EIEF WORKING PAPER SERIES

IEF

Einaudi Institute for Economics and Finance

EIEF Working Paper 17/09
July 2017

Strategic Entry and Potential Competition: Evidence from Compressed Gas Fuel Retail

by

Giulia Pavan

(Toulouse School of Economics)

Andrea Pozzi

(EIEF and CEPR)

Gabriele Rovigatti

(University of Chicago Booth School of Business)

Strategic Entry and Potential Competition: Evidence from Compressed Gas Fuel Retail*

Giulia Pavan
Toulouse School of Economics

Andrea Pozzi
EIEF and CEPR

Gabriele Rovigatti
University of Chicago - Booth School of Business

July 4, 2017

Abstract

We study the effect of competition on preemption incentives. An unexpected change in regulation in the Italian retail market for compressed natural gas fuel allows us to identify the potential entrants to the market and creates exogenous variation in their number. We document that areas with a larger pool of potential competitors experience faster entry. We provide evidence suggesting that this occurs because facing a higher number of potential entrants raises firms' incentives to preempt.

JEL classification: L12, L22, L81

Keywords: preemption, potential entrants, retail fuel market

*We thank for useful comments Jaap Abbring, Victor Aguirregabiria, Bettina Drepper, Pierre Dubois, Glenn Ellison, Philipp Schmidt-Dengler, Alan Sorensen, Frank Verboven, Biliana Yontcheva and especially our discussants Marit Hinnosaar, Jean-François Houde and Hidenori Takahashi as well as participants in the seminars at the Università di Roma "Tor Vergata" and CREST and in the following conferences: MaCCI Annual Conference 2016, IIOC 2016, EARIE 2016, Jornadas de Economia Industrial 2016, Vilfredo Pareto Workshop 2016, Toulouse Winter IO seminar 2017. All errors are our own.

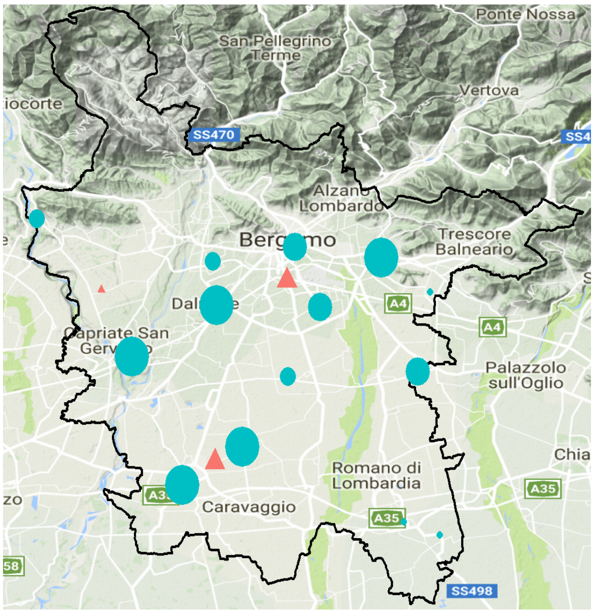
1 Introduction

In this paper we study how competition affects the incentive to preempt entry in a new and growing industry. In the presence of preemption motives, firms in a duopoly act earlier than a monopolist would for fear of being displaced, leading to inefficiency (Fudenberg and Tirole (1985)). It seems natural to conjecture that increasing the number of competitors beyond two leads to even more hurried entry and further efficiency losses. However, the theoretical literature does not deliver an unambiguous predictions on the timing of entry in oligopoly games as a function of the intensity of competition (Shen and Villas-Boas (2010); Argenziano and Schmidt-Dengler (2014)): under certain conditions, the presence of an additional competitor may even delay entry (Argenziano and Schmidt-Dengler (2013)). Shedding light on the issue has proven equally hard on the empirical front. Dynamic entry games are computationally intensive and this has forced much of the extant literature (Schmidt-Dengler (2006); Igami and Yang (2016); Zheng (2016); Igami (2017)) to analyze only duopolies or oligopolies with a limited number of players. A further challenge to empirically documenting the relationship between competition and preemption is that it is difficult to capture the relevant dimension of competition. In fact, preemption incentives should respond to the number of potential entrants which, unlike actual entrants, are not typically observed.

We examine the effect of competition on preemption exploiting a novel dataset we assembled documenting the early years of the compressed natural gas (henceforth, CNG) retail fuel industry in Italy. CNG is a car fuel alternative to gasoline and diesel that can power cars designed or retrofitted to run on it. Italian legislation originally prevented filling stations selling gasoline and diesel from offering CNG, due to safety concerns. It also forbade selling CNG in establishments located near populated area or major roads. This confined the market for CNG to a small niche served by monofuel stations placed in hard to reach locations. The lifting of these constraints in the late 1990s\early 2000s brightened the prospects for the retail sale of natural gas attracting new entry into the market. In this scenario, existing filling stations already selling traditional fuels became the obvious candidates to enter the market due to the cost of adding a CNG pump being much lower than that of a greenfield entry with a new filling station and to regulation still controlling the number of new establishments that could be opened.

Figure 1 documents the evolution of CNG retail supply in the market of Bergamo, a large province in Northern Italy, and provides an example of the pattern we described. In 2005, shortly after the legislation change, there were very few station offering CNG, marked by red triangles. These establishments were all monofuel and their distance from main roads (inversely proportional to the size of the markers on the map) was significant. By 2015, the situation has dramatically changed. First, a number of new stations offering CNG (marked by blue circles) has entered, raising the total number of establishments

Figure 1: Market evolution: An example



Notes: The figure displays the location of filling stations offering CNG in the Local Market Area of Bergamo at the end of 2015. The red triangles denote stations that were already offering CNG by the end of 2005; the blue circles mark the locations of stations that started offering CNG between 2006 and 2015. The size of the station markers is inversely proportional to the distance of the station from a main road.

distributing natural gas by a factor of four. This confirms that market expansion has been large and quick after the change in legislation. All the new entrants are multifuel stations, meaning that they sell natural gas along with gasoline and diesel. In all cases in which we have information on entry mode,[1] the new entrants in the CNG segment were pre-existing fuel stations which added a CNG pump. Finally, the location of the new entrants is different from that of the old monofuel supplier: they operate in areas closer to major roads.

Although Bergamo is one of the largest markets in our sample, the pattern of entry observed there is representative of what happens to the market for retail CNG after the legislation regulating its distribution changes. Hence, we submit that using the number of existing filling stations in a market is a natural way to measure the number of potential entrants. Moreover, we argue that this measure of competition can be plausibly considered exogenous to demand for CNG. In fact, gasoline and diesel filling stations had opened well before the change in the legislation on CNG distribution occurred and without anticipating the shift in regulation. Therefore, they would not have factored expectations on CNG profitability in their location choice.

We identify the effect of competition on preemption by estimating a market-level Cox model for the entry process of filling station in the CNG market where the hazard rate

[1]For six out of the thirteen new entrants in the market, the data do not allow us to say definitively whether the station was operating before starting to sell CNG or whether it was a greenfield multifuel entrants.

depends on the number of potential entrants in the market. Our main findings is that, controlling for a number of market characteristics, entry occurs significantly faster in areas with a higher number of potential entrants. Moving a market from the bottom to the top tercile of the distribution of the number of potential entrants raises by ten times the hazard that a filling station chooses to enter the CNG market in that area.

Since our approach does not rely on a model and preemption incentives are not observables, we cannot definitively tie the correlation between potential competition and rate of entry to preemptive motives. However, we offer several pieces of evidence pointing to the fact that our main result does discend from a positive relationship between potential competition and incentives to preempt entry. First, a calibration of stations' profits performed using data from industry sources suggests that early entry could not yield positive static profits in most of the markets in our sample. This implies that allowing for a dynamic element is important to rationalize the entry pattern we observe in the data. Next, we perform a test à la Ellison and Ellison (2011) verifying that the effect of potential competition on the speed of entry fades away in the markets where we expect the incentive to preempt to be lower. Finally, we document that unbranded filling station have exogenously higher incentive than branded pumps to introduce CNG and, therefore, the presence of unbranded pumps among the competitors of a station represents a shift in the incentive to preempt. We show that establishments facing more unbranded stations among their competitors have a higher chance of early entry in the CNG market.

This paper contributes to the literature on preemption by providing empirical evidence of its relationship with the intensity of competition. Our unique setting allows us to produce direct evidence of preemption-related phenomena with a reduced form approach similar to that employed in a recent literature focussing on preemption and entry deterrence (Ellison and Ellison (2011); Yang (2015)). We complement structural studies analyzing preemption games (Schmidt-Dengler (2006); Gil et al. (2015); Igami and Yang (2016)) that, due to their computational burden, cannot handle competition between a large number of players. Since exploiting the regulation change we can identify the potential entrants in the CNG market, our analysis also relates to the scant literature providing evidence on the effect of the threat of entry on firms' strategies (Goolsbee and Syverson (2008); Illanes and Moshary (2015)).

The rest of the paper is structured as follows. In section 2 we describe some institutional details of the retail CNG fuel market and present the dataset we constructed to study it. In section 3, we estimate the effect of competition on the speed of entry in the CNG market and in section 4 we document that our main result derives from an increase in the intensity of preemption in markets with more potential entrants. Section 5 concludes.

2 Background and data

In this section we provide information on the market for retail CNG in Italy and highlight some of the institutional features which we exploit to identify the effect of potential competition on the incentive to preempt. We also introduce a novel dataset we compiled listing the universe of the filling stations active in Italy with detailed information on location, fuels offered and year of entry.[2]

2.1 The retail CNG market in Italy

Natural gas consists mainly of methane, a hydrocarbon originated from the decay of organic compounds in absence of oxygen. In its compressed form the gas can be used as fuel in the automotive industry. Although cars able to run on CNG are typically more expensive than gasoline\diesel powered ones, CNG is cheaper than both gasoline and diesel and it has lower impact in terms of greenhouse emissions and environmental footprint.

As of 2016, the International Association for Natural Gas Vehicles (NGV Global) reports that Italy is the top EU country and among the top 10 worldwide for both stock and flow of circulating CNG cars. They accounted for 5.3% of all new cars registrations in the country in 2014, with the bulk of these purchases being represented by vehicles intended for private or commercial use rather than public transportation vehicles. On the supply side we count, as of June 2015, over 1,000 filling stations offering natural gas. Most of the stations with CNG pumps are directly linked to the gas pipelines grid, which is owned and operated by a state-controlled regulated company (SNAM Rete Gas), and buy the fuel from a number of distributors (Estra, Edison and Engie Italia among the others). Unlike the gasoline and diesel ones, the retail CNG market does not present a high degree of vertical integration: Eni is the only company that both sells CNG wholesale and operates filling stations. It is also the only player with significant stakes in both the traditional fuels and the CNG markets.

Our identification strategy exploits the institutional features of the Italian market for retail fuels of the early 2000s, the key one being the evolution of the tight regulation of entry in the CNG market. In fact, such regulation shifts represented a major factor fueling the recent market expansion. Due to the risk of explosion linked to the distribution of natural gas, until the late 1990s Italian regulation did not allow to sell CNG in establishments close to main roads or densely populated areas. It also imposed costly technical requirements for stations offering CNG jointly with other fuels.[3] Therefore, the supply of CNG occurred exclusively in monofuel stations placed in hard to reach locations. This

[2] Details on the data collection and variable construction are discussed in Appendix A

[3] These restrictions are listed in several pieces of legislation, including the D. M. 8 Giugno 1993 "*Norme di sicurezza antincendi per gli impianti di distribuzione di gas naturale per autotrazione*".

narrowed the market: CNG customers had to value the convenience of the fuel enough to justify incurring substantial travel costs to refill.

In time, CNG pipelines, pumps and tanks experienced technological improvements that made them safer to operate and the regulations were progressively lifted starting from the early 2000s.[4] The implications of this regulatory shift were twofold. First, it increased the incentive of filling stations to offer natural gas, since it was now possible to sell it in more appealing locations. Second, it cut entry costs for existing stations, which were allowed to distribute CNG by adding a pump instead of building a brand new dedicated station. It also meant that pre-existing gasoline and diesel stations became the subjects more likely to enter the CNG market.[5]

Figure 2 describes the evolution of the CNG market in Italy. The left panel shows how the Italian CNG market started experiencing significant growth, both in the stock of circulating CNG cars and in the quantity of CNG consumed, in the aftermath of the regulatory changes. The expansion is particularly rapid in the years after 2007, in part thanks to the fact that after 2000 the largest car manufacturers started gradually to introduce models specifically designed to run on CNG.[6] The right panel of Figure 2 displays the evolution of the number of CNG filling stations in Italy between 2005 and 2015, distinguishing between monofuel and multifuel establishments, and attests to the impact of the regulatory change on the supply of CNG. First, the number of stations selling CNG more than doubles in the time span. Second, whereas in 2005 the stock of filling stations selling natural gas is almost equally split between stations that sell only CNG and multifuel stations, by 2015 stations selling CNG alongside gasoline and diesel represent 80% of the supply.

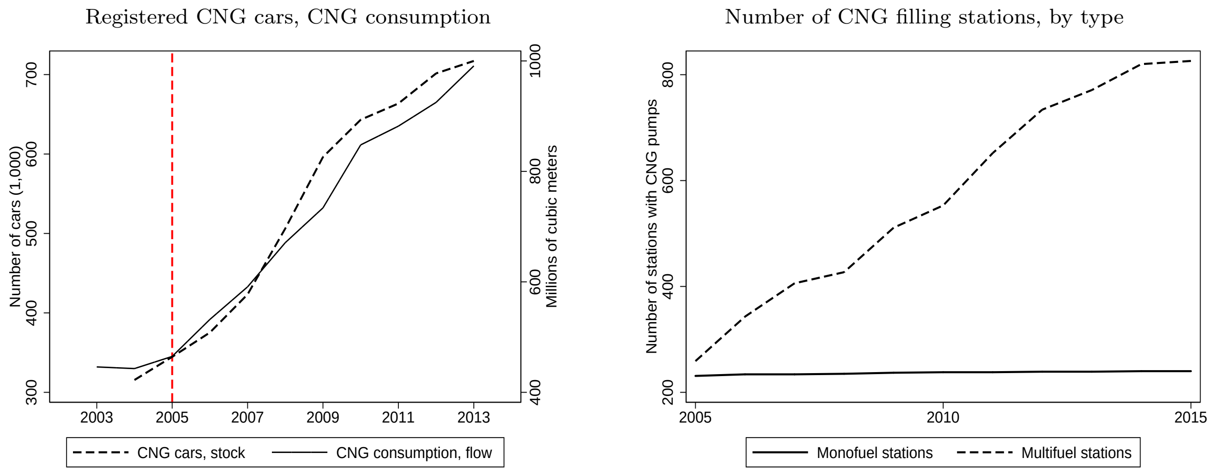
A second institutional detail we take advantage of concerns the different types of ownership\management contracts that can be observed in Italian filling stations. Some filling stations are owned and operated by the refiners themselves (company operated stations); these represented 8% of the filling stations in Italy in 2015. 74% of the stations are owned by a station manager who signs long term agreements for provision of fuel with a particular refiner (franchising stations). Company operated and franchising stations are commonly referred to as "branded" station because they sell fuels associated with a specific refiner's brand. A significant fraction of Italian filling stations (18%) are instead operated by independent owners who buy fuel from refiners without any long term contract and sell it as "unbranded" (independent or "white pumps"). This taxonomy is significant from

[4] The original piece of legislation is the D.M. 28 Giugno 2002 "*Norme di prevenzione incendi per la progettazione, costruzione ed esercizio degli impianti di distribuzione stradale di gas naturale per autotrazione*", modified in 2006, in 2008 and again in 2014.

[5] On top of being more costly, greenfield entry was also discouraged by the law regulating opening of new filling stations in Italy until 2008. More details on entry regulations are provided in Appendix A.

[6] Until the early 2000s, most CNG cars were regular vehicles retrofitted to run on natural gas. The retrofitting, however, leads to undesirable reductions in trunk or seat space due to the necessity of placing a gas tank on the car. Moreover, it can lead to void the warranty of the vehicle.

Figure 2: EVOLUTION OF THE CNG MARKET



Notes: The figure on the left shows the evolution of the demand for CNG in Italy. The solid line refers to the millions of standard cube meters of CNG sold in Italy in the year as reported by *Assogasmetano*, the association of Italian automotive CNG distributing firms. The dashed black line tracks the stock of CNG powered cars circulating in Italy computed from the car registry database maintained by ACI. The red vertical dashed line marks the beginning of the time span we consider in our analysis. The figure on the right tracks the number of filling stations that sell natural gas, distinguishing between stations that sell exclusively compressed gas (monofuel) and stations that sell both compressed gas and gasoline\diesel (multifuel). These figures are computed based on the database of filling stations maintained by *Prezzibenzina.it*.

the point of view of the incentives to introduce a CNG pump at the station. Company operated and franchised stations have strong ties to refining companies who have interest in hampering the growth of the natural gas market to protect gasoline and diesel from its competition. This represents a constraint in their decision to offer CNG which is instead not faced by independent stations.

2.2 Data

We combine data from multiple sources to build a novel panel dataset containing information on the universe of the filling stations operating in Italy between 2005 and 2015. The bulk of the information comes from data provided by the website *Prezzibenzina.it*, a search engine reporting fuels prices at each Italian filling station using information posted and updated by either customers or filling station managers and then verified by the staff. We observe the location and characteristics (type of fuels sold, brand, whether it is a franchising, etc.) of each station. We also know the year in which the station entered the *Prezzibenzina.it* database and the year in which it was first reported to be selling natural gas, if ever. Since the data are user-reported, coverage was initially limited when the website went online in 2004 and it became progressively more complete as the it grew popular. By the end of 2009 *Prezzibenzina.it* contained verified information on nearly every Italian filling station. This implies that there is measurement error in data concerning whether a station was active in the years 2005-2008: a station appearing in the database in 2007 may be a new entrant or a pre-existing station on which nobody had submitted information before.[7] In order to limit the impact of the incomplete coverage in the 2005-2008 years, we have used paperback guides as well as databases compiled by other websites (*Metanoauto.com* and *Ecomotori.net*) to check and integrate the information from *Prezzibenzina.it*. Furthermore, since entry in the fuel retail market was strictly regulated in Italy until recently, the complete snapshot of filling stations in 2009 we obtain from the *Prezzibenzina.it* database should resemble closely the situation at the beginning of our sample (2005).

Table 1: Market characteristics

	Mean	**St. Dev.**	**5th pctile**	**25th pctile**	**50th pctile**	**75th pctile**	**95th pctile**
N. of filling stations	29.5	68.58	2	6	14	32	85
N. of stations with CNG (2005)	0.75	1.165	0	0	0	1	3
N. of stations with CNG (2009)	0.89	1.918	0	0	0	1	4
N. of stations with CNG (2015)	4.58	5.596	0	1	2	5	17
Share monofuel stations	0.014	0.106	0	0	0	0	0
Share white pumps	0.15	0.135	0	0	0.13	0.21	0.39
Circulating cars	62,525	169,175	4,528	12,963	27,876	55,609	179,633
Circulating CNG cars	1,061	2,627	4	24	145	895	5,357
Population (2011)	95,287	19,8077	7,890	23,177	46,777	91,948	292,748
Average yearly income (€)	19,894	2,386	16,309	18,154	19,877	21,581	24,100

Notes: An observation is a Local Market Area in a year. All the statistics refer to the year 2009, unless otherwise specified. Information on the number of filling stations, the number of filling stations offering CNG, the share of monofuel and independent stations are obtained by *Prezzibenzina.it* and validated and integrated with information from printed guides. Data on the total number of circulating cars and on the number of circulating CNG cars were provided by ACI. Data on population come from the 2011 Census of Italian population conducted by ISTAT. The average income in a market is calculated based on income tax data collected by the Italian Ministry of Finance.

In addition to the census of Italian filling station, we purchased data on the yearly stock of circulating cars between 2005 and 2014 at the municipality level by type of fuel from the vehicle registration database maintained by the Italian Association of Car Owners (ACI). This source is highly accurate since vehicle registration is mandatory in Italy. Finally, we collected information on population for each market from the 2011 Census conducted by the Italian National Institute of Statistics (ISTAT) and average market income for years 2005-2014 for all the markets in our sample.

Our market definition throughout the study is a Labour Market Area (henceforth,

[7]In Appendix A we present a detailed discussion of the issue, describing the steps we took to improve the quality of the data and presenting a validation exercise which compares our dataset with official sources.

LMA). A LMA is a geographical aggregations constructed by ISTAT based on the analysis of reported households commuting patterns so that people living in a LMA are likely to work within its boundaries. The choice of defining a market as a LMA has been adopted by other studies analyzing retail sectors in Italy (Magnolfi and Roncoroni (2016)) and serves particularly well our purposes: if individuals are primarily commuting within a LMA, the stations they can potentially refill at will also lie within its boundaries.

In Table 1 we present the descriptive statistics for the main variables for the LMAs in our sample. The Italian territory is partitioned into 611 LMAs with an average population of 90,000. There is, however, substantial heterogeneity in the size of the LMAs: some only include 4,000 people and the largest ones have up to 4,000,000 inhabitants. Similar cross-sectional variation can be observed for the number of filling stations operating in the market and the share of independent stations, which are absent in some markets while representing almost 40% of the supply in others. The table also provides a sense of the growth of the CNG supply presenting a snapshot of the number of stations offering CNG at three points in time. In 2005, the CNG retail market is still just a niche and it grows slowly until 2009 when growth accelerates.

3 Potential competition and preemption

To quantify the effect of the intensity in potential competition on the incentives for entry preemption, we estimate a duration model whose hazard depends on a measure capturing the strength of potential competition in a market. Although our data contain information at the station level, we elect not to model the timing of individual filling stations' entry in the CNG market. Instead, we aggregate data at the market level and analyze the hazard of the event that a CNG pump is installed by any of the stations located in the LMA. This is an easier outcome to study than the timing of adoption of an individual firm, which would require us to model complex strategic interaction among competitors (Schmidt-Dengler (2006)). At the same time, the speed of adoption at the market level is informative on the effect we wish to estimate. In Appendix B we show that under mild assumptions on the nature of heterogeneity across stations, aggregating the information at the market level would not impact the inference on the relevance of potential competition for preemption behavior.

Our identification strategy exploits the changes occurred in the early 2000s in the legislation regulating the opening of CNG filling stations. Under the new rules, it became possible to sell CNG close to major roads and populated areas and jointly with traditional fuels. This led to an increase in the supply of CNG, driven in large part by gasoline filling stations expanding their offer by adding CNG pumps. Hence, the number of stations selling gasoline in a LMA seems a reasonable proxy for the set of firms that could potentially enter the CNG market. Two LMAs with similar characteristics but a different number of gasoline and diesel stations operating within their boundaries will entail a different level of potential competition for firms pondering entry in the CNG market. Since the change in legislation was unforeseen, the option of selling natural gas could not have been anticipated by refiners and station managers at the time of the station opening. Therefore, the market potential for CNG should not have factored in the decision of opening a station in a particular LMA. In other words, the variation in the number of active diesel\gasoline stations across LMAs is plausibly exogenous to demand for CNG.[8]

We estimate a Cox proportional hazard model where a failure occurs when a station located in the LMA starts to distributing CNG. Seeing as some markets experience entry of more than one station in the CNG market at different points in time, we allow for multiple failures. We measure the strength of the potential competition in a market using the number of firms that could decide to offer CNG, which we assume to be the number of gasoline stations active in a market in 2009 and not yet selling natural gas. Ideally, we would want to use the number of stations active at the beginning of our sample span to construct our measure of potential competition. In fact, the more time elapses from the year in which the CNG regulation was eased, the more we risk picking up entry endogenous to that change. However, as we explain in Appendix A, data on the number of gas station collected by *Prezzibenzina.it* are likely to exhibit higher measurement error at the beginning of our sample. Therefore, we use the number of gasoline stations active in 2009, the first year in which the *Prezzibenzina.it* database is fully verified, to identify potential entrants. There we also show that entry of gas stations was still heavily regulated in the early 2000s so that their number is stable between 2005 and 2009. Hence, using the 2009 market structure instead of the 2005 one should not dramatically impact our results. Moreover, all the results are robust to using the 2005 data on the number of gas stations, as can be seen in Appendix C.3.

We model the impact of potential competition on the hazard by creating dummies for the terciles of the cross-market distribution of the number of potential entrants in 2009. Markets in the first tercile count no more than 8 potential entrants; markets in the second tercile have between 9 and 21 potential entrants; markets in the top tercile have 22 potential entrants or more. We expect the threat of preemption by competitors to be increasing in the number of potential entrants, although not necessarily in a linear fashion. The spline structure allows for flexibility in recovering this effect. In our baseline specification we parametrize the hazard as follows:

$$\lambda(t, X) = \exp(\beta_0 + \beta_1 \mathbb{1}\{\text{2ndTercile}\} + \beta_2 \mathbb{1}\{\text{3rdTercile}\} + \beta_3 \text{Population} + \beta_4 \text{Avg\_Income}) \cdot \lambda_0(t)$$

[8] This assumption would not hold if demand for diesel\gasoline and CNG were highly correlated. In Appendix C.1 we show that, once we control for observable measures of market size, this is not the case.

Besides the dummies capturing the effect of potential competition, we control for the size of the market using the population of the LMA, as customary in the entry literature (Bresnahan and Reiss (1991b)), as well as for the average income in the LMA. We also include a set of dummies for the five macro regions of Italy: Northwest, Northeast, Center, South and Islands.

In our specification, the measure of potential competition is not scaled for market size: we use the number of potential competitors and not, for instance, the number of potential competitors per person in the market. Market size is instead separately controlled for. We do so because market size can have nonlinear effects in dynamic entry models (Ellison and Ellison (2011); Gil et al. (2015)). For example, in highly populated areas it can be hard to preempt or deter entry because of the sheer size of the market. Blocking subsequent entry can instead be possible in narrower markets.[9] Scaling potential competition by market size would risk confounding the variation in this measure, mixing the change in the strength of the incentive to enter early due to competition with that due to the feasibility of preemption. Instead, our identification of the impact of potential competition comes from variation in the number of potential entrants for given market size, isolating the two effects. Market size and potential competition are obviously correlated but not to the point of not being separately identifiable. The distribution of the market size in the different terciles of the measure of potential competition have substantial overlap.

The hazard ratios from the baseline specification are reported in the first column of Table 2. We find that the rate of entry rises almost five times for markets with intermediate as opposed to low levels of potential competition. A shift from low to high potential competition induces a tenfold increase in the hazard that a station in the market will start offering CNG. This result is neatly displayed in the left panel of Figure 3, showing the survival function for markets with different levels of potential competition. At any point in time, the probability of not experiencing introduction of a CNG pump by any stations in the LMA is a monotone function of potential competition. The survival function for markets with the lowest number of gas stations lies above those for markets in the top two terciles of such distribution and the survival function for the top terciles is lowest. We find a strong and monotonic effect of potential competition on the speed of entry in the CNG market also when we refine our measure of the intensity of competition among potential entrants by classifying markets using the quintiles of the distribution of the number of potential entrants in the LMA[10] (column (2) and right panel of Figure 3) or when we assume a quadratic relationship between the hazard and the intensity of potential competition as in Gil et al. (2015) (column (3)).

[9]This argument motivates the test for entry deterrence developed in Ellison and Ellison (2011), which we will replicate for the case of preemption in Section 4.

[10]The quintiles of the distribution of potential entrants are 5, 10, 17 and 33.

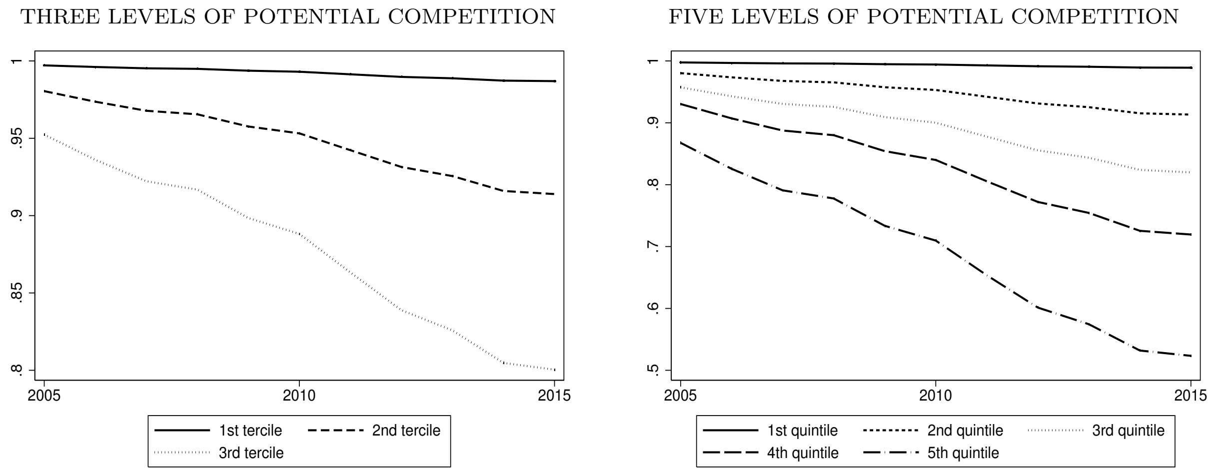
Table 2: Effect of potential competition on preemption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
𝟙 {2nd Tercile} Potential Entrants	4.923*** (0.000)			2.839*** (0.000)	4.675*** (0.000)	4.794*** (0.000)	9.357*** (0.000)	5.095*** (0.000)
𝟙 {3rd Tercile} Potential Entrants	10.655*** (0.000)			3.539*** (0.000)	10.346*** (0.000)	9.124*** (0.000)	12.222*** (0.000)	10.792*** (0.000)
𝟙 {2nd Quintile} Potential Entrants		7.520*** (0.000)						
𝟙 {3rd Quintile} Potential Entrants		12.677*** (0.000)						
𝟙 {4th Quintile} Potential Entrants		19.324*** (0.000)						
𝟙 {5th Quintile} Potential Entrants		36.610*** (0.000)						
Num Potential Entrants			1.018*** (0.000)					
(Num Potential Entrants)^2			1.000*** (0.000)					
Long Run CNG Demand					1.000*** (0.000)			
Population (100,000)	1.094*** (0.000)	1.090*** (0.000)	0.904*** (0.002)			1.073*** (0.000)	1.273* (0.093)	1.144*** (0.000)
Log(Gasoline and diesel cars)				2.083*** (0.000)				
Average Income (1,000 €)	1.111*** (0.000)	1.062** (0.045)	1.169*** (0.000)	0.972 (0.363)	1.200*** (0.000)	1.092*** (0.000)	0.966 (0.637)	1.083** (0.020)
Observations	1206	1206	1206	1206	1206	1206	479	1206
LMAs	611	611	611	611	611	611	404	611

Notes: The table reports the hazard ratios from a Cox proportional hazard model. The controls for potential competition are the terciles of the distribution of the number of potential entrants in the market in 2009, with the bottom tercile as the excluded group. We consider as potential entrants all the diesel\gasoline filling stations active in the market. In column (2) we measure differences in potential competition more finely using the quintiles of the distribution of the number of potential entrants and in column (3) we make a parametric assumption on the effect of competition assuming a quadratic relationship. Columns (4) uses the number of circulating diesel and gasoline cars in 2005 as a controls for market size. In Column (5) market size is proxied by fitted demand for CNG cars if the density of CNG filling stations were the same as that of LPG filling stations. This figure is obtained using demand estimates in Pavan (2015). In Column (6), we allow markets to have different baseline hazards depending on the number of CNG entry events they experienced (0 CNG incumbents, 1 CNG incumbent, 2 CNG incumbents, etc.) and in column (7) we only include LMAs until they experience the first failure (i.e., entry of the first station serving CNG). Column (8) reports results of a shared frailty model at LMA level. All the specifications include macro areas (Northwest, Northeast, Center, South and Islands) fixed effects. The robust standard errors are clustered at the LMA level (LMAs can appear multiple times if they experience entry by more than one filling station). The p-values of the coefficients are reported in parentheses. ***: p<.01, **: p<.05, *: p<.1.

The results are also robust to different proxies for the size of the market. In column (4) we replace population with the number of traditional fuel (gasoline and diesel) cars

Figure 3: SURVIVAL FUNCTION: THE EFFECT OF POTENTIAL COMPETITION



Notes: The left figure portrays the survival functions implied by the estimates in column (1) of Table 2. The right figure portrays the survival functions fitted using the estimates in column (2) of Table 2.

registered in the LMA in 2005. Since entry decisions are made by stations with expectations of future demand for CNG in mind, in column (5) we use the demand model estimated in Pavan (2015) to construct a proxy for such expectations. We compute the "long run" demand for CNG cars in the LMA as the number of CNG cars that would be demanded if the infrastructure serving natural gas cars were the same size as that for liquified petroleum gas ones (LPG), another green fuel whose diffusion was not restricted by regulation and whose supply already reached a more homogenous coverage of the Italian territory.

It is natural to think that the presence of incumbents in the CNG market would reduce the payoff from entry for the potential entrants. Hence, the second entry would be slowed down with respect to the first. However, in games with more than two potential entrants this intuition does not necessarily hold and a variety of predictions can be obtained on the effect of the number of incumbents on the entry rate (Reinganum (1981); Argenziano and Schmidt-Dengler (2013, 2014)). This raises a potential concern on whether the contemporaneous market structure of the CNG market can confound our assessment on the effect of potential competition. We propose two exercises to dispel this notion. First, in column (6) we estimate a conditional risk set model were the observations are stratified based on the number of prior entry events experienced. This means that we allow markets at risk of experiencing the first entry to have a different baseline hazard from LMAs at risk of a second, third, etc. entry occurrence. To identify this model, we exploit the fact that some LMAs in our sample see multiple entry events and that there are markets with stations already serving CNG at the beginning of our data in 2005. In column (7), instead, we estimate the model only on the sample of LMAs at risk of experiencing the first adoption of a CNG pump by a station. Therefore, this specification removes the effect of heterogeneity in market structure and documents the impact of potential competition conditional on no prior entry having occurred in the CNG market. In both cases the results of the baseline specification are qualitatively confirmed.

The identification of all the specifications presented up to this point exploits cross-sectional variation across LMAs. We take markets with similar size but with different number of potential entrants and assess whether the rate at which they experience entry in the CNG market differs. In column (8), we exploit the fact that in a number of markets we observed multiple entry events by estimating a Cox model with a LMA specific frailty accounting for the effect of market specific unobserved heterogeneity on the hazard. Reassuringly, our results confirm the qualitative and quantitative findings of the baseline specification.

4 Evidence of preemptive behavior

The evidence we presented in Section 3 states that markets with an exogenously higher number of potential entrants have a higher hazard to register entry into CNG at any point in time. This finding is consistent with the existence of a preemption race in the developing market for retail CNG where the presence of a large number of firms that could enter puts additional pressure on each station considering to sell natural gas and accelerates entry. However, a positive relationship between the number of potential entrants and entry rates needs not to imply the presence of preemptive behavior. The same correlation could arise even in a static entry model like the one introduced in Bresnahan and Reiss (1991a) where there is an unobserved (to the econometrician), firm specific component of the profit function. Then, a larger number of potential entrants would imply more draws from the distribution of such idiosyncratic shock and, therefore, a higher chance of observing a firm with a draw good enough to justify entry in the CNG market on the basis of static profits alone.

Obviously, determining the mechanism inducing the correlation between potential competition and speed of entry is important to understand its implications. If it is due to preemption, we could worry about an increase in the losses induced by inefficiently early entry. If instead it results from heterogeneity across filling stations, the acceleration of entry would not be harmful. In fact, it could even be beneficial in scenarios, like the one we are analyzing, where "supply creates demand" (Agarwal and Bayus (2002)). Unlike structural approaches that can overcome the unobservability of the preemption incentives relying on the model to isolate them, there is no obvious way of linking the reduced form evidence on the rate of entry we presented in the previous section to preemption motives. Therefore, in this section we present several pieces of evidence that strongly suggest that such link does exist.

The standard conditions leading to preemption entail either high fixed or marginal costs or limited demand. In that case, entrants are unlikely to make non negative profits. Entry, however, could still be profitable in a dynamic perspective, if costs are believed to be falling rapidly and\or demand is expected to grow fast. The early years of the CNG retail markets fit well this description: the cost of setting up a CNG pump are high and, most important, the share of circulating CNG cars is still limited. To provide a sense of the profitability of the CNG market in the years covered in our sample, we perform a calibration of the filling station profit function. We focus on the profits of a firm entering as a monopolist since this is the most stringent test of profitability. Profits are usually declining in the number of incumbent firms, whereas fixed costs are not. Hence, a market not profitable for a monopolist will also be not profitable for any other market configuration.

The post entry profits accruing to a firm from adding CNG pump in market m take the same functional form as in Pavan (2015):

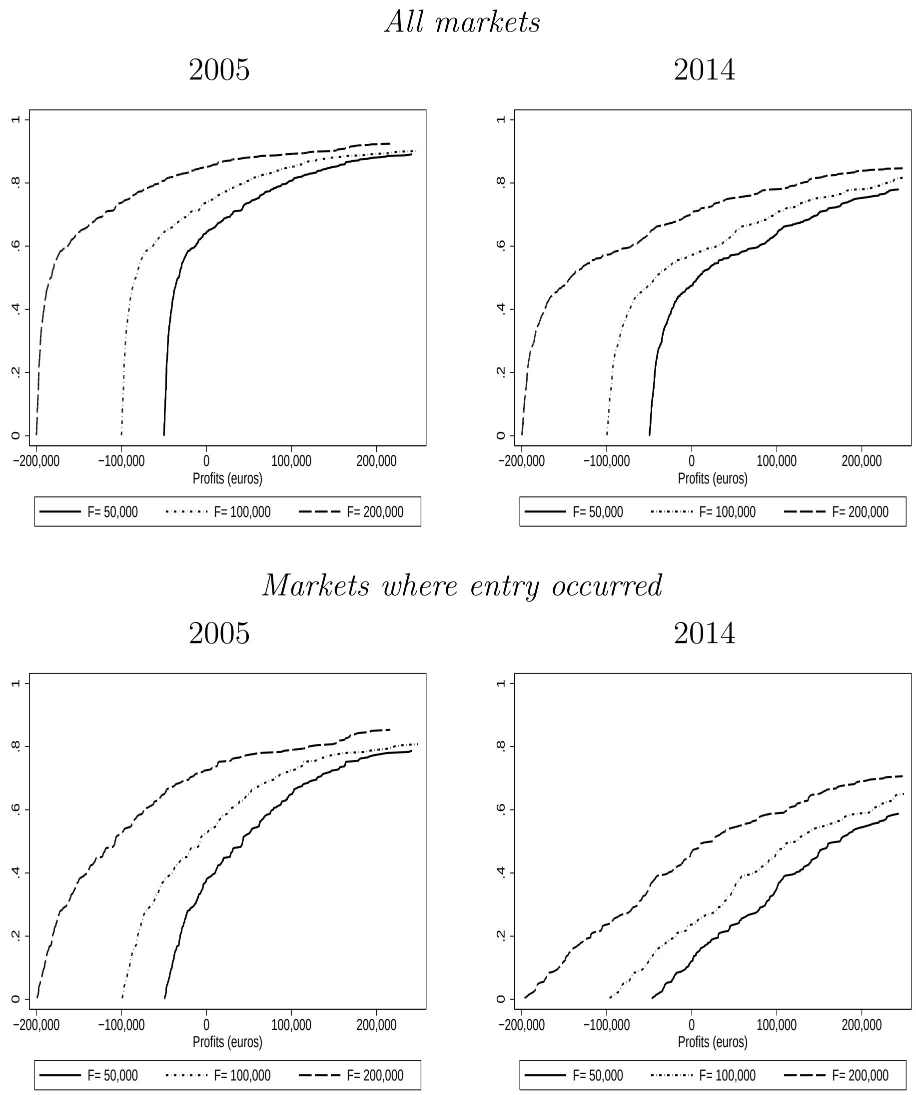
$$\Pi_{im} = (p - c)\ (k_m \cdot Q_m) - F_m + \varepsilon_i \tag{1}$$

where $(p - c)$ is the markup and k_m is the average yearly consumption of fuel by a CNG car. Q_m is the stock of circulating cars consuming CNG and F is the fixed cost of entering the CNG market. Pavan (2015) consulted industry sources to obtain national averages for markups in 2012 (0.15 €$/m^3$) and fuel consumption (Kg of gas consumed yearly by the average car in the region where the LMA is located). We use our data on circulating car by type of fuel to impute market specific demand (Q_m).[11]

In Figure 4 we plot the cumulative distribution of the calibrated static profits for a representative monopolist with a profit function as the one in equation 1. We only calibrate the market specific component of the profit function, that is we do not simulate the idiosyncratic profitability shock ε_i and profitability only differs across LMAs. This means that we capture cross-market variation in the profits accrued to a representative filling station entering market m as a monopolist. Each plot contains three distributions, obtained assuming different values for the fixed cost of entry (€50,000 , €100,000 and €200,000). The top two panels display the distribution for all the markets in our sample at two points in time: before (2005) and after (2014) demand for CNG took off.

[11]The profit function refers only to profits originated from CNG sales. We disregard the spillover of increaesed traffic through the station on revenues from other goods on sale at the station (food, newspapers, etc.). Only a fraction of the gas stations in our sample sell merchandise other than fuel and the impact on this business due to offering CNG is likely to be quite limited.

Figure 4: Static monopoly profits from entry in the CNG market



Notes: Each plot represents the distribution of the static profits of a representative monopolist across LMAs under different assumptions on the level of the entry fixed costs. The top row figures portray the distribution at the beginning of our sample (2005, left) and at the end of our sample (2014, right) for all the markets in our data. The figures on the bottom row display the same objects but consider only the subset of markets where we observe at least one station offering CNG by the end of our sample span. The profits are calibrated using the parametrization in Pavan (2015).

The results of this exercise can only provide suggestive evidence but it is nevertheless striking that in 2005, in the immediate aftermath of the lifting of the CNG sale restrictions, over 60% of the markets were statically unprofitable even under the most conservative assumption on the level of the CNG installation fixed costs. This figure is lower but still considerable if we limit our attention to markets where we do observe stations offering CNG in the data (bottom two plots). Given that the level of the calibrated profits implied by the common component of the profit function is often negative, it is hard to justify early entry in the CNG market as due to static positive profits. This would in fact require rather large and positive station specific profitability shocks. At the same time, we observe that market conditions evolve rapidly. By the end of our sample period, the half of the markets where we observe entry would be profitable for a representative monopolist even if entry cost were at the high end of the range of values we considered. This can rationalize a preemption entry strategy where the station earns negative profits upon entry but it secures a spot in the market and eventually obtains positive profits.

We obtain further evidence in support of the presence of preemption in our data by adapting the strategy Ellison and Ellison (2011) originally developed to test for entry deterrence behavior. The test is based on the premise that incentives to preempt should vary depending on the size of the market in which firms operate. In large markets, the number of firms that can be accommodated in equilibrium is high with respect to the number of potential entrants. Therefore, early entry does not impact future competition and we should not observe preemption in this context. Instead, in markets of intermediate size the number of firms that can operate in equilibrium is more likely to be lower than the number of potential entrants, giving rise to incentives for preemption. Finally, the prediction for the behavior we should observe in small markets is ambiguous. In fact, a small market where no firm has yet entered could be either too small to ever sustain even a single firm or just large enough to accommodate one firm. In the former case, there would be no preemption race no matter how many potential entrants are in the market. In the latter case, the gains from preemption are the highest: the first firm to open a CNG gas pump will secure a position as a monopolist.[12]

Table 3 presents estimates of the same Cox duration model analyzed in Table 2 performed separately for three subsamples corresponding to the top, middle and bottom tercile of the cross sectional distribution of market size, proxied by population. Since the number of observations in each population tercile is limited, we can no longer flexibly retrieve the effect of potential competition. We implement a parsimonious specification analogue to the one of column (3) of Table 2 where the hazard is a quadratic function of the number of potential entrants and average household income serves as a proxy for market size.[13] We can interpret the coefficients on the number of potential entrants in the different columns of Table 3 as the impact on the hazard of observing a station introducing CNG in a market given by an increase in the intensity of potential competition conditional on the size of the market. The result we obtain is consistent with the predictions of an entry preemption model and in line with those delivered by a similar exercise performed in Gil et al. (2015). Although our measure of potential competition is significant for small, intermediate and large markets, the effect is weaker in large markets where we would not

[12]The prediction for the incentive to preempt in small markets differs from that derived by Ellison and Ellison (2011) for entry deterrence. In the case of entry deterrence no action should be observed in small markets because the limited size of the market would naturally protect the incumbent from the threat of newcomers.

[13] Our findings are robust to using alternative variables in this role such as the stock of circulating traditional cars or the long run demand for CNG cars.

Table 3: Timing of entry, nonlinear effect in market size

	Small markets	Intermediate markets	Large markets
Num. Potential Entrants	1.426*	1.154*	1.009***
	(0.076)	(0.054)	(0.000)
(Num. Potential Entrants)^2	0.994	0.996	1.000***
	(0.518)	(0.101)	(0.000)
Average Income (1,000 €)	0.834	0.951	1.053
	(0.242)	(0.436)	(0.167)
Obs	228	332	646
LMAs	204	204	203

Notes: The table reports the hazard ratios from a Cox proportional hazard model. The controls for potential competition is a quadratic function of the number of potential entrants in the market measured as the number of gasoline stations active in the market in 2009. The model is estimated on a different subsample in each column. In the first column (*Small markets*), we use markets in the first tercile of the distribution of population; in the second column (*Intermediate markets*) we use market in the second tercile and in the third column (*Large markets*) LMAs in the top tercile. All the specifications include macro areas (Northwest, Northeast, Center, South and Islands) fixed effects. The robust standard errors are clustered at the LMA level (LMAs can appear multiple times if they experience entry by more than one filling station). The p-values of the coefficients are reported in parentheses. ***: p<.01, **: p<.05, *: p<.1.

expect the preemption motive to play a role.

The last bit of evidence tying entry behavior observed in our data with preemption motives exploits the presence of independent gasoline stations in the Italian retail fuel market. As we explained in Section 2, the majority of the gasoline stations in Italy are branded pumps, controlled directly or indirectly by refining companies. However, about 10% of the stations are "white pumps". These stations are independently operated by individual entrepreneurs who purchase the fuel they sell on the wholesale market without agreements tying them to any particular oil company. This distinction is important because branded and unbranded stations have different incentives with respect to entering the CNG markets. Oil companies may be concerned about contributing to fostering the availability of CNG, which provides an alternative to gasoline. Hence, the pumps they control directly and their franchisees may be wary of selling CNG. White pumps instead should be willing to distribute CNG as long as the cannibalization with gasoline is not so high that fixed cost of adding the CNG pump cannot be recovered. In short, for branded stations there is a force tempering the desirability of entering the CNG market which is not present in the decision of unbranded stations.

The difference in the incentives to enter the CNG market between branded and unbranded stations provides us with a shifter of the incentives to preempt alternative to the sheer number of potential competitors. In fact, even for given number of potential entrants, in LMAs where the share of white pumps is larger there is a higher risk of being beaten to the market by a competitor. Therefore, the incentive to preempt should be stronger.[14] With respect to the specification in equation 1, this shifts the main identifying assumption from the *overall* number of filling stations being uncorrelated to CNG profitability to the number of *unbranded* stations in a market being orthogonal to it. Therefore, we no longer require, for instance, that demand for regular fuel and CNG is uncorrelated conditional on observables but only that white pump profitability in the gasoline and CNG segment is. In Appendix C.2, we document that the location of white pumps does not appear to reflect any differences in expected profitability across the two segments with respect to branded gas stations.

Unlike the entry of new stations, which we have documented to be rare and heavily regulated during most of our sample span, there are no restrictions to the decision of a station to change its status from branded to unbranded or vice versa. In fact, the number of unbranded stations has started to grow rapidly since the beginning of 2008.[15] This means that it is important for us to identify the status of each station around the time of the change in the legislation about CNG distrbution. Otherwise, we risk picking up spurious variation in station branded status due to the fact that stations wanting to sell CNG may decide to become unbranded to avoid opposition from oil companies franchisors. Fortunately, the *Prezzibenzina.it* dataset reports information on past brand status of each filling station, which allows us to distinguish branded station and white pumps since the beginning of our sample.[16]

In Table 4, we use the fraction of white pumps in a market as a measure of intensity of potential competition to replicate our baseline exercises from Table 2.[17] Even with this alternative identification approach, our results are qualitatively confirmed. The main difference with the baseline estimates is that the effects are smaller in magnitude, but still quite sizeable, and the clear monotonicity of the impact of potential competition has vanished. We estimate a positive effect on the speed of entry for markets that are not in the bottom of the distribution in terms of share of white pumps operating. This effect is, however, similar for markets around the average and markets in the right tail of the distribution.

[14] We could have pursued a similar strategy using filling stations operated or franchised by Eni, an oil company that also extracts natural gas. In that case, the presence of Eni pumps in a market would have signalled a higher likelihood of entry since Eni would have benefited its natural gas business by favoring the diffusion of CNG at the retail level. However, the number of Eni filling stations is not large enough to give us sufficient power to undertake this approach.

[15] The delay in the switching to independent retailing, with respect to the year in which the regulation of CNG distribution changed, can be explained by the fact that, even absent regulation, franchisee have to wait until the end of their contract with a refiner before they can become white pumps.

[16] The information on the old brand has been validated using printed guides. More details on this can be found in Appendix A.

[17] We cannot replicate specification using a quadratic function (column (3) of Table 2) as the distribution of the share of white pumps has a significant mass at zero and it is close to zero in several other markets. This makes the linear and the quadratic part highly collinear. For the same reason, in column (2) we use the quartiles of the distribution of this variable (instead of the quintiles as we did for the number of potential entrants in column (2) of Table 2). The first two quintiles of the distribution of the share of white pumps coincide and are equal to zero.

Table 4: Effect of competition by white pumps on preemption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
𝟙 {2nd Tercile} % White Pumps	2.952*** (0.470)		1.509** (0.243)	3.088*** (0.484)	2.444*** (0.352)	1.694 (0.637)	2.743*** (0.435)
𝟙 {3rd Tercile} % White Pumps	2.540*** (0.362)		1.641*** (0.225)	2.623*** (0.381)	2.141*** (0.282)	1.588* (0.428)	2.422*** (0.348)
𝟙 {2nd Quartile} % White Pumps		3.169*** (0.636)					
𝟙 {3rd Quartile} % White Pumps		3.215*** (0.474)					
𝟙 {4th Quartile} % White Pumps		2.188*** (0.341)					
Long Run CNG Demand				1.000*** (0.000)			
Population (100,000)	1.083*** (0.016)	1.082*** (0.016)			1.070*** (0.013)	1.318* (0.194)	1.168*** (0.038)
LOG(gasoline and diesel cars)			2.198*** (0.142)				
Average Income (1,000 €)	1.219*** (0.030)	1.204*** (0.029)	0.984 (0.033)	1.307*** (0.031)	1.179*** (0.026)	1.089 (0.076)	1.212*** (0.042)
Observations	1202	1202	1202	1202	1202	475	1202
LMAs	607	607	607	607	607	400	607

Notes: The table reports the hazard ratios from a Cox proportional hazard model. The variables proxing for the intensity of potential competition are the terciles of the distribution of the share of white pumps over the total number of gasoline stations active in the LMA in 2009, with the bottom tercile as the excluded group. The status of a pump as branded\unbranded is established using "past brand" information in *Prezzibenzina.it* and looking at printed guides for the year 2005. In column (2) we measure differences in potential competition more finely using the quartiles of the distribution of the share of white pumps. Columns (3) and (4) experiment with alternative controls for the size of the market using the number of circulating diesel and gasoline cars in 2005 and the long run demand for CNG car respectively. The latter is calculated using the estimtes from the demand model in Pavan (2015), assuming that the density of CNG stations were as high as that of LPG stations in 2015. In Column (5) we allow markets to have different baseline hazards depending on the number of CNG entry events they experienced (0 CNG incumbents, 1 CNG incumbent, 2 CNG incumbents, etc.) and in column (6) we only includes LMAs until they experience the first failure (i.e., entry of the first station serving CNG). Column (7) reports results of a shared frailty model at LMA level. The robust standard errors are clustered at the LMA level (LMAs can appear multiple times if they experience entry by more than one filling station). The p-values of the coefficients are reported in parentheses. ***: p<.01, **: p<.05, *: p<.1.

As a last exercise, we assess whether the share of white pumps among the competitors a station is facing has any impact on its decision to preempt. Since the most salient difference between branded and unbranded stations when it comes to the adoption of CNG is the lower disincentives for white pumps to add the fuel to their offer, variation in this variable speaks directly to preemption incentives and provides evidence that dynamic considerations are driving entry choices in the market for CNG.

In Table 5, we estimate a set of station level linear probability models to characterize the entry behavior of establishments in our sample. We exclude from the analysis the monofuel CNG stations that were operating since before the change in the regulation for the distribution of CNG: for those the choice of entering the market had surely not to do with preemption. In the first column of the table, the dependent variable is a dummy taking value 1 if the station starts selling CNG before the end of our sample period, and

Table 5: Station level probability of preemption.

	Entrant	Early Entrant		First Entrant	
𝟙{White Pumps among competitors}		0.007***		-0.020	
		(0.002)		(0.102)	
Share white pumps among competitors			0.014***		-0.073
			(0.004)		(0.135)
𝟙{White Pumps Among competitors}*White pump		-0.009*		-0.148	
		(0.005)		(0.167)	
Share white pumps among competitors * White pump			0.007		0.014
			(0.009)		(0.226)
White Pump	0.048***	0.029***	0.025***	0.228**	0.179*
	(0.004)	(0.003)	(0.003)	(0.096)	(0.093)
Observations	12415	12415	12415	308	308
R^2	0.02	0.02	0.02	0.29	0.28

Notes: The table reports estimates of a linear probability models with different dependent variables. In the first column, the dependent variable is an indicator for whether the station has begun selling CNG by the end of our sample span (2014). In the second and the third column, the dependent variable is a dummy taking value 1 if the station has begun selling CNG before 2007. In the fourth and fifith column, the dependent variable is an indicator for whether the station was the first establishment to sell CNG in the LMA. An observation is a filling station and we consider the universe of filling stations active in Italy except the monofuel CNG pumps, all of which had entered the market before the shift in the regulation on the distribution of CNG. The number of competitors faced by a station is defined as the number of active filling stations within a circle of a certain radius centered in the location of the station in question. The dimension of the radius is 1Km for stations located in urban markets, 2Km for stations sited in suburban markets and 4Km for stations in rural areas. In all specification we control for LMA fixed effects, station location (urban, suburban or rural) fixed effects and distance of the station from the closer main road. We also include dummies for the number of stations in each LMA that were already selling CNG in 2005. The standard errors of the coefficients are reported in parentheses. ***: p<.01, **: p<.05, *: p<.1.

0 otherwise. This exercise formally validates the premise of our identification using the share of white pumps as a shifter of the intensity of the potential competition by checking whether white pumps are more likely to distribute CNG. Beyond the dummy indicating that the station is a white pump (*White pump*), we control for LMA fixed effects, for the type of the area in which the station is located (urban, suburban or rural) and for the distance of the station from the closest main road. Finally, we account for the number of incumbents in the CNG market in the LMA at the beginning of our sample (2005) with a full set of dummies. We find that white pumps are over twice as likely to sell CNG by 2014 as branded stations.

The remaining columns in Table 5 assess the impact of the share of white pumps among a station's competitors on its decision to preempt. Since our level of observation is now a station, we have to introduce a criterion to identify the competitors of each filling station. We do so by drawing a circle of a certain radius around each station and considering as the station's competitors all the filling stations active within that radius. To account for the fact that the catchment area of a station depends on the density of population in the location where the station operates, we allow for circles of different radius to be drawn in different types of locations. In urban areas, we count as competitors of a station all the establishments operating in the 1Km (0.62 miles) radius circle around it; in suburban areas, we stretch the radius to 2Km and in rural locations we set the radius of the circle to 4Km.[18]

We experiment with two different definitions of preemptive behavior on the part of the filling station. First, we use as dependent variable a dummy that takes value 1 if the stations entered "early". We define a station selling CNG as an early entrant if it started doing so before the take off in demand displayed in the left panel of Figure 2. Therefore the "early entrant" dummy takes value 1 for stations that were selling CNG before 2007. Then, we explore a much more stringent definition of preemption by using as dependent variable of the linear probability model a "first entrant" dummy that takes value 1 if the station was the first to sell CNG in the LMA in which it is located. We regress both these indicators of preemptive behavior on the same set of controls introduced in the specification in the first colum of the tables as well as on two different variables that capture variation in the incentive to preempt. In one specification, we use a dummy variable signaling stations that have at least one white pump among their competitors. In an alternative exercise, we exploit the intensive margin of the presence of white pumps and use the share of competitors of a station that are white pumps. Since we documented that the entry behavior of branded and unbranded pumps is different, we control for the branded status of the station and we also let the effect of preemption incentives differ for branded station and white pumps by interacting the dummy variable for presence of unbranded competitors (or the share of unbranded competitors) with the station brand status. Finally, in all specifications we include LMA fixed effects.

The results, reported in Table 5, support the notion that the early years of the CNG retail market were characterized by preemptive behavior. When we use early entry as a sign of preemption, we find that the probability of engaging in preemptive behavior goes up by almost 30% for branded stations that face at least one white pump among their competitors. On the intensive margin, one standard deviation increase of a branded station's competitors represented by white pumps is associated with a 2% increase in the probability of early entry. The interaction terms suggest that white pumps do not react to the extensive margin of competition by fellow unbranded stations but do respond as branded stations to increases in the share of unbranded stations among their opponents. We do not find any significant evidence of preemption motives when we define as preemption the decision to enter the CNG market first. However, this is a stricter notion of

[18]Results are robust to changes in the definition of the catchment areas. For instance, we have experimented with homogeneous catchment areas of radius 1Km or 2Km for urban, suburban and rural markets; with reducing the catchment area of rural markets to a circle of 3Km radius and with reducing both the catchment areas of suburban and rural market setting radii of 1Km and 2Km, respectively. Results are available upon request.

preemption that severely limits our market size: we can have only as many observations as markets.

5 Conclusions

In this paper we have studied the role of competition in shaping the timing of entry in a young and rapidly growing industry. We exploited a shift in legislation that allowed filling stations selling traditional fuels to start selling compressed natural gas. This legislative intervention triggered both expansion in the market for retail CNG and identified existing stations as the main candidates to enter the market.

We showed that, controlling for the size of the market, the rate of entry in the newborn CNG retail market is significantly faster in areas with a larger number of potential entrants. We argued that this is due to the threat of potential competition speeding up the preemption race. First, we documented that static profits for early entrants in the CNG market are likely to be negative, suggesting that dynamic considerations were probably behind the entry decisions we observe. Next, we performed an adaptation of the Ellison and Ellison (2011) test for preemptive behavior finding that the effect of competition on the speed of entry is only present in markets where preemption is viable. Finally, we showed that firms facing as competitors a larger number of unbranded stations, which have an exogenously higher probability of entering the CNG market, are more likely to enter early.

References

Agarwal, R. and B. L. Bayus (2002). The market evolution and sales takeoff of product innovations. *Management Science 48*, 10424–1041.

Argenziano, R. and P. Schmidt-Dengler (2013). Competition, timing of entry and welfare in a preemption game. *Economic Letters 120*, 509–512.

Argenziano, R. and P. Schmidt-Dengler (2014). Clustering in n-players preemption games. *Journal of the European Economic Association 12*(2), 445–460.

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

Bresnahan, T. F. and P. C. Reiss (1991b). Entry and competition in concentrated markets. *Journal of Political Economy 99*(5), 977–1009.

Ellison, G. and S. Ellison (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics 3*(1), 1–36.

Fudenberg, D. and J. Tirole (1985). Preemption and rent equalization in the adoption of a new technology. *Review of Economic Studies 52*, 383–401.

Gil, R., J.-F. Houde, and Y. Takahashi (2015). Preemptive entry and technology diffusion in the market for drive-in theaters. Working paper.

Goolsbee, A. and C. Syverson (2008). How do incumbents respond to the threat of entry? Evidence from the major airlines. *Quarterly Journal of Economics 123*(4), 1611–1633.

Igami, M. (2017). Estimating the innovator's dilemma: Structural analysis of creative destruction. *Journal of Political Economy 125*, 798–847.

Igami, M. and N. Yang (2016). Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada. *Quantitative Economics 7*(2), 483–521.

Illanes, G. and S. Moshary (2015). Estimating the effect of potential entry on market outcomes using a licensure threshold. Kilts Booth Marketing series No. 2-013.

Magnolfi, L. and C. Roncoroni (2016). Political connections and market structure. Working paper.

Pavan, G. (2015). Green cars adoption and the supply of alternative fuels. Working paper.

Reinganum, J. F. (1981). Market structure and the diffusion of new technology. *Bell Journal of Economics 12*(2), 618–624.

Schmidt-Dengler, P. (2006). The timing of new technology adoption: The case of MRI. Working paper.

Shen, Q. and J. M. Villas-Boas (2010). Strategic entry before demand takes off. *Management Science 56*(8), 1259–1271.

Yang, N. (2015). Non-monotonic entry and preemptive motives. Working paper.

Zheng, F. (2016). Spatial competition and preemptive entry in the discount retail industry. Working paper.

Appendix - Not for publication

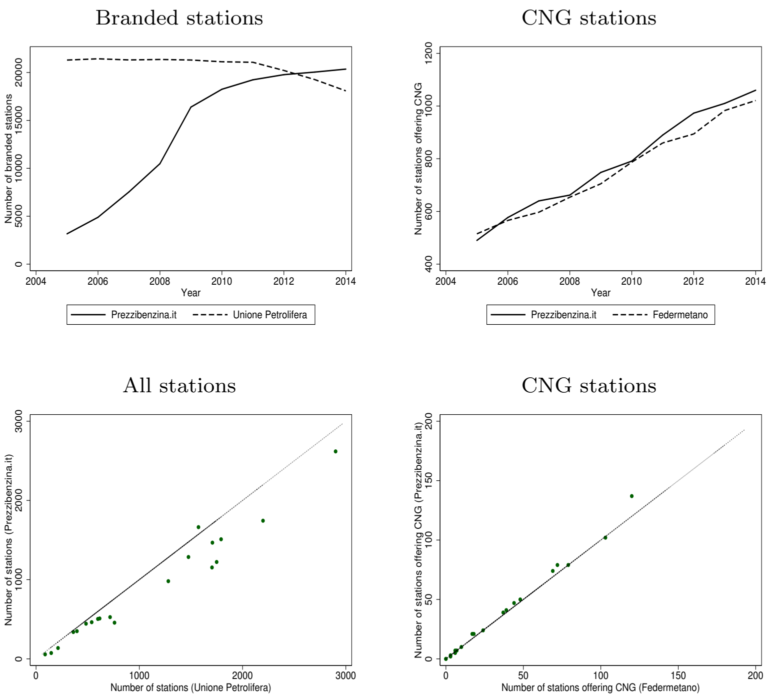
A Data

We compile a dataset with information on the universe of Italian filling stations relying on the data from *Prezzibenzina.it*, a search engine for retail fuel prices. The website was founded in 2004 and today is the most complete database on the Italian fuel retail industry. The information provided to us include name of the station, whether it is a branded station -and if so which one is its brand- or an independent white pump, its address, the complete set of fuels offered and the year the station first appeared in the database. The information on the website is reported by users or stations managers and verified by staff. Therefore, it is likely that in the first few years of activity *Prezzibenzina.it* was not covering the universe of Italian stations. The consensus is that only in 2009 *Prezzibenzina.it* reached wide enough diffusion to be confident that it listed information on nearly all the active gas station in Italy.

Number of potential entrants. The incomplete coverage of the *Prezzibenzina.it* database in the early part of our sample creates two potential problems. First, it introduces measurement error in the number of stations. In particular, for establishments entering the *Prezzibenzina.it* sample before 2009 we cannot be sure whether they are truly new entrants or existing stations that had not been reported to the website before. Second, for the same reason it may generate error in the year in which a station started selling CNG.

We have taken several steps in order to validate and integrate the information in *Prezzibenzina.it* and to minimize the extent to which these issues affect our results. First, for each region-year pair we have compared the overall number of stations obtained based on data from the website with the same figure as reported by several other sources producing official aggregate statistics on the retail fuel sector: the Italian Competition Authority, Unione Petrolifera (the association of the Italian oil companies), Federmetano and Assogasmetano (associations of distributors of methane). The evidence from the comparison is reported in Figure A.1. In the top left plot, we display the time series for the number of branded filling station active in Italy as reported by Unione Petrolifera and as resulting from the *Prezzibenzina.it*. The former is an association of all the major oil companies and should, therefore, have accurate measures of the establishment run or franchised by all the major brands. The underestimation of the number of active station is severe in the first few years of activity of *Prezzibenzina.it* but measurement error reduces rapidly. By the late 2000s, the total number of stations reported by *Prezzibenzina.it* and Unione Petrolifera is nearly identical. The bottom left plot shows that the *Prezzibenzina.it* database is fairly accurate even at a more disaggregate level by 2009. It displays the correlation between the total number of stations *Prezzibenzina.it* reports and the one obtained from the data by Unione Petrolifera region by region for the year 2009.[19] Most regions fall on or close to the 45 degrees line, implying that the website data are consistent with a complete administrative source.[20]

Figure A.1: Data validation



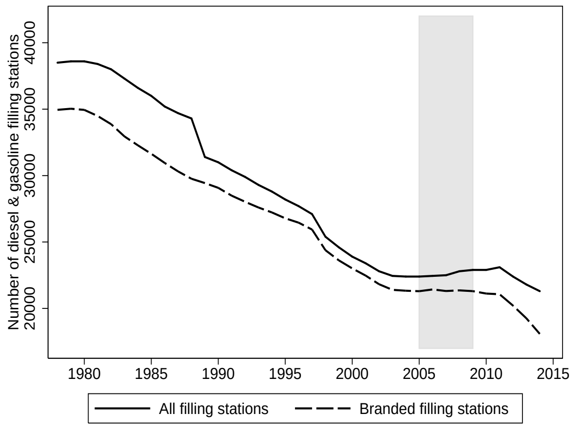
Notes: The plots on the left column compare figures on the number of filling stations obtained from *Prezzibenzina.it* with those reported by Unione Petrolifera. The top left plot displays the time series for the number of branded filling stations; the bottom left plot shows the correlation between the total number of stations reported by the two sources in 2009 region by region. The plots in the right column perform a similar comparison for the number of stations offering CNG. Here we compare the data obtained from *Prezzibenzina.it* and integrated with the printed guides with those reported by Federmetano. The dotted line in the bottom plots is the 45 degrees line.

Our validation exercise brings good news for the quality of the *Prezzibenzina.it* data but it also confirms that they are not reliable in their early years. Therefore, in the main analysis we decided to use the 2009 number of gasoline station to ensure that measurement error in the number of potential entrants variable is minimized. This choice has the obvious upside of having us rely on an accurate measurement of all the stations active at the cost of having to use a snapshot at a later date than the one we would ideally

[19]The report by Union Petrolifera does not provide regional data separately for branded and unbranded stations. Therefore, the region-by-region comparison can only be performed using the overall number of stations.

[20]Starting with the year 2015 the Ministry of Economic Development has launched the website *Osservaprezzi*, similar to *Prezzibenzina.it* but relying on a duty established by law for station owners to report their prices to the website. Therefore, *Osservaprezzi* surely spans the universe of Italian filling stations. We cross-checked the records of *Prezzibenzina.it* with those from *Osservaprezzi* for the year 2015, finding no missing stations in the *Prezzibenzina.it* data (which are actually in several cases more accurate in the geo-referencing of the station than the Ministry data).

Figure A.2: Number of active filling stations in Italy, 1978-2014



Notes: The figure plots the time series for the number of filling stations active in Italy. It is constructed using data reported by Unione Petrolifera. The series for the number of branded filling stations is based on actual data that Unione Petrolifera collects from its members. The data on the total number of filling stations is imputed using an estimate for the number of active unbranded stations. In the figure, the grey-shaded interval covers the period between 2005 (the year in which we would ideally want to measure the number of active gas stations and 2009 (the year in which we have an accurate measure at the LMA level in the *Prezzibenzina.it* data).

want to. However, a brief overview of the evolution of the retail fuel sector in Italy can help assuage concerns potentially raised by our strategy. As summarized in a report of the Italian Competition Authority (Provvedimento n. 9636 del 7 Giugno 2001, *"Ristrutturazione della rete carburanti"*), the Italian network of filling stations developed mostly between 1930 and 1970 when little legislative control was in place. From 1970 to 1989 the opening of new gasoline stations was regulated with the explicit goal of discouraging new entry. Between 1989 and 1998 the regulation was made even more stringent allowing the opening of a new gasoline station only if two others had previously shut down in the same region. Entry was fully deregulated only with the Law n.133 in 2008. This implies that: 1) there is some inertia in the number of stations that makes the status quo in 2009 a good proxy for the situation in 2005; 2) any adjustment would go in the direction of reducing the number of active station, therefore minimizing the risk that using the 2009 data leads us to pick up entry endogenous to the change in the legislation on CNG distribution. To validate this claim, in Figure A.2 we show the evolution of the total number of filling stations in Italy since the late 1970s based on data from Unione Petrolifera. The trend is indeed steeply declining until the early 2000s, when it becomes flat. In particular, it is encouraging that at the aggregate level there is not a large difference between 2005, the year in which we would want to measure the number of active stations to construct a variable proxing the intensity of potential competition, and 2009, the year we actually use to perform such task. Arguably, there is not much of a difference even between 2002, when the first change in the legislation on the distribution of CNG occurred, and 2009.

CNG distribution. On the front of the potential threat to the reliability of the year in which a station has introduced CNG among the fuels it sells, we were able to solve the problem altogether by cross-checking and completing the *Prezzibenzina.it* data with information from paperback guides[21], lists of CNG filling stations provided by some regions (for instance, Piemonte and Lombardia) and from lists found on other websites (*Metanoauto.com* and *Ecomotori.net*). The information collected allowed us to construct, for each year starting with 2005, the complete set of stations offering CNG in each market. This is witnessed by the plots on the right column of Figure A.1: the data obtained from *Prezzibenzina.it* and integrated with the printed guides match nearly perfectly those collected by the association of Italian methane distributors both for the times series of the aggregate number of stations offering CNG and for region-by-region counts for the year 2009.

White pumps. Late entrants in the *Prezzibenzina.it* database are recorded with their contemporaneous brand status. When we are able to pre-date the year in which a station was active using information from alternative sources, this leaves the problem of establishing its status in the years before it appeared in *Prezzibenzina.it*. We exploit the fact that the *Prezzibenzina.it* data contain a "Past brand" variable which we assume represents the status of the station before the current one. We use the information on the past brand to impute the branded status of the station in 2005. We have validated this assumption by checking with paperback guides. For the stations on which we find information in the 2005 guide, there is extremely high correlation between the brand (or the unbranded status) reported in the variable "Past brand" and that recorded in the guide.

B From firm to market level hazards

In our analysis of the speed of the preemption race, we focus on the hazard of entry of a CNG pump in a market rather than modeling the timing of CNG introduction by individual filling stations. This choice significantly simplifies the analysis. Predicting when a particular filling station may decide to install a CNG pump requires delving into the strategic considerations characterizing between stations competition; whereas the timing of introduction of a new CNG pump by any firm in the market can be modeled abstracting from that. To understand the implication of our approach, we propose below a simple model of station level hazard of introducing CNG to the market and show the assumptions under which it maps in the market level framework we estimate.

Let us start from a competing risk model where the failure (i.e. the introduction of a

[21]Main sources: "GPL & Metano. Atlante Stradale d'Italia" (ITER) editions 2005, 2008, 2009, and 2012); "Guida Metano. Atlante Stradale d'Italia" (Belletti) editions 2007 and 2010); "Guida GPL & Metano per auto" (Egm) edition 2006.

CNG pump in the market) can be due to the decision of any of K gas stations active in market m. The hazard for station k is:

$$\lambda_k(X_{km}) = \lim_{dt \to 0} \frac{P(t \le T \le t + dt, k \text{ enters}|T \ge t, X_{km})}{dt}$$

The overall hazard rate for observing any entry in the market it

$$\lambda(X) = \sum_{k=1}^{K} \lambda_k(X_{km})$$

where we have assumed for simplicity that there are no ties (i.e. no multiple stations installing a CNG pump in the same market in the same year).

In this formulation, the covariates shifting the hazard are both market and station specific. If we assumed that firms are homogeneous and station specific characteristics do not affect the hazard of entry, the station level hazard would be the same for all the stations in the same market and the competing risk model would reduce to the duration model we estimate where there is only one type of risk: the entry by any of the K stations in the market.

The assumption of homogeneity of the potential entrants follows into the footsteps of the early entry models (Bresnahan and Reiss (1991b,a)) which inferred firm profitability of small scale retail businesses abstracting from the geographical dimension and other firm specific profit shifters. This simplifying assumption is certainly not met exactly in our data: we document ourselves and exploit for identification the fact that unbranded gas stations are more likely to install a CNG pump. However, even if station specific characteristics exist that shift the propensity to enter the market, the estimates of our parameters of interest would not be affected as long as these characteristics are not systematically correlated with the number of potential entrants in a market. Namely, the impact of potential competition on the entry rate we estimate will still be the correct one if the share of "white pumps" in a market is not related to the strength of its potential competition.

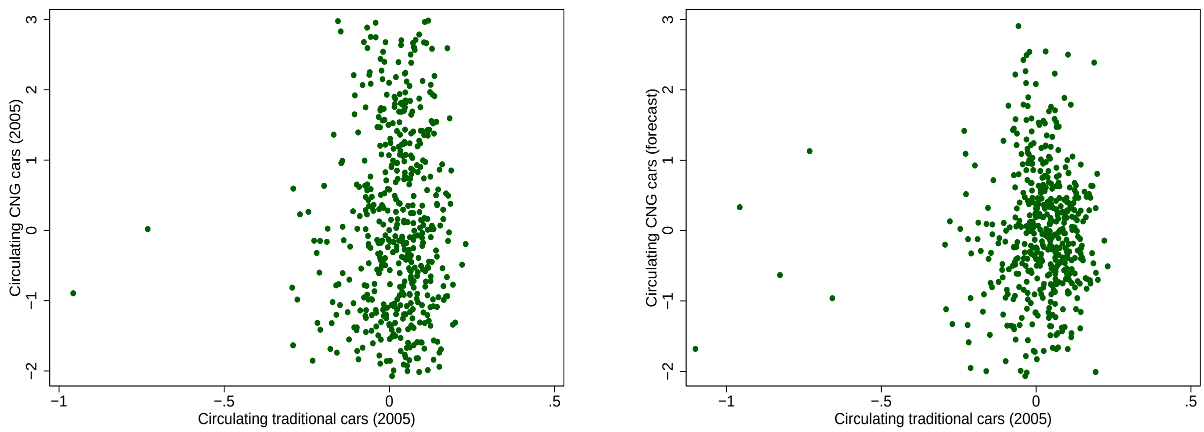
C Other tables and graphs

C.1 Correlation of gasoline and CNG demand

Figure C.1 displays the correlation between demand for traditional and natural gas fueled cars, providing some evidence in support of the main identification assumption behind the results in Table 2. In the figure on the left, we plot the residuals from a regression of the logarithm of the total number of cars running on gasoline and diesel in an LMA in 2015 on market population and income against the residual of an analogous regression whose dependent variable is instead the logarithm of the stock of vehicles running on

CNG in the LMA in 2015. The correlation between the two is low: CNG vehicles are differently popular in markets with a comparable number of registered gasoline\diesel cars per capita. Since station managers should base decisions on expected demand when the market is fully developed, which may not be the case in 2005, we repeat the exercise using a measure of long run demand for CNG car, instead of the current one. Long term demand for CNG cars is constructed using demand estimates from Pavan (2015) to obtain a fitted value for methane cars if the filling station infrastructure for CNG were as developed as that for Liquified Petroleum Gas (LPG). We infer from it that unobservables driving profitability of the CNG market are not perfectly predicted by demand for traditional fuels.

Figure C.1: Demand for CNG and traditional fuels



Notes: Each dot in the plots represents a Local Market Area. The figure on the left displays on the x-axis the residuals of a regression of the logarithm of the number of registered diesel and gasoline powered cars in the LMA in 2005 on the log of Population in the LMA and on the log of the average income in the LMA. The y-axis displays the residuals of a regression of the logarithm of the number of registered CNG powered cars in the LMA in 2005 on the log of Population in the LMA and on the log of the average income in the LMA. The figure on the right has the same variable on the x-axis but it reports instead on the y-axis the predicted long-run number of CNG powered cars in the LMA obtained using the demand model in Pavan (2015). Observations below the 1st percentile and above the 99th percentile are excluded.

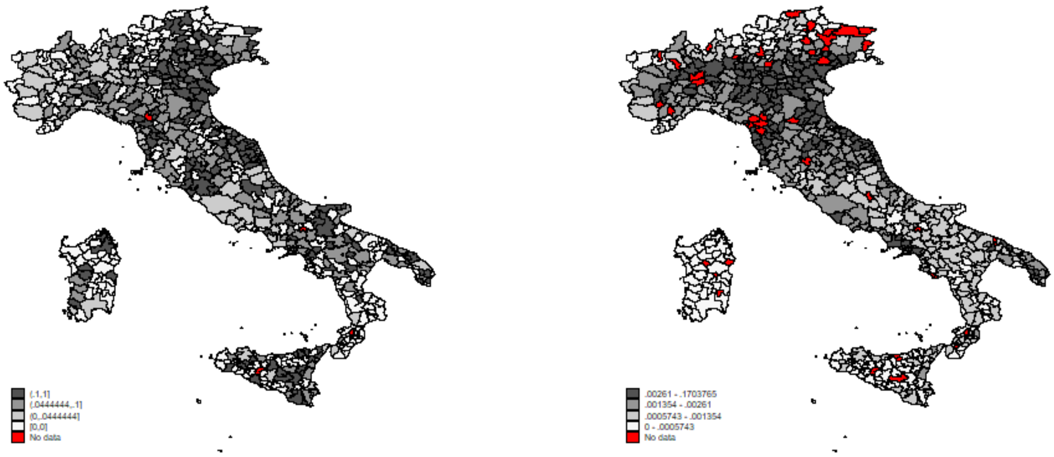
C.2 Correlation of white pump location and CNG demand

The identification strategy exploiting variation in the share of white pumps among a station's competitors hinges on the assumption that the prevalence of unbranded establishments in a particular market is uncorrelated with the profitability of CNG. Here we present some evidence that the correlation between the share of white pumps and early CNG adoption is not due to sorting of independent stations into markets where the prospects for CNG distributors are better.

In Figure C.2 we contrast two maps of Italy: one (on the left) distinguishing LMAs by the quartile of the distribution of the share of white pumps to which they belong (with darker shaded areas implying higher shares); the other (on the right) doing the same for the long run market share of CNG cars, obtained as we just described in section C.1. It is immediate that the correlation between the two is far from perfect. Long run demand for CNG is strong almost exclusively in the northern areas, whereas there are markets with high prevalence of independent pumps also in the South and in the Islands. In Figure

Figure C.2: LOCATION OF WHITE PUMPS AND CNG PROFITABILITY

Share of white pumps | *Long run market share of CNG cars*



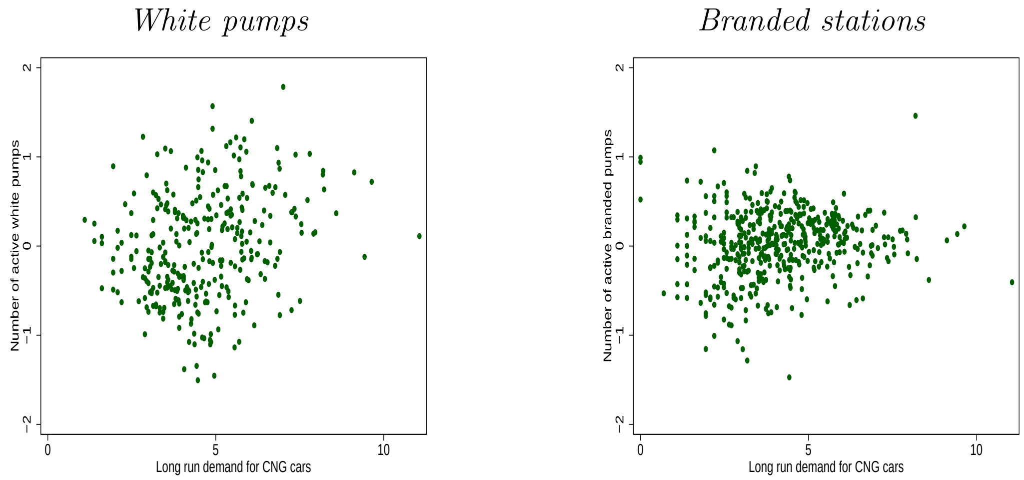
Notes: The map on the left displays LMAs belonging to different quartiles of the distribution of the share of white pumps in the market. The map on the right displays LMAs belonging to different quartiles of the distribution of the long run demand for CNG cars, constructed using estimates from the demand model in Pavan (2015).

C.3, we provide more formal evidence that unbranded stations do not seem to locate in areas providing a profitability advantage for CNG comercialization. In the plot on the left, we show the correlation between long run demand for CNG and the residuals from a regression of the logarithm of the number of white pumps on LMA population. In the plot on the right, we do the same but use the number of branded pumps in the LMA as dependent variable of the regression. In both cases, we fail to detect a sistematic correlation between CNG profitability and the location of the filling stations, no matter their branded status. This support our assumption that decision on locations were taken at time of entry without station managers anticipating that CNG distribution would become an option.

C.3 Robustness

Table C.1 replicates Table 2 constructing the number of potential competitors using the stations that had entered the *Prezzibenzina.it* database by 2005. We could not replicate the specification in columns (2) of Table 2 since the number of potential competitors according to the 2005 data is lower and the first two quintiles of the distribution of the number of competitors are identical and equal to 0. Table C.2 also replicates the

Figure C.3: Correlation between branded status and CNG profitability



Notes: Each dot in the plots represents a Local Market Area. The figure on the left displays on the x-axis the predicted long-run number of CNG powered cars in the LMA obtained using the demand model in Pavan (2015) and on the y-axis a the residuals of a regression of the logarithm of the number of white pumps in the LMA in 2009 on the logarithm of Population in the LMA. The figure on the right is analogous except that the dependent variable of the regression is the logarithm of the number of branded pumps in the LMA in 2009. Observations below the 1st percentile and above the 99th percentile are excluded.

baseline results in Table 2 excluding from the sample all the large markets (i.e., LMAs with population above 800,000).

Table C.1: EFFECT OF POTENTIAL COMPETITION ON PREEMPTION. STATIONS ACTIVE BY 2005

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\mathbb{1}$ {2nd Tercile} Potential Entrants	1.722***		1.152	1.660***	1.639***	2.095***	1.714***
	(0.312)		(0.192)	(0.305)	(0.264)	(0.365)	(0.286)
$\mathbb{1}$ {3rd Tercile} Potential Entrants	3.150***		1.300	3.114***	2.550***	4.299***	2.980***
	(0.616)		(0.236)	(0.611)	(0.445)	(0.745)	(0.540)
Potential Entrants		1.009*					
		(0.005)					
(Potential Entrants)^2		1.000***					
		(0.000)					
Long Run CNG Demand				1.000***			
				(0.000)			
Population (100,000)	1.089***	1.109***			1.073***	1.118***	1.174***
	(0.015)	(0.033)			(0.012)	(0.022)	(0.039)
Log(Gasoline and diesel cars)			2.209***				
			(0.141)				
Average Income (1,000 €)	1.198***	1.282***	0.985	1.298***	1.163***		1.193***
	(0.035)	(0.039)	(0.036)	(0.035)	(0.029)		(0.044)
Observations	1206	1206	1206	1206	1206	893	1206
LMAs	611	611	611	611	611	611	611

Notes: The table reports the hazard ratios from a Cox proportional hazard model. The controls for potential competition are the terciles of the distribution of the number of potential entrants in the market in 2005, with the bottom tercile as the excluded group. We consider as potential entrants all the diesel\gasoline filling stations active in the market. In Column (2) we make a functional form assumption making the hazard rate depende on a quadratic function of the number of potential competitors. Columns (3) uses the number of circulating diesel and gasoline cars in 2005 as a controls for market size. In Column (4) market size is proxied by fitted demand for CNG cars if the density of CNG filling stations were the same as that of LPG filling stations. This figure is obtained using demand estimates in Pavan (2015). In Column (5), we allow markets to have different baseline hazards depending on the number of CNG entry events they experienced (0 CNG incumbents, 1 CNG incumbent, 2 CNG incumbents, etc.) and in column (6) we only include LMAs until they experience the first failure (i.e., entry of the first station serving CNG). Column (7) reports results of a shared frailty model at LMA level. All the specifications include macro areas (Northwest, Northeast, Center, South and Islands) fixed effects. The robust standard errors are clustered at the LMA level (LMAs can appear multiple times if they experience entry by more than one filling station). The p-values of the coefficients are reported in parentheses. ***: p<.01, **: p<.05, *: p<.1.

Table C.2: Effect of potential competition on preemption. No large LMAs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}$ {2nd Tercile} Potential Entrants	5.005*** (0.000)			2.720*** (0.000)	5.057*** (0.000)	4.869*** (0.000)	9.357*** (0.000)	5.022*** (0.000)
$\mathbb{1}$ {3rd Tercile} Potential Entrants	9.233*** (0.000)			3.089*** (0.000)	10.950*** (0.000)	8.328*** (0.000)	12.222*** (0.000)	8.998*** (0.000)
$\mathbb{1}$ {2nd Quintile} Potential Entrants		7.748*** (0.000)						
$\mathbb{1}$ {3rd Quintile} Potential Entrants		12.937*** (0.000)						
$\mathbb{1}$ {4th Quintile} Potential Entrants		19.591*** (0.000)						
$\mathbb{1}$ {5th Quintile} Potential Entrants		30.886*** (0.000)						
Num Potential Entrants			1.040*** (0.000)					
(Num Potential Entrants)^2			1.000*** (0.000)					
Long Run CNG Demand					1.000*** (0.000)			
Population (100,000)	1.317*** (0.000)	1.277*** (0.000)	1.265* (0.088)			1.263*** (0.000)	1.273* (0.093)	1.387*** (0.000)
Log(Gasoline and diesel cars)				2.201*** (0.000)				
Average Income (1,000 €)	1.032 (0.374)	0.999 (0.974)	1.029 (0.414)	0.985 (0.655)	1.121*** (0.001)	1.024 (0.426)	0.966 (0.637)	1.020 (0.593)
Observations	1148	1148	1148	1148	1148	1148	479	1148
LMAs	604	604	604	604	604	604	404	604

Notes: The table reports the hazard ratios from a Cox proportional hazard model. We refine the main sample to exclude large LMAS (population above 800,000). The controls for potential competition are the terciles of the distribution of the number of potential entrants in the market in 2009, with the bottom tercile as the excluded group. We consider as potential entrants all the diesel\gasoline filling stations active in the market. In column (2) we measure differences in potential competition more finely using the quintiles of the distribution of the number of potential entrants and in column (3) we make a parametric assumption on the effect of competition assuming a quadratic relationship. Columns (4) uses the number of circulating diesel and gasoline cars in 2005 as a controls for market size. In Column (5) market size is proxied by fitted demand for CNG cars if the density of CNG filling stations were the same as that of LPG filling stations. This figure is obtained using demand estimates in Pavan (2015). In Column (6), we allow markets to have different baseline hazards depending on the number of CNG entry events they experienced (0 CNG incumbents, 1 CNG incumbent, 2 CNG incumbents, etc.) and in column (7) we only include LMAs until they experience the first failure (i.e., entry of the first station serving CNG). Column (8) reports results of a shared frailty model at LMA level. All the specifications include macro areas (Northwest, Northeast, Center, South and Islands) fixed effects. The robust standard errors are clustered at the LMA level (LMAs can appear multiple times if they experience entry by more than one filling station). The p-values of the coefficients are reported in parentheses. ***: p<.01, **: p<.05, *: p<.1.