

Capacity choices in liberalised electricity markets[☆]

Fidel Castro-Rodriguez^a, Pedro L. Marín^b, Georges Siotis^{c,*}

^a Departamento de Fundamentos de Análisis Económico and RGEA, Facultad de Económicas, Universidad de Vigo, Spain

^b Departamento de Economía, Universidad Carlos III de Madrid, Spain and CEPR

^c Departamento de Economía, Universidad Carlos III de Madrid, Spain, European Commission, and CEPR

ARTICLE INFO

Article history:

Received 20 June 2008

Accepted 28 January 2009

Available online 14 April 2009

Keywords:

Electricity

Capacity

Long-run investment

ABSTRACT

This paper addresses the issue of investment in electricity generation in the context of a liberalised market. We use the main results derived from a theoretical model where firms invest strategically to simulate the Spanish electricity system with real-world data. Our results indicate that, under reasonable parameter constellations regarding the number of agents, the level of capacity resulting from private decisions falls well short of the social optimum. Last, we show that two regulatory mechanisms that have been used to generate additional incentives for private agents to install capacity (capacity payment and price-adder) are ineffective and/or prohibitively costly.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Following England and Wales' lead, many countries have embarked on a process of liberalisation of their respective electricity sectors. There are two central aspects associated with this liberalisation process. The first is that, in most of these markets, liberalisation has led to the establishment of wholesale spot markets for electricity (usually known as an *electricity pool*). The second is that the investment decisions are no longer centrally planned, as agents are free to choose the level of new generation capacity.

The central issue analysed in our paper regards the performance of a stylised electricity system in which agents are free to choose their capacities (a long-run decision) before competing to supply energy in a pool. More precisely, we use the central results derived from a theoretical model of strategic investment decisions

[☆] This research was partially financed by Unión Eléctrica Fenosa (UEF), whose financial support is gratefully acknowledged, and whose staff greatly helped us navigate through the technological and regulatory intricacies of the electricity sector. Eduardo Bazán Navarro, José María Chamorro, María Angeles de Frutos, Carlos Hervés, José-Luis Moraga and Antonio Romero-Medina, seminar participants in Lausanne and EDF (Paris), as well as participants to EARE and the Jornadas de Economía Industrial provided very insightful suggestions and comments. Castro-Rodriguez gratefully acknowledges financial support from the Spanish Ministry of Science and Innovation and FEDER through projects SEJ2005-07913 and ECO2008-05771. Siotis gratefully acknowledges funding from Ministerio de Educación y Ciencia (SEJ2007-66268).

* Corresponding author at: DG COMP (CET), 70 Rue Joseph II, B-1049 Brussels, Belgium. Tel.: +32 2 2952380.

E-mail address: siotis@eco.uc3m.es (G. Siotis).

¹ This paper was written before Siotis joined the European Commission and the views expressed are those of the author and do not necessarily reflect those of DG COMP or the European Commission.

and provide simulations of the main results using real-world Spanish data. Our ultimate goal is to gauge the magnitude of the effects that have been identified theoretically (such as the extent of underinvestment in capacity) in the context of a model that embodies many of the industry's idiosyncrasies.

From a theoretical perspective, there are contributions that deal with the issue of long-run investment in generation capacity in the context of liberalised markets. For instance, Gabszewicz and Poddar (1997) present a model with demand uncertainty and analyse a symmetric duopoly where firms first decide on capacity and then compete in the product market. They show that the equilibrium price and quantities coincide with a certainty equivalent Cournot game. A recent paper by Grimm and Zoettl (2006) provides a general treatment of these issues. They present various two-stage games (capacity choice followed by production) in the context of demand uncertainty. In particular, they show that with optimal regulation at the production stage, total investment is even lower than in the Cournot market game (Grimm and Zoettl, 2006, p.3). von der Fehr and Harbord (1997) also present a two-stage model of long-run investment choices where firms first choose capacity and then compete to supply energy. The novelty of their approach is that the second stage is modelled as a non-discriminatory multiple-unit sealed bid reserve auction. They demonstrate that, under competitive conditions, a no-intervention private outcome will yield an insufficient level of installed capacity. In a similar vein, Fabra et al. (2008) analyse how the design of the spot market affects market performance through its impact on investment decisions. They show that under reasonable assumptions regarding the distribution demand, a discriminatory auction yields better results in terms of investment incentives.

In our paper, the central concern is to better capture the industry's specificities for the purpose of carrying out simulations

with Spanish data. More precisely, our simulations are based on a model that contains the following ingredients: demand uncertainty, the existence of different types of consumers, strategic investment at the capacity stage, competitive pricing in the pool, and random rationing of demand when the system is short of capacity. Introducing these real-world features adds an unduly large amount of algebra. Thus, in this paper, we limit ourselves to presenting our model's architecture, while all the formal derivations and proofs have been relegated to a web Appendix available at: <http://www.eco.uc3m.es/siotis/investigacion.html>.

While this paper shares the same motivation and adopts similar modelling choices as those of von der Fehr and Harbord (1997) and Grimm and Zoettl (2006), the approach differs from theirs in a number of respects. For instance, our contribution distinguishes between consumers that can undertake demand-side bidding and those that do not, reflecting the fact that most consumers are not exposed to real-time price signals. Also, we explicitly allow for random rationing leading to blackouts (in the papers mentioned above, there always exists a price that clears the market). Last, the simulations illustrate the extent to which generators can extract rents by investing strategically, even if they behave competitively at the time of supplying energy. These simulations also permit an assessment of two regulatory mechanisms used to generate additional incentives for private agents to install and maintain capacity (capacity payment and price-adder).

Our simulation results indicate that, under reasonable parameter constellations regarding the number of agents, the level of capacity resulting from private decisions falls well short of the social optimum. In addition, the potential for extracting rents via strategic investments decisions in capacity is large, even under the assumption of a competitive spot market. Last, the two regulatory mechanisms analysed in this paper (capacity payment and price-adder) are both ineffective and very costly.

The paper is organised as follows. Section 2 provides a non-technical version of the model's architecture. Section 3 presents the data, describes how we build our simulations, and provides the results. Section 4 models two regulatory mechanisms that have been used to generate additional incentives for private agents to install and maintain capacity (capacity payment and price-adder). Section 5 discusses some policy implications and concludes.

2. Building blocks of the theoretical model

In this section, we present a non-technical version of the theoretical model that forms the basis of the simulations. The extensive version has been relegated to a web Appendix that can be found at: <http://www.eco.uc3m.es/siotis/investigacion.html>.

We distinguish between two types of hourly demand. The first type acquires energy through demand-side bidding, while the second type obtains electricity by paying a pre-determined fixed price tariff.² The first group represents the demand for energy that stems from the bids that electricity suppliers make on behalf of price-sensitive demand that can be adjusted within the hour, as well as the energy used by pumped-storage.³ In practice, this

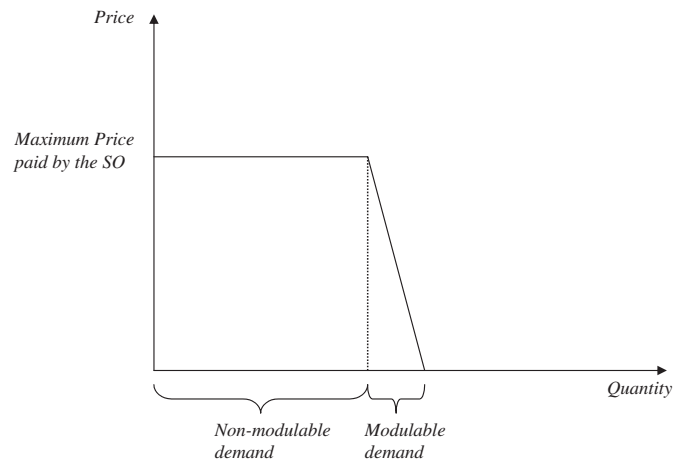


Fig. 1. Aggregate demand.

represents a small share of total demand at any given price. For that segment, we say that demand is modifiable.

The bulk of actual consumption consists of demand that is non-modifiable in a given hour; that is, the demand that stems from those consumers that pay a pre-determined tariff.⁴ As this group does not receive real-time price signals, their consumption is insensitive to the pool's price. This means that, when tight capacity drives up price, demand curtailment only affects the modifiable group. When price is equal or above the level that drives modifiable demand to zero, the system faces the prospect of total collapse, as non-modifiable consumers do not react to price (and therefore, do not adjust their consumption). Then, it is the system operator that decides to cut customers off when the system is short of capacity to avoid a collapse of the entire system. Given the system operator cannot differentiate these consumers according to their individual valuation of each unit of electricity at each moment in time, it cuts them off randomly. We assume that the maximum price that triggers this rationing on the part of system operator is the average value that non-modifiable consumers paying a tariff give to one unit of electricity net of transportation, distribution, and administrative costs.⁵ We are thus assuming that the system operator acts as the average consumer that does not receive real-time price signals.

Fig. 1 represents the aggregate hourly demand that participates in the wholesale market, which is made up of two segments. The first is horizontal and represents non-modifiable demand (reflecting their average valuation of one unit of electricity). The second segment stems from modifiable demand and is downward sloping. For the sake of computational simplicity, we assume that the average value that non-modifiable consumers give to one unit of electricity is equal to the maximum willingness to pay for electricity of modifiable consumers.⁶

We also take into account the variability of demand that stems from real-time fluctuations, as well as mid- to long-term uncertainty that stems from the business cycle. Concerning the uncertainty related to hydrological conditions, we assume

² Note that a consumer may use both channels simultaneously to satisfy two different parts of her demand.

³ Electricity suppliers are agents that buy electricity in the pool, pay a fee for its transmission, and sell it to a final consumer. Pumped-storage refers to a common practice, whereby dams use electricity to pump-up water into the reservoir. Thus, the amount of pumped-storage is at its maximum when spot prices are at their minimum (the reservoir can be refilled cheaply), and zero when prices are highest (the cost of pumping water upstream is larger than the expected revenue).

⁴ What we call non-modifiable load represents the part of aggregate demand that cannot be adjusted within the hour, either because consumers do not receive price signals, or because technological conditions impede them to do so. For instance, residential customers pay a pre-determined tariff per KWh and do not receive within the hour price signals. Examples of technological constraints involve some refrigeration operations, or metallurgical production.

⁵ The web Appendix provides a precise characterisation of the price that triggers rationing.

⁶ The web Appendix provides a precise characterisation of the demand stemming from these two consumer groups and explains how they are technically aggregated to construct total hourly demand.

peak-load shaving, i.e., hydrocapacity is dispatched when demand is at its highest and it is most valuable (this encompasses a monopoly situation, centralised planning, and perfect competition; see García-Díaz and Marín (2003) for a discussion of this issue). Thus, the aggregate demand curve utilised in the simulations corresponds to the residual demand for conventional thermal generators (total demand minus the part covered by dams and renewable energy sources that operate under a special regime).

On the supply side, we assume that all capacity is characterised by a constant marginal cost technology (we have combined cycle gas turbines, CCGT, in mind). This yields a right angle aggregate marginal cost schedule, with the kink being determined by the level of installed capacity. This implies that once the level of available capacity is reached, the marginal cost becomes infinite, which reflects the fact that supply cannot be expanded in the short run. Because of maintenance work and unexpected outages, only a fraction of total installed capacity is available at any point in time.

To compute the unitary fixed cost associated with one unit of new capacity, we assume a fixed lifetime for each unit of capacity and distribute the latter's fixed costs among the MWh that are expected to be produced, which itself depends on aggregate capacity. Clearly, a plant's utilisation is inversely related to aggregate capacity as more installed capacity implies a lower probability of being dispatched.

2.1. Characterisation of the socially optimal level of capacity

The social optimum corresponds to a level of capacity that would be chosen by a benevolent regulator that maximises joint surplus, i.e., the sum of producers' and final consumers' surpluses net of capacity costs, with equal weight given to both groups.

Socially efficient use of installed capacity implies marginal cost pricing as long as the capacity constraint is not binding. When capacity is insufficient to cover demand at a price equal to variable unit cost, the market price is raised along the demand curve so as to reduce consumption until it no longer exceeds available capacity (efficient rationing). When prices are equal or above the level that drives modifiable demand to zero, further price increases do not result in reductions in consumption, as the remaining non-modifiable consumers are completely insensitive to price.

As a consequence, the system operator has to blackout a proportion of non-modifiable consumers to maintain the system's integrity. Thus, our model features rolling black-outs (when an entire geographical area is left without energy for a finite period of time).⁷

This implies that social welfare boils down to consumers' surplus whenever available capacity is sufficient to cover demand (as marginal cost pricing drives producers' surplus to zero). When capacity is not sufficient to cover demand at price equal to marginal cost, the price increases along the aggregate demand curve until capacity is sufficient to cover demand. In these circumstances, social welfare is the sum of consumers' and producers' surplus. When some consumers have to be rationed, social welfare corresponds to producers' surplus, irrespective of the realisation of demand. In all cases, gross social welfare is reduced by capacity costs.⁸

⁷ Recall that the price that triggers random rationing on the part of system operator is the average value that non-modifiable consumers give to one unit of electricity.

⁸ Note that our definition of welfare probably yields a lower bound of real-world socially optimal capacity. For instance, we have assumed that the benevolent social welfare maximiser is risk neutral, a condition unlikely to be met in practice. Also, there may be non-linearities in the social cost associated

2.2. Capacity investments in decentralised markets

To derive the level of capacity corresponding to a situation in which firms are completely free to choose the amount of capacity to install, we consider a finite number of symmetric firms, and represent their decisions as a two-stage game. During the first stage, agents decide on the level of capacity to install, a decision that we consider to be a long-run one. During stage two, firms supply energy competitively.⁹ This second characteristic is very convenient from the perspective of analytical tractability, as it allows us to work with more than two agents in a scenario characterised by demand uncertainty. Since agents act non-strategically during the second stage, our model is close, but not equivalent, to a single period game in which firms have to simultaneously decide on their capacity and output. The difference arises from the existence of demand uncertainty: as firms may turn out to be unconstrained if demand is low (and cannot extract rents since we impose marginal cost pricing during the second stage), investment decisions will be affected during the first stage.

The profits obtained by each firm depend not only on its capacity but also on aggregate capacity. In particular, a firm's profit (gross of capacity costs) is nil when capacity is sufficient to cover demand at price equal to marginal cost. In other cases, firms' gross unitary profit depends on the difference between price and marginal cost.

3. Simulations for the Spanish electricity market

Given the theoretical model's complexity, explicit solutions for decision variables cannot be obtained. Thus, we had to solve the model numerically in to quantify the gap between the social and private outcomes in real-world economies.¹⁰ The data that we have used for our simulations pertain to the hourly Spanish electricity market. As will be seen below, the data necessary to simulate the model are readily available. Consequently, our model could be easily tailored to the specific situation of other electricity systems.

3.1. Data

3.1.1. Distribution of demand

To measure demand variability, we obtained data on the Spanish hourly distribution of demand for 2005. This information is made available by the market operator, and can be downloaded from the web (OMEL, www.omel.es). Hourly demand was updated with projections of future demand up to 2010 provided by the system operator, Red Eléctrica de España (REE).¹¹ Given the time required to build new capacity and the fact that long-term

(footnote continued)

with repeated black-outs. For instance, electricity shortages could produce dynamic costs, particularly in countries or regions that heavily rely in foreign direct investment, a condition that is clearly met in the Spanish case.

⁹ It is plausible to argue that the assumption of competitive bidding during the second stage is unrealistic, as firms would exercise their short-run market power. The counterargument is that firms' bidding behaviour in the spot market has been under constant regulatory supervision since the inception of liberalisation. In the UK, the pool's perceived failings in this respect led to the introduction of NETA. In Spain, the regulator systematically requests explanations whenever there is a price spike, even when the increase is modest. This assumption is also consistent with a situation where firms strategically withhold part of their capacity and bid the remainder at marginal cost. We are grateful to an anonymous referee for making this point.

¹⁰ Details can be found at: <http://www.eco.uc3m.es/siotis/investigacion.html>.

¹¹ The projections provided by REE oscillate within some interval that we have used to introduce demand uncertainty.

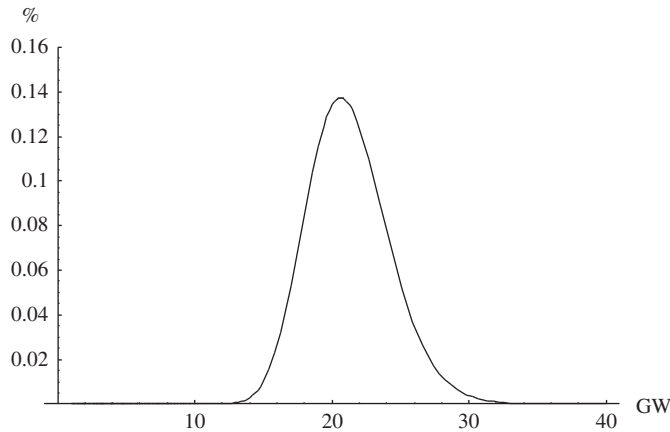


Fig. 2. Probability density function of thermal demand.

demand projections are less precise, we focus on the year 2008 in our simulations.

As mentioned before, we assume that renewable energy and hydrogeneration capacity are exogenously determined, and that no new dams will be built in the foreseeable future. We net out hydrocapacity in the following manner. First, we use historical data on rainfall since the last dam started producing energy to obtain the average, minimum, and maximum hydrocapacity in a year. Since electricity from hydrosources can be stored, hydro-power will be sold when demand (and prices) is highest. We apply the usual peak-load shaving technique, that is hydrocapacity is distributed to serve the hours of peak demand till the entire hydrocapacity is used-up. In doing so, we took into account the relevant technical restrictions such as the fact that there is a maximum (minimum) of MWh that can be (have to be) dispatched during an hour. Combining this information with the predictions of demand provided by REE, we obtained the probabilistic distribution of hourly net thermal demand for the year 2008, ranging from a minimum of 11.45 GWh to a maximum of 30.04 GWh. Adjusting a lognormal distribution to the data yielded an average hourly thermal demand of 21.26 GWh with a standard deviation of 2.90 GWh (see Fig. 2).

3.1.2. Average valuation for one unit of electricity

The flat segment of our demand schedule represents the average valuation of one unit of electricity on the part of non-modulable consumers. To the best of our knowledge, the Spanish regulator has never publicly determined an average valuation for 1 MWh. To approximate this value, we have computed annual GDP per MWh; the results are presented in Table 1. With this information, we have imposed a value of 3005 euros per MWh for our simulations. As shown in Table 2, our average valuation is very close to the values that exist in countries that specifically incorporate the value of lost load (VOLL) in their regulatory systems.¹² We further assumed that in the presence of excess demand, firms would be able to increase prices up to that average valuation.¹³ Alternative scenarios where firms' prices are capped below the average valuation are presented in Section 4, where we model existing regulatory mechanism to simulate investment in generation capacity.

¹² This value is consistent with the conclusions reached after the 1977 New York blackouts.

¹³ We carried out simulation for alternative values of the average valuation (VOLL). The results are of the same order of magnitude for reasonable values of that parameter (e.g., between €180 and €3005). These additional results are available upon request.

Table 1

Quarterly Spanish GDP per MWh, 1999.

Quarter	€/MWh
First	2926
Second	3300
Third	3055
Fourth	3065
Average for 1999	3087

GDP from *Boletín Trimestral de Coyuntura*, INE, March 2000, no. 75.

Table 2

Value of lost load in Euros/MWh.

Country	1995	1996	1997	1998	1999
Victoria (Australia)	2788	2979	3273	2825	3028
England and Wales	2704	2810	3488	3716	3924

Exchange rate: IMF, June 2000.

Note: Average exchange rate for the year

3.1.3. Average modulable demand

The slope of the downward segment accounts for the fact that some customers reduce their consumption when price increases. We assumed that pumped-storage would be at its maximum when prices are zero, and would disappear when prices reach their maximum. We then augmented the demand for hydro-pumping with that derived from large interruptible consumers.¹⁴ We assume that in an hour of average demand, customers would be able to reduce their consumption within the hour by 5% if prices were to reach their maximum. Finally, we assume that the slope of our demand schedule remains constant throughout the year.¹⁵

3.1.4. Variable costs

On the supply side, we assume that all investment consists of combined cycle gas turbines which is deemed to be the most efficient today.¹⁶ We assume a unit variable cost of 40 €/MWh.¹⁷

3.1.5. Capacity costs

In Spain, the smallest sized of CCGT plant is of 387.1 MW, which can be built at an investment cost of €379.191 millions. To obtain the unitary capacity cost, the generating unit's total fixed costs are distributed among the MWh that the plant is expected to produce during its entire lifetime. This implies that the expected unitary fixed cost is lower when the plant is expected to be dispatched a larger number of hours. For example, assuming an average level of utilisation of 7500 h per year during 20 years, a plant's expected production stands at 58,065 GWh, and the resulting unitary capacity cost is 6.53 €/MW.

¹⁴ A large consumer is defined as one with an aggregate annual demand greater or equal to 1 GWh.

¹⁵ Variations in this parameter (e.g., between 1% and 50%) do not change the essence of the results; however, increasing the proportion of interruptible load (modelled as an increase in the slope of the inverse demand function) narrows the wedge between the proportion of unsatisfied demand associated with the social optimum and the private outcome. These additional results are available upon request.

¹⁶ All new non-renewable capacity in Spain (recently installed and projected) uses this technology.

¹⁷ Variations in this parameter have been analysed but do not substantially change our results.

3.1.6. Availability ratio

To take into account temporary outages resulting from maintenance work, we apply the average availability ratio in Spain, which stands at 92.5%.

3.1.7. Number of firms

Regarding the number of firms active in the market, we present simulation results associated with three scenarios. The first scenario corresponds to a situation in which we assume free entry. Apart from integer problems, the upper bound of active players is determined by the necessity to cover fixed costs and the minimum size of a generating plant (387.1 MWh). Combining these restrictions yields a maximum of 78 generators. Clearly, this first scenario is highly unlikely to materialise in practice. In the real world, additional constraints to entry and/or capacity expansion exist, that is to say, there are important barriers to entry. For instance, the number of locations for generating units is limited by building and environmental restrictions, as well as by transmission constraints. Therefore, we also analyse two additional scenarios.

The second scenario contemplates a very large increase in the number of active participants in the Spanish market in the near term, with 20 firms. The third pertains to the number of established, non-atomistic generators that compete to supply energy in the Spanish pool, that is six firms.¹⁸ If the recent experience is any guide, considering 20 firms most probably overestimates the number of actual firms that will be active in Spain. Liberalisation started in January 1998, and new entry has been very limited to date.¹⁹ Compared to the larger England market where liberalisation started much earlier and where the regulator ordered de-concentration, 20 appears as a very high number. Thus, we believe that considering an interval for the number of firms between six and 20 firms encompasses all realistic situations.

The data used in the simulations are summarised in Table 3.

3.2. Results

3.2.1. Capacity choices

Our main results are reported in Table 4. The social optimum yields a level of 30,680 MW of conventional thermal capacity for the year 2008.²⁰ As can be deduced from Table 4, the socially optimal level of capacity does not cover peak demand. More precisely, a level of capacity of 32,450 MW would be required to eliminate all shortages. This result is due to the high variability of demand. In particular, peak demand only materialises during a small number of hours in the year (that is, it is a low probability event in the demand distribution), which implies that cost of incremental capacity necessary to cover that demand is significantly higher than the incremental social welfare. Obviously, this result is conditional on our conservative definition of the welfare function that does not include the negative externalities that are usually associated with black-outs.

¹⁸ There are a number of very small (atomistic) generators, such as microdams and generation for self-consumption in some industries. These additional players are marginal and do not exert any influence on price.

¹⁹ In fact, were it not for the blocking of various mergers, the current number of large active firms would possibly be smaller.

²⁰ This corresponds to a total of approximately 61,400 MW of installed capacity if we include hydrocapacity and renewable energy sources (Red Eléctrica de España, 2006). The latter operate under very specific rules, and are not exposed to market signals (i.e., renewable capacity is set exogenously). In addition, availability (which is typically very low) is beyond the firms' control, as it depends on climatological factors such as sunshine and wind.

Table 3

Values of parameters.

Parameter	Base scenario
Lognormal distribution of hourly thermal demand	
Minimum (GWh)	11.447
Maximum (GWh)	30.035
Average (GWh)	21.255
Standard deviation (GWh)	2.90
Average modulable demand in one hour (%)	5%
Average valuation for one MWh (€/MWh)	3005
CCGT generating plant characteristics	
Minimum size (MW)	387.1
Variable cost (€/MWh)	40
Fixed cost (millions of euros)	379.191
Lifetime operation (years)	20
Average availability ratio	92.5%
Number of firms	6, 20, 78
Capacity payment (€/MWh)	13.25

The socially optimal level of installed capacity would result in a tiny amount of unsatisfied demand at marginal cost pricing (0.004% of total annual demand). Thus, while the gap between these two scenarios in terms of installed capacity is non-negligible (1770 MW, corresponding to almost 6% of the socially optimum level of capacity), the difference with respect to unsatisfied demand is small.²¹ This is due to the fact that the number of peak-demand hours is very limited (i.e., a very low probability is associated with that event occurring).

Under the first decentralised scenario, which corresponds to a free-entry long-run equilibrium with 78 generators, the privately installed level of capacity falls short of the social optimum, but the gap is small. A decentralised outcome would yield 30,250 MW, which represents 98.6% of the socially optimal capacity. Under these circumstances, only 0.006% of total annual demand would be left unsatisfied, a figure very close to that pertaining to the social optimum. However, as mentioned above, we feel that this scenario is unlikely to materialise in practice, as it would involve a 13-fold increase in the number of generators.

Under the second scenario with 20 symmetric generators, a decentralised outcome yields a capacity of 28,340 MW, which represents 92.4% of our thermal social optimum. Thus, a situation with 20 market participants would involve some brownouts and limited random rationing of demand. The finding is that, even with a substantial number of generators (20), a private outcome still falls short of the social optimum by a non-negligible margin in terms of unsatisfied demand (0.043%, a 10-fold increase as compared to the social optimum). With 20 firms our results imply shortages during peak-demand hours.²²

Under the third scenario, simulations clearly indicate that a decentralised outcome would be sub-optimal. Generators would only install 22,400 MW, which represents 73.0% of the conventional thermal capacity that maximises social welfare. This would imply shortages in many hours of the year, and would result in 4.71% of total annual demand being unsatisfied. As such, this result is not surprising. In our closed oligopoly with six firms, a profit maximising strategy consists in maintaining capacity shortages for the purpose of increasing prices.

²¹ Note that, in terms of demand coverage, the social optimum that our model generates is very close to the simple rule that involves covering peak demand.

²² These hours are clustered during two winter months (December and January) and July.

Table 4
Installed conventional thermal capacity.

	Social optimum	Capacity to cover peak demand	Capacity, number of firms: 78	Capacity, number of firms: 20	Capacity, number of firms: 6
Capacity (MW)	30,680	32,450	30,250	28,340	22,400
% of social optimum	100	105.8	98.6	92.4	73.0
Unsatisfied demand (%)	0.004%	≈ 0.00%	0.006%	0.043%	4.705%

Note: The figures pertain to conventional thermal capacity. This corresponds to a total of approximately 61,400 MW of installed capacity if we include hydrocapacity and renewable energy sources (2006 data).

Table 5
Results with respect to the number of firms.

Number of firms	Capacity (MW)	Expected Consumption per hour (MWh)	Decentralised welfare as a % of total surplus	Consumer surplus as % of decentralised surplus	Producer surplus as % of decentralised surplus
6	22,400	19,668	94.689	50.40	49.60
20	28,340	21,007	99.862	95.53	4.47
40	29,760	21,052	99.984	98.47	1.53
60	30,110	21,059	99.994	98.95	1.05
78	30,250	21,061	99.997	99.12	0.88

Intuitively, agents install a level of capacity that increases the likelihood of being constrained in low-demand states. In other words, imposing competitive behaviour at the second stage does not impede oligopolistic firms from extracting supra-competitive rents. This is due to the fact that firms act strategically when deciding upon the capacity to install. This is in line with the result of Grimm and Zoettl (2006, p.3) mentioned in the Introduction.

Socially sub-optimal outcomes are common in oligopolistic industries. These results show that in a fully deregulated electricity industry with six firms, the gap between the social optimum and a private outcome is large.

3.2.2. Welfare analysis

We next turn to the welfare analysis associated with each of these scenarios. We show how total surplus is divided between consumers and producers in each case. Table 5 presents total surplus derived from a decentralised outcome as a proportion of maximum achievable welfare. It also indicates how the surplus is divided between the two groups.

With six firms, a decentralised outcome achieves 94.7% of maximum welfare, with consumers receiving one slightly more than half of total surplus (50.4%). With 20 firms, consumers receive 95.5% of total surplus, with the latter representing 99.9% of achievable surplus. Last, with 78 firms the decentralised outcome delivers almost 100% of achievable welfare (99.997%), with consumers receiving 99.1% of it.

These results indicate that welfare increases non-linearly (at a decreasing rate) with the number of firms. This derives from two important characteristics of real-world electricity markets. First, as shown in Fig. 2, the standard deviation of the entire distribution is low, resulting in a clustering of demand realisations around the mean (22,255 MW), meaning that extreme realisations are rare events. This implies that once installed capacity is greater than demand's mean expected realisation, extra capacity results in limited increases in consumption, thus generating small increments in welfare. As aggregate installed capacity (22,400 MW) in a decentralised market with six agents is larger than the mean demand realisation, an increase in the number of competitors results in a small increment in expected consumption, and therefore, of expected welfare. Second, the demand schedule is made up of a flat part and of a very steep downward

sloping segment, with the latter accounting for 5% of total demand in our central simulations. This implies that, for most demand realizations, lower prices due to increases in capacity result in a transfer from generators to consumers, but not large welfare gains. In our setting, the main effect of increasing the number of agents is to transfer surplus from producers to consumers (see Table 5).

We interpret these findings as indicating that, in the absence of large-scale entry, some form of regulatory intervention will have to be maintained in to protect consumers and to avoid capacity shortages.²³

4. Regulatory mechanisms

Several forms of regulation specially aimed at ensuring an adequate level of supply have been used. In this section, we analyse two of them: capacity payments and a price-adder. The manner in which these mechanisms are introduced in our model is described in the web Appendix (available at <http://www.eco.uc3m.es/siotis/investigacion.html>).

4.1. Capacity payment

Conceptually, capacity payments consist in paying a monetary amount to generation units that have declared their availability (i.e., have made supply bids), irrespective of whether they are actually dispatched or not. In practice, there exist some conditions that must be met before a plant is deemed to be available.

Since our simulations pertain to Spain, we model a capacity payment as it exists in that country. Concretely, the amount that each firm receives is determined by a two-step procedure. First, the total amount to be paid to firms is obtained by multiplying a monetary amount (capacity payment in terms of euros per MWh available) by the system's total demand. Second, that total amount

²³ We carried out an extensive sensitivity analysis with respect to the underlying parameter values (e.g., alternatives values for maximum price, demand elasticity and degree of modularity, variable, and fixed costs). These additional results are of the same essence as those presented in the text and are available upon request.

is proportionally distributed among generators according to their declared availability.

4.2. Price-adder

Under such a regulatory system, a monetary amount is determined on the basis of the reserve margin, and paid to all the units that have been declared available, irrespective of whether they are dispatched or not. In particular, we introduce the mechanism following the rules that existed in England and Wales prior to the introduction of the NETA in early 2001. Under that system, a capacity charge is added to the price of energy. This charge is defined as $CC = LOLP(VOLL - P)$ where *LOLP* stands for loss of load probability and is defined as the probability of demand being greater than available capacity, *VOLL* stands for value of lost load and *P* is the price used to determine the amount paid to generators.²⁴ If the generating unit is effectively despatched, the price *P* plugged into the formula is the spot price, and if the unit is not despatched, the *P* applied is the bid price made by that particular plant.

4.3. Simulation results in the presence of capacity payment and price-adder mechanisms

Under a scenario of total deregulation, a capacity payment involves a prohibitive cost. Indeed, even with 20 firms, it requires a very large transfer to generators to induce them to install the optimal level of capacity. The intuition for this result is simple. The social optimum corresponds to the intersection between the industry's marginal costs schedule and aggregate demand. The private outcome obtains as a consequence of firms equating their marginal cost to the marginal revenue derived from the residual demand they each face. Thus, to induce private firms to install the socially optimal level of capacity, it is necessary to have a capacity payment large enough so that the perceived private marginal revenue shifts and intersects with marginal cost at a capacity level that corresponds to the social optimum.²⁵

As mentioned above, Spain operates a system of capacity payments, where a monetary amount is paid for each MWh made available to the system (but not necessarily dispatched), as long as the generating unit is made available 480 h per year. This compensation is determined for six daily time intervals that correspond to different levels of demand; the maximum is set at 13.25 € per MWh available for the peak-demand interval. Our results indicate that such a mechanism is ineffective to induce a socially optimal level of capacity. The Spanish capacity payment system is really a means to avoid the closure of existing old plants. However, it does not provide sufficient incentives to install new capacity.

Turning to a price-adder, it emerges that this mechanism is also ineffective in a fully liberalised market. The reason is simple: firms have a double incentive to reduce capacity. On the one hand, it allows them to charge high prices throughout the year. On the other hand, maintaining capacity tight increases *LOLP*, which in turn inflates the amount that agents receive as a capacity charge.²⁶ The counterpart of a high *LOLP* is extremely large payments to generators.

These simulations clearly indicate that, in this oligopolistic context, both the two regulatory mechanisms that are analysed

are ineffective, and in case of the capacity payment, prohibitively costly.

4.4. Simulation results in the presence of a price cap coupled with capacity payments or a price-adder

In most markets, liberalisation has been incomplete in the sense that price intervention has been maintained in the form of regulated tariffs and/or price caps. We next analyse how the two regulatory mechanisms perform in the presence of a price cap. In case of a capacity payment, the introduction of a cap does not change the essence of the results: lowering the cap reduces installed capacity, and the magnitude of capacity payments required to induce the socially optimal level of capacity remains prohibitive. Table 6a shows how installed capacity changes with the level of the cap and simulates the effect of the maximum capacity payment currently in force in Spain (13.25 € per MWh made available) for the case of 20 firms. As can be readily seen, the increase in installed capacity induced by the existing payment achieves is minimal. As a direct corollary, it is possible to deduce that the necessary transfer to achieve the social optimum remains prohibitive, even in the presence of a cap.

In case of the price-adder, the simulation results are somewhat more encouraging. The price-adder somehow compensates for the reduction of privately installed capacity induced by the price cap. Table 6b indicates that a cap of €180 per MWh (note that, while the price is capped at €180, the *VOLL* remains equal to €3005) reduces decentralised capacity as a proportion of the optimum by about 10 percentage points. However, the price-adder does induce a substantial amount of extra investment in capacity and almost completely neutralises the effect of the price cap.

Price caps are widely used to protect consumers from excessive prices. Our simulation results indicate that this may be a wise choice in a situation characterised by a small number of generators. The downside of this measure is that it reduces incentives to invest in capacity. From that perspective, a price-adder is quite successful in mitigating this adverse effect, while a capacity payment remains inadequate.

5. Conclusions

We center our attention on capacity choices in the context of a two-stage game, in which firms first choose capacity and then compete in the product market. We explicitly take into account the industry's idiosyncrasies, such as the level of uncertainty surrounding supply and demand, the near-impossibility to store electricity, as well as the technical restrictions that characterise electricity generation. The various simulations encompass a fairly wide spectrum of outcomes. Our main finding is that a deregulated market will result in underinvestment in generation capacity. Since, short-run competition is easier to monitor for the regulator, we believe that capacity choices may become the main instrument through which generators will attempt to exercise market power. This suggests that there is a need for some regulatory mechanism that provides the right incentives to install a socially desirable level of capacity. In that respect, the two mechanisms that we analysed fail to deliver the desired outcome.

The main policy implication of our results is that de-concentration would be highly desirable in electricity markets characterised by a reduced number of players. If this option is not available (for instance, because of legal constraints or security of supplies considerations), then lowering entry barriers ought to be a priority. For instance, increasing the transmission capacity between different systems could, in principle, mitigate the exercise of market power. Still, it remains the case that, as long

²⁴ In our central simulations, we use the maximum price accepted by the system operator as *VOLL*.

²⁵ For instance, in the extreme case of a monopolist facing a linear demand curve, the required transfer is twice as large as the total competitive surplus.

²⁶ See Wolak and Patrick (1997) for an extensive treatment.

Table 6

Price cap (€/MWh)	Decentralised outcome without capacity payment (MW)	% Optimum capacity	Decentralised outcome with capacity payment (MW)	% Optimum capacity
(a) Effect of a capacity payment of 13.25 € per MW available, with 20 firms				
3005.05	28,340	92.37	28,367	92.46
1800.00	28,150	91.75	28,185	91.87
600.00	27,220	88.72	27,315	89.03
180.00	25,430	82.89	25,715	83.82
Price cap (€/MWh)	Decentralised outcome without a price adder (MW)	% Optimum Capacity	Decentralised outcome price adder (MW)	% Optimum Capacity
(b) Effect of price-adder, with 20 firms				
3005.05	28,340	92.37	28,978	94.45
1800.00	28,150	91.75	28,978	94.45
600.00	27,220	88.72	28,892	94.17
180.00	25,430	82.89	28,822	93.94

as the market remains concentrated, regulatory intervention remains necessary to ensure some degree of consumer protection.

Appendix A. Supporting Information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.enpol.2009.01.038.

References

- Fabra, N., de Frutos, M-A., von der Fehr, N-H., 2008. Investment incentives and auction design in electricity markets, CEPR Discussion Papers # 6626.
- García-Díaz, A., Marín, P.L., 2003. Strategic bidding in electricity pools with short-lived bids: an application to the Spanish market. *International Journal of Industrial Organization* 21 (2), 201–222.

- Grimm, V., Zoettl, G., 2006. Capacity choice under uncertainty: the impact of market structure," Working Paper Series in Economics, University of Cologne (available at: <<http://www.wiso.uni-koeln.de>>).
- Gabszewicz, J., Poddar, S., 1997. Demand fluctuations and capacity utilization under duopoly. *Economic Theory* 10, 131–146.
- Red Eléctrica de España, 2006. El sistema eléctrico español. Informe 2006. available at: <http://www.ree.es/sistema_electrico/pdf/infosis/Inf_Sis_Elec_REE_2006.pdf>.
- von der Fehr, N-H., Harbord, D.C., 1997. Capacity investment and competition in decentralized electricity market, Department of Economics Working Paper, University of Oslo.
- Wolak, F.A., Patrick, R., 1997. The impact of market rules and market structure on the price determination process in the England and Wales Electricity Market, POWER Working Paper PWP-047, University of California Energy Institute.

Web Sites

- OMEL (Spanish market operator): <www.omel.es>.
- Red Eléctrica de España (REE, Grid Operator): <www.ree.es>.