OLIVEIRA, F.S. and COSTA, M.L.G. 2018. Capacity expansion under uncertainty in an oligopoly using indirect reinforcement-learning. *European journal of operational research* [online], 267(3), pages 1039-1050. Available from: https://doi.org/10.1016/j.ejor.2017.11.013

Capacity expansion under uncertainty in an oligopoly using indirect reinforcement-learning.

OLIVEIRA, F.S. and COSTA, M.L.G.

2018

Capacity Expansion under Uncertainty in an Oligopoly using Indirect Reinforcement Learning

Fernando S. Oliveira, NUS Business School, National University of Singapore, Singapore, bizfmdo@nus.edu.sg

Manuel L. G. Costa, Center for Economics and Finance, Faculdade de Economia, Universidade do Porto, Portugal, mlcosta@fep.up.pt

---

Abstract: We model capacity expansion in oligopolistic markets, with endogenous prices, under uncertainty, considering multiple production technologies. As this environment is complex, capacity expansion is the outcome of a learning process by individual. We propose indirect reinforcement learning to model the interaction between price determination and capacity decisions, in the context of an oligopoly game. We apply our model to the analysis of the Iberian electricity market, considering multiple technologies, focusing on how subsidies, the price of CO_2 emission and gas affect the capacity expansion policies.

Keywords: OR in Energy; Capacity Expansion; Computational Learning; Electricity Markets; Oligopoly.

---

1. Introduction

The capacity expansion problem concerns the determination of the size, location and timing of buying additional capacity. It inherently occurs in a context of uncertainty and incomplete information. For example, in the liberalized electricity industry, the access to cheaper gas (in the UK in the 1990s and the USA in the 2000s) and governmental subsidies (e.g., feed-in tariffs) for renewable energy sources (hydro, solar and wind power) have led to surges of capacity expansion in some technologies, with an unforeseen decrease in the revenue of other technologies.

We analyze the capacity expansion problem in an oligopolistic market with endogenous prices, dealing with the option to invest (and divest) in different technologies, non-linear and stochastic production functions, lumpy capacity expansion decisions (with a binary structure and discontinuous profit functions), in a multi-period planning horizon, and multiple players. Given the very large number of variables involved in computing the solution of the game, the integer nature of some of the variables, and the non-linear demand and cost functions, the possibility of multiple equilibria renders the problem not amenable to an analytical solution.

In modeling such a complex environment, we need to raise a few questions. How can real world capacity expansion decisions be modeled when a full optimization process is not computationally tractable? Is it possible to construct an alternative model of capacity expansion based instead on a learning process? But if so, how can a learning model be designed and adapted in order to address our problem, and how to implement it attending to the fact that the number of opportunities to interact with the environment and receive feedback from the decisions made is relatively limited? Finally, in the case of capacity expansion when several firms are learning simultaneously is it possible for equilibrium to emerge?

In order to address these questions (and this is our main methodological contribution) we develop the indirect reinforcement learning algorithm as the cornerstone of a dynamic multi-period model that simultaneously computes the capacity expansion decisions by firms, and price determination in an oligopolistic industry. The algorithm allows individual firms to learn how much to produce and when, and how, to expand capacity, in the context of price and cost uncertainty and, therefore, providing a better understanding of real world capacity expansion problems. Each firm learns the value of a possible capacity decision by observing the evolution of the market prices and how these may be affected by the different options available. This learning process takes place in an environment where the investment decisions occur a limited number of times during the planning horizon. The algorithm is explained in detail in section 3. We analyse its main properties, showing that the learned behavior converges to equilibrium.

As said before, given the nature of our capacity expansion problem, the analytical option is inadequate, and instead a learning model is proposed in order to generate equilibrium predictions. It should, however, be noticed that both methods require a vast degree of information, in particular information about the interaction game as well as about expected prices. In this regard, while analytical methods subsume that agents are capable of solving the model and of anticipating the possible equilibria and outcomes, in the learning method adopted here the possible equilibrium outcomes are constructed as the result of a learning dynamics, in the course of which agents interact repeatedly.

We apply the indirect reinforcement learning algorithm to the analysis of an electricity market as this is a particularly challenging industry: there are several production technologies (coal, hydro, natural gas, nuclear, solar and wind power) all of which have their own technical constraints; regulation is important (and subsidies crucial to the survival of many firms); the production costs are uncertain; there is an oligopolistic market power by incumbent firms. We present a case study on the Iberian electricity market for two major reasons: it is a liberalized market in which all the information is publicly available and it is one of the electricity markets in the world with the largest percentage of generation by renewable sources.

Specifically, we explore different scenarios for the level of subsidies, CO_2 emission prices, and for the price of natural gas. Moreover, we model the interaction between the electricity pricing and the capacity expansion (and divestment) decisions, over 20 years, considering multi-stage capacity expansion, and demand and cost uncertainty. The initial endowments of capacity per player, the demand and cost parameters, and the CO_2 prices are taken as exogenous, whereas the electricity prices, generation, and the investment decisions by each player are endogenous to the model.

Next, in section 2 we present a review of the literature on the capacity expansion problem focusing, particularly, on the electricity industry. In section 3 we develop the indirect reinforcement learning algorithm. In section 4 we present the model of the interaction between pricing and capacity expansion in oligopoly. In section 5 we illustrate the workings of the model in the Iberian Electricity Market. Section 6 concludes the article.

2. Literature Review on Capacity Expansion

As this article addresses the capacity expansion problem, in an oligopolistic market, under uncertainty and incomplete information, with an application to a liberalized electricity market, the literature review is focused on the oligopolistic capacity expansion problem with learning. The classic capacity expansion literature has considered the size, timing and location of additional capacity: the initial model was deterministic and focused on minimizing the discounted cost of all the required expansions to meet a known demand by a single producer. Extensions to this problem have considered multiple products, multiple periods, the investment timing, and demand uncertainty, e.g., Van Mieghem (2003) and Julka et al. (2007). The classic capacity expansion problem in electricity markets has also been attempted from the perspective of a single producer using large-scale mixed-integer linear programming (e.g., Chaton and Doucet, 2003; Kazempour et al., 2011).

The impact of learning on the optimal policy is an important issue in the capacity expansion literature, and it addresses two main aspects. A first one regards how learning about demand, over time, impacts the optimal policy (e.g., Burnetas and Gilbert, 2001). A second aspect regards the impact of learning by doing on production costs and performance in general (e.g., Bolton and Katok, 2008) and, in some cases, it analyzes when learning justifies keeping production in-house, instead of outsourcing it (e.g., Anderson and Parker, 2002; Gray et al., 2009; Xiao and Gaimon, 2013). Miller and Park (2005) have recognized the importance of the learning process in the multi-stage capacity expansion problem, using statistical learning in the context of real options with a single decision maker, to the valuation of uncertain irreversible investments.

The capacity expansion issue has been analyzed with endogenous pricing, in a monopoly with demand uncertainty (Bish and Wang, 2004). Goyal and Netessine (2007) have considered the capacity expansion game in the context of technology choice, in a duopolistic industry with stochastic price-dependent demand. The problem of investing in electricity generation has also been addressed using the Cournot model (e.g., Pineau and Murto, 2003; Murphy and Smeers, 2010; Pineau et al., 2011; Filomena et al., 2014; Bunn and Oliveira, 2008, 2016).

Real option games are an important approach used to analyze the capacity expansion problem in oligopolistic industries (e.g., Smit and Trigeorgis, 2004; Fang and Whinston, 2007;

Chevalier-Roignant et al., 2011). In the context of electricity markets, Siddiqui and Takashima (2012) have compared the impact of capacity switching options under monopoly and duopoly in deregulated industries, taking into account lumpy capacity expansions, and showing that the duopolist value function is always lower and decreasing with uncertainty. Siddiqui and Fleten (2010) have analyzed the options of a firm choosing between alternative technologies that do not use fossil fuels when compared to adopting a newly developed unconventional alternative. Also Fleten and Nasakkala (2010) have analyzed capacity expansion in gas-fired electricity plants, proposing a new method to compute the upper and lower bounds on the plants' value.

The setup of the capacity expansion problem is quite complex since we are in the presence of non-linear and stochastic production functions coupled with incomplete information, lumpy capacity expansion decisions, a multi-period planning horizon, interactions between the decisions of different players in the industry (with the possibility of multiple decision profiles). In order to deal with such complexity, simulation methods have been used to address this problem, namely discrete-event simulation (e.g., Franke et al., 2006), system dynamics (e.g. Georgiadis and Athanasiou, 2013), and agent-based simulation (e.g., Bunn and Oliveira, 2008, 2016; Ehlen et al., 2007; Anderson and Cau, 2009; Kimbrough and Murphy, 2013). In fact, currently, there is no optimization solver able to address a problem as the one defined in this paper; the ones currently used to solve capacity expansion problems in games require linear production and demand functions, continuous decision variables, and known distributions for the uncertainty parameters and, most importantly, are limited by the size of the problem it can solve due to memory constraints and computational limitations that still exist.

In this article, we use agent-based simulation as a framework for the analysis of the capacity expansion problem as this modeling technique can handle discrete and decentralized decisions, non-linear costs, isoelastic demand, in multi-period horizons. This is an important feature of the agent-based models as it allows a much better description of the problems faced by the electricity market. We, therefore, adhere to the bounded rationality paradigm in explaining investment in electricity markets as the perfect-rationality models of investment suffer from serious problems in the simplifications required to parameterize the model (as summarized in Oliveira, 2008): the linear demand assumption, which is intended to be an approximation of the actual unobserved demand function, leads to very large errors in forecasting production and prices (e.g., Caginalp, 2005) and to systematic underestimation of prices, when production deviates from the expected levels. Coupled together with capacity constraints (e.g., Puu and Norin, 2003) and, possibly, non-linear cost functions, price caps, and subsidies (e.g., Gabriel et al., 2013, Ch. 8), the non-linear demand leads to the possibility of multiple-equilibria, generalized Nash equilibria, and non-linear dynamics which adaptive methods are more suitable to analyze: in practice, when using linear demand and linearized costs functions the researchers need, typically by trial and error, to set up the parameters so that the solution falls within an acceptable range: this would correspond to pre-define the solution before the experiment is run. Non-linear optimization problems on the other hand need to set up the model so that it is solvable, i.e., the model converges to a "possibly local" equilibrium. Adaptive methods (as used in the multi-armed bandit problems, and employed in Puu and Norin, 2003) have the advantage of allowing a better description of the problem being analyzed without the need to pre-impose a desirable solution.

Regarding the problem of learning in environments in which there are a limited number of interactions between the agents, the first articles to discuss this were Bunn and Oliveira (2008, 2016) in the context of the trading of electricity generation plants. They modeled agents that were able to learn the probability of trading an asset, in a given auction, and used dynamic programming (with a finite planning horizon, which was used to capture bounded rationality) to build the optimal trading plans for the different firms, and, most importantly, still used the Cournot model of competition to endogenize the electricity prices and plant valuations. In our article, we extend their work by analyzing the investment problem and, crucially, by abandoning the Cournot model of electricity pricing and allowing the agents to learn both the plant valuations and the endogenous electricity prices. To this effect we have developed a new algorithm, the indirect reinforcement learning.

The indirect reinforcement learning algorithm consists in a learning process whereby each agent, when interacting repeatedly with the environment, arrives at an optimizing outcome individually, given the data and uncertainties that define the problem as well as, in the present situation, the effect of their capacity decisions on future prices. Therefore, we use the *n*-armed model to find stationary states of learning processes, as it allows the study of complex problems for which the analytical optimization is too cumbersome. Consequently, learning has no real time interpretation but, as a method, it is instrumental to the task it is meant to accomplish, that is, to find stationary states of the capacities decision problem.

3. An Indirect Reinforcement Learning Algorithm for Capacity Expansion Games

We now present the *n*-armed bandit algorithm (e.g., Sutton and Barto, 1998), a repeated game in which at each iteration *t* each player chooses a given decision to maximize profit. Given the repeated nature of the game the player can learn, over time, the expected value of choosing each one of the possible decisions. The *n*-armed bandit algorithm has been used to model electricity markets, namely by Bunn and Oliveira (2003) in order to study the ability of British Energy and

AES to profitably manipulate market prices, in the context of a court case brought by the energy regulator in the UK electricity market.

Let $\pi_t(a)$ stand for the expected profit from a decision *a*, at iteration *t* (*a* stands for a decision, for example, to invest or to retire an electricity plant of a given technology). This means that, at any given iteration, for all the technologies available and installed capacity, the firm decides which ones to shut down and in which ones to invest. These are binary decisions. Furthermore, let u_{t+1}^{a} stand for the profit received, at iteration *t*+1, by taking decision *a*. An agent computes an estimate of the expected value of *a*, at iteration *t*+1, $\pi_{t+1}(a)$, using (3.1), with the initial value $\pi_{t=1}(a) = u_{t=1}^{a}$.

$$\pi_{t+1}(a) = \pi_t(a) + \alpha\left(u_{t+1}^{a} - \pi_t(a)\right), \ \forall a . \tag{3.1}$$

Equation (3.1) represents an exponential smoothing of past rewards with a weight-factor α, the learning rate, such that $0 < \alpha \leq 1$. This is an exogenous parameter that characterizes the learning process used by a given firm. As the number of repetitions of the game converges to infinity the true value of decision *a* is learned.

However, this algorithm has an unsatisfactory behavior when modeling capacity expansion. The problem arises from the fact that by deciding an agent has a non-negligible effect on his environment and on his expected payoffs. Whereas in a standard *n*-armed bandit game the choices of the player do not change the rewards received, the same is not true in a capacity expansion game. For example, a capacity expansion in a nuclear electricity plant will most likely have an impact on prices and on the rewards of future capacity expansions. In order to model capacity expansion as an evolutionary learning process, we are required to develop a reinforcement learning algorithm which takes this fact into consideration. According to the model properties of the *n*-armed bandit algorithm in equation (3.1), in order to allow for a firm to learn the true value of a decision we need to give the algorithm time to converge. Therefore, a way to extend the *n*-armed bandit algorithm to model decisions that lead to a structural change is to introduce a mechanism that controls for equilibrium, so that a firm only takes effectively an action after equilibrium occurs.

Let us now see how the *n*-armed bandit algorithm can be extended in order to deal with structural changes. Following the *n*-armed bandit algorithm, at iteration $t+1$, using equation (3.1), an agent estimates the value of a capacity expansion opportunity a, $\pi_{t+1}(a)$. We control for the equilibrium of the learning algorithm using equation (3.2), which smoothed the changes in $\pi_{t+1}(a)$, and in which $\Delta\pi_t(a)=\pi_t(a)-\pi_{t-1}(a)$; $\Delta\pi_t(a)$ stands for the change in the expected value of decision a, at iteration t. Let $W_t(a)$ stand for the estimate of the change in expected value of decision a. The variable $W_t(a)$ enables the detection of structural changes in the value of assets and the initial value is $W_{t=1}(a)=u^a_{t=1}$. We further decide when the estimation of the value of a given decision is correct enough for a new decision to be taken. In the case of the capacity expansion problem, we need to estimate the value of the capacity expansion close enough in order for the firm to choose this capacity expansion when the value is positive.

$$W_{t+1}(a)=W_t(a)+\alpha[\Delta\pi_t(a)-W_t(a)],\ \forall a. \tag{3.2}$$

Let δ represent the maximum valuation error, an exogenous parameter (close to zero) delimiting the neighborhood within which an estimated value is considered to be correct, as described by equation (3.3), in which u represents the true value we are estimating, i.e., it is the profit received by executing decision *a* when $u_t^a = u_{t+1}^a = ... = u$. δ defines the level of certainty required by a firm before investing in (or retiring) a given asset. In standard discounted cash flow analysis, a firm invests if $\pi_t(a)$ is positive. However, it is well known from the real options literature that there is a value in delaying capacity expansion, a fact captured by this decision rule, that by delaying capacity expansion a firm obtains more information, allowing for a better valuation. In this case, the lower δ the more conservative is a firm (as it demands better forecasts before investing). In Proposition 3.1 we derive the formula to compute the profit π_t, from which we conclude that in a stationary environment the firm is able to better approximate the correct value of a given action by interacting with the environment an infinite number of times.

$$\left|\frac{u - \pi_t}{u}\right| \leq \delta \qquad (3.3)$$

Proposition 3.1: *Let* $\pi_{t=0} = 0$, a*fter t iterations the learned value of* π_t *is* $\pi_t = u\left(1 - (1-\alpha)^t\right)$. [The proof is in the appendix.]

Given that a firm cannot wait an infinite time to evaluate a decision, using the result in Proposition 3.1 we can now derive a first requirement for testing the convergence of the agent's estimate of the value of decision *a*, as presented in Corollary 3.1, which follows immediately by replacing $\pi_t = u\left(1 - (1-\alpha)^t\right)$ into (3.3).

Corollary 3.1: *The estimate of the value of decision a converges if and only if* $(1-\alpha)^t \leq \delta$.

From Proposition 3.1 and Corollary 3.1 we can now derive a condition under which the estimated value of a decision *a* is correctly evaluated for a given market structure, Proposition 3.2. This Proposition shows that the estimated change in the value of the capacity expansion needs to be close to zero (and that the lower the learning rate the closer to zero the percentage change is required to be) for a firm to approximate the value of a decision *a*, for a given decision *a*. This is a rule based on the statistical significance of the estimated value associated to a capacity expansion: if the value created is not significantly different from zero there is no statistical evidence that the investment creates value. An important outcome of this proposition is that it clearly states that it is not only the expected profit that determines the timing of triggering a decision but also, most importantly, the reliability of the estimates of the value associated to that decision.

Proposition 3.2 (Capacity expansion rule): *A firm invests if the estimated value of the capacity expansion is positive and if* $\left|\frac{W_t}{\pi_t}\right| \leq t\alpha^2\delta$. [The proof is in the appendix.]

As referred to above, a second problem with the straightforward use of the *n*-armed bandit algorithm is that, to model capacity expansion, we need to forecast the impact of a marginal capacity expansion on the current price. As the *n*-armed bandit algorithm only looks at past rewards it is not adequate to forecast the future in our case, as a capacity expansion may lead to a structural change in the stream of profits received by an agent. In the following section this price effect is incorporated.

4. A Model of Learning for Capacity Expansion in Oligopolistic Markets

In this section we present an evolutionary model of capacity expansion in oligopolistic markets. We consider the possibility of capacity expansion in *J* different production technologies: each firm *i* can own a portfolio of several production technologies (and several production plants of the same technology), which is adjusted over time. Regarding the objective of endogenous price formation in an oligopoly, we use quantities produced as the decision variable to capture the interaction between the different producers in the oligopoly, as producers decide how much to sell, as the learning process evolves, at each iteration, from each one of their plants, at any given time period.

It is well known that uncertainty impacts the investment decisions (e.g., Plambeck and Taylor, 2013). For this reason, in our model we consider demand uncertainty and marginal cost uncertainty, as described next, and production uncertainty from intermittent energy sources and regulatory uncertainty, as described in section 5.

For each demand block h of the year (an aggregation of many trading hours with similar levels of demand) we have a price. In this context, model the total demand D_{hg}, at each demand block h, in year g, at iteration t, as isoelastic, $D_{hgt} = a_{hgt} + k_h P_{hgt}^{-e_h}$; a_{hgt} and k_h are parameters used to adjust the level of demand to the price P_{hgt}, and e_h is the price elasticity of demand. We model demand uncertainty in the demand intercept, regarding changes both from year to year (index g) and from iteration to iteration (index t); this uncertainty is modeled by a normal distribution, $a_{hgt} \sim N\left(\overline{a_{hg}}, \sigma_{hg}^2\right)$, the mean and variance of which are allowed to be time dependent. It then follows that the electricity market price at block h, in year g, and iteration t, P_{hgt}, is computed using equation (4.1), in which Q_{hgt} is the total quantity sold in the market.

$$P_{hgt} = \left(\frac{k_h}{Q_{hgt} - a_{hgt}} \right)^{e_h^{-1}} \qquad (4.1)$$

Therefore if, in the short-term, the aggregated level of production determines the prices, in the long-term the level of installed capacity crucially affects them, as this constrains ongoing production levels. Existing firms carry out capacity expansion. Total production costs include (*i*) a variable component, which is a direct function of the produced quantities (especially fuel cost), (*ii*) a fixed operational component, which is independent of the level of production, and (*iii*) capacity expansion costs, which are sunk in the short-term but crucial for the capacity expansion and retirement decisions.

In equation (4.2) we represent the company *i*'s long-term profit function to be maximized, at a given iteration *t*, where: *h* is an index for a demand period, in a total of *H* blocks of demand per time period *g* (a period *g* is typically a year and includes many trading hours); ρ is a discount factor; C_{jgt} is the marginal production cost of technology *j*, in year *g,* at iteration *t*; Q_{ijhgt} stands for *i*'s the total production, from technology *j,* at block *h*, year *g* and iteration *t*; F_{jgt} is the fixed cost, in year *g* and iteration *t*, and per unit of capacity, of technology *j*; and K_{ijgt} is the total installed capacity of technology *j*, by company *i,* in period *g* and iteration *t* (these fixed costs include both capacity expansion and operational costs). Firm *i*, at iteration *t*, learns the value of the level of production from each technology, and the level of capacity required in each year, by comparing the profits at iteration *t*-1 with the profits at iteration *t*.

$$u_{it} = \sum_{g=1}^{G} \rho^{g} \left[\sum_{j=1}^{J} \left[\sum_{h=1}^{H} \left(P_{hgt} - C_{jgt} \right) Q_{ijhgt} - F_{jgt} K_{ijgt} \right] \right] \tag{4.2}$$

Note that in (4.2) the parameters for the fixed cost F_{jgt} are, in general, time dependent; the marginal costs C_{jgt} are also time dependent and, in the analysis in this article, are considered stochastic. The specific processes followed by the marginal costs are described in section 5. The capacity per technology is time dependent and, thus, available capacity is stochastic; as for technologies such as hydro, solar and wind, capacities change in a stochastic way, the stochastic processes used to represent the available capacities are also left to section 5. Given the complexity of this problem that involves the conjoint solution of production and capacity expansion decisions during the entire planning horizon, by all the player in the industry, we numerically approximate local-Nash equilibria using Monte-Carlo simulation, based on the indirect reinforcement learning algorithm presented in section 2.

This Monte-Carlo simulation, which can be used to compute the value of a given policy in the presence of uncertainty, is a very intensive search procedure. Another reason to use Monte-Carlo simulation to compute the possible equilibria of the system is that oligopolistic industries are characterized by the presence of multiple equilibria and a very large state space. These are local-Nash equilibria, as they are computed using reinforcement learning, which ensures local optimality.

The simulation starts with the parameterization of: a) total capacity per company and technology and discrete investment possibilities; b) investment, fixed and variable costs for each technology; c) load duration curve for the typical year and the parameters for the isoelastic demand; d) subsidies for the different technologies, if these exist, and the CO_2 emission prices; e) learning rate and maximum valuation error used in the indirect reinforcement learning; f) number of years in the planning horizon; g) length of the warm-up period, during which the firm learns how much to sell in the initial market structure. In sections 4.1 and 4.2 we describe the Monte-Carlo simulation, which is summarized at the end of the section, in Table 4.1.

4.1 Learning the Quantity to Produce

In this subsection we analyze how firms can learn, in an evolutionary model, the quantities to sell in a market. The quantities sold, at any given demand block, are learned by each agent, for each technology, for any given demand block *h*. The proportion of the total capacity to be sold in each market period, B_{ijhgt}, is computed using (4.3.a), for each agent *i* and technology *j,* at block *h*, and time period *g* and iteration *t*, in which B^{+}_{ijhgt} stands for the company's *i*'s best indirect reinforcement learning estimate of the optimal proportion of capacity from technology *j*, at time g, block *h*, and iteration *t*, computed using (4.4). The ε_{ijhgt} stands for the exploration the agent may attempt at hour *h* (and year *g*), at iteration *t*, for technology *j*, by choosing to sell a proportion of capacity different from the estimated optimal, this choice being made with a small probability (usually below 10%), e.g., Sutton and Barto (1998). Finally, the quantity sold is computed using equation (4.3.b).

$$B_{ijhgt} = B^{+}_{ijhgt} + \varepsilon_{ijhgt} \tag{4.3.a}$$

$$Q_{ijhgt} = B_{ijhgt}.K_{ijgt} \tag{4.3.b}$$

Equation (4.4) describes how a firm learns the optimal proportion of available capacity to sell, from technology *j*, in demand block *h*, at year *g* and iteration *t*, using indirect reinforcement learning. α_t is the learning rate at iteration *t*, and Ψ_{ijhgt} stands for an indicator variable that equals 1 if the profit of firm *i*, at period *g*, and demand block *h*, increased, -1 if the profit is lower, and zero if there was no change in profit, at iteration *t,* when compared to iteration *t*-1. We require this indicator function of profit in order to learn the optimal quantities to sell, when using indirect reinforcement learning. The learning rate is larger at the start of the simulation in order to allow the firms to better explore the space of possible pricing policies. As described in (4.5), the learning rate is $1/t$, if the simulation is in the warm-up period ($\text{t} \le t_w$), and $\underline{\alpha}$ thereafter.

$$B^{+}_{ijhg(t+1)} = B^{+}_{ijhgt} + \alpha_t \Psi_{ijhgt}\left[B_{ijhgt} - B^{+}_{ijhgt}\right] \tag{4.4}$$

$$\begin{cases} \alpha_t = 1/t & \text{if} \quad \text{t} \le t_w \\ \alpha_t = \underline{\alpha} & \text{if} \quad t > t_w \end{cases} \tag{4.5}$$

First, in step 1.a), each firm decides the proportion of the capacity to sell at each one of the blocks of demand, during the entire planning horizon. These proportions have two components, an optimal ratio estimated from past-experience, B^{+}_{ijhgt}, and a second part which is the firm's attempt to explore improvements to the current policy, ε_{ijhgt}. The algorithm proceeds in step 1.b) with the calculation of the equilibrium prices, P_{hgt}, for each one of the *h* demand blocks in the planning horizon, by using equation (4.1), an approach similar to the one used in Kazempour et al. (2011). In step 1.c), for each firm, technology, block demand, year and iteration, we compute the profit gained and compare it with the one in the previous iteration to compute the indicator function Ψ_{ijhgt} used to update the quantities to sell in step 1.d). In step 1.d) the proportions of available capacity to sell are updated taking into account their previous level, the change considered, and the indicator function that evaluates the effect of that change on profit. If the change leads to an increase in profit and the quantity is above (below) average, the decision of quantity to sell is updated higher (lower); if the change leads to a decrease in profit and the quantity was above (below) average, the decision of the quantity to sell is decreased (increased).

4.2 Learning when to invest in capacity (or to retire capacity)

When considering the option to invest (or retire), for each type of plant, an agent computes the marginal value of a given investment (or retirement) during the lifetime of the plant. Then, only investments (retirements) with expected positive value are considered. An agent invests in (retires) a given technology if the value of the post-investment (post-retirement) portfolio of plants is higher than the current value of the portfolio of plants. Let u_{ijt} stand for the estimate of the profit of agent *i*, from technology *j,* at iteration *t,* without considering the option to invest (or retire) plants; u_{ijt}^{+} represent the profit of agent *i*, from technology *j*, at iteration *t,* after he exercises the option to invest (or retire) a plant of technology *j*; π_{ijt}^{+} represent the expected value of the investment in (or the retirement of) technology *j*, by agent *i,* at iteration *t*; $\Delta\pi_{ijt}^{+}$ represent the change in the expected value of an investment (or retirement) in technology *j*, at iteration *t,* by agent *i.* The agent learns when to invest (retire) by using equations (4.6)-(4.8).

$$\pi_{ij(t+1)}^{+} = \pi_{ijt}^{+} + \alpha_t \left(\left(u_{ijt}^{+} - u_{ijt} \right) - \pi_{ijt}^{+} \right), \ \forall j . \tag{4.6}$$

$$W_{ij(t+1)}^{+} = W_{ijt}^{+} + \alpha_t \left(\Delta\pi_{ijt}^{+} - W_{ijt}^{+} \right), \ \forall j . \tag{4.7}$$

$$\left| \frac{W_{ijt}^{+}}{\pi_{ijt}^{+}} \right| \le \alpha_t^{\,2} \delta \tag{4.8}$$

First, each firm calculates the smoothed estimates of the change in value associated with investing (divesting) in a given technology, $W_{ij(t+1)}^{+} = W_{ijt}^{+} + \alpha_t \left(\Delta\pi_{ijt}^{+} - W_{ijt}^{+} \right)$. This estimate is a guide to the volatility of the market structure, as when there is a new investment (divestment), or a change in the pricing policies, these smoothed estimates change, whereas when the market structure is stable the smoothed values converge. After convergence the players are able to assess if a given investment (divestment) should take place. The question then becomes to determine when a given smoothed value has converged. Using the indirect reinforcement learning presented in section 2, the threshold can be identified using the inequality $\left| \frac{W_{ijt}^{+}}{\pi_{ijt}^{+}} \right| \le \alpha_t^{\,2} \delta$. When this inequality is met then a given investment (divestment) is triggered if, as described in section 2.2, $\pi_{ijt}^{+} > 0$ and (4.8) is verified. This decision rule is an application of the discounted cash flow approach in which the value of a project depends on all the additional cash flows $\left(u_{ijt}^{+}-u_{ijt}\right)$ that follow from the project during its entire time horizon (Brealey and Myers, 1991, p. 96).

However, it is well known that this simple rule may be incomplete if the firm can delay the decision. The real options approach to project valuation shows that the option to delay a decision can have a positive value (e.g., McDonald and Siegel 1986; Pindyck 1991). This positive value can be obtained when during the delay period the firm receives information that corrects its valuation. This is calculated by using (4.8). By investing, a player changes the value of the capacity expansion for other players.

By using a real-options approach an agent can delay the capacity expansion in order both to diminish uncertainty regarding the behavior of his rivals and to use an adjusted estimate of the value of a possible capacity expansion decision. This is implemented by equations (4.5)-(4.8), with the advantage that in this evolutionary algorithm complete knowledge of the value of capacity expansion decisions is not assumed, instead these are learned by interacting with the environment. Table 4.1 presents a summary of the simulation used to solve the evolutionary capacity expansion model.

A stationary state is reached when, for every player and technology, the estimated value of an investment (retirement) is not positive and it is accurate, i.e., when at iteration *t*: $\pi_{ijt}^{+}\leq 0$ and equation (4.8) is verified. There is a *stationary state* when no agent wishes to change his portfolio of plants. This outcome of the learning algorithm is robust from a learning perspective and, therefore, it qualifies as equilibrium given that, when this state is reached, the learning agents would not change decisions. In fact, if the case is that every agent stops changing his capacity choice, the case must be that he does not perceive any possibility of benefiting from any potential decision he might consider; if all the agents attain this position, this is a stationary state, and it qualifies as a stable equilibrium state.

Table 4.1: Simulation of the Evolutionary Capacity Expansion Model

Initialize: a) capacities per company and technology; b) costs of the different assets; c) demand function and the load duration curve; d) subsidies and CO_2 emission prices; e) learning parameters; f) planning horizon; g) warm-up period length.

While last iteration is not reached:

1.) Every firm learns the initial offer quantities, per technology, for each demand block.

1.a) Compute the quantity sold per player and demand block of the planning horizon

$$B_{ijhgt} = B^{+}_{ijhgt} + \varepsilon_{ijhgt} .$$

$$Q_{ijhgt} = B_{ijhgt} . K_{ijgt}$$

1.b) Compute the prices per demand block for the planning horizon

$$P_{hgt} = \left(\frac{k_h}{Q_{hgt} - a_{hgt}} \right)^{e_h^{-1}}$$

1.c) Compute the profits, per player, and technology, for every block of demand

$$u_{ijhgt} = \left(P_{hgt} - C_{jgt} \right) Q_{ijhgt} - F_{jgt} K_{ijgt}$$

1.d) Learn the quantity to be sold at every demand block in the planning horizon

$$B^{+}_{ijhg(t+1)} = B^{+}_{ijhgt} + \alpha_t \Psi_{ijhgt} \left[B_{ijhgt} - B^{+}_{ijhgt} \right]$$

2.) Every firm learns when to invest/divest in each one of the technologies.

2.a) Compute the present value of the planning horizon profits, per player *i*, and technology *j*, at iteration *t*, $u_{ijt} = \sum_{g=1}^{G} \rho^g \left[\sum_{h=1}^{H} \left(P_{hgt} - C_{jgt} \right) Q_{ijhgt} - F_{jgt} K_{ijgt} \right]$

2.b) Compute the present value of the profit for an investment/divestment, per player *i* and technology *j*, at iteration *t*,

$$u^{+}_{ijt} = \sum_{g=1}^{G} \rho^g \left[\sum_{h=1}^{H} \left(P^{+}_{hgt} - C_{jgt} \right) Q^{+}_{ijhgt} - F_{jgt} K^{+}_{ijgt} \right]$$

2.c) Update the value of investing/divesting, for every year in the planning horizon, for every technology

$$\pi^{+}_{ij(t+1)} = \pi^{+}_{ijt} + \alpha_t \left(\left(u^{+}_{ijt} - u_{ijt} \right) - \pi^{+}_{ijt} \right)$$

2.d) Decide capacity expansion timing

$$W^{+}_{ij(t+1)} = W^{+}_{ijt} + \alpha_t \left(\Delta \pi^{+}_{ijt} - W^{+}_{ijt} \right)$$

$$\left| \frac{W^{+}_{ijt}}{\pi^{+}_{ijt}} \right| \le \alpha_t^2 \delta$$

End while

5. Computational Experiments in the Iberian Electricity Market

In this section, we use the indirect reinforcement learning algorithm in the analysis of capacity expansion (and contraction) in the Iberian electricity market. The generation structure of this liberalized electricity market is summarized in Table 5.1 (the sources of which are REE, 2012; REN, 2012). In the "Other Operators" column we aggregate the many small generators in the Iberian Peninsula, all of which are considered as part of a competitive fringe, behaving as price takers.

Table 5.1: Capacity Installed in Spain and Portugal, Dec. 2011 (MW).

	Iberdrola	Endesa	Gas Natural Fenosa	EDP	Other Operators
Nuclear power	3373	3681	595	156	
Coal	1253	5294	2048	2640	3681
Fuel oil	157	520	157	1822	1706
Hydro	8619	4716	1901	5441	4557
CCGT	5893	3799	6998	3737	10369
Wind power	5875		1061	3143	14194
Solar power					15005

Our estimates of the generation and capacity expansion costs per technology, and CO_2 emissions are presented in Table 5.2 (which were assumed exogenous and deterministic in the model) and were based on Brinckerhoff (2011), MacDonald (2010), Ove Arup & Partners (2011), European Commission (2011), and Lenzen (2008). ASC with FGD represents the Advanced Super Critical (coal) technology with Flue Gas Desulphurisation; CCGT is the Combined Cycle Gas Turbine; OCGT stands for Open Cycle Gas Turbine; PWR is the Pressurized Water Reactor. In the case of hydro the number of available capacity expansion opportunities is limited: industry projections allow for a growth of capacity in the Iberian hydro system of less than 10 TW for the next 20 years.

We have modeled cost uncertainty. The marginal costs were assumed constant and known for all technologies, except for CCGT and coal: these two marginal costs were modeled using a normal distribution with standard deviation of 60% and 54% of the expected cost for CCGT and coal, respectively, and with a correlation coefficient of 93%, reflecting the properties of the historical data for gas and coal prices in Europe.

Table 5.2: Fixed operating and capacity expansion costs, and marginal costs, and CO_2 emissions, per technology.

Technology	Fixed Capacity expansion Cost (€/kW)	Fixed operating Costs per year (€/kW)	Economic Life (years)	Marginal Cost (€/MWhe)	Capacity (MW)	g(CO^2e) per kWhe
CCGT	770.9	31.6	30	27.5	850	577
Coal ASC-FGD	1893.2	49.1	35	30.5	1600	955
Nuclear PWR	3491.5	86.7	40	6.7	3300	62.5
Onshore wind	1751.8	39.4	24	0	100	21
Offshore wind	3273.1	87.8	24	0	200	21
Hydro >5 MW	2658.8	62.2	41	0	>5	15
Solar PV >50 MW	2475.6	24.2	25	0	>50	106

We have modeled production uncertainty for the renewable energies (hydro, solar, and wind). Solar power only runs during the day, and its production depends on the luminosity at any given time; wind power production depends on the wind speed and duration, which is very unstable throughout the year, and the time of the day, and it is a function of the location of the power plant; hydro generation is also dependent on the time of the year, and of the different uses the water may have (for irrigation, for example). In Table 5.3 we present the load factors per month for hydro, solar and wind power plants: this represents, given the historical data (REE, 2015; REN, 2015) for the Iberian Peninsula, the probability that a given MW of capacity is available at any given month and, therefore, the available capacity for these technologies is stochastic. It is assumed in this analysis that the available hydro capacity in a given month is independent from the previous month, as these load factors, which are historically used by the decision makers already take into account the probability of draughts in the following months.

Table 5.3: Load Factors per Type of Renewable Energy and per Month.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Load Factors (%)	Hydro	26	30.4	22.9	21	19	15	6.3	3.1	4.2	9.6	15	20.6
	Solar	9.8	16.6	20	22.7	25.6	25.2	27.4	35.7	20.1	18.9	14.4	9.4
	Wind	32	31	30	25	15	22	30	19	15	27	33	29

Finally, in order to model the evolution of the market we need to capture the profile and the expected behavior of demand. The typical demand in the Iberian electricity market is summarized in Figure 5.1 by the load duration curve for one year of hourly data since November 1st until October 31st. In it we have aggregated the hourly demands into 26 blocks, one per each different demand level. The demand per block follows a normal distribution centered on the demand reported in Figure 5.1 with a standard deviation of 3% of the respective mean.

In this study we have focused on the current technologies and the market structure to which the simulations converge represent the equilibrium given the current production costs. We also analyze the case of a significant drop in the price of natural gas in section 5.4. The structure of the model is flexible enough to allow for the inclusion of technological evolutions such as better and cheaper storage of electricity, which would make wind and solar plants to behave more like thermal plants as the firms can decide when to discharge the batteries. However, in order to include the use of storage in an interesting way, and incorporate the added flexibility of selling at any given demand block, several adaptations would be required: first, in order to capture the details of the timing for charging and discharging the batteries, the aggregation of demand in the load duration curve by blocks, that is used here, would have to be changed to an hourly load duration curve; second, a much finer representation would be needed of the technical features associated with the scheduling of the different types of plants such as ramp-rates or start-up costs (which are captured in the operational costs); third, the load factors (see Table 5.3) would have to be modified. That is to say, the inclusion of the use of storage in the system would have to capture salient features that the model in its present form cannot represent.

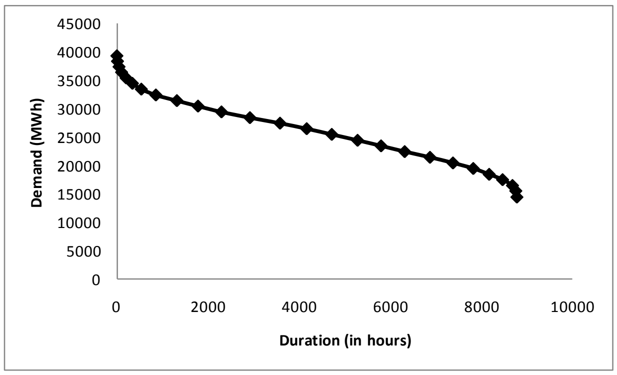

Figure 5.1: Load duration curve in the Iberian electricity market from November to October.

In all the simulations we have used a planning horizon of 20 years with 1 million iterations each in order to achieve convergence to the stationary state, the warm-up period was set at 25000 iterations, and the maximum valuation error was set at a very small value (0.001) in order to ensure that capacity expansion decisions are taken only when the valuation has converged. The results in sections 5.2-5.4 are reported for the stationary state.

5.1. Model Calibration and Sensitivity Analysis

In this subsection we assume a full subsidy scenario, with no carbon tax. We attempt to describe the procedure followed in order to calibrate the model. First, we have run the model with a fixed capacity for each player (fixed at the current level), in which the players learn the pricing policy only, and we have attempted different combination of minimum learning and exploration rates, and a level of aggregation of the load duration curve, so to ensure the model would converge to prices close to the ones in the Iberian market.

We then carried out an analysis of the sensitivity of the long-term solution, taking into consideration the initial parameters defined for the pricing policy. In Figure 5.2 we report the evolution of the total capacity installed in the system for the different scenarios, over 1 million iterations, in which Base represents the scenario with a minimum learning rate of 0.05, an exploration rate of 0.1 and discount of 0.9. All the other simulations use the same parameters as in the base scenario except for the specific parameter tested, Disc. 0.98 (discount factor 0.98), Disc. 0.95 (discount factor 0.95), Exp. 0.05 (exploration rate of 0.05), Exp. 0.2 (exploration rate of 0.2), Learn 0.2 (minimum learning rate of 0.2). The very large number of iterations (in which the firms observe the electricity prices, evaluate the different technologies and decide if the timing for a (d)investment has arrived) is required so that, given the small learning rate, the value of the different plants is approximated with high accuracy, avoiding needless investment and divestment decisions that, given the chosen parameters, are kept to a minimum, reducing the buyer's regret. As people learn continuously, every time they observe a price, which in many electricity markets happens at least 48 times a day, this algorithm is a first approximation to this learning process.

As can be seen from the results, these parameters have almost no influence on the final result as the total capacity converged to values around 145.5 GW (the difference between the different simulations is not statistically significant). The total quantities per technology, not reported, are also very similar in the different scenarios. (The model was implemented using Vensim, and in the simulations we used a desktop with 32GB of RAM and a 3.5GHz processor; the computational time for a 1 million iterations simulation was on average about 18 hours.)

Additionally, we have also tested the impact of different fixed operational and investment costs on the outcomes of the model. The results were as expected. The agents always invest more in a technology when the costs decrease significantly in comparison with the other. This behavior, however, is non-linear, as it may not appear for "small" cost reductions.

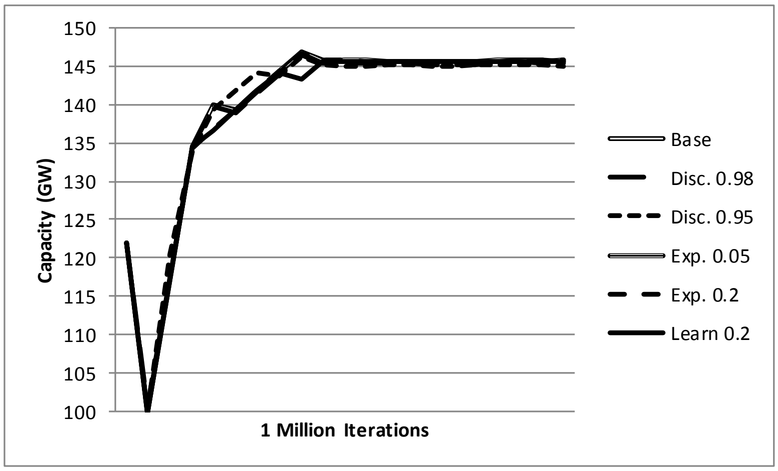

Figure 5.2: Capacity installed (GW) as a function of the discount factor, exploration rate and learning rate.

5.2. The Impact of Subsidies and Regulation on Market Structure

One of the features of the Iberian electricity market in the last decade is the presence of subsidies to renewable electricity generation. Under this regime these subsidies assume the form of a feed-in tariff (e.g., Garcia et al., 2012) as set by the Royal decree (Real Decreto) 1578/2008 for photovoltaic electricity and by the Royal decree 661/2007 for other renewable technologies: just until 2012 there was a feed-in tariff of 291 Euros/MWh for solar power, 79 Euros/MWh for wind power and 84 Euros/MWh for small hydropower. Other subsidies to power generation firms have been in place, namely compensation for lack of expected returns in free market supply and capacity payments (with effect mainly on gas, cogeneration and hydro plants). However, due to the economic crisis, these subsidies have been cancelled for all new renewable projects, and measures are being taken to remove or substantially reduce subsidies currently paid to the existing generation plants. For this reason, in this section, we discuss the impact of dropping these subsidies on the production structure of the Iberian electricity market.

In Table 5.4 we present the total capacity in the system, the capacity per technology, the utilization ratio and the average price (at the end of the simulation) as a function of the level of subsidies per MWh. In order to make the analysis more compact we present the subsidies as a proportion of their current level, for example, 75% means that the subsidies are decreased by 25% when compared to the base case, and zero means that there are no subsidies. A decrease in the subsidies leads to a reduction in the total capacity (the total capacity in the system is halved when the subsidies are removed). There is also an increase in the average price as the level of subsidies is reduced (the largest average price is about 44 Euros/MWh when the subsidies are decreased to about 25% of their current level), but the average price drops when there are no subsidies. Moreover, the subsidy increases the overall cost of electricity for society (as can be seen in column Price + Subsidy): with the current subsidy levels, the total subsidy per MWh is larger than the price paid for non-renewable electricity, but this is the scenario with the lowest CO_2 emissions. Therefore, the scenario with no subsidies reduces the price+subsidy by almost 50%, but, again, the CO_2 emissions will be five times higher.

This result is counter-intuitive. One would expect that the removal of the subsidies would lead to less capacity and higher prices (but prices rise by less than the loss in the subsidy). What we observe is that the removal of the subsidy is followed by capacity contraction and price reduction (when compared to the subsidies in the percentiles 25 to 75). The price reduction can be explained by two major factors: first, the price signals work better when there are no feed-in subsidies as, in this case, the generators can see clearly the effect of price changes on profits;

second, there is a selection effect, in the longer term, when there are no subsidies, as the most efficient technologies expel the inefficient ones from the market, and the price is set by these less expensive technologies. The lowest price, nonetheless, is achieved with the full subsidy and a very large capacity expansion.

Table 5.4: Capacity per technology as a function of the level of subsidies.

	Total Capacity Per Technology (GW)							Price +	Capacity	Utilization	CO_2 Emis.
Subsidy	CCGT	Coal	Hydro	Nuclear	Solar	Wind	Price (E/MWh)	Subsidy (E/MWh)	(GW)	Ratio (%)	(%)
100	0	0	35	17	36	57	26	56	145	21	20
75	19	0	25	8	24	14	38	48	91	31	98
66	18	0	25	8	24	14	41	48	89	34	94
50	21	0	5	8	24	14	42	46	72	39	107
33	22	0	5	8	24	14	43	45	70	40	112
25	22	0	5	8	24	14	44	45	68	44	112
0	20	0	5	11	7	12	29	29	55	56	100

Furthermore, there is a clearly and significant non-linear relationship between the level of subsidies and CO_2 emissions (reported in the last column as a percentage of a base case): the total emissions are 20% with the full subsidy from what they would be in the no-subsidy case. More important, the reduction to 75% on the level of subsidies has a very important impact (almost 5 times higher) on the level of emissions as it allows CCGTs to become profitable once again.

Table 5.4 contains also a summary of the evolution of the technological structure as a function of the subsidies. First, coal is never an option for generation in the Iberian electricity market, as this is an old technology that was installed before the access to cheaper natural gas was available and before the dash for the renewables, and it was mainly used as a ramp-up service at peak times. With the current level of subsidies (100%) only nuclear and the green technologies survive. Nonetheless, just a reduction of the subsidies to 75% of the current value leads to a significant change in investment opportunities, with the capacity expansion in CCGT.

From Table 5.4, we can also see that a subsidy of 25% of its current level is enough to ensure a capacity expansion in solar power equal to the one attained with larger subsidies. Overall it seems that hydro and solar-power are the two technologies that most benefit from the subsidies. Wind only benefits from the increase in subsidies from 75% to 100% of their current level as it is a very resilient technology, surviving even in the absence of subsidies. The utilization ratio (total production per year over the total installed capacity) decreases directly with the increase in the subsidies, attaining its lowest level (21%) with the current subsidy and its largest level (56%) when there is no subsidy. The major explanation for this result is the inefficiency generated by the subsidies that give incentives for capacity expansion in technologies with very low utilization ratios.

The effect of regulation *versus* competition on performance is summarized in Table 5.5, comparing the oligopolistic outcomes and the marginal cost price paradigm result (from perfect competition or in regulated monopoly, PC - Reg.Mon.). The regulated monopoly (perfect competition) prices are equal to the long-run marginal cost of the marginal technology at any given demand block, for every year in the horizon, and are significantly lower than the oligopoly ones. As reported in Table 5.5, subsidies have an important role on the reduction of CO_2 emissions, which are only 6% (under marginal cost pricing) and 13% (under oligopoly) of the total emissions when there are no subsidies under perfect competition. However, one should be aware that these significant reductions on CO_2 emissions are ultimately subsidized, at a significant cost for society, as summarized in column Price+Subsidy.

Table 5.5: Capacity per technology as a function of the level of subsidies and Market Structure.

Subsidy	Industry	Total Capacity Per Technology (GW) CCGT	Coal	Hydro	Nuclear	Solar	Wind	Price (E/MWh)	Price + Subsidy (E/MWh)	Capacity (GW)	Utilization Ratio (%)	CO_2 Emis. (%)
100	Oligopoly	0	0	35	17	36	57	26	56	145	21	13
0	Oligopoly	20	0	5	11	7	12	29	29	55	56	63
100	PC - Reg. Mon.	0	0	26	0	36	24	11	73	86	32	6
0	PC - Reg. Mon.	31	0	24	28	0	9	20	20	92	28	100

5.3. The Impact of the CO_2 Emission Permits on Capacity Expansion

Another interesting issue that may influence the long-term behavior of electricity markets is the CO_2 emission price (e.g., Garcia et al., 2012; Ambec and Crampes, 2012). In the previous section we have assumed, correctly, that this price is negligible at the moment. In this section we test the possible effects of CO_2 emission prices on the capacity expansion (and contraction) policies. We assume a base case with no subsidies, as this is the most likely future of the industry in two countries having very strained government budgets. In the history of the European Trading Scheme the CO_2 emission prices have fluctuated strongly, they have peaked at about 30 Euros per ton in 2005 and have been as low as 0.1 Euros per ton in 2007; in 2017 the CO_2 emission prices is ranging from about 3 to 8.5 Euros per ton. For this reason, in this section, we test the impact of the increase of the CO_2 emission price per ton from 0 Euros to 50 Euros/ton. In Table 5.6 we present the total capacity in the system, capacity per technology, utilization ratio, and prices, as a function of the CO_2 price (Euros/ton).

It is evident that the total capacity in the system increases with the CO_2 emissions price and that there is a significant jump in total capacity when this price increases from 10 to 20 Euros/ton. The reason for this increase in capacity is the need to replace the fossil fuel technologies by greener technologies that require more spare capacity with a lower utilization ratio. As presented in Table 5.6, the CCGT capacity decreases as a function of CO_2 emission prices, being replaced by hydro, solar and wind power (the increases of which more than compensate, in total capacity, the decrease in CCGT) and the utilization ratio falls sharply with the increase in the CO_2 price.

Table 5.6: Capacity per technology, total capacity, utilization ratio and electricity price as a function of the CO_2 Price (Euros/ton).

	Total Capacity Per Technology (GW)									
CO_2 Price (Euros/ton)	CCGT	Coal	Hydro	Nuclear	Solar	Wind	Price (Euros/MWh)	Capacity (GW)	Utilization Ratio (%)	CO_2 Emis. (%)
0	20	0	5	11	7	12	29	55	56	100
10	19	0	10	11	4	14	36	59	49	95
20	17	0	24	11	8	14	46	75	39	87
30	13	0	25	11	15	15	50	77	36	70
40	7	0	25	14	15	15	54	78	36	44
50	7	0	25	14	15	17	64	78	34	44

Regarding the impact of CO_2 prices on the average electricity price, we have: the successive increases in CO_2 price always translate into increases in the electricity price. There is a clear negative relationship between CO_2 emission prices and the total CO_2 emissions: the increase in CO_2 price up to 40 euros/ton is effective in reducing emissions to 44% of the state with no CO_2 emission markets.

5.4. The Impact of a Possible Gas Revolution on Capacity Expansion

In this section we simulate the possible impact of a strong decrease in gas prices, due to the exploration of shale gas, or due to the importation of cheap gas, in Portugal and Spain. We consider four scenarios: full gas price (Gas Price Index 100), a reduction of the gas price to half of its current level (Gas Price Index 50), with full Subsidy (100) and without Subsidy (0). In all scenarios we assume the CO_2 emissions price is zero. The results are summarized in Table 5.7.

As can be seen from Table 5.7 a fall in the gas price to half would have a very strong impact on the capacity investment (contraction) decisions, with a significant (even though less spectacular) impact on prices. When considering feed-in subsidies, a reduction of 50% in the gas price leads to a replacement of nuclear by CCGT power plants; the total capacity in the system, and the utilization ratio, remain roughly the same, and the average price+subsidy falls by 4 Euros/MWh. In the absence of subsidies, the only technology surviving would be gas, as the CCGTs would replace the nuclear as well as the green technologies.

Under no subsidy scenario the CCGTs replace all other technologies with a much lower cost for society but larger CO2 emissions. The government would be under pressure to subsidize the green technologies so that the CO_2 emission targets are met. This would mean that the more efficient the fossil fuel technologies become, the more the government may need to spend on subsidizing the green technologies. As can be seen in the last column of the table, even if the government keeps the full subsidies (second row), a reduction of gas prices by 50% leads, in our simulations, to a level of emissions 5 times larger than the one with the current prices (first row).

Table 5.7: Capacity per technology, total capacity, utilization ratio and electricity price as a function of the subsidy and gas price.

		Total Capacity Per Technology (GW)										
Subsidy	Gas Price Index	CCGT	Coal	Hydro	Nuclear	Solar	Wind	Price (E/MWh)	Price + Subsidy (E/MWh)	Capacity (GW)	Utilization Ratio (%)	CO_2 Emis. (%)
100	100	0	0	35	17	36	57	26	56	145	21	5
100	50	18	0	35	0	36	57	21	52	146	22	25
0	100	20	0	5	11	7	12	29	29	55	56	26
0	50	83	0	0	0	0	0	23	23	83	35	100

6. Conclusion

In this paper we have analyzed an industry in which each firm possibly owns a portfolio of several technologies and is characterized by different cost structures and discrete capacity expansion decisions. The firms decide in an incomplete information environment and need to learn the optimal pricing policies and the timing (and type) of capacity expansion (contraction) by interacting with other firms. The primary contribution is methodological. We have proposed the indirect reinforcement learning algorithm to study the capacity expansion (and contraction) problem in a multi-period oligopolistic industry with uncertainty regarding demand, prices, and costs, studying the conditions under which equilibrium can emerge. The major advantage of our approach is the ability to address real world problems more concretely and to have a quantitative analysis of the possible capacity expansion decisions and predict the impact of regulatory regimes on the firms' profitability.

Our contribution to policy making in electricity markets stems from the analysis of the dynamics of investment, regulation and pricing, considering multi-technologies and market power, with twenty years planning horizon for the Iberian electricity market. This is an unprecedented study, the first one to our knowledge, which contributes to the policy discussion in the Iberian Peninsula in several ways: 1) We have analyzed how the subsidies to renewable electricity affect the pricing and investment strategies of the firms; these subsidies are currently under revision and are a major source of political discussion in Portugal and Spain. 2) We have addressed the distortions created by oligopolistic pricing in the pricing and investment strategies. This again, is a hot policy issue not only in the Iberian countries but in Europe, where the privatization of the former public utility companies is being challenged and object of political discussions and judicial investigations due to allegations of corruption by some of the major incumbent firms. 3) We have analyzed the impact of the CO_2 emissions policies in electricity pricing and investment, which have been a major component of the European engagement in tackling global warming with very dismal results so far. 4) Finally, we have analyzed the impact of the envisioned dash for fracking gas in Europe. This is a very complex topic of discussion, as, for example, in France such activity has just been outlawed, and apparently justly so, as supported by our results, which clearly illustrate the possible perverse effects of such a revolution. In summary, the model produced interesting results, most of which not derivable without the full detail of the complex model presented in this article, based on indirect reinforcement learning.

Concretely, first, we have found that subsidies to green technologies tend to increase total capacity in the system, lowering the utilization ratio and that they tend, on average, to decrease electricity prices; a surprising result was that with no subsidy, prices are lower than the ones in the scenarios with a subsidy percentile between 25 and 75. These subsidies are, nonetheless, indispensable to the reduction of CO2 emissions. Moreover, the impact of the subsidy on the total emissions is non-linear; in order to have a significant impact the subsidy needs to be large enough to render the CCGTs unprofitable. Second, both the subsidies to green technologies and an increase in CO_2 emission prices lead to a higher proportion of green technologies; in fact, when CO_2 emission prices rise above 20 Euros/ton there is a sharp increase in electricity prices. Third, we have also simulated the gas price “revolution”, concluding that a drop in 50% on the gas price would lead to a replacement of nuclear by CCGT (in the case with subsidies) and to an electricity sector fully dominated by CCGTs (in the absence of subsidies), leading to a very high increase in CO_2 emissions. Therefore, an increase in efficiency of the CCGTs may require the government to spend more money in subsidizing the green technologies. This seems a counterintuitive result, as the private bonanza of gas abundance would translate itself into the public misery of increased subsidies in already impoverished governments.

From a technical perspective, the major limitation of this type of work is that, due to the very large number of variables, and the presence of stochastic parameters, we cannot derive the equilibrium of the electricity model analytically and the numerical solution of the problem, based on Monte-Carlo simulation, takes a very large number of iterations to converge. From a conceptual perspective, the indirect-reinforcement learning algorithm proposed in this article postulates the existence of a learning process between pricing and investment decisions, nonetheless, such an interaction was not based on empirical data (which at the moment is not available) nor was it based on laboratory simulations that would be able to test if this postulate is validated by experiments with human subjects. For these reasons, the laboratory testing of the indirect reinforcement learning algorithm, as used with other learning models which have been extensively used both in the economics (e.g., Axelrod, 1980; Roth, 1988; Smith, 1994; Banks et al., 1997; Costa-Gomes et al., 2001; Le Coq and Orzen, 2006; Rendell et al., 2010; Anderson, 2012) and electricity markets (e.g., Contreras et al., 2002; Chen and Wang, 2007) literatures, would seem a worthy future work capable of validating the conditions under which the algorithm is a good approximation of human behavior.

As future work, we would like to use the current market model to analyze different policies for reduction of CO_2 emissions in a way that can be sustained by the taxpayer. We would also like to extend the current model to consider an hourly representation of the load duration curve which would be able to take into account the technical constraints associated with managing the different technologies and to better address the inclusion of electricity storage which is due to be a crucial feature of the development of electricity markets in the not so distant future.

Acknowledgments

We are very thankful to the Editor, José F. de Oliveira for his guidance, and four anonymous reviewers for their detailed comments, all of which have been instrumental in improving this article. We are thankful to REN, Rede Eléctrica Nacional, Portugal, for their help in gathering some of the data used in our study. All the results reported in this article and any possible errors or typos remaining are entirely of our responsibility.

References

Ambec, S., C. Crampes. 2012. Electricity provision with intermittent sources of energy. Resource and Energy Economics, 34 (3): 319-336.

Anderson, C. M. 2012. Ambiguity aversion in multi-armed bandit problems. Theory and Decision, 72: 15-33.

Anderson, E. A. Jr., G. G. Parker. 2002. The effect of learning on the make/buy decision. Production and Operations Management, 11 (3): 313–339.

Anderson, E. J., T. D. H. Cau. 2009. Modeling implicit collusion using coevolution. Operations Research, 57 (2): 439-455.

Banks, J., M. Olson, D. Porter. 1997. An experimental analysis of the bandit problem. Economic Theory, 10: 55-77.

Axelrod, R. 1980. Effective choice in the prisoner's dilemma. Journal of Conflict Resolution, 24 (1): 3-25.

Bish, E. K., Q. Wang. 2004. Optimal investment strategies for flexible resources, considering pricing and correlated demands. Operations Research, 52 (6): 954-964.

Bolton, G. E., E. Katok. 2008. Learning by doing in the newsvendor problem: A laboratory investigation of the role of the experience and feedback. Manufacturing & Service Operations Management, 10 (3): 519-538.

Brealey, A. Richard, S. C. Myers. 1991. Principles of Corporate Finance, 4^{th} ed., New York: McGraw-Hill.

Brinckerhoff, P. 2011. Electricity Generation Cost Model - 2011 Update Revision 1. Prepared for the Department of Energy and Climate Change, August, London.

Bunn, D. W., F. S. Oliveira. 2003. Evaluating individual market power in electricity markets via agent-based simulation. Annals of Operations Research, 121 (1-4): 57-77.

Bunn, D. W., F. S. Oliveira. 2008. Modeling the impact of market interventions on the strategic evolution of electricity markets. Operations Research, 56 (5): 1116-1130.

Bunn, D. W., F. S. Oliveira. 2016. Dynamic capacity planning using strategic slack valuation. European Journal of Operational Research, 250 (3): 914-924.

Burnetas, A., S. Gilbert. 2001. Future capacity procurements under unknown demand and increasing cost. Management Science, 47 (7): 979–992.

Caginalp, G. 2005. Nonlinear price evolution. Quarterly of Applied Mathematics, LXIII, 4: 715-720.

Chaton, C., J. A. Doucet. 2003. Uncertainty and investment in electricity generation with an application to the case of Hydro-Québec. Annals of Operations Research, 120: 59-80.

Chen, H., X. Wang. 2007. Strategic behavior and equilibrium in experimental oligopolistic electricity markets. IEEE Transactions on Power Systems, 22 (4): 1707-1716.

Chevalier-Roignant, B., C. M. Flath, A. Huchzermeier, L. Trigeorgis. 2011. Strategic investment under uncertainty: A synthesis. European Journal of Operational Research, 215: 639-650.

Contreras, J., A. J. Conejo, S. de la Torre, M. G. Munoz. 2002. Power engineering lab: Electricity market simulator. IEEE Power Engineering Review, 22 (2): 223-228.

Costa-Gomes, M., V. P. Crawford, B. Broseta. 2001. Cognition and behavior in normal-form games: An experimental study. Econometrica, 69 (5): 1193-1235.

Ehlen, M. A., A. J. Scholand, K. L. Stamber. 2007. The effects of residential real-time pricing contracts on transco loads, pricing, and profitability: Simulations using the N-ABLE™ agent-based model. Energy Economics, 29: 211-227.

European Commission. 2011. Proposal for Directive of the European Parliament and of the Council on energy efficiency and repealing Directives 2004/8/EC and 2006/32/EC. COM(2011) 370 final 2011/0172 (COD).

Fang, F., A. Whinston. 2007. Option contracts and capacity management - Enabling price discrimination under demand uncertainty. Production and Operations Management, 16 (1): 125-137.

Filomena, T. P., E. Campos-Nanez, M. R. Duffey. 2014. Technology selection and capacity investment under uncertainty. European Journal of Operational Research, 232: 125-136.

Fleten, S.-E., E. Nasakkala. 2010. Gas-fired power plants: Investment timing, operating flexibility and CO2 capture. Energy Economics, 32: 805-816.

Franke, C., B. Basdere, M. Ciupek, S. Seliger. 2006. Remanufacturing of mobile phones - Capacity, program and facility adaptation planning. Omega, 34 (6): 562–570.

Gabriel, S.A., A. J. Conejo, J. D. Fuller, B. F. Hobbs, C. Ruiz. 2013. Complementarity Modeling in Energy Markets. Springer Science + Business Media New York, p . 629.

Garcia, A., J.M. Alzate, J. Barrera. 2012. Regulatory design and incentives for renewable energy. Journal of Regulatory Economics, 41, 315-336.

Georgiadis, P., E. Athanasiou. 2013. Flexible long-term capacity planning in closed-loop supply chains with remanufacturing. European Journal of Operational Research, 225: 44-58.

Goyal, M., S. Netessine. 2007. Strategic technology choice and capacity investment under demand uncertainty. Management Science, 53 (2): 192-207.

Gray, J. V., B. Tomlin, A. V. Roth. 2009. Outsourcing to a powerful contract manufacturer: The effect of learning-by-doing. Production and Operations Management, 18 (5): 487–505.

Julka, N., T. Baines, B. Tjahjono, P. Lendermann, V. Vitanov. 2007. A review of multi-factor capacity expansion models for manufacturing plants: Searching for a holistic decision aid. International Journal of Production Economics, 106 (2): 607–621.

Kazempour, S. J., A. J. Conejo, C. Ruiz. 2011. Strategic generation investment using a complementarity approach. IEEE Transactions on Power Systems, 26 (2): 940-948.

Kimbrough, S. O., F. H. Murphy. 2013. Strategic bidding of offer curves: An agent-based approach to exploring supply curve equilibria. European Journal of Operational Research, 229: 165-178.

Le Coq, C., H. Orzen, 2006. Do forward markets enhance competition? Experimental evidence. Journal of Economic Behavior & Organization, 61 (3): 415-431.

Lenzen, M. 2008. Life-cycle energy balance and greenhouse gas emissions of nuclear energy: A review. Energy Conversion & Management, 49 (8): 2178-2199.

MacDonald, M. 2010. UK Electricity Generation Costs Update. June, Brighton.

McDonald, R., D. Siegel. 1986. The value of waiting to invest. Quarterly Journal of Economics, 101 (4): 707-728.

Miller, L. T., C. S. Park. 2005. A learning real options framework with application to process design and capacity planning. Production and Operations Management, 14 (1): 5-20.

Murphy, F. H., Y. Smeers. 2010. On the impact of forward markets on investment in oligopolistic markets with reference to electricity. Operations Research, 58 (3): 515-528.

Oliveira, F. S. 2008. The value of information in electricity investment games. Energy Policy, 36: 2364–2375.

Ove Arup & Partners. 2011. Review of the Generation Costs and Deployment Potential of Renewable Electricity Technologies in the UK. Study Report. Prepared for the Department of Energy and Climate Change', prepared by Ove Arup & Partners Ltd., October, London.

Pindyck, R. 1991. Irreversibility, uncertainty, and investment. Journal of Economic Literature, 29 (3): 1110-1148.

Pineau, P.-O., P. Murto. 2003. An oligopolistic investment model of the Finnish electricity market. Annals of Operations Research, 121: 123-148.

Pineau, P.-O., H. Rasata, G. Zaccour. 2011. Impact of some parameters on the investments in oligopolistic electricity markets. European Journal of Operational Research, 213: 180-195.

Plambeck, E. L., T. A. Taylor. 2013. On the value of input efficiency, capacity efficiency, and the flexibility to rebalance them. Manufacturing & Service Operations Management, 15 (4), 630-639.

Puu T., A. Norin. 2003. Cournot duopoly when competitors operate under capacity constraints. Chaos, Solitons & Fractals, 18: 577-592.

REE, Red Eléctrica de España. 2012. El Sistema Eléctrico Español, Madrid.

REN, Rede Eléctrica Nacional. 2012. National Electricity System, Lisbon.

Rendell, L., R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, K. N. Laland. Why copy others? Insights from the social learning strategies tournment. Science, 328: 208-213.

Roth, A. 1988. Laboratory Experimentation in Economics: A Methodological Overview. The Economic Journal, 98: 974-1031.

Siddiqui, A. S., S.-E. Fleten. 2010. How to proceed with alternative energy technologies: A real options analysis. Energy Economics, 32: 817-830.

Siddiqui, A. S., R. Takashima. 2012. Capacity switching options under rivalry and uncertainty. European Journal of Operational Research, 222: 583-595.

Smith, V.L. 1994. Economics in the laboratory. The Journal of Economic Perspectives, 8 (1): 113-131.

Smit, H.T. J., L. Trigeorgis. 2004. Strategic Investment: Real Options and Games. Princeton University Press.

Sutton, R. S., A. G. Barto. 1998. Reinforcement Learning: An Introduction. MIT Press.

Van Mieghem, J. A. 2003. Capacity management, investment, and hedging: Review and recent developments. Manufacturing & Service Operations Management, 5 (4): 269-302.

Xiao, W., C. Gaimon. 2013. The effect of learning and integration investment on manufacturing outsourcing decisions: A game theoretic approach. Production and Operations Management, 22 (6): 1576-1592.

APPENDIX

Proposition 3.1: Let $\pi_{t=0} = 0$, *after t iterations the learned value of* π_t *is* $\pi_t = u\left(1-(1-\alpha)^t\right)$.

Proof: From equation (3.1) it follows that the value of π_t is equal to the sum of the *t* terms of a geometric progression and, therefore, we can generalize to $\pi_t = (1-\alpha)^t \pi_{t=0} + \alpha u \sum_{j=0}^{t} (1-\alpha)^j$. Then, for any given *t,* the value π_t is equal to $\pi_t = (1-\alpha)^t \pi_{t=0} + \alpha u \frac{1-(1-\alpha)^t}{1-(1-\alpha)}$. As $\pi_{t=0} = 0$, then we obtain $\pi_t = u\left(1-(1-\alpha)^t\right)$.

Proposition 3.2 (Capacity expansion rule): *A firm invests if the estimated value of the capacity expansion is positive and if* $\left|\frac{W_t}{\pi_t}\right| \le t\alpha^2\delta$.

Proof: If, at each iteration *t*, we have $u_{t+1}^a = u$, then the change in the expected value of the profit, at iteration *t*, is equal to $\Delta\pi_t = u(1-\alpha)^t \alpha$. This is shown by a simple iterative application of (3.1). From (3.1) we know that $\pi_{t+1} = \pi_t + \alpha\left(u_{t+1}^a - \pi_t\right)$. Since the process is stationary we have $\pi_{t+1} - \pi_t = \alpha(u - \pi_t)$. The value of π_t is equal to the sum of the *t* terms of a geometric progression, $\pi_t = u\left(1-(1-\alpha)^t\right)$. Therefore, we have $\Delta\pi_t = \alpha\left[u - u\left(1-(1-\alpha)^t\right)\right]$ and hence $\Delta\pi_t = u(1-\alpha)^t \alpha$. Then, by replacing $\Delta\pi_t$ in equation (3.2) we get $W_{t+1} = W_t + \alpha\left[u(1-\alpha)^t \alpha - W_t\right]$. Next, through an iterative process of replacement of W_t in W_{t+1}, W_{t+2}, …, we get $W_t = tu\alpha^2(1-\alpha)^t$. Moreover, as in a stationary process, given enough time, i.e., as the maximum number of iterations $T \to +\infty$, (3.1) can estimate the correct value of π: $\pi_t + (1-\alpha)\pi_{t-1} + (1-\alpha)^2 \pi_{t-2} + ... + (1-\alpha)^T \pi_{t-T} = \sum_{j=1}^{T} \left(j\alpha u (1-\alpha)^{j-1} \right) = u$. The proof is by induction. From equation (3.1) we know that $\pi_t = \alpha u + (1-\alpha)\pi_{t-1}$. Therefore, for *j*=1 we have $\pi_t + (1-\alpha)\pi_{t-1} = \alpha u + (1-\alpha)\pi_{t-1} + (1-\alpha)\pi_{t-1} = \alpha u + 2(1-\alpha)\pi_{t-1}$. Similarly, for *j* =2 we have $\pi_t + (1-\alpha)\pi_{t-1} + (1-\alpha)^2 \pi_{t-2} = \alpha u + 2\alpha u(1-\alpha) + 3(1-\alpha)^3 \pi_{t-3}$, and for *j* = 3 we have $\pi_t + (1-\alpha)\pi_{t-1} + (1-\alpha)^2 \pi_{t-2} + (1-\alpha)^3 \pi_{t-3} = \alpha u + 2\alpha u(1-\alpha) + 3\alpha u(1-\alpha)^2 + 4(1-\alpha)^4 \pi_{t-4}$.

Hence, we can generalize for *j* = *T* for

$$\pi_t + (1-\alpha)\pi_{t-1} + (1-\alpha)^2 \pi_{t-2} + ... + (1-\alpha)^T \pi_{t-T} = \sum_{j=1}^{T} \left(j\alpha u (1-\alpha)^{j-1} \right) + (T+1)(1-\alpha)^T \pi_{t-T}.$$

As $\lim_{T \to +\infty} (T+1)(1-\alpha)^T \pi_{t-T} = 0$ we obtain

$\pi_t + (1-\alpha)\pi_{t-1} + (1-\alpha)^2 \pi_{t-2} + ... + (1-\alpha)^T \pi_{t-T} = \sum_{j=1}^{T} \left(j\alpha u (1-\alpha)^{j-1} \right)$. Finally, we have

$\lim_{T \to +\infty} \sum_{j=1}^{T} \left(j\alpha u (1-\alpha)^{j-1} \right) = \alpha u \lim_{T \to +\infty} \sum_{j=1}^{T} \left(j (1-\alpha)^{j-1} \right) = u$. Therefore, the ratio between the change in the forecast and the current forecast as *t* increases to infinity converges on $\frac{W_t}{\pi_t} = t\alpha^2 (1-\alpha)^t$ which, eventually, converges on zero. Next, by multiplying both terms of equation Corollary 3.1 by $t\alpha^2$ we get $t\alpha^2 (1-\alpha)^t \leq t\alpha^2 \delta$ and, therefore, we show that the equilibrium condition for the extended *n*-armed bandit algorithm is $\left| \frac{W_t}{\pi_t} \right| \leq t\alpha^2 \delta$ (the absolute value is used so that the condition can be applied to both positive and negative expected values).