The italian electricity market and consumers' information

Sonia Petrini - 983149, Matteo Biglioli - 938199

Abstract

The aim of this paper is to understand what are the main drivers of italian consumers' choice of entering the free electricity market. More than 20 years after the beginning of the liberalization process, in italy still coexist a free and a regulated retail market for energy, and a considerable portion of domestic users still provisions itself on the latter. We thus estimate the influence of both potential savings and consumer's information on the probability of entering the free market, exploiting a novel panel dataset, and two different definitions of information. Our findings have significant policy implications, as they provide evidence of the final users' insufficient knowledge of the market, which represents a crucial barrier to the complete realization of the liberalization of the italian free market of energy.

1 Introduction

1.1 Liberalization of the european electricity market

In the scope of its integration and harmonization process, in 1996 the European Union laid the grounds for the constitution of a common European energy market, in which competition could flourish. Directive 96/92/CE [1] set the starting point for this objective, to be realized through both a widespread process of liberalization and the unbundling of the value chain.
Hitherto, in the typical MS, generation, transmission, and distribution of such commodities were entrusted to a state-owned company; since the citizens have to always be guaranteed with the provision of energy, the presence of the State was justified by the essential nature of the good.
The liberalization process allowed to dismantle this system of vertically integrated local monopolies, introducing the space for competition where possible, and keeping under regulation those components of the value chain for which the presence of economies of scale explained the existence of natural monopolies. The focus of our main analysis will be the electricity market, in which the transmission and distribution segments are held under regulation, while the upstream and downstream stages of generation and retail have been opened to competition, with all its benefits.
Liberalization is in fact realized in a cost-efficiency perspective: with more players contending their share of the market for the homogeneous good electricity, each of them is motivated to provide the best service at the most convenient tariff, thus fostering innovation and price reduction.

1.2 Italian Experience

In Italy, Directive 96/92/CE was implemented in 1999 through the Dlgs 79/99 [2], the so called *decreto Bersani*: with this legislative decree the new actors of the new free market (FM) were appointed, and the bases for its actual creation were laid. In fact, in order to set up a competitive space the first necessary step was the decommissioning of the plants possessed by Enel, the italian state-owned company that was in charge of the electricity provision to that date[3]. The early 2000s were thus devoted to the redistribution of the incumbent monopolist's productive capacity, with the subsequent creation of new smaller companies. Finally, in 2007 the so-called *Bersani bis* (Law 03/08/07 n. 125 [4]) inaugurated the novel electricity market.

However, while giving the possibility to all end customers to enter the newly born FM, as a temporary provision a portion of such market was kept under "Greater Protection" (GP). The rationale of this duality was that of ensuring a gradual and progressive transition, allowing those domestic and industrial consumers who had not chosen their seller yet to be guaranteed with the supplying of electricity in a safeguarded regime, at economic (prices) and contractual conditions defined by the Authority. In the italian framework such role is embodied by ARERA[1] (Autorità di regolazione per Energia Reti e Ambiente), which revises the economic conditions on a quarterly basis, according to variations in the wholesale market price.

Originally the date for the extinction of the GP regime was set to 2016 for domestic consumers, but after being first delayed to 2018, such event has been continuously postponed over the years. During the past month, a new decree has settled the new date to January 2023 [5], while small firms who were under GP are in an entirely FM since the beginning of this year. In addition to the possibile inequalities arising among domestic and industrial consumers, this scenario raises concerns with regards to the feasibility of an harmonious shift to a completely liberalized energy market.

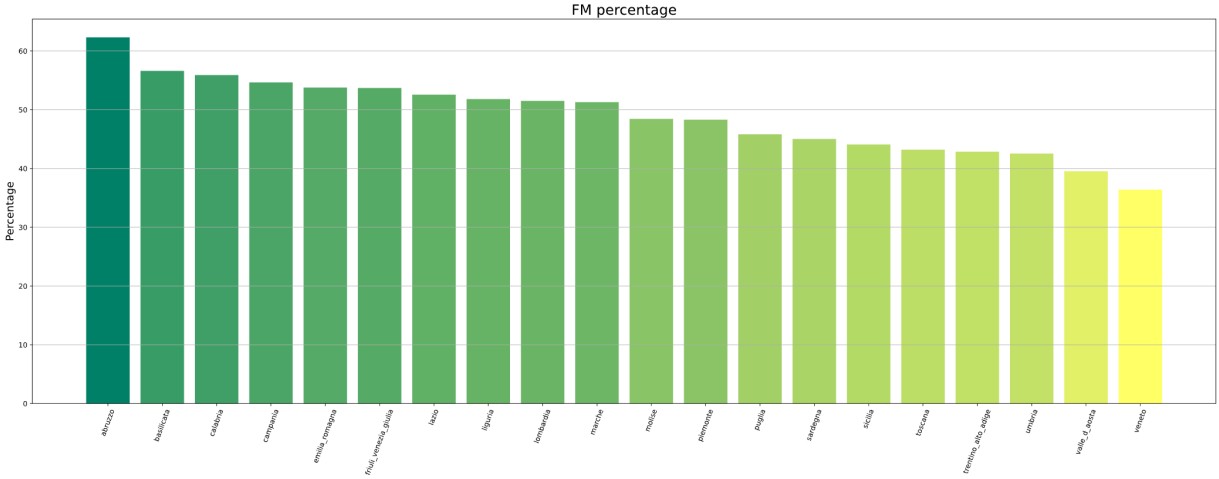

Figure 1: Free Market percentage in 2019.

Across these years the market has been persistently perceived as not ready for this change, both on the supply and on the demand side. Concerning the latter, the most recent data provided by ARERA show that - after more than 20 years since the advent of the european liberalization - 44,68% of domestic pick-up points are served under protected conditions, while this value decreases to 32,4% when considering non domestic users.
A 2019 survey [6], carried out by Research Hub on behalf of ARERA to investigate the state of the liberalization of the energy market in Italy, found that 87,5% of individuals don't know anything at all (48,6%) or know little (38,9%) about the FM for electricity and gas. Moreover, when asked about the usage of comparison websites, 86,4% of the sample declared not to employ them when acquiring information about existing offers; among those who do use such websites, only 7,9% refers to *Portale Offerte*[2], the tool specifically designed by the Authority to guide and orient the final customers in their choice of the supplier.

Thus, among domestic consumers there still seems to be in place a widespread issue of misinformation concerning the energy market, an issue that has not yet been solved since the first postponement in 2015, which was indeed advocated for by sixteen users associations concerned with the possible aggravation of households' energy expenditures [7].
In fact, exacerbating this insufficient information phenomenon, on the supply side the competitive framework sought by the liberalization process does not seem to have been completely realized yet. In such a

[1] https://www.arera.it/it/index.htm
[2] https://www.ilportaleofferte.it/portaleOfferte/

competitive scenario, the free electricity market would be a space with numerous players, each of them able to set a competitive price according to demand and supply dynamics, and to propose a variety of services and custom offers and discounts. Under these conditions, on the FM, end customers would be charged tariffs that are lower than those established by the Authority, in virtue of the benefits derived from the presence of competition. In practice, since 2007 several investigations have shown that the average consumer would spend more in the FM than under GP. In 2011 the price paid by domestic clients was 12,8% higher compared to the benchmark set by ARERA; in 2015, the users associations that called for the first postponement had moved their claims basing on the 20% increase in households' bills they expected from the abolition of the protected regime [7].

The main purpose of this paper will be to investigate the elements upon which the choice of entering the FM depends, in order to understand the direction in which policies could be implemented to favour the pending transition to an entirely liberalized energy market.
Our research inquiry is thus to understand what explains the probability of entering the italian FM for electricity; the focus will be set on domestic users, since they appear to be the most vulnerable category. Given the framework depicted above, our expectation is that information will play a preponderant role in the consumers' decisions.

2 Model and Methodology

To address our question, we will exploit a novel Panel dataset composed of N=20 regions and T=8 years, from 2012 to 2019, which will be extensively described in the next section.
The basic assumption underlying our model is that of a rational consumer facing the choice between remaining in her current market (GP), or entering a new one (FM).
Economic theory requires that the rationality of an agent will lead her to maximize her utility; thus, the choice will depend on the savings she expects to obtain on the new market, and on the level of information she has about such a market and its structure, basing on which she can derive her expectations. Having as dependent variable the share of individuals that choose to enter the FM in each region (r) and year (t), we will run a Fixed Effects (FE) Regression Model based on the following specification:

$$P\_FREE_{rt} = \alpha SAVINGS_{rt} + \beta INFORMATION_{rt} + \gamma CONTROLS_{rt} + \phi_r + \tau_t + \epsilon_{rt} \qquad (1)$$

where $SAVINGS_{rt}$ is a variable accounting for the expected monetary saving that consumers would derive from entering the competitive free market, $INFORMATION_{rt}$ aims at proxying and quantifying the level of customers' information, and r and t are respectively entity and time fixed effects. We include these last two elements to account for the portion of variability in P_FREE_{rt} which is due to those unobserved variables that are fixed across entities or across time. Thus, r will absorb the structural differences between regions that are not controlled for through the regressors, while t will incorporate the shocks that affect all the entities within a given year.

We will run two configurations of this model, exploiting two different definitions of our variable of interest, $INFORMATION_{rt}$, to check the robustness of our results.
First we will adopt a generalized definition of information based on the regional spread of newspapers (variable $NEWSPAPERS_{rt}$). Our expectation is that, by reading local or national news about generic issues, one will have a higher probability of learning about the structure and development of the electricity FM, and thus to take advantage of its benefits.
Second, a definition specific to the economic sector will be employed (variable $ALTROCONSUMO_{rt}$); in this case the variable will be defined through the percentage number of online searches of keywords related to *Altroconsumo*, the most well known Consumers' Association in Italy. Again, we expect that those individuals visiting the website of this Association will gain a relevant amount of information about the functioning of the FM, and will thus be more prone to enter it.

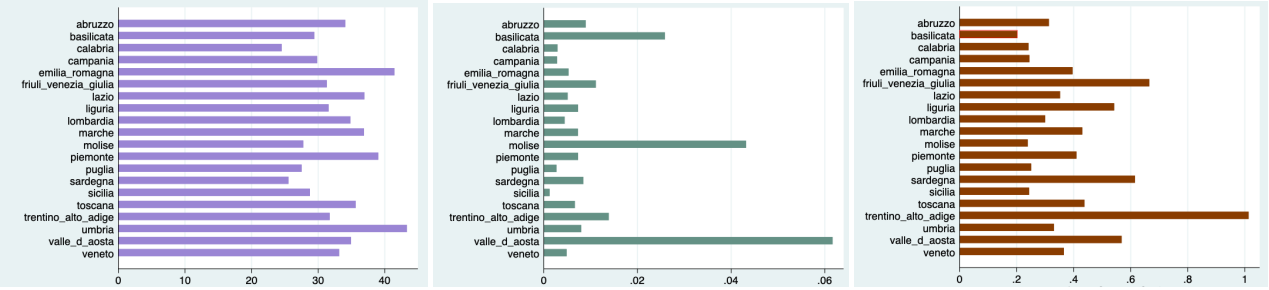

Figure 2: Entity averages of P_FREE, ALTROCONSUMO and NEWSPAPERS.

In addition to providing a robustness check, this double specification allows to address a possible selection bias. An investigation on the importance of local level information in Italy, performed by AGCM[3] (Autorità Garante della Concorrenza e del Mercato) in 2017, reveals the preferred means of information by region [8]. The newspaper is among the first five positions in each region, but we cannot exclude a correlation between the regions in which it holds the first position and those displaying highest shares of domestic users in the FM. This could inflate the coefficient on information defined as average newspaper spread, without however completely overturning the results. For this reason, to confirm our feelings and test the robustness of our results, we also run the second specification of the regression.

To deal with confounding, a set of demographic controls is also included: we recognize that features such as education, mean age, and population density may affect both the level of information within a region and the probability of entering the FM, thus we add them to our regression specification in order to avoid endogeneity problems.
Moreover, to heed the possible barrier to entry represented by internet access, we also control for the share of families with internet connection; the easiest and more practical way to access and compare the FM offers is in fact through online portals and sites, and virtually all of them can now be activated in the same way.

For what concerns $SAVINGS_{rt}$, we do not expect to find a significant coefficient on it; as already anticipated, the actual convenience of the FM is not all that clear and consistent. We get a clearer picture on this matter from the Price Index constructed by Selectra[9], a web company specialized in offers' comparison and consumers' protection. This Index, which estimates the mean price proposed on the FM, appears to have been on average in line with the regulatory benchmark price in 2020, alternating periods of positive and negative differentials. However, when giving a closer look, Selectra finds that the prices offered on the FM display a great range of variation: hence, if the customer is well informed and is able to identify the best available offers she will actually be able to save money (34.88 €/MWh for the 20% best ones, compared to 52.55 €/MWh paid for the average offer).
Unfortunately, this also holds for the reverse: if a customer enters the FM but she does not have the tools to understand it properly, she will be likely to fall victim of aggressive commercial behaviour, and to subscribe to an offer which is much more expensive than the GP price (68.16 €/MWh for the 20% worst offers in 2020).

The offers exploited in our analysis lead to mean prices that are consistent with this scenario, as shown in Figure 3, considering that they are taken from a previous period in time:

[3]https://www.agcm.it/

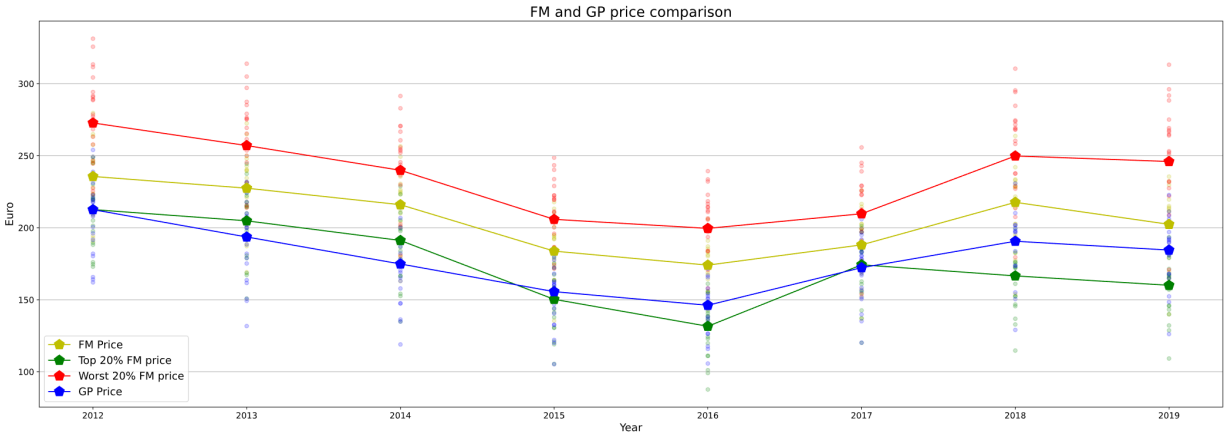

Figure 3: FM and GP price comparison

The positive differential between FM and GP prices was also detected in ARERA's 2020 Annual Report [10]. The explanation suggested by the Authority is based on the different types of tariffs available on the FM, which are based on fixed or variable prices, and on the balancement of procurement costs with discounts on the raw material, or additional non-monetary benefits related to services. The total benefit of these price components is difficult not only to quantify, but also to compare across offers, especially considering that the above mentioned survey by Research Hub learned that 71% of the surveyed population could not distinguish the basic cost components of the bill.

This brings under consideration the matter of unobservable variables: if we could quantify all such discounts, cross-selling promotions, maintenance and assistance services, and guarantees that are absent on the GP, we would certainly obtain a greater saving rate on the FM, even when considering the average offer. Nevertheless, a handful of surveys and investigations from the beginning of the liberalization to more recent years, depicts an energy consumer which is not able to effectively and entirely understand her bill, and move fluently across the ever increasing number of offers. Thus, we believe that there are very few chances that, even if we managed to account for such unobservables, this would make any difference in our results, as they are strictly related to information, and as such they contribute to the reinforcement of our hypothesis.

For all these reasons, we have chosen to compute the savings variable considering only the top 20% offers available on the FM (variable $BEST\_SAVINGS_{rt}$). This will help us understand if, when the consumer can actually expect a cost-saving from entering the free market, such saving is a significant explanatory factor in her choice to leave the GP, or information is still more relevant. By using this subset of offers we can also in part proxy for the unobservables described above.
In conclusion, the models that we are testing are the followings:

$$P\_FREE_{rt} = \alpha BEST\_SAVINGS_{rt} + \beta NEWSPAPERS_{rt} + \gamma CONTROLS_{rt} + \phi_r + \tau_t + \epsilon_r t \quad (2)$$

$$P\_FREE_{rt} = \alpha BEST\_SAVINGS_{rt} + \beta ALTROCONSUMO_{rt} + \gamma CONTROLS_{rt} + \phi_r + \tau_t + \epsilon_r t \quad (3)$$

3 Data

For the purpose of our analysis, we collected a novel observational panel dataset for the period 2012-2019, over a cross-section of 20 regions (for a total of 160 observations). All variables are gathered on a yearly basis and grouped by region.
Since this paper addresses the share of the market populated by domestic consumers, every dataset in this paper concerns only that type of user (i.e. when we refer to "total", we mean "total domestic").

Our primary data source are the relations posted annually by ARERA[4], which provide a deep data insight on the status of services in the italian energy market. From annual relations 2013 to 2020 we extracted two main kind of datasets:

1. Volumes and Pick-up-Points of both the total market and the GP segment.
2. Market composition, more precisely the number of operators and the C3 index.

We furtherly processed the data related to the former topic to compute the variables that will be used in the regression described in Section 5.

3.1 P_FREE

This is the dependent variable of our regression, defined as the probability of entering the FM of a domestic consumer in region r and in year t. It is computed as percentage of Pick-up-Points in the FM over the total Pick-up-Points.

3.2 SAVINGS

This variable is constructed as the difference in the expected spending on the GP market and the FM (i.e. the savings expected when entering the free market). To build this variable, we gathered data concerning the average consumption per region, and the average prices offered on both markets.

3.2.1 Consumption

med_tut_cons is the mean yearly consumption for the typical consumer in the GP market, computed as the ratio between the volumes and the number of Pick-up-Point of each region and year.

3.2.2 Prices

1. Prices in the GP market are provided by ARERA[5] and are revised on a quarterly basis.
2. Prices in the FM should be provided in the transparency section of Portale Offerte[6], that is the official website owned by ARERA to help consumers navigate offers in the FM. The problem was that the archive only dates back to 2019, and does not report any offer before that year. To overcome this limitation we scraped a website, SOSTariffe[7], which provides all offers historically available on the FM. Finally, to compute the FM price we took the averages of all the offers *avaiable* across the considered years, where we assumed that an offer can be attributed *avaiable* to a certain year if its availability overlaps the year for at least 60 days.

It should be noted that the website we scraped does not report the starting date of any offer but only its ending date; to overcome this constraint we computed the average duration for all the offers stored in Portale Offerte - which is 9 months - and we applied this value to each of the SOSTariffe offers.

We know that prices are not homogeneous through all consumers, but differ based on multiple variable factors. Because of the scope of this paper, which encompasses the ordinary consumers, we set those variables to their typical values as follows:

- Peak Power: 3 kW
- Consumption distribution: 40% day - 60% night and weekend
- The household is the residency of the holder of the bill.

[4]https://www.arera.it/it/relaz_ann/20/20.htm
[5]https://www.arera.it/it/ dati/condec.htm
[6]https://www.ilportaleofferte.it/portaleOfferte/
[7]https://www.sostariffe.it/energia-elettrica/archivio/

From ARERA's relations, we also learned that the only component of the bill that can vary through different operators is the one related to procurement costs, i.e. the actual "raw material" electricity, which accounts for 40% of the bill on average. The remaining cost components, which are constituted by transportation and system charges, and taxes levied by the State, are homogeneous across all consumers in both markets as a general rule. For this reason we only took into account the "raw material" component and the basic discounts offered by operators of the free market, at the same time excluding those offers which promised a null cost on such headings, while charging an extra price on the procurement side.

3.3 Variable Construction

We then computed the actual yearly expected expenditure for consumers in different regions using the average consumption computed at the previous step. The main complexity encountered in this calculation regards the multiple components in which the consumers bill is splitted in: both markets provides multiple ranges of consumption with different tariffs that must be taken into account.
E.g. for a yearly household consumption (in 2012) of 3000 kWh the bill will be computed as follow:

$$\begin{aligned}
\text{First range:} \quad & 1800 \text{ kWh} \times 0.1001865 \text{ €/kWh} = 180.33 \text{ €} \\
\text{Second range:} \quad & (2640 - 1800) \text{ kWh} \times 0.1038365 \text{ €/kWh} = 87.22 \text{ €} \\
\text{Third range:} \quad & (3000 - 2640) \text{ kWh} \times 0.1077965 \text{ €/kWh} = 38.81 \text{ €} \\
& \text{Fixed cost:} \quad 16.47 \text{ €}
\end{aligned}$$

$$\text{Total: } 180.33 \text{ €} + 87.22 \text{ €} + 38.81 \text{ €} + 16.47 \text{ €} = 322.83 \text{ €} \tag{4}$$

We furtherly processed the dataset to compute the savings-related variables that will be used in the regression described in Section 5:

1. **savings_avg** is the average value that a consumer could save by accepting an offer on the FM, compared to the regulatory benchmark.

2. **savings_best_20_perc** is the average value that a consumer in the GP market would have saved if she could access the best 20% FM offers.

3.4 INFORMATION

We then computed the variables related to consumers' **information** using two different kind of data:

1. **diff_media** NEWSPAPERS is the average yearly spread of newspapers in each region. This variable is based on the dataset provided by ADS[8], which collects data regarding the number of copies sold by the main newspapers in every regions, and is computed as the ratio between the average number of copies sold by all relevant newspapers in a given region and its population. It should be noted that to assess the relevance of the newspapers, we excluded from the dataset all newspapers that sold a number of copies less than 0.05% of the region's population.

2. **n_search_altroconsumo_perc** ALTROCONSUMO is the percentage number of Google searches for keywords related to *Altroconsumo*. This variable, based on the dataset provided by Google Trends[9], is computed as the ratio between the number of Google searches for keywords related to *Altroconsumo* in a given region and its population.

3.5 CONFOUNDERS

Finally, we collected different control variables to address **endogeneity problems** in our regression. We selected different regional indexes provided by Istat[10] such as:

[8] www.adsnotizie.it
[9] trends.google.com
[10] https://www.istat.it/it/archivio/16777

- **total population** by region.
- **scholarization index** by region, computed as the ratio between the number of people with at least a high school degree in a given region and its population.
- **internet access index** computed as the percentage number of families that declare to have internet access.
- **mean age** by region.

All of the variables in our panel are summarized in Table 1.

Name	Description	Source
region	region name	DELETE
year	year	DELETE
perc_tut	percentage of Pick-up-Points in the Greater Protection market	Annual Relation ARERA
perc_lib	percentage of Pick-up-Points in the Free market	Annual Relation ARERA
pdp_dom_tot	total number of domestic Pick-uop-Points	Annual Relation ARERA
pdp_dom_tut	number of domestic Pick-uop-Points in the Greater Protection market	Annual Relation ARERA
vol_dom_tut	total consumption of consumers in the Greater Protection market	Annual Relation ARERA
cons_medio_dom_tut	average consumption of consumers in the Greater Protection market	Annual Relation ARERA
prezzo_mercato_libero_sos_tariffe_60_gg	average price in the Free market	SOStariffe
prezzo_mercato_libero_sos_tariffe_best_20_perc	average price of the best 20% offers the Free market	SOStariffe
prezzo_mercato_libero_sos_tariffe_worst_20_perc	average price of the worst 20% offers the Free market	SOStariffe
prezzo_mercato_tutelato	price in the Greater Protection market	ARERA
diffusione_media	average spread of newspapers	ADS
n_search_altroconsumo	number of Google searches for keywords related to *Altroconsumo*	Google Trends
perc_internet	percentual number of families that declare to have an internet access	ISTAT
population	population	ISTAT
region_area	region surface area	ISTAT
pil	region pil	ISTAT
n_operators	number of operators in the Free Market	Annual Relation ARERA
c3	C3 index in the Free Market	Annual Relation ARERA
edu_diploma_4_5_a	number of people with at least an advanced high school degree	ISTAT
edu_diploma_4_5_a_perc	percentage of people with at least an advanced high school degree	ISTAT
eta_media	average age	ISTAT
switching_perc	DELETE	DELETE
dens_pop	DELETE	DELETE

Table 1: Panel variables.

4 Descriptive Statistics

In the following section we briefly comment the descriptive statistics of our panel's variables, represented in Figure 4.

	P_FREE	price_GP	mean_price_FM	best_price_FM	worst_price_FM	NEWSPAPERS	perc_edu_diploma_4_5_y	mean_age
count	160.000000	160.000000	160.000000	160.000000	160.000000	160.000000	160.000000	160.0000
mean	32.904095	174.877490	205.635250	172.345687	239.619750	0.408072	19.315889	45.2750
std	10.749550	30.451233	31.997389	34.155360	38.758526	0.202671	1.919695	1.6226
min	13.714286	105.260484	131.700000	87.700000	153.850000	0.154700	14.727876	40.9000
25%	24.552518	152.604218	183.515000	146.620000	212.510000	0.265325	18.064969	44.2000
50%	31.473748	175.280251	202.825000	172.225000	237.675000	0.356250	19.241009	45.4000
75%	40.749730	196.850436	229.297500	197.752500	268.625000	0.488950	20.068977	46.3250
max	62.322275	254.025607	279.660000	249.050000	331.300000	1.081200	24.058334	49.2000

Figure 4: Descriprive statistics.

As discussed before, the share of population in the FM is extremely low, as suggested by the value related to the 75% quantile (40.75%).

The statistics related to the different prices confirm the finding shown in Figure 3, showing prices in the FM to better the GP tariffs only in the top 20% offers, and, even in that case, by an insignificant margin.

$NEWSPAPERS_{rt}$ and $perc\_edu\_diploma\_4\_5\_y$ present a not so comforting situation, with at most a 1.08% newspapers' spread (meaning that the average copies sold account for 1% of the region's population) and a 24.06% education index (meaning that only 24% of the region's populaton holds at least an high school degree).

5 Empirical Analysis

To perform our analysis, we first exploit a Hausman test to verify whether the FE Model is indeed the right choice, or a Random Effects Model would be more suitable; we can reject the hypothesis that the difference in the regression coefficients are not systematic, thus confirming our Fixed Effects choice.
Then, we include Entity and Time fixed effects and test their joint significance separately, obtaining a p -value of 0.000 for both; we hence introduce them in the model.
Since we are dealing with panel data, we need to be aware of the possible issues that can arise with the OLS assumptions for what concerns the variance of the error term and the cross-sectional correlation. Moreover, given that the sample is not large, we cannot rely on tests based on asymptotic properties such as the Breusch Pagan test.
For this reason, we test for heteroskedasticity by running our FE model, saving the residuals, and then regressing the residuals' squares over both the predictors multiplied by their coefficient and their squares. The joint significance of such regressors and a high R^2 (0.4649 for $NEWSPAPERS_{rt}$ and 0.3910 for $ALTROCONSUMO_{rt}$) imply that the main regression's independent variables help explain the variance of the residuals (proxied by the residuals' squares), thus heteroskedasticity can bias our estimations, and we have to take it into account.
To verify the presence of correlation within the cross-section, we again run our FE model and perform a Pesaran test, with the Null Hypothesis being cross-sectional independence; we cannot reject the Null in either model configuration (p -value = 0.3070 for $NEWSPAPERS_{rt}$ and p -value = 0.4854 for $ALTROCONSUMO_{rt}$).
We thus conclude that we don't need a clustered variance estimator since there is no significant dependence between regions, but we will use the (robust) specification to obtain heteroskedasticity-robust standard errors and inferences.
The significance of each variable will be evaluated at a 95% confidence level.

5.1 Information as Newspapers' spread

We thus run our first model design, with the general definition of Consumer's Information.

P_FREE	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
BEST_SAVINGS	.1271363	.0716573	1.77	0.092	-.0228442	.2771168
NEWSPAPERS	10.0816	4.407407	2.29	0.034	.8567924	19.30641
perc_internet	-.1021206	.1076947	-0.95	0.355	-.3275282	.123287
dens_pop	-.1421658	.1255019	-1.13	0.271	-.4048443	.1205128
mean_age	-4.861637	2.301819	-2.11	0.048	-9.6794	-.0438749
perc_edu_diploma_4_5_y	-.9863871	.5252684	-1.88	0.076	-2.085786	.1130122
TFE2	4.619404	1.505889	3.07	0.006	1.467542	7.771266
TFE3	9.534331	2.388197	3.99	0.001	4.535777	14.53288
TFE4	17.31153	2.382441	7.27	0.000	12.32503	22.29804
TFE5	23.05858	3.135035	7.36	0.000	16.49688	29.62028
TFE6	27.66838	3.775748	7.33	0.000	19.76565	35.57112
TFE7	38.68238	4.916926	7.87	0.000	28.39114	48.97363
TFE8	46.19044	5.381453	8.58	0.000	34.92693	57.45395
_cons	280.1457	107.998	2.59	0.018	54.10325	506.1882
sigma_u	19.156979					
sigma_e	1.8506299					
rho	.99075406	(fraction of variance due to u_i)				

Figure 5: Regression output.

5.1.1 BEST_SAVINGS$_{rt}$

As foreseen, even when considering the best available FM offers, the expected savings do not play a significant role in explaining a consumer's probability to enter the free market. Taking into account the considerations made in the previous sections, there are some intersecting explanations for this:

- **Inconsistency in the sign of savings**: our analysis, in alignment with the above-mentioned market investigations, suggests that the competitive market is actually not ready yet to provide consistently lower tariffs with respect to the GP benchmark. Moreover, when this differential is indeed positive, it amounts on average to 2,53 euro; the entity of such savings does not seem to be large enough to justify the transition to the new market.
- **Consumers' difficulty in actually grasping the potential savings**: even when the prices set on the FM are actually lower than those on GP, users are not generally able to identify and exploit this saving opportunity. This derives from the joint effect of a lack of awareness with respect to the bill's cost components, and the complexity of the actual total price computation and comparison, given by the variety of discounts and services offered on the FM.

Figure 6: Average and best savings.

5.1.2 NEWSPAPERS$_{rt}$

Focusing the attention on our first definition of the Information variable we can observe that it displays a significant coefficient and a positive effect. Thus, our hypothesis about the overriding role of information was in fact correct; once again, we interpret this result as the interaction of a plurality of effects:

- **More informed customers are more likely to enter the free market**: by reading newspapers, customers learn about any sort of issues, thus they are also likely to acquire information regarding the electricity market, its regulation, mechanisms, and the opportunities offered by competition.
- **People who read newspapers look for more information i.e. are more proactive**: if newspapers are widely spread in a region, its inhabitants are likely to be curious and prone to look for Information, thus they probably spend more time in comparing offers and in searching the more suitable for their consumption. This results in a higher probability of both entering the FM, and actually finding a money-saving offer; if this is true, they also have a higher probability of retaining in such a market.

In addition, the magnitude of the Information effect is quite noticeable: a 1% increase in a region newspapers' spread would lead to a 10% increase in the probability of entering the FM. Still, this is a reasonable influence, as the average spread is 0,41%, and the datapoint displaying the highest value is Trentino Alto Adige in 2015, with a newspaper's spread equal to 1,08%; a percentage point increase is thus a significant change, which explains the large effect on the dependent variable.

5.1.3 CONFOUNDERS

As mentioned in the previous section, to account for possible endogeneity problems we include a set of confounders, and we plot the correlation matrix to check whether they show any particularly strong correlation. Our original set includes: internet access, population density, GDP (pil), mean age, and education. We are concerned that there may be a relevant association between population density and GDP, and between age and education.

	perc_i~t	pil	dens_pop	mean_age	perc_e~y
perc_inter~t	1.0000				
pil	0.2602	1.0000			
dens_pop	0.1007	0.7608	1.0000		
mean_age	0.4050	-0.1511	-0.2436	1.0000	
perc_edu_d~y	0.0449	-0.0793	-0.0284	0.3370	1.0000

Figure 7: Correlation matrix.

We do indeed find a 76% correlation within the first variables couple, and we thus decide to only retain population density to avoid collinearity. On the other hand, the correlation between age and education is not high enough to cause issues, therefore we opt for keeping both variables as they account for a different sort of confounding information.
In our regression output, mean age is the only control displaying a significant coefficient; however, not including the remaining confounding variables would result in an inflated coefficient on $NEWSPAPERS_{rt}$, as it would absorb the effect of such omitted factors.
For what concerns the effect of the mean regional age, a 1 year increase would lead to a 0,05% decrease in the probability of entering the FM; the sign of this influence is coherent with our expectations, as older people are less likely to both have an internet access and have the skills to exploit it. Since the more effective and rapid way of discovering and comparing offers on the FM are websites, this aspect plays a significant role in explaining the variability in P_FREE_{rt}.

5.2 Robustness Check

In order to verify that our main analysis results are not driven by the way in which we defined our variable of interest, as a robustness check we run the second specification of our model, considering the number of online searches for *Altroconsumo* over the population.
In this way, we contemplate two different aspects:

1. **an alternative information tool**: web searches rather than newspapers
2. **a specific type of information**: such variable accounts for the consumers' interest in market dynamics and in the services related to their tutelage

We introduce the same set of confounders and we make a comparison between the two models.

Variable	NEWSPAPERS	ALTROCONSUMO
BEST_SAVINGS	-.12713628*	-.11308479
NEWSPAPERS	10.081602**	
perc_inter~t	-.10212058	-.0952968
dens_pop	-.14216577	-.12831195
mean_age	-4.8616373**	-5.1213562*
perc_edu_d~y	-.98638712*	-.8177204
TFE2	4.619404***	4.3083352***
TFE3	9.534331***	8.8757921***
TFE4	17.311532***	16.152568***
TFE5	23.058579***	21.645726***
TFE6	27.668384***	26.519424***
TFE7	38.682385***	36.809914***
TFE8	46.19044***	44.19433***
ALTROCONSUMO		11.53517**
_cons	280.14572**	290.67333**
N	160	160
r2_a	.96718876	.96610553
rmse	1.7260168	1.7542766
rho	.99075406	.98865521
corr	-.8420886	-.81288793

legend: * p<.1; ** p<.05; *** p<.01

Figure 8: Newspapers and altroconsumo regression output.

This alternative regression design leads to the same results: $BEST\_SAVINGS_{rt}$ is not significant, while the Information variable remains at a 5% significance. Once again, the more the inhabitants of a region search for online market-related information, the more likely they will be to leave the protected regime to enter the FM.
Noticeably, the two variables $NEWSPAPERS_{rt}$ and $ALTROCONSUMO_{rt}$ display an effect of comparable size (10,1% and 11,5%) on the dependent variable; however, we have to consider that the mean value of ALTROCONSUMO is 0.012%, with a maximum of 0.14%. As such, a percentage point increase would mean a radical change in the region's habits, which would justify the great increment in P_FREE_{rt}.
The confounder for mean regional age is still significant, but at the higher 10% threshold.

We can gain further knowledge about both models by looking at their descriptive parameters, which are remarkably similar:

- **R^2 adjusted (r2_a)**, despite not being the most reliable indicator of goodness of fit, is considerably high: this suggests that our basic assumption about the consumer's rationality is confirmed, as our savings and information regressors (plus their confounders) explain almost 97% of the variability in the dependent variable.
- **Root Mean Square Error (rmse)** is sufficiently low, suggesting a good fit for both models.
- **Rho**, the *intraclass variability*, is very high: this statistics represents the share of variability in P_FREE_{rt} which is due to unobserved heterogeneity across panels; this means that the majority of the residual variance is due to differences between regions. This makes sense, as it is well known that there still exists a large number of structural differences across italian regions, especially concerning the comparison between north and south. These differences - when observable - are useful in the explanation of various microeconomic phenomena, and reflect on a vast number of social, political, and economic factors.
- **Corr**, the correlation between the entity error terms and the regressors, is different from zero: this is exactly what we would expect in a FE model.

6 Conclusions

In this paper, we gave an insight on the conditions of the electricity sector in Italy, focusing our attention upon domestic customers and their preparedness to enter a thoroughly liberalized market. After many de-

lays, the extinction of the GP regime should happen in 2023, but as reported many times by ARERA and many consumers' associations, Italy is not ready for a complete removal of such a market yet. Our analysis is perfectly aligned with such a concern: first, the prices set in the liberalized portion of the market are not as competitive as it would be desirable after such a long time span after the beginning of the liberalization process.
Secondly, even when an actual convenience with respect to the regulatory benchmark can be appreciated, this is too difficult for consumers to detect. Such difficulty derives both from the ever increasing number of offers presented on the FM, each with specific and diverse services and promotions, and from the complexity of the bill's composition, upon which italian consumers do not have adequate knowledge yet. Regarding the former, a further aspect to be considered is competition's other side of the coin. While providing consumers with a wide gamut of choice, in such a poor information setting this variety often has the negative effect of discouraging them from the idea of entering the FM because they cannot orient themselves. Even worse, this could lead to consumers entering the market in view of a generic saving ideal, but then choosing a more expensive offer, because of their scarce knowledge of the framework. These preliminary results lead to our study's main finding: independently of the savings they could make by entering the FM, the paramount element to consider when evaluating the customers' tendency to switch to the new market is the knowledge they have about it.
As economic theory suggests, Information is key for a good functioning of a competitive market; thus, until this gap will not be filled, the european liberalizing objective is unlikely to be harmlessly completed. In the light of these findings, we would advise the Authority to focus more of her efforts on the consumer's Information level, rather than on the structure of the market; despite her endeavor to create a competitive setting for sellers to operate in, these attempts remain vain if the demand side is not adequately stimulated and informed.
Indeed, ARERA's constitutive law (481/95) states her obligations not only towards the promotion of competition, but also towards the spread of transparent information and the ex ante protection of the consumer; one of the principal concerns of the Authority should in fact be that of guiding the end users in a conscious choice.
A second policy suggestion could be related to the simplification of the bill's readability; as stated in the introduction, this remains in fact a huge obstacle between the consumers and the savings they can possibly derive from competition. Actually, ARERA is already working in this direction: in the last decade she managed to switch her complex system, based on consumption bands, to a more understandable scheme with few variable costs and a larger fixed component.
For the scope of this project we restricted ourselves to analyze only two possible definitions of the information variable. A possible further implementation of this analysis could concern different additional definitions, with the aim of taking into account all the possible means of Information available to the consumers (e.g. shares of national/regional TG and different keywords for google searches). In conclusion, even if the share of domestic consumers entering the FM is growing over time, this growth is not proceeding at the expected pace; among the main challenges to face in the promotion and concretization of the competitive market, providing users with adequate knowledge is certainly the most outstanding.

References

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31996L0092

[2] https://www.camera.it/parlam/leggi/deleghe/99079dl.htm

[3] https://www.autorita.energia.it/allegati/relaz_ann/00/cap2400.pdf

[4] https://www.camera.it/parlam/leggi/07125l.htm

[5] https://www.ilsole24ore.com/art/bollette-fine-maggior-tutela-slitta-2023-ecco-cosa-succedera-consumi-famiglie-e-imprese-ADsWLvLB

[6] https://www.arera.it/allegati/eventi/190220_indagine.pdf

[7] https://www.corriere.it/economia/15_febbraio_16/no-all-abolizione-mercato-tutelato-rischio-aumenti-20percento-bolletta-f14abb16-b5e7-11e4-bb5e-b90de9daadbe.shtml

[8] https://www.agcom.it/documents/10179/8144896/Documento+generico+27-07-2017/3fd21601-4abd-4f2b-9d1a-063bd0c745ec?version=1.0

[9] https://selectra.net/energia/guida/mercato/prezzo-indice-selectra#:~:text=Il%20Prezzo%20Indice%20Selectra%20%C3%A8,delle%20offerte%20presenti%20nel%20mercato

[10] https://www.arera.it/it/relaz_ann/20/20.htm