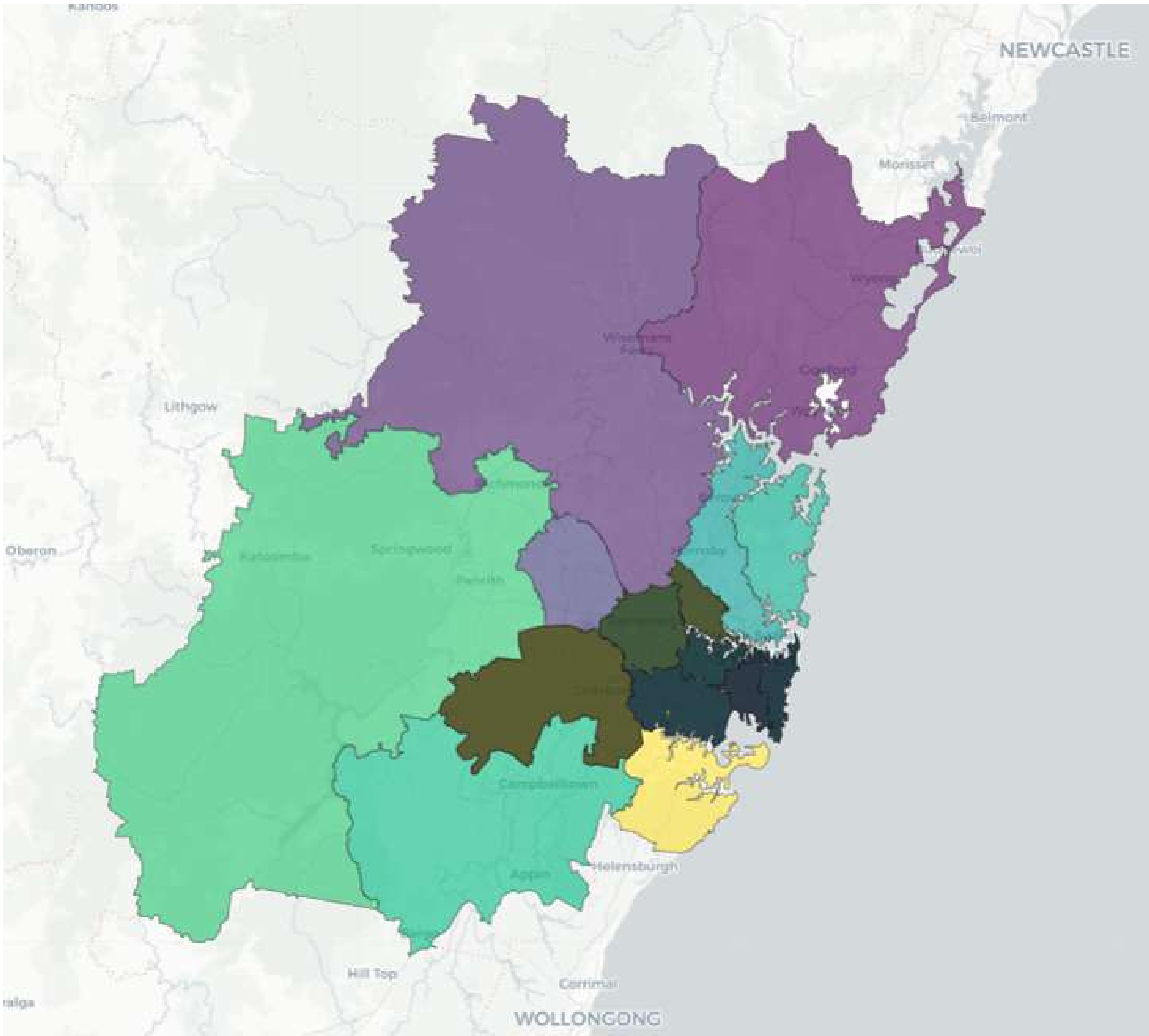

Figure 5: SA4s as isolated markets

Google Maps station views: an example

Figure 6: Example station before upgrade

Figure 7: The station after upgrade

References

Albek et al. (1997). "Government-assisted oligopoly coordination? A concrete case" The Journal of Industrial Economics.

Andrews I, Gentzkow M, Shapiro JM. (2017) "Measuring the Sensitivity of Parameter Estimates to Estimation Moments" Quarterly Journal of Economics.132 (4) :1553-1592.

Ater, I. and Rigbi, O. (2017). "The effects of mandatory disclosure of supermarket prices"

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. The RAND Journal of Economics, 25(2), 242–262.

Berry, Steven, et al (1995). "Automobile Prices in Market Equilibrium." Econometrica, vol.63, (4).

Byrne, D. P. and de Roos, N. (2019). "Learning to Coordinate: A Study in Retail Gasoline". American Economic Review, 109 (2): pp. 591–619.

Brown, Zach Y. (2019): "An Empirical Model of Price Transparency and Markups in Health Care", Revise resubmit, Journal of Political Economy in progress

Brantly Callaway, Pedro H.C. Sant'Anna. (2021) "Difference-in-Differences with multiple time periods". Journal of Econometrics

Donald W. K. Andrews. (1988). "Chi-Square Diagnostic Tests for Econometric Models: Theory. Econometrica, 56(6), 1419–1453.

Donald W.K. Andrews. (1988). "Chi-square diagnostic tests for econometric models: Introduction and applications",Journal of Econometrics

Goeree, M.S. (2008), Limited Information and Advertising in the U.S. Personal Computer Industry. Econometrica, 76: 1017-1074.

Houde, Jean-Franfcois. (2012). "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline" American Economic Review, 102(5)

Luco, F. (2019). "Who Benefits from Information Disclosure? The Case of Retail Gasoline." American Economic Journal: Microeconomics, 11, (2).

Martin, S. (2023). "Market transparency and consumer search - Evidence from the German retail gasoline market". RAND Journal of Economics

Mollgaard, H. Peter and Overgaard, Per Baltzer (2001) : "Market transparency and competition policy", Rivista di Politica Economica 91: 11-58.

Montag, F and Winter, C. (2020) "Price Transparency Against Market Power", Available at SSRN: https://ssrn.com/abstract=3256476

Nash et al. (2012)"Enhancing the Convergence Properties of the BLP (1995) Contraction Mapping". Vives Discussion Paper

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. Journal of political Economy, 110(4), 705-729.

Schultz, C. (2005), "Transparency on the consumer side and tacit collusion", European Economic Review, 49, (2).