ÉCOLE DES PONTS ET CHAUSSÉES

SUPPLY FUNCTION EQUILIBRIA ON THE ELECTRICITY MARKET

THÈSE

pour l'obtention du titre de Docteur en Sciences Économiques

présentée et soutenue publiquement le XXXXXXXX

par

Alexis Bergès

Jury

XXXX, XXXX XXXXXX XXXXXXXX

XXXX, XXXX XXXXXX XXXXXXXX

XXXX, XXXX XXXXXX XXXXXXXX

XXXX, XXXX XXXXXX XXXXXXXX

Remerciements—Acknowledgements

Dedication

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXX

Contents

Remerciements—Acknowledgements **i**

General introduction **1**

Summary **5**

Chapter 1: Dynamics of the Electricity Day-Ahead Market : Supply Function Equilibria and Ramping Costs **9**

Appendix **47**

Chapter 2: Methodological tools for non parametric functional data evaluation and weather data usage **57**

Appendix **89**

References **95**

List of Figures **97**

General introduction:

Summary

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 1

Dynamics of the Electricity Day-Ahead Market : Supply Function Equilibria and Ramping Costs

1.1 Introduction

1.1.1 Litterature review

The electricity markets have flourished in Europe during the 1990s during the wave of privatisation. The argument for their creation was one of competition, that was supposed to bring lower prices to the end consumer of electricity.

An important specificity to the economics of electricity is that electricity cannot be stored in large amounts, which in turn implies that at every moment production and consumption have to match. This means that in order to have a working electric grid, that is one that can produce electricity for high levels during the winter and lower levels in summer, one has to have production units ready to be turned on if the demand is high enough, but turned off otherwise. This in turn means that although their existence is required, it is difficult to see how to marginal cost pricing can cover their investment costs, which has been a long running argument in the litterature [Boiteux, 1960]. For this reason, from the very beginning the issue of the market design was deemed to be crucial to insure that the wished for outcome of the privatization wave came to fruition [Green, 1991].

Most countries having open the production of electricity to competition have implemented day-ahead markets. As said above, the production and the consumption have to match constantly. The very short term matching is done by automating tiny adjustments around what a producer is already producing in order to match the fluctuating consumption. To plan which plant should be online at which hour of the day however, the day-ahead markets come in. The idea is that producers and big consumers of electricity (either for themselves, or as aggregators of the individual consumptions) are asked to bid demand or supply functions respectively. The market operator then aggregates the demand and supply curves, which yields an equilibrium giving the price and quantities to be produced for each producer.

There has been an active litterature trying to model and measure the market power of oligopolists on these newly created markets [Green and Newbery, 1992, Newbery, 1998, Green, 1999]. The models have mainly been based on Klemperer and Meyer 1989's Econometrica founding paper about supply function equilibria [Klemperer and Meyer, 1989] (henceforth known as KM).

This paper builds upon previous results about competition in supply schedules without uncertainty [Grossman, 1981], which yielded a very high multiplicity of equilibria. KM add a key ingredient : uncertainty about the demand schedule facing the suppliers. This addition reduces greatly the multiplicity, and adds more structure to it, although in this framework there is still a continuum of Nash equilibria, which are always pinned between Cournot and Bertrand outcomes.

Groundbreaking and fertile, the original model by KM studied how demand uncertainty collapses dramatically the set of available supply function equilibria to a well defined continuum when contrasted to the case of competition in supply schedules without uncertainty [Grossman, 1981]. These equilibria are always pinned between Cournot and Bertrand outcomes. This continuum collapses further to a single Nash equilibria by considering an infinite support of demand shocks. All of these equilibria are ex-post optimal, meaning that changes in anticipated demand shocks do not impact the actual solutions, but only the parts of the solutions that are actually explored as shocks realize. In this setup markets are always efficient, a very strong result.

The electricity markets litterature has embraced this framework because it is considered to capture some of the structure at play in the electricity markets : the producers do not know what demand they are going to face when they choose their supply schedule, the demand side is considered much less sophisticated than the supply side, and their demand schedules can therefore be considered to some extent as being exogenous. Some have argued that the schedules submitted in the real markets are discrete and that this discrete nature makes their modelling as continuously differentiable schedules is both incorrect and yields different results from discrete ones [von der Fehr and Harbord, 1993]. However recent results suggest that with a sufficient amount of steps both approaches converge [Holmberg et al., 2008], and indeed we see that recent implementations of the market rules increase the number of steps allowed for a single bid, and consider that these points are linearly joined instead of stepwise.

One of the mose striking aspects of the supply function equilibria approaches is, as was alluded to above, the multiplicity of Nash equilibria. This result has been generally viewed as the source of the danger of tacit collusion in electricity markets : if there is a continuum of nash equilibria, repeated interactions are feared to be conducive to a convergence of bidding strategies towards the most profitable equilbria [Bolle, 1992].

Furthermore, these models abstract away some of the details of the actual markets, reason for which authors which try and evaluate the market power of producers on the electricity markets view their endeavour as painting the situation with an optimistic brush [Green and Newbery, 1992].

Here we will tackle the points raised in the last two paragraphs to some extent. We propose to consider a technical reality of the operating of power plants : their cost structure is history-dependant, more precisely, producing a quantity q_1 does not entail the same cost if the previous quantity produced was already q_1 or if the previous quantity was different from it. Raising or decreasing production in and of itslef imply costs. By introducing these costs we aim to produce a model capable of capturing more precisely the competition that arises in the electricity markets, and in so doing we will show that the continuum of equilibria caracteristic of supply function equilibria under uncertainty collapses to unique equilibria, which in turn allows us to comment on the question of tacit collusion.

1.1.2 The day-ahead markets

On the electricity day ahead markets, producers are generally required to submit supply schedules once a day for all the auctions taking place during the next day. The APX (England) and the EPEX (Austria, France, Germany and Switzerland) markets allow hourly auctions [APX, 2017, EPEX, 2015], and EPEX allows for bids comprising up to 256 price quantity combinations, effectively approximating smooth supply functions. Producers can submit different supply schedules for each individual auction, but every bid must be placed at the same time one day in advance for each block of 24 hours. Customers go through the same process and submit their demand schedules, then the market operator matches supply and demand for each auction. Producers thus have to submit schedules facing uncertain demand, this is the reason for the popularity of supply function equilibria approaches to the electricity market.

However, on this market, bids change from auction to auction. From the point of view of KM's model, this should happen only through a coordination of agents agreeing to collectively swap from one Nash equilibrium to another in the available continuum. Describing these dynamics, however, is increasingly important as the energetic mix is bound to include an increasing fraction of renewables. Power production can be separated in two classes: dispatchable and non-dispatchable technologies. Nuclear, coal, land-fill gas or hydroelectric power generation are mainly dispatchable as one can actually choose their level of production whereas the two rising stars of renewable energy, namely wind and solar, are non-dispatchable: they react to weather conditions. Having these technologies in the mix introduces uncertainty on the production side, which comes down to dispatchable units facing a more uncertain residual demand [Boyle, 2007]. In this paper we want to explore how to model these dynamics.

Electricity production faces very specific technological constraints. These constraints, generally labelled as ramping costs, vary across production technologies, and have yet to be captured in a model. We propose to do so by introducing a multivariate cost function, depending as always on the quantity produced, but also on the rate at which production varies: $C(S, \frac{dS}{dt})$. We call this class of cost functions dynamic cost functions.

All power plants face maintenance costs. However part of these maintenance costs are induced by the dynamics of production, and can be seen as ramping costs. More precisely, whatever the production technology, fluctuations in production are costly. Indeed, they imply fluctuations in the temperature of the core of the power plant, thus dilation and contraction cycles of the different parts, which cause wear and tear. The industry is aware of these effects [GE, 2015], some B2B companies even specialize in minimizing the related long term costs. For example, Wartsila Power Plants, a supplier of power plants and tools to forecast long term costs, explains in a white paper [Arima, 2012]:

> Increased variability in net load demand means that dispatchable generating units have to ramp considerably more steeply and deeper than traditionally, thus increasing wear and tear to components.

We are going to model these ramping costs through a dynamic cost function, increasing in the absolute value of its second argument : any change in production is costly. This paper will focus on the implications of considering this type of ramping costs. Other types of ramping costs exist, for example startup costs, but they will not be studied in this paper.

These effects cannot be captured by traditionnal cost functions depending on the level of production alone. One needs to take into account the actual path leading to a given quantity produced. This implies that we need to impose structure on the dynamics of the system while retaining uncertainty, the key ingredient of KM's paper. To do so, we use stochastic dynamics.

This seemingly small addition to KM's framework has a lot of implications on the results obtained. The solutions are not ex-post optimal anymore, allowing to account more satisfactorily for the dynamics of optimal supply schedules, and our solutions are unique, even for bounded demand shocks. We also define a novel selection rule to choose from KM's continuum of equilibria. Finally these results open the possibility to distinguish intraday and day-ahead markets.

In section 1.2 we will present a heuristic approach to get the intuition of the model. Then, in section 1.3 we will introduce the mathematical tools needed to use stochastic dynamics in this context, in section 1.4 and section 1.5 we will solve the monopoly and the symmetric oligopoly cases while considering that producers have information about the overall distribution of shocks during the day, but do not have information about differences in the shocks at different dates. Finally in section 1.6 we will discuss the dynamic variation of the optimal bids, while sections **??** and 1.8 will respectively cover a few implications of these results and conclude the paper.

1.2 Heuristic Description of the Model

In this section the essence of the model is presented before introducing the proper mathematical tools needed to treat this problem rigorously in the next section. It is thought of as an overview of the mathematical methods that are going to be used, as a way to give a sense of the intent of the modelling choices.

As in KM's setup, the aim is to model an oligopoly facing uncertain demand, taken as exogenous. Before the demand shocks are realized, each firm needs to commit on a strategy. Firms also incur costs that not only depend on the level of production but also on the evolution of the production given its anterior level produced.

More formally, the producer, as in KM, faces uncertain demand, $D(\theta, p)$, with θ a stochastic shock to the demand and p the price. We add to that both ramping costs and uncertain dynamics of demand. As we want to keep the key ingredient of KM, the introduction of uncertainty, but take into account the dynamics of this uncertainty, of these demand shocks, we need to add more structure.

Consider the following notation, where $\theta(t)$ denotes the value of the stochastic shock at time t, whereas Θ denotes the family of all available time trajectories of our demand shocks.

In the real market, bidders submit a finite number of bids once a day, and face the ramping costs inter-period, that is, when production has to be adjusted to reach the subsequent market outcome. The first bit of structure we introduce is that we are going to assume that time is continuous. The second is that ramping costs are incurred continuously, and can be thought of as costs depending on the variation of production over time. Finally we consider that bidders are allowed to submit a different supply schedule for every point in time between 0 and T. This amounts to being asked to submit a surface of strategy in the price-quantity-time space for the next day.

The producer maximises her expected profits, and we consider here the simplest case in which the distribution of shocks is static, that is that the distribution of probability of shocks does not depend on time, and the producer is asked by the market operator to submit the same supply schedule for every point in time a day in advance. In an oligopoly, the program maximised by producer i is therefore:

$$\max_{S_i(p)} \mathbb{E}_\Theta \left[\int_0^T \left(p(\theta(t)) S_i(p(\theta(t))) - C\left(S_i(p(\theta(t))), \frac{dS_i(p(\theta(t)))}{dt} \right) \right) dt \right] \tag{1.1}$$

with $p(\theta(t))$ the price given the demand shock $\theta(t)$ at date t, $S_i(\cdot)$ the supply schedule of producer i and $C(\cdot,\cdot)$ the dynamic cost function. Note that the price depends on t only through $\theta(t)$, i.e. a given level of demand shock implies a given price.

The goal of this section is to provide a first run through of the model, therefore we will not describe here the conditions that must be verified by the different terms of the model. We will simply assume that the dynamic cost function is additively separable between a static and a ramping term, $C(S_i, \frac{dS_i}{dt}) = C_s(S_i) + C_r(\frac{dS_i}{dt})$, and that the demand shocks θ are bounded in $[\underline{\theta}, \overline{\theta}]$. Lastly we require the ramping term $C_r(\cdot) = \frac{\gamma}{2}(\cdot)^k$ for clarity, and $k \geq 2$ an integer. We distribute the expectation operator and write that $\frac{dS_i}{dt} = \frac{dS_i}{dp}\frac{dp}{d\theta}\frac{d\theta}{dt} = S_i' \cdot \dot{p} \cdot \frac{d\theta}{dt}$, with X' the derivative of univariate function X with respect to its argument, $\dot{X} = \frac{dX}{d\theta}$.

With this setup, by distributing the expectation operator over all possible trajectories of shocks, we are able to rewrite the problem without having time t appear explicitly. This point is crucial, as it is what will let us use mathematical tools that will yield our unicity results. The maximisation program can indeed be written as follows :

$$\max_{S_i(p)} \int_{\underline{\theta}}^{\overline{\theta}} f(\theta) \left(p(\theta)S_i(p(\theta(t))) - C_s(S_i(p(\theta(t)))) - \frac{\gamma}{2}\left(S_i' \cdot \dot{p}\right)^k \mathbb{E}_\Theta \left[\left(\frac{d\theta}{dt}\right)^k \middle| \theta \right] \right) d\theta \quad (1.2)$$

with $f(\theta)$ the distribution of shocks, and γ the ramping cost parameter capturing the magnitude of the ramping costs. The expected value on the trajectory of shocks of any of the terms above that only depend on $\theta(t)$, that is the value of the shock at a point in time, can be rewritten simply as an integral over the possible values of the shock.

We are left with $\mathbb{E}_\Theta \left[\left(\frac{d\theta}{dt}\right)^k \middle| \theta \right]$ as the only term that depends on the trajectory of shocks. Take for granted that this term can only depend on θ for now, this result will be defended properly in the next section.

Note now that producer i faces a residual demand so that $S_i(p(\theta(t))) = D(\theta, p(\theta(t))) - S_{-i}(p(\theta(t)))$ which depends only on θ and p, t does not intervene directly, with S_{-i} the aggregate supply schedule of all the other producers, taken as given by producer i. This implies that the integrand in eq. 1.2 depends only on three variables: θ, p and $\dot{p}$. The maximisation pogram is therefore equivalent to an Euler-Lagrange problem, a very well described mathematical object: $\max_p \int \mathcal{L}(\theta, p, \dot{p}) d\theta$.

The information obtained from taking the first-order condition of an Euler-Lagrange problem yields a second order differential equation as well as two boundary conditions: $\frac{\partial \mathcal{L}}{\partial p} = \frac{d}{d\theta}\frac{\partial \mathcal{L}}{\partial \dot{p}}$ and $\frac{\partial \mathcal{L}}{\partial \dot{p}}\Big|_{\underline{\theta}} = \frac{\partial \mathcal{L}}{\partial \dot{p}}\Big|_{\overline{\theta}} = 0$. This is why we obtain unique solutions: if the boundary conditions are not verified there exists profitable deviations.

In less mathematical terms, taking ramping costs into account as specified above means that for a given level of shock, the producer not only cares about the optimal level of production for this shock, but also about the optimal slope of the supply schedule at this level of production. Effectively, this means that optimal levels of production cannot be chosen independently for different level of shocks as is the case in KM, thus shrinking the continuum of equilibria. The boundary conditions' argument explains why the continuum not only shrinks, but collapses to a unique equilibrium.

Note that if the ramping cost parameter γ is taken equal to 0 we are back to KM's model: one doesn't care about the slope of the supply schedule anymore, and the problem comes down to a pointwise maximisation which therefore yields ex-post optimal equilibria. We want to stress that this means that it is not sufficient to specify the dynamics of the shocks to obtain a supply function model that would react to these dynamics, one needs to take into account ramping costs.

The maximisation program 1.2 is a heuristic description of the situation. We want to model the stochastic nature of demand and of its dynamics. We do this by using Itō processes, a class of stochastic processes built through brownians, to describe the stochastic trajectory of the demand shocks with respect to time. The difficulty is that brownians are everywhere continuous but nowhere differentiable, therefore the way program 1.2 is written, with a term in $\frac{d\theta}{dt}$, is a shortcut.

In the next section we introduce the stochastic dynamics properly without using the concept of derivative.

1.3 Stochastic Dynamics

As described in the previous section, we consider that bidders submit surfaces, that is supply schedules for every point in time. The reason to describe a discrete dynamic market as a continuous one is that although discrete time is conceptually more easily understood, continuous time allows to use much more powerful mathematical tools and to obtain closed form solutions, which we think are crucial in gaining intuitive insights about these dynamics. Therefore we consider that demand fluctuates continuously and that ramping costs are incurred instantaneously. This approximation would need to be tested, although it should be noted that day ahead markets operate with hourly or half-hourly periods and producers are therefore facing a reasonable amount of periods each day.

We want our shock variable to evolve over time in a random fashion. The class of mathematical objects used to describe this are stochastic processes. The simplest stochastic process one can think of, and indeed the most important historically, is a Brownian motion process.

Unfortunately, Brownian processes are unbounded, and cannot therefore be used to describe the dynamics of the electricity market in which demand shocks, denoted $\theta(t)$, are bounded: there are no days for which demand is null nor are there days for which demand tends towards infinity. The structure to be imposed on the dynamics of the shocks has to imply bounded shocks.

1.3.1 The stochastic process

A richer set of stochastic processes is the set of Itō processes.

A simple Itō process one can consider that leads to bounded shocks is defined by the following stochastic differential equation (SDE):

$$d\theta(t) = -2\theta(t)dt + \sqrt{1-\theta(t)^2}dB_t \tag{1.3}$$

with B_t a brownian and dX an infinitesimal variation of quantity X.

Observe that this SDE is formed by a deterministic mean-returning term $-2\theta(t)dt$ and a bounded stochastic one $\sqrt{1-\theta(t)^2}dB_t$. As $\theta(t)$ approaches -1 or 1 the stochastic term goes to 0, thus $\theta(t) \in [-1, 1]$. Without loss of generality we can restrain ourselves to this special case. Other bounded supports, $\theta \in [\underline{\theta}, \overline{\theta}]$, can be captured through renormalisations of θ.

Such a stochastic process has a distribution of probability $f(\theta)$ given by Fokker-Planck's equation, easily solved here. In the general case of an Itō process given by SDE 1.4, one obtains in 1.5 the generic Fokker-Planck equation for its distribution of probability $f(\theta, t)$:

$$d\theta = \mu(\theta, t)dt + \sigma(\theta, t)dB_t \tag{1.4}$$

$$\frac{\partial}{\partial t}f(\theta, t) = \frac{\partial}{\partial \theta}(\mu(\theta, t)f(\theta, t)) + \frac{1}{2}\frac{\partial^2}{\partial \theta^2}(\sigma(\theta, t)^2 f(\theta, t)) \tag{1.5}$$

Here, for SDE 1.3, this yields that $f(\theta) = \frac{3}{4}(1-\theta^2)$ on $[-1, 1]$ and 0 elsewhere.

1.3.2 The ramping costs

In the rest of the paper we are going to consider quadratic ramping costs. More precisely we consider the costs induced by fluctuations in the production level. As described in the introduction, fluctuations imply increased wear and tear, whether the production is increasing or decreasing. In addition, these ramping costs are null in the absence of fluctuations. This means that they can be captured by a function $C_r(\cdot)$ verifying $C_r(0) = 0$, $C_r(\cdot) \geq 0$ and increasing in the absolute value of its argument. In the abscence of more detailed knowledge about the actual shape of these ramping costs, it seems reasonable to consider a quadratic cost function, that is the first term in a Taylor expansion of the actual real ramping cost function.

We cannot compute $\frac{d\theta}{dt}$ as it appears in Eq. 1.2, as a stochastic process, although everywhere continuous, is nowhere differentiable. The goal of this section is to express properly the maximisation program of the producer that we presented rapidly in Eq. 1.2, and most importantly, to introduce properly how we can work in continuous time with a cost function which depends on fluctuations, and fluctuations which are nowhere differentiable.

We are therefore going to first consider the discrete case of a random walk of timestep Δt which converges towards the Itō process 1.4, using the Euler-Maruyama approximation, a generalisation of the Euler method to stochastic differential equations. We consider a Markov chain Y defined as follows :

$$\Delta Y_n = Y_{n+1} - Y_n = \mu(Y_n, n\Delta t)\Delta t + \sigma(Y_n, n\Delta t)\Delta B_n \tag{1.6}$$

where $\Delta B_n = B_{(n+1)\Delta t} - B_{n\Delta t}$. These ΔB_n are *i.i.d.* normal random variables of mean 0 and variance Δt. Note that as Δt is taken towards 0, this Markov chain converges towards its underlying stochastic process defined by eq.(1.4).

The ramping costs are taken as quadratic in the variation of the production, and also depend on a ramping cost parameter $\Gamma(\Delta t)$, that is the cost per unit of quadratic variation at horizon Δt, so we compute the following quantity :

$$\mathbb{E}\left[\frac{\Gamma(\Delta t)}{2}\cdot\left(\frac{Y_{n+1}-Y_n}{\Delta t}\right)^2\middle| Y_n\right]=\frac{\Gamma(\Delta t)}{2}\cdot\frac{\sigma(Y_n,n\Delta t)^2}{\Delta t} \tag{1.7}$$

For this quantity to converge to a finite value when the Markov chain is taken towards its underlying stochastic process we have to consider that for small enough timescales, the ramping cost parameter $\Gamma(\Delta t)$ is linear in Δt, i.e. $\Gamma(\Delta t)=\gamma\Delta t+o(\Delta t)$. Mathematically, if $\Gamma(\Delta t)$ had a slower than linear relationship at small timescales, the ramping costs would diverge, and if it was faster they would converge to 0. A physical constraint, namely thermal inertia, ensures that the ramping cost parameter does actually behave in this way[1].

Consider for now that the mean function μ and the variance function σ from eq. 1.4 do not depend on time explicitly and are therefore written $\mu(\theta)$ and $\sigma(\theta)$. Consider now a transformation $T(\cdot)$ that we apply to the Markov chain Y. Then:

$$\mathbb{E}\left[\frac{\Gamma(\Delta t)}{2}\cdot\left(\frac{T(Y_{n+1})-T(Y_n)}{\Delta t}\right)^2\middle| Y_n\right]=\mathbb{E}\left[\frac{\Gamma(\Delta t)}{2}\cdot\left(\frac{T(Y_{n+1})-T(Y_n)}{Y_{n+1}-Y_n}\cdot\frac{Y_{n+1}-Y_n}{\Delta t}\right)^2\middle| Y_n\right] \tag{1.8}$$

And in the limit where the markov process Y converges towards the Itō process θ of equation 1.4:

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\frac{\Gamma(\Delta t)}{2}\cdot\left(\frac{T(Y_{n+1})-T(Y_n)}{\Delta t}\right)^2\middle| Y_n\right]=\frac{\gamma}{2}\cdot T'(\theta(t))^2\cdot\sigma(\theta)^2 \tag{1.9}$$

We apply this result to the problem at hand, that is that we evaluate the ramping

[1] Ramping costs come from thermal fluctuations in the core of the plant. Therefore we have to describe how temperature responds to fluctuations in production. Thermal inertia acts as a low pass filter, meaning that it smoothes out fluctuations on short timescales. Think about heating a saucepan full of water: although lighting the stove is almost instantaneous, the temperature of the water being heated increases only progressively, in an exponential fashion that is therefore linear in time for short timescales.

costs in the case where the demand shocks are given by eq. 1.3:

$$\lim_{\Delta t \to 0} \mathbb{E}\left[\frac{\Gamma(\Delta t)}{2} \cdot \left(\frac{\Delta S_i(p(\theta(t)))}{\Delta t}\right)^2 \middle| \theta(t)\right] = \frac{\gamma}{2} \cdot S_i'(p(\theta(t)))^2 \dot{p}(\theta(t))^2(1-\theta^2) \tag{1.10}$$

with X' the derivative of quantity X with respect to its argument and $\dot{X}$ its derivative with respect to θ. Note that we considered here that the variance term $\sigma(\theta) = 1 - \theta^2$ depends only on θ and not explicitly on t, which in turn implies that the strategy S_i does not depend explicitly on t either.

Let us consider the case where the strategy and the variance depend explicitly on time, and are thus written $S_i(p(\theta(t), t), t)$ and $\sigma(\theta, t)$ respectively. By using a first order expansion as before, the ramping cost function can be approximated as follows:

$$\begin{aligned} \lim_{\Delta t \to 0} \mathbb{E}\left[\frac{\Gamma(\Delta t)}{2}\left(\frac{\Delta S_i(p(\theta(t), t), t)}{\Delta t}\right)^2 \middle| \theta(t)\right] &= \lim_{\Delta t \to 0} \mathbb{E}\left[\frac{\gamma}{2}(\partial_1 S(p(\theta(t), t), t)\partial_1 p(\theta(t), t))^2 \frac{\Delta \theta^2}{\Delta t} + \mathcal{O}(\Delta t)\right] \\ &= \frac{\gamma}{2}(\partial_1 S(p(\theta(t), t), t)\partial_1 p(\theta(t), t))^2 \sigma(\theta, t)^2 \end{aligned} \tag{1.11}$$

with $\partial_i X$ the partial derivative of quantity X with respect to its i^{th} argument. See Annex. 1.A for a more details on this derivation.

Now, we can write down the instantaneous expected value of the profit of producer i if the demand shock is $\theta(t)$, $\pi_i^e(t, \theta(t))$, that is the profit that one expects to obtain when demand is at $\theta(t)$ given the expected value of the ramping costs:

$$\pi_i^e(t, \theta(t)) = p(\theta(t), t)S_i(p(\theta(t), t), t) - C_s(S_i(p(\theta(t), t), t)) - \frac{\gamma}{2}\partial_1 S_i(p(\theta(t), t), t)^2 \partial_1 p(\theta(t), t)^2 \sigma(\theta, t)^2 \tag{1.12}$$

Lastly we have to write down the expected profit for a day's worth of submitted strategies. Let us consider that the chosen unit of time is the day. Therefore, the total expected profit Π_i^e writes:

$$
\begin{aligned}
\Pi_i^e &= \int_0^1 \mathbb{E}_{\theta(t)}[\pi_i^e(t,\theta(t))]dt \\
&= \int_0^1 \int_{\underline{\theta}}^{\overline{\theta}} f(\theta,t)\Big[p(\theta,t)S_i(p(\theta,t),t) - C_s(S_i(p(\theta),t)) \\
&\quad -\frac{\gamma}{2}\partial_1 S_i(p(\theta,t),t)^2\partial_1 p(\theta,t)^2\sigma(\theta,t)^2\Big]\, d\theta dt
\end{aligned} \tag{1.13}
$$

1.3.3 Discussion of the approximations

We want a tractable mathematical formulation of the dynamic problem faced by producers on the electricity market. To achieve this we seek to describe the discrete real life problem by an approximated continuous one. We first use two technological facts: fluctuations in production are costly and these costs decrease linearly in time for short timescales. We then rely heavily on first order expansions of the different terms we have to compute.

1.3.4 The maximisation program

Here, we consider that the dynamics of demand shocks are given by eq.(1.3), and that therefore $\sigma(\theta,t)^2 = \sigma(\theta)^2 = (1-\theta^2)$.

We now introduce the different conditions that have to be satisfied by the various terms in this problem. First, on most electricity markets, schedules must be increasing, therefore here we take $S_i'(\cdot) \geq 0$. Second, the aggregate demand is non negative as consumers do not have production facilities at their disposal: $D(\theta(t), p(\theta(t))) = \sum_i S_i(p(\theta(t))) \geq 0$. Last, we consider that the shocks θ are ordered so that the demand is increasing in θ, i.e. $\frac{\partial D}{\partial \theta} \geq 0$, and that the price has to weakly increase with the shocks, i.e. $\dot{p} \geq 0$. Our initial stochastic maximisation program can thus be rewritten as a regular optimal control problem:

$$\max_{S_i(p)} \int_{-1}^{1} f(\theta) \left(p(\theta) S_i(p(\theta)) - C_s(S_i(p(\theta))) - \frac{\gamma}{2}(1-\theta^2) \left(S_i'(p(\theta))\dot{p}(\theta)\right)^2 \right) d\theta \tag{1.14}$$

$$\begin{aligned} s.t. \qquad & S_i'(\cdot) \geq 0 \\ & \dot{p} \geq 0 \qquad (1.15) \\ & D(\cdot,\cdot) \geq 0 \end{aligned}$$

$$\tag{1.16}$$

The next section solves this problem for a monopoly.

1.4 The Monopoly

Let us consider that the aggregate demand is linear, that is:

$$D(\theta(t), p(\theta(t))) = a\theta(t) + b - p(\theta(t))$$

with a and b parameters taken to describe any bounded support of shocks given the stochastic dynamics introduced in the previous section for which $\theta \in [-1, 1]$. Here $(a\theta + b) \in [b-a, b+a]$.

In a monopoly situation we have $S = D(\theta(t), p(\theta(t)))$, therefore the constraints reduce to:

$$\dot{p}(\theta) \in [0, a], \text{ and } p(\theta) \leq a\theta + b$$

where $\dot{X}$ corresponds to $\frac{dX}{d\theta}$.

Consider in addition that the static cost function is also quadratic: $C_s(S_i) = \frac{\lambda}{2} S_i^2$.

The maximisation program is rewritten as:

$$\max_{p(\cdot)} \int_{-1}^{1} f(\theta) \left(p(\theta)(a\theta + b - p(\theta)) - \frac{\lambda}{2}(a\theta + b - p(\theta))^2 - \frac{\gamma}{2}(1-\theta^2)\left(a - \dot{p}(\theta)\right)^2 \right) d\theta \tag{1.17}$$

$$\text{s.t.} \qquad \dot{p}(\theta) \in [0, a]$$

$$p(\theta) \leq a\theta + b$$

1.4.1 Results

Proposition 1.4.1 *The solution exists, is unique, and has the following form:*

$$\forall \theta \in [-1, 1] \; p^*(\theta) = a\frac{4\gamma + 1 + \lambda}{4\gamma + 2 + \lambda}\theta + b\frac{1 + \lambda}{2 + \lambda} \tag{1.18}$$

The optimal schedule is parametrised by θ so that $S(p(\theta))$ is formed by the points of coordinate $(a\theta + b - p(\theta), p(\theta))$. Its equation is given by:

$$S^*(p) = \frac{1}{4\gamma + 1 + \lambda}\left(p + \frac{4\gamma}{2 + \lambda}b\right) \tag{1.19}$$

Proof See annex 1.B. ■

We present in Fig. 1.1(a) the results obtained for increasing values of the ramping cost parameter γ, starting at $\gamma = 0$ in black and moving progressively from black to blue to red to green.

As expected, adding these inertial costs narrows down the domain of attainable quantities produced, as a larger quantity domain implies larger incurred ramping costs.

More interesting is the way the quantity domain is narrowed down. The domain of prices increases conversely, so that the solutions are steeper than the traditional monopoly situation, bringing the schedules ever closer to a Cournot-like situation. In addition, the optimal supply schedules do not depend on a, the parameter determining the width of the possible shocks, but only on b which defines the average value of the shocks.

One can then study the comparative statics when the values for a and b are varied, as illustrated in Fig. 1.1(b). In particular, if we consider an increase in a without changing b, the solution is represented by the same "master" function, but the explored region expands. On the other hand, if we consider a fixed a but an increasing b, the explored length is kept constant, but the optimal schedule is translated towards the north-east region of the plane as expected intuitively: more demand implies a given mix between higher quantities and prices, which is given by the direction of the vector of translation. Note that the independence of the solution on variations of a comes from the fact that we are considering comparative statics, which is very different from dynamically evolving values of a and b, case which will be treated in detail in section 1.6.

Lastly, note that all schedules cross at a single point. These quadratic ramping costs imply a symmetric deformation of the solution without ramping costs. The limit of extremely high ramping costs is a Cournot-like schedule, i.e. a vertical one, taken at this crossing point.

1.5 The Symmetric Oligopoly

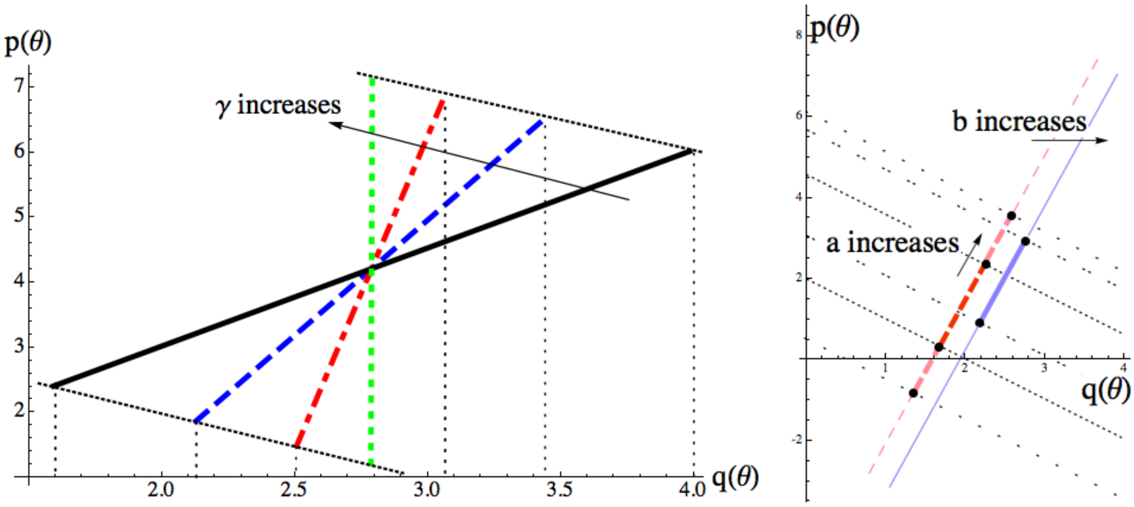
We keep the same linear demand specification as in the monopoly, therefore, with n competitors one has to consider the residual demand faced by each producer:

$$S(p(\theta)) = a\theta + b - (n-1)S(p(\theta)) - p \tag{1.20}$$

$$S(p(\theta)) = \frac{a\theta + b - p}{n} \tag{1.21}$$

$$S'(p(\theta)) = \frac{a - \dot{p}}{n\dot{p}} \tag{1.22}$$

$$S''(p(\theta)) = -\frac{a\ddot{p}}{n\dot{p}^3} \tag{1.23}$$



(a) $S^{-1\,*}(p)$ (b) $S^{-1\,*}(p)$

Figure 1.1: **(a)** Four optimal supply schedules are plotted. In black (full line) $\gamma = 0$. As γ increases we transition from the black curve to the blue curve (large dashes), then the red curve (mixed dashes) and then finally for $\gamma \to \infty$ to the green one (small dashes). The range of production is highlighted for each curve through the thin vertical dotted lines.
(b) The thin black dotted lines represent the extremal demand functions given a and b, i.e. $D(\underline{\theta}, p)$ and $D(\overline{\theta}, p)$. From ... to ... b is kept fixed while a is increased, and from ... to ... a is kept constant while b is increased. In red (dashed) the solution for a given value of b. As a increases, the solution widens from the thick deep red region to the thick light red one. In the case for which a is kept constant and b is increased the solution shifts from the dashed deep red region to the full thick blue one.

For concision, we drop the explicit dependencies of the different functions on their arguments in the following equations; $f(\theta)$, $p(\theta)$ and $S(p(\theta))$ will be noted f, p and S respectively. The maximisation program now writes:

$$\max_{p(\cdot)} \int_{-1}^{1} f\left(p(a\theta+b-p-(n-1)S)-\frac{\lambda}{2}(a\theta+b-p-(n-1)S)^2-\frac{\gamma}{2}(1-\theta^2)\left(a-\dot{p}(1+(n-1)S')\right)^2\right)d\theta \tag{1.24}$$

$$s.t. \qquad \dot{p} \in [0, a]$$

$$p \leq a\theta + b$$

with, as before, $\dot{X} = \frac{dX}{d\theta}$ and X' is the derivative of function X with respect to its argument.

Results

Proposition 1.5.1 *The solution exists, is unique, and has the following form:*

$$\forall \theta \in [-1, 1], \; p^*(\theta) = aK_1\theta + bK_2 \tag{1.25}$$

with

$$K_1 = \frac{n\sqrt{(4\gamma + \lambda + n)^2 - 4n + 4} - (4\gamma + \lambda + n)(n - 2)}{2(4\gamma + \lambda + 2n)} \tag{1.26}$$

$$K_2 = \frac{\lambda(n - 1) + K_1(\lambda + n)}{(\lambda + n)(n - 1) + K_1(\lambda + 2n)} \tag{1.27}$$

and the supply schedule has the following expression:

$$S^*(p) = \frac{1}{n}\left(p\left(\frac{1}{K_1} - 1\right) + b\left(1 - \frac{K_2}{K_1}\right)\right) \tag{1.28}$$

Proof See Annex 1.C. ■

Proposition 1.5.2 *The slope of the supply schedule is increasing with γ and the schedule is shifted to the right of the plane (q, p) as γ increases. This is to say that the schedule rotates around a point in the positive quadrant of the plane.*

Proof See Annex 1.D. ■

We are now going to focus on the graphical representation of these solutions. As in the monopoly case we obtain unique solutions of increasing steepness in the ramping cost parameter γ. When the ramping costs increase, it becomes more and more costly to allow for a large domain of potential quantities to be produced.

The black curve in Fig. 1.2 corresponds to the limit solution when $\gamma \to 0$, for which the problem gets closer to that of KM, i.e. no ramping costs. Note that as long as $\gamma \neq 0$ the solutions are unique. This contrasts with the case of $\gamma = 0$ which is the model presented in KM, for which there is a continuum of solutions. There is no smooth transition between our sets of solution : when considering ramping costs, there is a single Nash equilibria, even in the limit of small such costs.

Secondly, in their paper, Klemperer and Meyer show that in the limit of a diverging upper bound for their shocks, their continuum of solutions converges towards a unique solution. Our unique solution in the limit of small ramping costs is the same as that ok KM in the limit of infinite support of demand shocks.

Proposition 1.5.3 *When $\gamma \to 0$, the solution remains unique and converges towards the linear schedule available in KM's set of solutions, that is the same schedule selected with KM's selection rule obtained when considering an infinite support for the shocks.*

Proof It is straightforward to check that K_1 and K_2 have the same values as KM for $\gamma \to 0$.
More intuitively the argument is as follows. When $\gamma \to 0$, with $\gamma > 0$, we retain a unique solution although the problem itself converges towards that of KM. We should select an equilibrium present in KM's continuum. When KM take the limiting case of an infinite support of shocks they select a unique equilibrium. In our case we can do the same thing by taking $a \to \infty$. In the limit, our solution being in their set which converges to a unique equilibrium, those two selected equilibria should be equal. Now note that our solution does not depend explicitly on a so that when the support is finite, we still select the same equilibria out of what is now a continuum of equilibria in KM's framework. ■

Intuitively, as we take γ to 0 we come closer to the situation captured in KM, but as long as $\gamma > 0$, the producer still faces ramping costs, and therefore converges towards the only linear schedule available in KM's set, as shown in Fig. 1.2, in which we plot our solutions on top of KM's solution set in order to clarify the comparison.

Note that it isn't possible to transition smoothly from our model to that of KM, although they are obviously closely related. Indeed, $\forall \gamma > 0$, our model yields unique

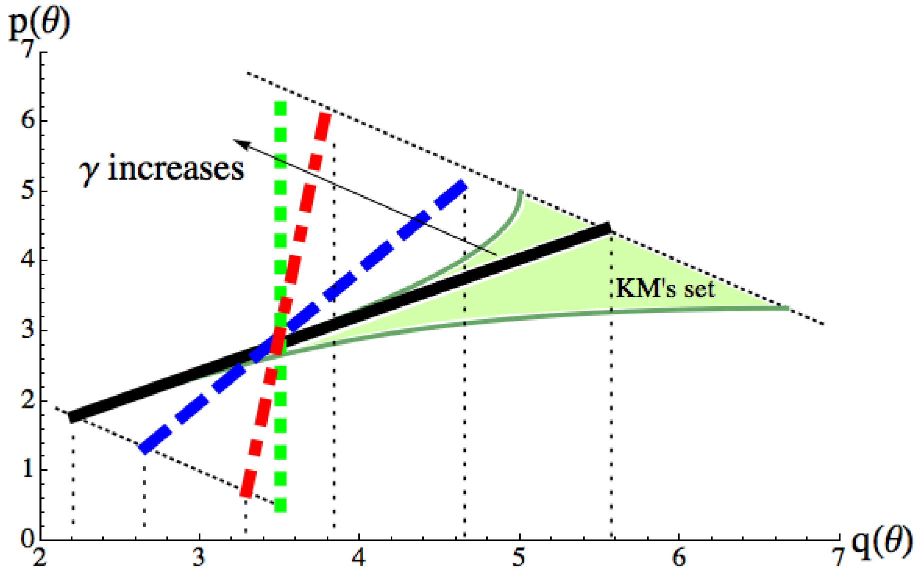


Figure 1.2: This graph plots $S^*(p)$ for different values of the ramping cost parameter, and compares them to the set of equilibria obtained in KM's framework. Four optimal supply schedules are plotted. The black curve (full line) corresponds to the case where $\gamma \to 0$. As before, as γ increases the optimal schedules get steeper and steeper until in the limit of $\gamma \to \infty$, the optimal schedule attains a vertical slope. In addition, we show the set of available equilibria in KM's model in light green, and the extremal demand schedules in dashed black.

solutions, but for $\gamma = 0$ we return to KM's model for which there is a continuum of equilibria. There is an intrinsic discontinuity between these two models, namely, the correspondence $\Gamma(\gamma)$ associating the set of equilibria to the symmetric oligopoly problem obtained for a given value of the ramping cost parameter γ is not lower hemicontinuous at $\gamma = 0$.

In addition to proposing a way to take into account dynamic technological constraints, our model provides a selection rule to choose from the continuum of equilibria described in KM's seminal work, i.e. the solutions' stability to ramping costs.

We have here a model which solutions depend on the distribution of shocks, therefore we are able to capture the interday variation of bids by assuming that the distribution of shocks varies from day to day. In this case, there exists only one symmetric equilibria each day, function of the distribution of shocks.

1.5.1 Discussion

This result sheds some light on one of the questions that the electricity market literature focuses on.

Accounting for ramping costs induces a collapse of the equilibria set from a continuum to a unique element.

Most of the tacit collusion concern that is present in the literature is based on the existence of a continuum of solutions. This continuum is thought as being conducive of tacit collusion because the electricity market entails repeated interactions between producers. In this case, producers can be feared to be able to learn to pick the most profitable Nash equilibria.

Our result implies this pathway for tacit collusion is not available anymore. With only one Nash equilibria at any given time no learning can bring about tacit collusion.

We are also able to account for negative prices which was impossible in the previous framework. Such negative prices are actually observed, although rarely, on the market: producers prefer to subsidize consumption instead of decreasing production by a lot. In our framework, if the ramping costs are large enough, and the demand shocks can reach a small enough value, our solutions can yield negative values : the equilibrium price might even be below the marginal cost of production, understood here as $\partial_q C$ which by definition does not capture our ramping costs.

In the next section we are going to present how to capture richer dynamics, and especially how the surface of bids should evolve with time when the producers have information about the anticipated variation of shocks during the day.

1.6 Dynamic behavior of the bids

The classical supply function equilibria models, as described before, yield a continuum of Nash equilibria, and each one of those equilibria is ex-post optimal. This a very strong result that we are going to take some time to describe and comment.

Consider for a moment that firms competing in supply schedules reach one of the many possible Nash equilibria under such a setup, and that they commit to their schedules. Now consider that the firms face a succession of demand shocks, and that this yields a succession of market outcomes. As the Nash equilbria are ex-post optimal, it means that given the strategies played by the other firms, no firm has any regrets concerning its strategy. Knowing about the realized demand shocks does not imply any willingness to change strategy as long as other firms keep their strategies fixed, and as long as the support of shocks is not reduced ot a point (one could think of observed realisations of shocks as helping to narrow down the expected range of shocks without implying a pinpoint accuracy).

A corollary to this observation is that the distribution of anticipated shocks does not play any role in KM's paper, apart from its bounds. Knowing that the demand shocks are going to be drawn from distributions of high or low values does not affect the willingness to play a given strategy, as long as the support does not evolve. The little role that is played by information about shocks in KM's paper is even more counter-intuitive : to a certain extent, information about demand shocks gives rise to a larger continuum of solutions. Indeed, if one compares the equilibria available to firms for a given support $\{\theta\}_1 = [\underline{\theta}_1, \overline{\theta}_1]$, noted ${S^*}_1$, to those obtained for a support strictly included in the first one $\{\theta\}_2 = [\underline{\theta}_2, \overline{\theta}_2] \subset \{\theta\}_1$, noted ${S^*}_2$, then the set of equilibria will be larger in the second case, in the sense that ${S^*}_1 \restriction_{\{\theta\}_2} \subset {S^*}_2$ (where $\restriction_{\{\theta\}_2}$ denotes that the supply functions are restricted to values over $\{\theta\}_2$).

However, actual firms bidding on the electricity markets are known to be actively engaged in forecasting the future demand levels in order to build their strategies. Bids that we can observe on the electricity markets change from hour to hour even when demand does not vary enough to warrant a change of online plants, a consideration that could explain some of the supply schedules variations.

The general interpretation of KM's paper when applied to electricity markets is that for some unknown underlying process, strategies converge towards different equilibria of the set of available equilibria from hour to hour. One can note that the general intuition for strategies converging towards Nash equilibria in the first place is through either a high degree of sophistication on the part of firms, or through a more organic learning process. Neither of these two explanations can account for frequent switching from one Nash equilibria to another, out of a myriad of available options, wihtout considering some communication among firms. Furthermore, if such communication existed, it should be expected to yield the most profitable equilibria out of the available lot.

We think that this strand of argument trying to explain bids' dynamics in the light of the supply function equilibria framework is unsatisfying and we argue that forecasting demand becomes important for firms when one considers dynamic effects, that is effects that are history dependent, of which ramping costs which we model in this paper are an instance (one can think of start-up and shut-down costs as another instance of such dynamic effects).

The model described in the previous section doesn't account for these hourly dynamics. Here we present a way to capture these intraday variations, by considering bids that depend continuously on the time t. We will show that our results imply that firms are not oblivious to information about the distribution of shocks anymore, and more than that, that their strategies directly evolve with the evolution of their knowledge about uncertain furture shocks.

1.6.1 The setup

Previously, the SDE (stochastic differential equation) defining the dynamics of the problem was written as:

$$d\theta(t) = -2\theta(t)dt + \sqrt{1-\theta(t)^2}dB_t$$

This specification implied a stochastic trajectory for the shocks, bounded by a constant envelope. That is to mean that, lacking any knowledge of the value of the shock at a point in time close to the period under consideration, the distribution of shocks does not depend on time.

To account for these intraday variations we are going to define a richer SDE.

SDEs have been well studied and as a consequence there exists a number of families of SDEs satisfying numerous characteristics [Hertzler, 2003]. The goal here is to find one SDE that will allow us to capture some of the dynamics of shocks and how this might influence strategies, while keeping it as simple as possible. Just as in the previous section, the first caracteristic that we want is to consider SDEs that imply a bounded support of shocks. This restricts our possible choice to four families out of the classical ones: Generalised Beta I, Beta, Power, Uniform. We also consider that a desirable property is that the distribution reaches 0 continuously at the bounds of its support. This restricts us further to only two families : Generalised Beta I and Beta. For tractability reasons we will focus here on the Beta family of SDEs, and more precisely on one of the simplest Beta SDE. However, we want to note that this choice stems from our focus towards solving analytically the problem at hand and obtain closed form solutions. If one were to try and estimate the distribution of shocks anticipated by firms from market data one might want to try and find which of the Beta or Generalised Beta I SDEs might match the distribution of errors between the published day -1 estimates for demand and the observed quantities.

Define the evolving envelope of shocks by two functions, $(\underline{\theta}(t), \overline{\theta}(t))$, respectively the lower and upper bounds of the shocks. These two functions, although very easy to comprehend, are not the most useful way to define the boundary. Instead we are going to use the average value of the shocks, and the half width of the envelope, $(\hat{\theta}(t), \omega(t))$. This means that $\underline{\theta}(t) = \hat{\theta}(t) - \omega(t)$ and $\overline{\theta}(t) = \hat{\theta}(t) + \omega(t)$. The only restriction we impose on the envelope is that we require it to be continuously differentiable, that is $(\hat{\theta}, \omega) \in \mathcal{C}^1(\mathbb{R})$.

Consider the following SDE which is the simplest Beta SDE that we can pick that still allows us to have a free choice of the bounds of shocks. For readability, we drop the explicit dependency of the different functions on time, that is $\theta(t)$, $\hat{\theta}(t)$ and $\omega(t)$ will be noted θ, $\hat{\theta}$ and ω:

$$\begin{aligned} d\theta = & \left[(\hat{\theta} - \omega - \theta) + \left(1 + \frac{1}{\omega}\frac{d\omega}{dt}\right)(\hat{\theta} + \omega - \theta) + \left(\frac{d\hat{\theta}}{dt} - \frac{d\omega}{dt}\right) \right] \cdot dt \\ & + \sqrt{\left(1 + \frac{1}{\omega}\frac{d\omega}{dt}\right)(\theta - \hat{\theta} + \omega)(\hat{\theta} + \omega - \theta)} \cdot dB_t \end{aligned} \tag{1.29}$$

The distribution of the shocks can be obtained through Fokker-Planck's equation 1.5 and we obtain:

$$f(\theta, t) = \frac{3}{4\omega(t)^3}(\theta(t) - \hat{\theta}(t) + \omega(t))(\hat{\theta}(t) + \omega(t) - \theta(t)) \tag{1.30}$$

In the following analysis, we are going to rely on the fact that the term $\left(1 + \frac{1}{\omega}\frac{d\omega}{dt}\right) > 0$. The justification for this inequaliy comes from the following remark: if one were to rescale time in the above equations, there wouldn't be any explicit change in the equilibrium distribution 1.30. The only effect that such a rescaling would play is in the variance of the Brownian term. In order to insure that our inequality is correct, one has to make sure that the variation of the envelope term occurs on longer timescales than the characteristic timescale of fluctuations in our problem, that is the timescale that fixes the rate at which information leaks out of the knowledge of the value of one shock at a given point in time. We are trying to capture the hourly changes in firms strategies when demand fluctuates at higher frequencies (think of the collection of individuals that choose to switch lights on or off at any given point in time in an entire country for instance). We therefore consider that this assumption is sound in this situation.

More formally, one can define τ a rescaling parameter allowing to change the rate at which the brownian process blurs information pertaining to an initial condition. We rescale time using this parameter, so that time t and the rescaled time t_r verify $t_r = \tau t$.

We can rewrite the above equations as :

$$
\begin{aligned}
d\theta = &\left[(\hat{\theta} - \omega - \theta) + \left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt_r}\right)(\hat{\theta} + \omega - \theta) + \tau\left(\frac{d\hat{\theta}}{dt_r} - \frac{d\omega}{dt_r}\right)\right] \cdot dt_r \\
&+ \sqrt{\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt_r}\right)(\theta - \hat{\theta} + \omega)(\hat{\theta} + \omega - \theta)} \cdot dB_{t_r}
\end{aligned}
\tag{1.31}
$$

and

$$
f(\theta, t_r) = \frac{3}{4\omega(t_r)^3}(\theta(t_r) - \hat{\theta}(t_r) + \omega(t_r))(\hat{\theta}(t_r) + \omega(t_r) - \theta(t_r)) \tag{1.32}
$$

By assumption, τ is small enough for the loss of information due to the stochastic nature of the process to be faster than the typical timescale of variation of strategies, therefore by hypothesis $\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt_r}\right) > 0$ is valid. We will drop this rescaled time index in the following sections as equations 1.30 and 1.32 are equal, it was just a temporary definition to justify the sign of the term that depends on the time derivative of the envelope. We will keep this τ parameter explicit however, in order to allow discussions differentiating effects related to the speed of variation of the envelope or to the relative timescales of this variation and the underlying stochastic process.

1.6.2 Results

Dynamics in the case of the Monopoly and of the oligopoly

We start by describing the dynamics of the monopoly case because the oligopoly case is not richer dynamically, but it is more complex to describe.

Our stochastic maximisation program can thus be rewritten as a regular optimal control problem as in section 1.4, but taking into account the time depency:

$$\max_{S_i(p,t)} \int_0^T \int_{\underline{\theta}(t)}^{\overline{\theta}(t)} f(\theta,t)\left(p(\theta,t)S_i(p(\theta,t),t) - C_s(S_i(p(\theta,t),t)) - \frac{\gamma}{2}\sigma(\theta,t)^2\left(S_i'(p(\theta,t),t)\dot{p}(\theta,t)\right)^2\right)d\theta dt \tag{1.33}$$

$$s.t. \qquad S_i'(\cdot) \geq 0$$

$$\dot{p} \geq 0 \tag{1.34}$$

$$D(\cdot,\cdot) \geq 0$$

$$\tag{1.35}$$

Proposition 1.6.1 *In the case of an envelope evolving with time, that is shocks belonging to the bounded support $[\hat{\theta}(t) - \omega(t), \hat{\theta}(t) + \omega(t)]$, there exists a unique optimal solution to the monopoly problem. It can be expressed as as surface in the price-quantity-time space:*

$$p^*(\theta(t),t) = \frac{4\gamma\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}(t)\right) + 1 + \lambda}{4\gamma\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}(t)\right) + 2 + \lambda} \cdot \theta(t) - \frac{1+\lambda}{2+\lambda} \cdot \hat{\theta}(t) \tag{1.36}$$

The corresponding optimal supply schedule writes as:

$$S^*(p,t) = \frac{1}{4\gamma\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}(t)\right) + 1 + \lambda}\left(p(t) + \frac{4\gamma\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}(t)\right)}{2+\lambda} \cdot \hat{\theta}(t)\right) \tag{1.37}$$

$\forall p(t) \in [p(\hat{\theta}(t) - \omega(t), t), p(\hat{\theta}(t) + \omega(t), t)]$

Proof See Annex 1.E. ■

Note that if $\frac{d\omega}{dt} = 0$ equations 1.36 and 1.37 are equal to equations 1.18 and 1.19 respectively as expected. Note also that the solution is exactly the same as in the static monopoly case in which one replaces the ramping cost parameter γ by $\gamma\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}(t)\right)$. This surprising fact, that our dynamic optimal strategy is simply the naive transcription of the static one with a specified dynamic stochastic process, can be understood as s consequence of the assumptions we have had to make in section 1.3.2.

In this section, in Annex. 1.A in which we develop the argument in more detail, and in section 1.6.1 we end up in effect making a scale separation argument : the ramping costs are completely driven by the very short term fluctuations, whereas the evolution of these ramping costs is driven by the longer timescale at which our information about the demand shocks evolves over time. This means that we make a version of what physicists call a quasi-static argument : because of this time-scale separation between what drives our ramping cost and our information about the shocks, we can effectively reason in two steps, first solving for the static situation, and then injecting naively the slow changes in the static results with confidence as to the validity of this approximation as long as the assumption about this separation of scale is verified.

The consequence of this is that we have a dynamic transcription of our static oligopoly of the same nature as for the monopoly above.

Proposition 1.6.2 *The solution exists, is unique, and has the following form:*

$$\forall\theta \in [-1, 1],\ p^*(\theta) = aK_1(t)\theta + bK_2(t) \tag{1.38}$$

with

$$K_1(t) = \frac{n\sqrt{\left(4\gamma\left(1+\frac{\tau}{\omega}\frac{d\omega}{dt}\right)+\lambda+n\right)^2-4n+4}-\left(4\gamma\left(1+\frac{\tau}{\omega}\frac{d\omega}{dt}\right)+\lambda+n\right)(n-2)}{2\left(4\gamma\left(1+\frac{\tau}{\omega}\frac{d\omega}{dt}\right)+\lambda+2n\right)} \tag{1.39}$$

$$K_2(t) = \frac{\lambda(n-1)+K_1(t)(\lambda+n)}{(\lambda+n)(n-1)+K_1(t)(\lambda+2n)} \tag{1.40}$$

and the supply schedule has the following expression:

$$S^*(p,t) = \frac{1}{n}\left(p\left(\frac{1}{K_1(t)}-1\right)+\hat{\theta}\left(1-\frac{K_2(t)}{K_1(t)}\right)\right) \tag{1.41}$$

Proof See Annex 1.F.■

1.6.3 Discussion

In both situations, the optimal supply schedule is shifted uniformly following the expected shock $\hat{\theta}(t)$, which is a rather intuitive result : if on average demand shifts upwards, the producers want to extract more profit and shift their supply curve accordingly, but there is no reason to change slope.

What is less trivial is the way the slope behaves. Let us focus on the monopoly result for a start. The slope is affected as if the ramping cost parameter was fluctuating with the relative change in the width of the bounds of the shocks (term in $\frac{1}{\omega}\frac{d\omega}{dt}$). The transition between a low uncertainty region to a higher uncertainty one behaves as if during the transient regime the ramping cost parameter had a higher value, implying a higher slope.

The optimal supply schedule depends on the relative rate of change of the width $\frac{1}{\omega}\frac{d\omega}{dt}$ and on the average shock $\hat{\theta}$. More precisely, with a constant width, the optimal supply schedule varies according to variations in the expected average value of the shocks. This is quite standard, if demand is higher, the price and quantities both increase, and here this increase occurs with a constant slope. The behavior of the supply schedule when the width varies is less trivial.

Remember that when describing the slope of the schedule, we are considering the plane $(quantity, price)$ while the schedule as defined by $S^*(p)$ represents the same curve but in the plane $(price, quantity)$. An increase in width is equivalent to a higher ramping cost parameter while a decrease in width is equivalent to a lower ramping cost parameter. These results are illustrated in Fig. 1.3.

To understand the economic intuition behind this result, consider first an increase in the width of the envelope at date t_1. Consider now one possible value of $\theta(t_1)$. At $t_1 + dt$, had the width been constant there would have been a given level of uncertainty about the values that $\theta(t_1 + dt)$, and thus the ramping costs, could have taken. If the width of the envelope is increasing then there is more uncertainty regarding the potential values that could be taken by $\theta(t_1 + dt)$, therefore more expected ramping costs incurred, and a higher slope to hedge these costs. On the other hand, when the width decreases, the situation is reversed. In that case, we move towards a situation in which there is less uncertainty about the ramping costs, so that the slope is smaller than for a constant envelope. This difference between increasing and decreasing width is illustrated by comparing the two regions of the envelope displayed in (full) black line in Fig. 1.3. In addition, when contrasting the left and the right side of the figure one sees that the change in the informativeness of the envelope is captured by the relative change of the width: for the same rate of change, if the width is larger (right) then the change in informativeness is smaller (the change in the area captured by the (dashed) red and (full) green arrows).

All of this reasoning applies to the dynamic oligopoly result as well as the effect can be understood in the same way as for the monopoly : changes in the width of the shock's bounds behave as if there was an effective dynamic cost that was higher than the baseline when information about the shocks is lost, and lower than the baseline when information is gained.

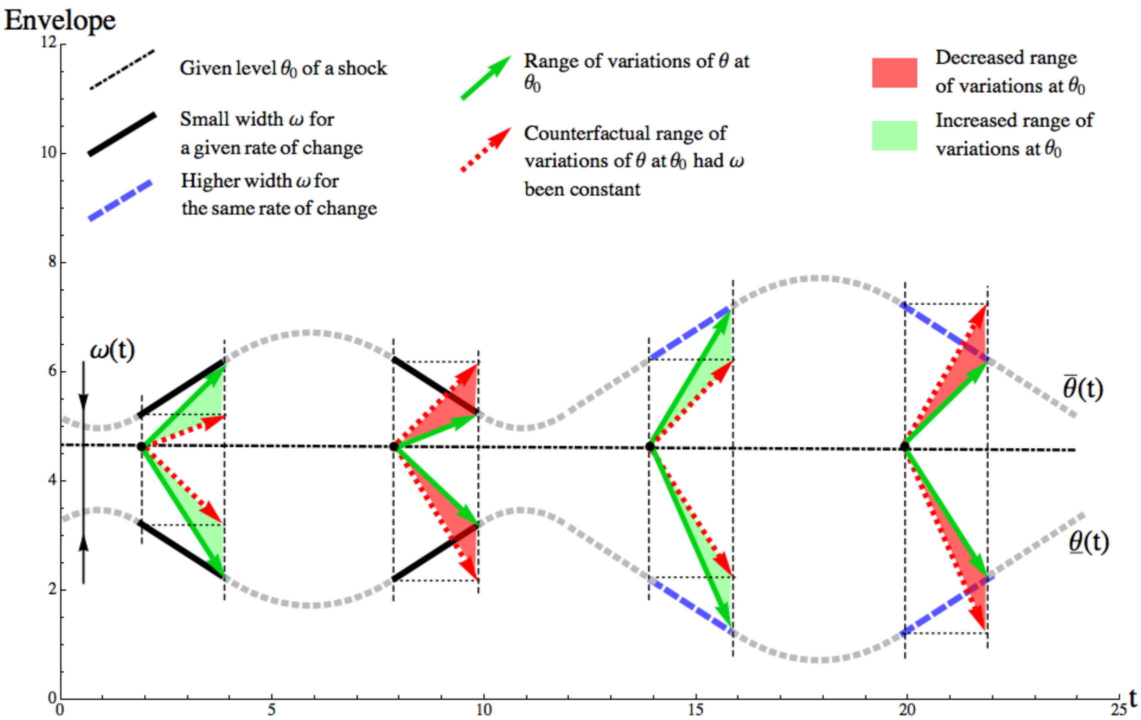


Figure 1.3: This graph plots an envelope of constant average value but varying width $\omega(t)$. By comparing regions of increasing or decreasing width, respectively the left or right side of a lobe, one sees that the informativeness of the envelope is being respectively reduced or increased with respect to a situation where the width would be kept constant. The change in informativeness is represented by the area between the (full) green arrows (observed level of informativeness) compared to the area between the (dashed) red arrows (level of informativeness had the width been constant). In addition, by comparing the left lobe to the right one, it is possible to see why the relative variation of the width, and not the absolute variation of the width, matters. For a larger width (right lobe) and the same rate of change in the width, there is less change in informativeness than for a smaller width (left lobe), i.e. the same rate of change matters less for the right lobe than for the left lobe.

1.7 Limits

This section aims at discussing whether or not one can consider that the mapping of these results on the real world is a set of non zero measure, to put it bluntly.

Further avenues of research would be to generalize our results to larger classes of demand functions. As hinted in the text of this chapter, a lot of effort has been invested towards this goal without results unfortunately. One could also solve the static case for different SDE's in order to test the sensitivity of our results on the underlying "mechanics" of the stochastic process. This has also been pursued without conclusive results : solving the optimization problem becomes quickly extremely difficult, as the second order differential equations exhibit poles, and divergences are difficult to cope with in optimization problems.

The nature of these avenues of research is testament to the fact that our results are obtained for a very narrow setting. However, although a healhty dose a skepticism as to the applicability of the closed form formulas is therefore warranted, I would like to argue that the results hint towards at least one more general takeaway message, namely the collapse of the set of equilibria.

This result stems from the nature of the mathematical problem and not from the way we set up the problem in order to maximize our chances of closed-form success per se. Therefore I think it hints towards possible more general results. The problem is the complexity of the maths as soon as one deviates from the simplest version of the problem presented here.

The question then becomes one of the method to employ to obtain those results. The brute force mathematical approach has proved too hard for the writer of these lines, but there is one tool that might prove useful : numerical simulations. One can solve the differential equations involved here numerically, check ex-post whether they satisfy the other conditions, and in so doing provide boundaries around possible solutions. If the unicity is a characteristic that is indeed more general than our model here, there is 0 probability of finding such a solution by the method proposed, quite literally. However providing such bounds, although not demonstrating the existence of a solution, could provide circumstantial evidence towards such a result.

More generally, I think that economics has not yet explored the full potential of numerical methods as a guide for theoretical results.

1.8 Discussion and Concluding Remarks

In this chapter we have introduced a supply function equilibria model of ramping costs under uncertainty.

By introducing technological constraints previously neglected we are able to take into account the effects of the dynamics of demand shocks on the supply function framework. We restrict ourselves to linear demand. The optimal supply schedules obtained are unique. This is a striking result when compared to traditional multiplicity of equilibria. Although we do not solve the model in the case of a general demand function (half a year was spent trying to find mathematical methods to tackle this, to no avail) we think that our results make a strong case for the reduction of the set of equilibria, in our case to a unique equilibrium, when taking into account dynamic effects, that is strategies that are history dependent.

We introduce a mathematical toolbox that was absent from this litterature in the past, and notably classes of stochastic differential equations that can be used to pick and choose processes yielding specific closed form distributions of probability of shocks at equilibrium.

Our methodology further introduces the notion of time-scale separation to our problem, which allows to transcribe quite simply static solutions to the case of dynamic envelopes of shocks, as long as the static case is solved for the same functionnal form of stochastic processes. In our case we focus our study to quadratic distributions, which we then extend to cope with any functional form for the time dependency of the envelope of shocks.

Our results are congruent with the economic intuition one can have about ramping costs : when they increase, the slope of the supply schedule increases in order to reduce the range of variation in production for a given range of variation of demand shocks.

Although mathematically more demanding than the traditional model by Klemperer and Meyer, we consider that this new model, while conceptually sparing (we only add ramping costs) allows for a richer, more realistic description of the electricity market, and opens new research avenues. It yields precise and testable predictions on the dynamics of the electricity market with tractable functional forms, at least in the linear demand case.

In addition, by explicitly modeling the dynamics, our work opens the possibility to explore interactions between intraday and day-ahead markets, markets that were indistinguishable in the previous framework : if solutions are ex post optimal, there is no need to create a second type of spot market, with a shorter time horizon, the bids of the previous day should suffice.

Further avenues of research would be to generalize our results to larger classes of demand functions. As hinted in the text of this chapter, a lot of effort has been invested towards this goal without results unfortunately. One could also solve the static case for different SDE's in order to test the sensitivity of our results on the underlying "mechanics" of the stochastic process. This has also been pursued without conclusive results : solving the optimization problem becomes quickly extremely difficult, as the second order differential equations exhibit poles, and divergences are difficult to cope with in optimization problems.

Finally, and more generally, we think that this concept of ramping costs, the fact that change is costly, is ubiquitous and could fuel interesting research into the dynamics of a large range of markets. Such avenues have been pursued in the case of stochastic optimal control, that is, instantaneous reactions to stochastic shocks. Here we are describing a market on which agents are forced to optimize in advance, so that they have to react to continuous changes in the anticipated shocks, but not the shocks themselves, which can be understood as stochastic optimization with periodic commitment.

Appendix

Appendix 1.A Proof of equation 1.11

We are here going to detail how we obtain the result in equation 1.11 on which the proofs of our dynamic results rely heavily. Recall that we are computing the continuous time limit of our ramping cost term which can be quite simply defined in the case of discrete dynamics but for which one has to work a bit more in order to cope with the non differentiable nature of stochastic processes.

We are therefore going to first consider the discrete case of a random walk of timestep Δt which converges towards the Itō process 1.4, using the Euler-Maruyama approximation, a generalisation of the Euler method to stochastic differential equations. We consider a Markov chain Y defined as follows :

$$\Delta Y_n = Y_{n+1} - Y_n = \mu(Y_n, n\Delta t)\Delta t + \sigma(Y_n, n\Delta t)\Delta B_n \tag{1.A.1}$$

We want to derive the following :

$$\lim_{\Delta t \to 0} \mathbb{E}\left[\frac{\Gamma(\Delta t)}{2}\left(\frac{\Delta S_i(p(\theta(t),t),t)}{\Delta t}\right)^2 \middle| \theta(t)\right] = \frac{\gamma}{2}(\partial_1 S(p(\theta(t),t),t)\partial_1 p(\theta(t),t))^2\sigma(\theta,t)^2 \tag{1.A.2}$$

Let us first compute the first order expansion of $\Delta S_i(p(Y_n, n\Delta t), n\Delta t)$, by assuming that both S_i and p are continuously differentiable with respect to their arguments:

$$\Delta S_i(p(Y_n, n\Delta t), n\Delta t) = \frac{\Delta S_i}{\Delta p}\left(\frac{\Delta p}{\Delta Y}\frac{\Delta Y}{\Delta t}\Delta t + \frac{\Delta p}{\Delta t}\Delta t\right) + \frac{\Delta S_i}{\Delta t}\Delta t + \mathcal{O}(\Delta t^2) \tag{1.A.3}$$

Using our differentiability assumption, note that the terms that do not depend on ΔY scale with Δt, and that the term depending on ΔY cannot be grouped in the same way, due to its stochastic nature, therefore:

$$\frac{\Delta S_i(p(Y_n, n\Delta t), n\Delta t)}{\Delta t} = \frac{\Delta S_i}{\Delta p}\frac{\Delta p}{\Delta Y}\frac{\Delta Y}{\Delta t} + \mathcal{O}(1) \tag{1.A.4}$$

$$\left(\frac{\Delta S_i(p(Y_n, n\Delta t), n\Delta t)}{\Delta t}\right)^2 = \left(\frac{\Delta S_i}{\Delta p}\frac{\Delta p}{\Delta Y}\frac{\Delta Y}{\Delta t}\right)^2 + C \cdot \frac{\Delta S_i}{\Delta p}\frac{\Delta p}{\Delta Y}\frac{\Delta Y}{\Delta t} + \mathcal{O}(1) \tag{1.A.5}$$

with C a term that does not depend on ΔY or Δt.

Now by considering the specification of our stochastic process we know that $\mathbb{E}\left[\frac{\Delta Y}{\Delta t}|Y_n\right] = \mu(Y_n, n\Delta t)$ and that $\mathbb{E}\left[\frac{\Delta Y}{\Delta t}^2\middle| Y_n\right] = \mu(Y_n, n\Delta t)^2 + \frac{\sigma(Y_n, n\Delta t)^2}{\Delta t}$. Using the fact that $\Gamma(\Delta t) = \gamma\Delta t + o(\Delta t)$ we obtain the result of equation 1.11.

Appendix 1.B Proof of Proposition 1.4.1

Define the following Hamiltonian:

$$\begin{aligned} H(p(\theta), \dot{p}(\theta), \mu(\theta), \theta) = f(\theta)\Bigg(& p(\theta)(a\theta + b - p(\theta)) - \frac{\lambda}{2}(a\theta + b - p(\theta))^2 \\ & -\frac{\gamma}{2}(1-\theta^2)\left(a - u(\theta)\right)^2\Bigg) + \mu(\theta)u(\theta) \end{aligned} \tag{1.B.1}$$

where $u(\theta)$ is the control variable defined through the following equation of motion: $u(\theta) = \dot{p}(\theta)$, $u(\theta) \in [0, a]$. We do not consider the non-negative demand constraint and will check ex-post that our solution verifies this condition.

Now note that:

$$\forall \theta \in (-1, 1), \quad \frac{\partial^2 H}{\partial p^2} = -(2+\lambda)f(\theta) < 0 \tag{1.B.2}$$

$$\frac{\partial^2 H}{\partial u^2} = -\gamma(1-\theta^2)f(\theta) < 0 \tag{1.B.3}$$

The Hamiltonian is therefore strictly concave in $p(\theta)$ and $u(\theta)$. Let $(p^*(\theta), u^*(\theta))$ be an admissible pair to the problem, that is a pair such that $u^*(\theta) = \dot{p}^*(\theta)$. If there exists a continuous and piecewise continuously differentiable function $\mu(\theta)$ such that:

$$\dot{\mu}(\theta) = -\frac{\partial H^*}{\partial p} \tag{1.B.4}$$

$$\mu(-1) = \mu(1) = 0 \qquad \text{in order for prices to be free at the boundaries} \tag{1.B.5}$$

$$\forall(\theta, u) \in [-1, 1] \times [0, a],\ \frac{\partial H^*}{\partial u}(u^*(\theta) - u) \geq 0 \tag{1.B.6}$$

with $\frac{\partial H^*}{\partial u} = \frac{\partial H}{\partial u}(p^*(\theta), u^*(\theta), \mu(\theta), \theta)$, then the Mangasarian sufficiency theorem ensures that $(p^*(\theta), u^*(\theta))$ is the optimal solution [Seierstad and Sydsaeter, 1987, p.105]. Let us check that eq. 1.18 defines the optimal solution.

Equation 1.B.4 defines $\mu(\theta)$ up to a constant. Through direct integration we obtain:

$$\mu(\theta) = 3a\left((2+\lambda)\frac{4\gamma + 1 + \lambda}{4\gamma + 2 + \lambda} - 1 - \lambda\right)(2\theta^2 - \theta^4) + const.$$

This expression is symmetric in θ therefore by choosing the adequate value for the constant, we ensure that eq. 1.B.5 is satisfied. The slope of the proposed p^* is in $[0, a]$ therefore eq. 1.B.6 requires $\frac{\partial H}{\partial u}$ to be null.

$$\forall\theta \in [-1, 1],\ \frac{\partial H}{\partial u} = 0 \implies \frac{d}{d\theta}\frac{\partial H}{\partial u} = 0$$

$$\text{i.e. } \dot{u}(\theta) = -\frac{4\theta}{1-\theta^2}(a - u(\theta)) - \frac{(1+\lambda)(a\theta + b)}{\gamma(1-\theta^2)} + \frac{(2+\lambda)p(\theta)}{\gamma(1-\theta^2)} \tag{1.B.7}$$

It is straightforward to see that the proposed solution satisfies this differential equation, thus we know that $\frac{\partial H}{\partial u}$ is a constant and as $\mu(-1) = 0$ it is in fact null. Lastly, we see that $p^*(\theta) \leq a\theta + b$.

The proposed $p^*(\theta)$ therefore defines the unique optimal supply function, i.e. the parametrized curve $(a\theta + b - p^*(\theta), p^*(\theta))$.

Appendix 1.C Proof of Proposition 1.5.1

As for eq. 1.24, for the sake of concision, we do not write the explicit depencies of the different functions on θ, thus $f(\theta)$, $p(\theta)$, $u(\theta)$, $\mu(\theta)$ and $S(p(\theta))$ will be written as f, p, u, μ and S respectively. Define the following Hamiltonian:

$$\begin{aligned} H(p,u,\mu,\theta) = f\Bigg(& p(a\theta+b-p-(n-1)S) - \frac{\lambda}{2}(a\theta+b-p-(n-1)S)^2 \\ & -\frac{\gamma}{2}(1-\theta^2)\left(a-u(1+(n-1)S')\right)^2\Bigg) + \mu u \end{aligned} \tag{1.C.1}$$

where u is the control variable defined through the following equation of motion: $u = \dot{p}$, $u \in [0,a]$. We do not consider the non-negative demand constraint and will check ex-post that our solution verifies this condition.

If a symmetric equilibria exists, eqs. 1.20 through 1.23 imply that the regular conditions for an admissible pair to be optimal write :

$$u = \dot{p} \in [0,a] \tag{1.C.2}$$

$$\partial_u H < 0 \implies u = 0 \tag{1.C.3}$$

$$\partial_u H > 0 \implies u = a \tag{1.C.4}$$

$$\begin{aligned} & \partial_u H = 0 \implies u \in [0,a] \text{ and} \\ \ddot{p} = & -\frac{4\theta(a-\dot{p})}{1-\theta^2} - \frac{\lambda(a\theta+b-p)}{\gamma(1-\theta^2)} - n\frac{\dot{p}(a\theta+b-2p)-a(n-1)p}{\gamma(1-\theta^2)(a(n-1)+\dot{p})} \end{aligned} \tag{1.C.5}$$

$$\dot{\mu} = -\partial_p H \tag{1.C.6}$$

$$\mu(-1) = \mu(1) = 0 \tag{1.C.7}$$

It is easy to check that $(K_1, K_2) \in (0,1)$ and that the solution 1.25 solves eq. 1.C.5 subject to the boundary conditions 1.C.7. The supply schedule is therefore also linear, with equation :

$$S(p) = \frac{1}{n}\left(p\left(\frac{1}{K_1}-1\right) + b\left(1-\frac{K_2}{K_1}\right)\right) \tag{1.C.8}$$

We can now use the Mangasarian theorem to obtain that our admissible pair is indeed solution, $H(p, u, \mu, \theta)$ being concave in (p, u) for linear supply schedules. However the Mangasarian cannot yield that this solution is unique because for a symmetric equilibria, if supply schedules are modified, the hamiltonian changes alongside and we are faced with a new maximisation program.

To obtain that the solution is unique we are going to show explicitly that no other candidate solution exists.

First, note that :

$$\begin{aligned}\dot{\mu} = -f\Bigg(& \frac{a\theta + b - 2p}{n} - a\frac{(n-1)p}{n\dot{p}} \\ & + \lambda\frac{a\theta + b - p}{n} \cdot \frac{a(n-1) + \dot{p}}{n\dot{p}} - \gamma(1-\theta^2)(n-1)\frac{a - \dot{p}}{n} \cdot \frac{a\ddot{p}}{n\dot{p}^2}\Bigg)\end{aligned} \tag{1.C.9}$$

If (p^*, u^*) maximises the program then the maximum principle implies that there exists a continuous and piecewise continuously differentiable function μ, as shown in [Seierstad and Sydsaeter, 19 Theorem 2 p.85]. This combined with the above equation implies that $\dot{p} \neq 0$ a.e.

Assume now a solution of the form $\forall\theta \in [-1, 1],\ p = a\theta + \beta$, by injecting this expression in eq. 1.C.9 there is no β such that the boundary conditions 1.C.7 are verified.

In addition:

$$\forall\theta \in (-1, 1),\ \frac{\partial^2 H}{\partial u^2} = -f\gamma(1-\theta^2)(1 + (n-1)S')^2 < 0 \tag{1.C.10}$$

The Hamiltonian is therefore strictly concave in u and $[0, a]$ is convex. These two properties yield that u^* is continuous, as shown in [Seierstad and Sydsaeter, 1987, Note 2.b. p.86]. We have proved the following result :

Lemma 1.C.1 *For any symmetric equilibrium $\exists A \subseteq [-1, 1]$ s.t. A is the union of segments of $[-1, 1]$ and $\forall\theta \in A$, $\partial_u H = 0$*

Assume the following hypothesis is true, $H_1 : \exists \theta_c \in (-1,1)$ s.t. $[-1,\theta_c] \subseteq A$, then knowing that $\dot{p} \in \mathcal{C}^0([-1,1],[0,a])$ we can rewrite differential equation 1.C.5 around the value $\theta = -1$ by defining $\theta = -1 + \epsilon$ with $\epsilon = o(1)$:

$$\frac{d^2 p}{d\epsilon^2} = \frac{C}{\epsilon} + o(1) \text{ with } C \neq 0 \text{ if } p(\theta) \neq aK_1\theta + bK_2 \tag{1.C.11}$$

This means that locally around -1, any solution to eq. 1.C.5 but solution 1.25 diverges. Hypothesis H_1 is therefore wrong and $\exists \theta_c \in (-1,1)$ s.t. $\forall \theta \in [-1,\theta_c]$, $\exists \beta$ s.t. $p(\theta) = a\theta + \beta$.

At θ_c we have $\partial_u H = 0$ and as $\dot{p}$ is continuous, $\dot{p}(\theta_c) = a$. For the solution to be interior we need $\ddot{p}(\theta_c) \leq 0$.

$$\partial_{\dot{p}} H(p,\dot{p},\mu,\theta_c) = 0 \Leftrightarrow \mu(\theta_c) = 0 \tag{1.C.12}$$

$$\ddot{p}(\theta_c) \leq 0 \Leftrightarrow b(1+\lambda) - \beta(n+1+\lambda) \geq na\theta \tag{1.C.13}$$

Straightforward computations show that both conditions are mutually exclusive, therefore there doesn't exist another candidate symmetric equilibria, and our solution is unique.

Lastly, to compute the optimal supply function, we inverse the optimal price in order to get the shock as a function of the price at the equilibrium, and we inject this expression in Eq. 1.21.

Appendix 1.D Proof of proposition 1.5.2

We want to prove that the slope of the supply schedules increases as the ramping cost parameter increases. As a reminder:

$$K_1 = \frac{n\sqrt{(4\gamma+\lambda+n)^2-4n+4}-(4\gamma+\lambda+n)(n-2)}{2(4\gamma+\lambda+2n)} \tag{1.D.1}$$

$$K_2 = \frac{\lambda(n-1)+K_1(\lambda+n)}{(\lambda+n)(n-1)+K_1(\lambda+2n)} \tag{1.D.2}$$

and the supply schedule has the following expression:

$$S^*(p) = \frac{1}{n}\left(p\left(\frac{1}{K_1}-1\right)+b\left(1-\frac{K_2}{K_1}\right)\right) \tag{1.D.3}$$

Let us study how K_1 varies with γ. Note first that if one defines $G = 4\gamma+\lambda+n$, then $\frac{\partial K_1}{\partial \gamma} = \frac{\partial K_1}{\partial G}\frac{\partial G}{\partial \gamma} = 2\frac{\partial K_1}{\partial G}$.

therefore the sign of $\frac{\partial K_1}{\partial \gamma}$ is given by that of:

$$\frac{\partial K_1}{\partial G} = \frac{\partial}{\partial G}\left[\frac{n\sqrt{G^2-4n+4}-G(n-2)}{2(G+n)}\right] \tag{1.D.4}$$

$$= \frac{(\sqrt{G^2-4n+4})((G+n)(nG-(n-2)\sqrt{G^2-4n+4})-(n\sqrt{G^2-4n+4}-G(n-2)))}{4(G+n)^2} \tag{1.D.5}$$

$$= \frac{(\sqrt{G^2-4n+4})(n^2G+4n^2-4n-n(n-2)\sqrt{G^2-4n+4})}{4(G+n)^2} \tag{1.D.6}$$

$$= \frac{(\sqrt{G^2-4n+4})(2G+4+(n-2)(G+4-\sqrt{(G+4)^2-8G-4n-12}))}{4(G+n)^2} > 0 \tag{1.D.7}$$

Therefore $\frac{\partial S^*(p)}{\partial \gamma} < 0$ which implies that schedules see their slope increase with γ in the plane (q,p).

We can perform the same type of computation for the ratio $\frac{K_2}{K_1}$, using the fact that

$\partial_\gamma K_1 > 0$:

$$\frac{\partial K_2/K_1}{\partial\gamma} = -\frac{\partial_\gamma K_1(K_1^2(\lambda+2n)(\lambda+n) + 2K_1(\lambda+2n)(n-1)\lambda + \lambda(\lambda+n)(n-1)^2)}{K_1^2((\lambda+n)(n-1) + K_1(\lambda+2n))^2} < 0 \tag{1.D.8}$$

This implies that the schedule is shifted to the right in the plane (q,p) when ramping costs increase.

Appendix 1.E Proof of proposition 1.6.1

Define the following Hamiltonian:

$$\begin{aligned} H(p(\theta,t),\dot{p}(\theta,t),\mu(\theta,t),\theta,t) = f(\theta,t)\bigg(&p(\theta,t)(\theta-p(\theta,t)) - \frac{\lambda}{2}(\theta-p(\theta,t))^2 \\ &-\frac{\gamma}{2}\sigma(theta,t)^2\left(1-u(\theta,t)\right)^2\bigg) + \mu(\theta,t)u(\theta,t) \end{aligned} \tag{1.E.1}$$

where $u(\theta,t)$ is the control variables defined through the following equation of motion: $u(\theta,t) = \dot{p}(\theta,t)$, $u(\theta,t) \in [0,1]$.

Note that the methods used previsouly generalise to multi-dimensional problems, and that here, our problem depends on θ and t instead of only θ as in the case of the static monopoly problem.

Further note that the problem does not depend on the time derivative of $p(\theta,t)$. This means that what would be a general Euler-Lagrange formulation expressed as $\frac{\partial\mathcal{L}}{\partial p} - \frac{\partial}{\partial\theta}\frac{\partial\mathcal{L}}{\partial\dot{p}} - \frac{\partial}{\partial t}\frac{\partial\mathcal{L}}{\partial\partial_t p}$, which is the equation that has to be solved for interior solutions, reduces to $\frac{\partial\mathcal{L}}{\partial p} - \frac{\partial}{\partial\theta}\frac{\partial\mathcal{L}}{\partial\dot{p}}$, where $\mathcal{L}(t,\theta,p,\dot{p}) = H(p,\dot{p},0,\theta,t)$. This is the exact same problem as before, with the only addition that our parameters now vary with t, but the partial differential equation is the same one as before.

Therefore the problem can be solved exactly as before by replacing the variance term by its new dynamic version, that is that it is as if the ramping cost parameter γ was replaced by $\gamma \cdot \left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}\right)$ in the static solution.

This can be seen by noting that $\sigma(\theta, t)^2 = \left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}\right)(\theta - \hat{\theta} + \omega)(\hat{\theta} + \omega - \theta)$ which has to fall back to the static case in the limit, therefore we see that we simply get an additional $\left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}\right)$ term that appears on the ramping cost term, that is that multiplies γ.

Appendix 1.F Proof of proposition 1.6.2

The exact same reasoning as the one in Annex 1.E applies here and we only have to take our static oligopoly result and replace γ by $\gamma \cdot \left(1 + \frac{\tau}{\omega}\frac{d\omega}{dt}\right)$ to obtain the dynamic results.

Chapter 2

Methodological tools for non parametric functional data evaluation and weather data usage

2.1 Introduction

In the last chapter, we have introduced a model of supply function equilibri under uncertainty that takes ramping costs into account and we derived solutions that depend on the information the firms have about the future demand at the time of bidding. Here, we will focus on introducing tools that will allow us to perform empirical analysis of the french day ahead market.

In this short chapter we develop a methodology to analyse data of two specific formats. The focus lies on the methodological details as well as evaluating the performance of our technique. The aim is first to extract points of interests from functional data in order to be able to compare function to one another across bids, and second to describe a methodology that will allow us take into account the uncertainty related to the weather. The economic interpretation is secondary in this chapter. Chapter 3 will use the methodology developed here for a case study of the French electricity market.

2.2 Point selection on functional data : a non parametric approach

Reduced form models often rely on exploiting market outcomes for their analysis, i.e. equilibrium prices and quantities, in order to identify the determinants of firm behaviour and test predictions of the theory. On a few markets, sufficient information is available to get around the problem of using endogenous equilibrium data. For example on the government bond markets, both the full aggregate demand and supply functions are observed. This market is of a specific type, it is a divisible goods auction (also called multi-unit or share auctions). These are auctions where multiple units of a good are sold in a single auction. The exact quantity is not predetermined, but endogenous and depends on the price. Furthermore, the auction format is more complex than for indivisible, single unit auctions and most notably requires that bidders simultaneously submit full bid functions for the goods, i.e. multiple price-quantity combinations at which each bidder is willing to buy or sell the goods. The market price and quantity are determined by the intersection of the aggregate demand and supply functions.

The aggregate bid functions are very rich in information and the reduced form models can be adapted to use this data. However, the literature on exploiting functional data is limited. This idea has been applied to investigate the determinants of demand bid functions in French government bond auctions [Préget and Waelbroeck, 2005]. They rely on the propositions first put forward in [Boukai and Landsberger, 1998] and [Berg et al., 1999], who identified that aggregate bid functions in divisible goods auctions follow an S-shaped curve that can be estimated by a logistic function. The fluctuations across auctions are claimed to be due to random shocks on the parameters of the estimated logistic function. The methodology is applied in [Özcan, 2004] to investigate the revenue superiority of the discriminatory price auction format over the uniform price auction format for the Turkish government bonds market.

More generally, their methodology consists of a two-stage regression. The first stage summarises the (presumably parametric) functional data of the aggregated demand function as parameters of an estimated smooth logistic function. The second stage reuses the information (concentrated in the estimated parameters) for cross-sectional analyses.

Although the auction mechanism is identical to that of the Treasury market and data availability is comparable, logistic function approach does not suit the context of the electricity market due to the strong heterogeneity of the bid functions and their deviations from such logistic shapes, as can be seen in the example of Figure 2.2.1.

The heterogeneity arises from the fact that the bid functions for the electricity auctions are much richer since we have multiple, strategic players on both the demand and

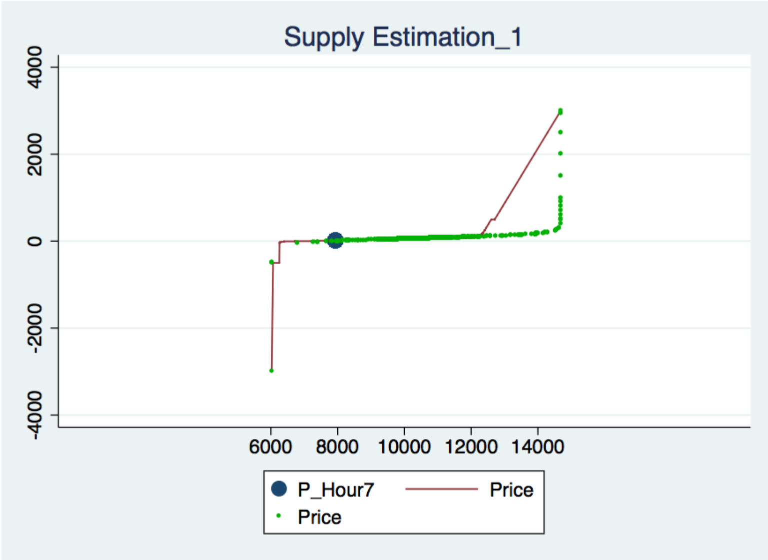


Figure 2.2.1: Example of an assymetric aggregate supply function. The x axis is the quantity in MWh, the y axis if the price in €. In red is the actual aggregate function, in green is an estimated logistic function showcasing the large discrepancies that can arise with this parametric approach. The blue point is the market outcome.

the supply side of the market (unlike the market of government bonds, where the supply is monopolistically determined by the Treasury itself). Furthermore, supplier bidding is strongly influenced by the underlying cost of the production technologies. The observed data is consequently less homogeneous and the fitting of the logistic model not convincing. Furthermore, the economic interpretation of the logistic function parameters is very difficult and reducing the whole bid function to two parameters of interest discards a lot of the original information of the bid functions. Finally, we are uncomfortable with the strong assumption of smooth underlying functions and want to circumvent the problems of fitting these.

Instead, we develop a non-parametric, functional data analysis approach to select comparable data points from the original bid functions. These selected points are comparable across repetitions of the market (i.e. auctions for different hourly contracts) and can then be used to run a cross-sectional reduced form model. The interest of this approach is threefold. First, it aims to use as much of the original information as possible without distorting it into parameters of a logistic function. Also, information about different parts of the bid function does not influence one another. Second, our approach is "scalable" and as many points as necessary can be extracted. The cross-sectional analyses are then conditioned on the type of comparable points selected. Third, while our analysis provides support for an underlying tri-linear or S-shaped functional form, we do not need to assume a specific functional form nor impose overly simplistic assumptions, such as symmetry of the functional forms, to ensure convergence of the estimator.

Here we present the methodology of our point selection and apply it to data from the French electricity market. For now, we ignore specificities of the market for the sake of concentrating on the methodology. We introduce very briefly the data and the market in section 2.2.1. For a full explanation of the data and the market, we refer the reader to chapter 3. In section 2.2.2, we explain the point selection algorithm. In section 2.2.3 we discuss the results of the methodology. Section 2.2.5 concludes.

2.2.1 Information about our data

Our methodology is general and can be applied to any market where the structure of data observations is similar. Here, we present and discuss the performance of the method- ology on data from the electricity market. For the purposes of this chapter we will focus only on the statistical properties of the data, not on the economic interpretation.

We apply our methodology to data from a divisible goods auction. In this auction, each buyer and seller submits a full individual bid function, i.e. a demand or a supply function, which consists of 2 to 256 monotone price-quantity combinations. The final bid function consists of these explicitly submitted bid points and all linearly interpolated points between them.

The market is cleared by computing the intersection of the aggregate demand and aggregate supply functions, which are each obtained by summing up all individual bid functions for the demand and supply side of the market respectively. In a uniform pricing format, the determined equilibrium price is applied to all units sold in that auction.

2.2.2 Point selection algorithm

To briefly fix ideas, let's assume that we are interested in a regression à la:

$$S' = \alpha + \boldsymbol{\beta X} + \epsilon$$

where S' is the steepness of the bid function, $\boldsymbol{X}$ the stacked vector of exogenous variables (not specified further here), α the regression constant, $\boldsymbol{\beta}$ the stacked vector of regression coefficients and ϵ the error term.

The information S' is drawn from the bid functions of a market, and varies along the different points of the bids.

For comparability, we require that a chosen point k from a supply function must be comparable to the k^{th} point from the supply functions of another auction. The same goes for chosen points of the demand functions. The reason for this assumption is that comparing those points accross auction allows us to describe how the functions, that is the aggregate strategies, change shape when our independant variables vary. Note that we do not impose comparability between a pair k of points from a supply and a demand function of the same auction.

Non-parametric technique to compare bid functions

Consider two demand functions (as shown in figure 2.2.2). We have to identify "features" of the different functions in order to determine which points can be compared to one another. We aim to reproduce the type of analysis that the brain performs automatically when faced with such curve: we clearly identify three regions of different slope, where the central region is less steep than the left and right regions.

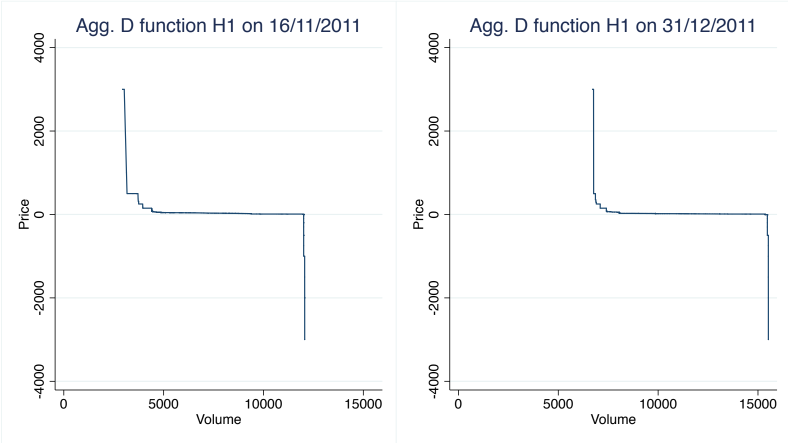


Figure 2.2.2: Comparison of two aggregate demand functions for the same hour

To recognise these features, we perform two successive kernel density analyses[1]. For details on the bandwidth and kernel selection as well as algorithm specificities, see appendix 2.A.1. This allows us to access estimates of the absolute values of the first and second derivatives of the demand functions as shown in graphs B and C of figure 2.2.3.

We are therefore able to identify the regions of very high curvature, which define the transition between the three characteristic regions of these functions. We assume that these maxima can be compared across different auctions. This hypothesis is commonly made in functional data analysis and known under the method of landmark registration [Silverman and Ramsay, 2005]. This has been applied in [Wölfing, 2013], chapter 4, to day-ahead electricity data, in order to identify the effect on fuel price shocks on supply curves. However, this landmark registration was applied in a parametric form : the regions of high steepness were identified as any part of the curve above 90€/MWh.

We can develop this method further and define intermediary points[2] that can again be compared to one another. This method allows to define as many points as needed, for

[1]Bandwidth in the first estimation = 45, bandwidth in the second estimation = 2, kernels: epanechnikov.

[2]As an example, we could extract those points corresponding to half the density value of the maximum density of the second order derivative. The four points selected (one for each monotone portion of the graph of second derivative estimates) would then correspond to those where the curvature of the function is halved. Together with the maximum, the additional point would contain information on the speed (radius of the curvature) at which the function changes.

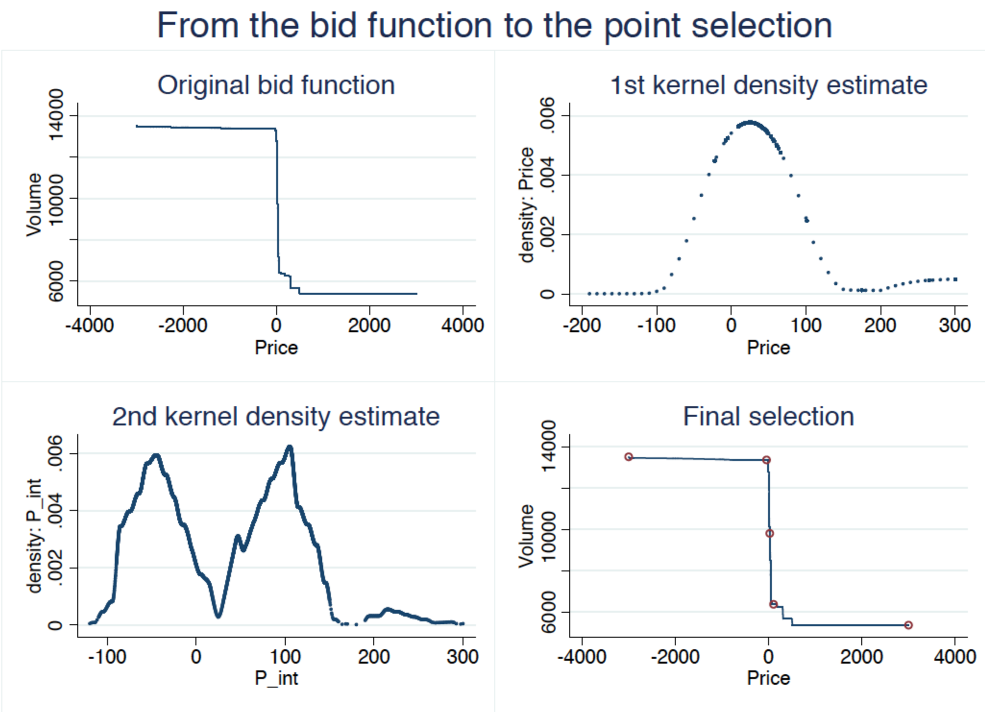


Figure 2.2.3: Steps of the point selection process

Top left (A): The full original aggregate demand bid function for hour 8 on 15.01.2011 in the quantity - price dimension. Top right (B): Kernel density estimates of the first derivative, zoomed on the relevant price range. Bottom left (C): Zoomed kernel density estimates of the second derivative. Bottom right (D): The full original bid function with the $K = 5$ selected points.

computational reasons we limit ourselves to $K = 5$ points[3].

Graph D of figure 2.2.3 visualises an original demand bid function and the selected points that we retain as an informative summary of the original curve. Once this work is done we are left with $K = 5$ points per observed aggregate function, those points defined in such a way that they can be compared from one auction to another.

In our setting, the selected points are the two end-points of the curves (where bidding is imposed by the auction rules at the minimum ($k = 1$) and maximum ($k = 5$) Price), the point corresponding to what can be thought of as the point of inflection (determined by the maximum of the first derivative, ($k = 3$) in the plane (p, q)) and the points sepa-

[3]The point selection algorithm took 2 weeks runtime to complete its task of selecting 5 points per function. Defining intermediary points would have taken disproportionately more time since many sorting and interpolation steps are necessary for each intermediate point.

rating the regions of high and low elasticity in price (determined by the maximum second derivatives to the left ($k = 2$) and right ($k = 4$) of the POI).

We described the technique here for the case of a demand function. The information measured at these points can thereby be compared across demand bid functions of different auctions. The method is used analogously for selecting comparable points on the supply function. We are hence able to extract slopes at these selected supply bid points, which are again comparable across auctions.

2.2.3 Results of the point selection methodology

Precision of point selection

We have selected $K = 5$ types of comparable points for each of the 37500 demand and supply functions present in our dataset. This section details the results of the point selection methodology and presents evidence why the point selection algorithm has produced comparable points reliably.

The graphs in figure 2.2.4 show the local density of selected points in the price - quantity space for the demand (left) and supply (right) curves. The fact that the groups of data points are disjoint from one another indicates that the points selected are distinctly different across groups.

In figure 2.2.4, selected points of type $k = 1$ manifest at the bottom of the graph with prices fixed at -3000€/MWh. Similarly, $k = 5$ points appear at the top of the graph with prices fixed at $+3000$€/MWh. The three distinct groups of data points refer to points of type $k = 4$, $k = 3$ and $k = 2$, respectively, when reading the zoomed, center part of the graph from top to bottom.

We note that the point selection for the demand curves has produced groups of points

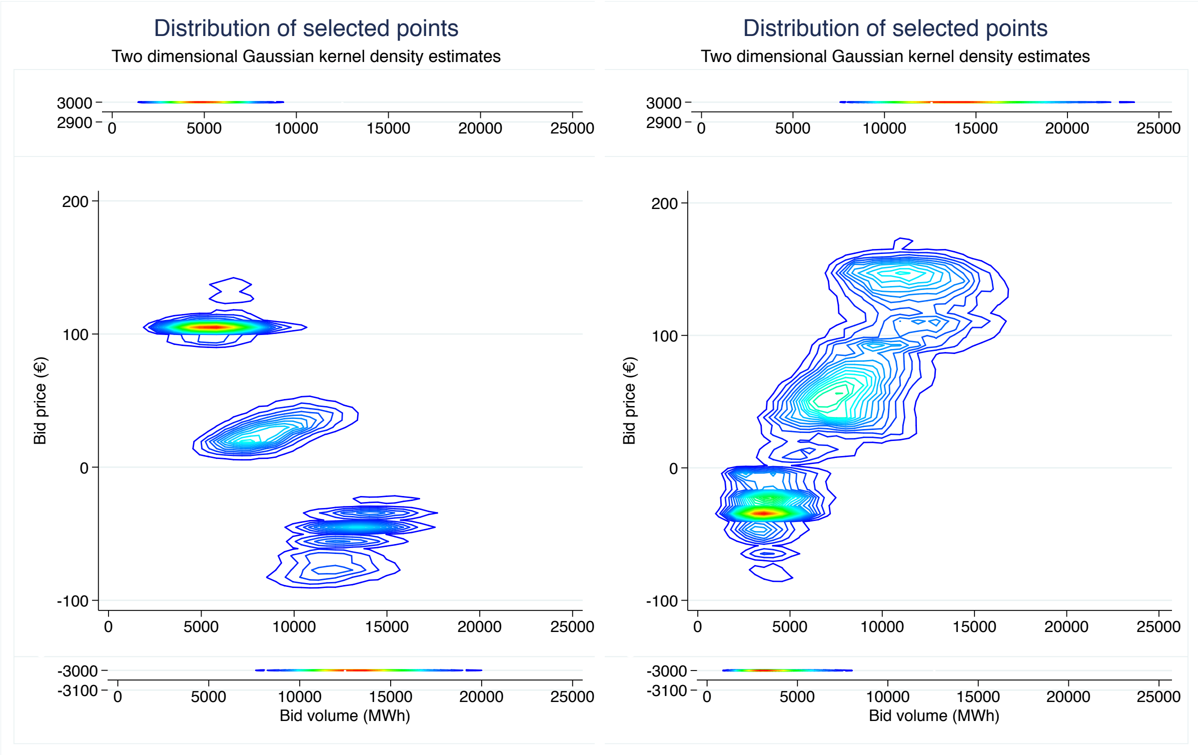


Figure 2.2.4: Heat map on selected, comparable demand and supply points

Note: Please note the discontinuity in the scale of the y-axis. The three seperate graphs are arranged to be understood as a single one. The warmer the colours of the heat map, the higher the frequency of selected price-quantity pairs. The colour legend is omitted for brevity, density changes between contours are of the order of 10^{-4}.

that are more distinct (and thus more robustly attributed to a certain type k) then for the supply function.

Our methodology only relies on assuming that the first derivative is uni-modal and that sufficient variation exists in the data to distinctly identify the regions of different slope. Overall, this is strong evidence that the algorithm is able to distinctly differentiate between points of different types.

2.2.4 Observations of bidding frictions

Distinct point selection is further supported by the evidence in figure 2.2.5. These graphs show the distribution in the price-quantity space of the selected points separately for the demand and supply function. Distinct clouds are an indication that selected points are different across types k.

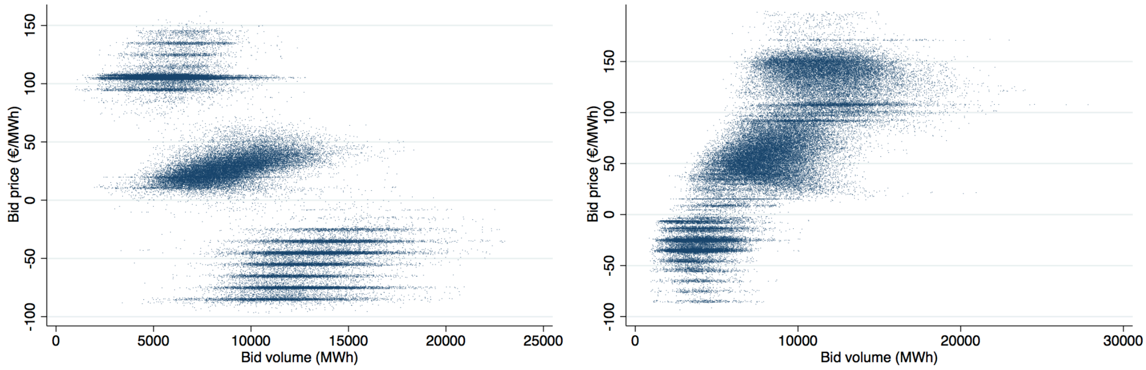


Figure 2.2.5: Distribution of selected demand (left) and supply (right) points

However, a feature of the graphs is striking: patterns (horizontal lines) seem to exist for the selected points of type[4] $k = 2$ and $k = 4$. Many selected points accumulate at certain prices of regular intervals of 10€/MWh, i.e. there seem to be focal price points for the bidders at the curvature points of the bid functions. The pattern is present for selected points of both the supply and demand functions, although the selected points from the supply function exhibit this pattern slightly less.

The points following the pattern (types $k = 2, 4$) represent the points of maximum curvature of the aggregate bid functions, i.e. the region where the aggregate bid function transitions from a price elastic center portion to the price inelastic extremities of the bid function.

Without prioritising any explanation[5], we acknowledge the existence of bid point patterns in the values (i.e. prices and quantities) of selected points.

[4]Types $k = 1$ and $k = 5$ do not exhibit variation in price, because bidding at the extreme prices of +-3000€/MWh is imposed by the auction rules. We thus neglect their analysis here.

[5] We do not investigate the origins of bidding frictions in this section, we focus purely on the methodology. For the electricity market, a few possible explanations are that (1) bid functions are driven by marginal costs consideration towards the extremes of the bid curve, (2) bidders bid coarsely since the have used up much of their bid point allowance (256 points) on the center portion of the curve, (3) bidders spend less effort on adequately bidding at extremes since the likelihood of the market outcome occurring at the extremes is much lower.

We are, however, interested in S', the slope at each selected point - an information measured at the selected point. We therefore investigate whether the values of the first derivative at the selected points display a pattern. Figure 2.2.6 shows the histograms of slopes of supply functions for the points $k = 2, 3$ and 4. No pattern in the values of the derivatives is apparent.

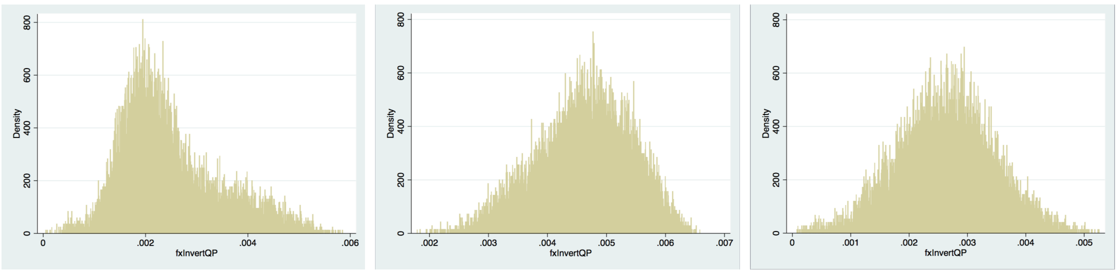


Figure 2.2.6: Histogram of slopes per point type

Note: Histograms of extracted slopes at points of type $k = 2$ (left), $k = 3$ (middle) and $k = 4$ (right).

Although values of the selected points are possibly biased due to focal price points, we do not observe patterns in the variable of interest (i.e. the first derivatives of the selected points) and deem the methodology adequate for our purposes.

Finally, we emphasize that the observed patterns are not caused by the point selection mechanism since the algorithm can only choose between explicitly bid points or linearly interpolated points, that could be part of a market equilibrium under the reigning price setting algorithm. The pattern arises from many horizontal steps occurring at the same prices in different auctions.

Value of selected points (determining K)

We remind the reader that the aim is to recover points that summarize well the behaviour of the full aggregate bid functions in different auctions. Our technique allows us to extract representative and comparable points across bid functions of different auctions. Form the selected points, we can also go back to infer the original bid function from which the points were selected. In order to evaluate the utility of our methodology, we investigate the added benefit of an additional point in our point selection.

By selecting $K = 5$ points per curve, rather than fewer points per curve, we are able to significantly reduce the degrees of freedom for inferring the original bid function. In other words, our information (as captured by the selected points) about the original bid function is more precise.

In order to investigate the marginal gain of information for additional points, we first define the mean registered curve. Consider a set of curves that each have N registered points. Take the average coordinates of every point across curves. Rescale linearly every curve by parts so that the registered points fall on their average[6]. Define the mean registered curve as the averaged rescaled curves. Now, separate the data into two groups : curves that are above or below this average curve. Take the averages of these two groups : this defines a measure of the variability of the curves around the total average which is able to capture asymmetries between the two groups.

Now that these quantities are defined, we can display how much information is captured by the successive addition of registration points for $K = 0$ to $K = 5$ points. We look at the decrease in uncertainty achieved by including an additional point, obtained using our technique. Figure 2.2.7 shows the mean registered curves (red lines) and the mean inferior or superior curves (pink shaded interval above and below the mean registered curves) as a function of the number of reference points.

We can see that as we include an increasing number of points the shaded areas shrink : this is a measure of how much of the information contained in the raw curves is captured by the registration points. We see that at 5 registration points, the shaded area is very

[6]We rescale all points between the reference points by a vector obtained as a linear combination of the displacement vectors of the closest reference points, of which weights are obtained as the inverse of the distance of the considered point to the enclosing reference points.

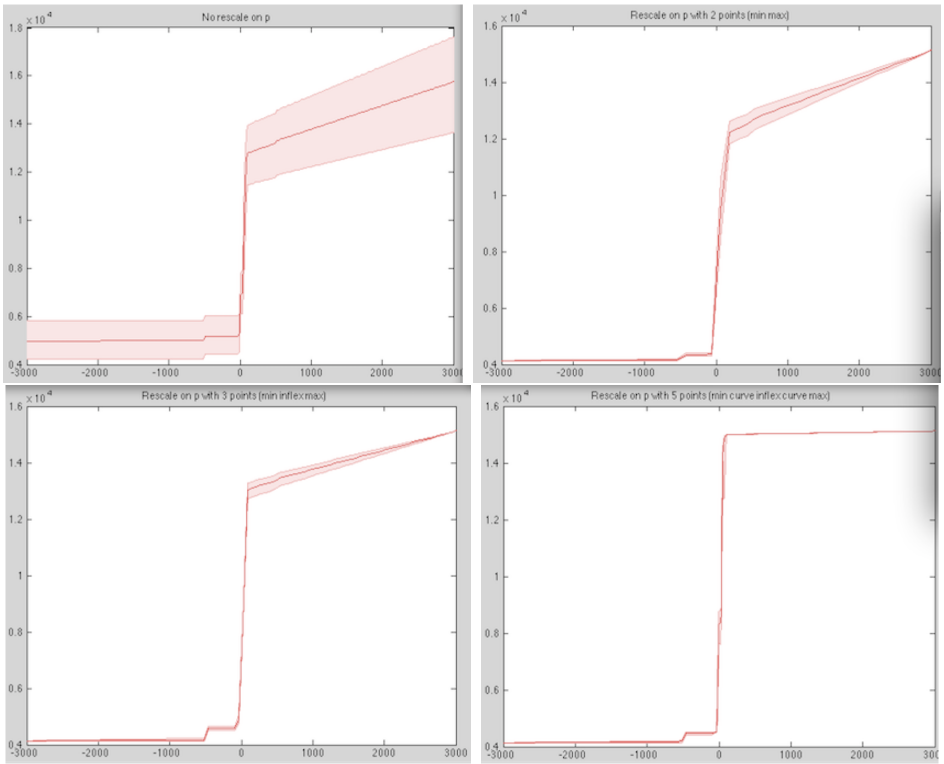


Figure 2.2.7: Error bars as a function of the number of extracted points

Note: The graphs represent the master curve with the error interval for inferring the original bid function, conditional on the number of extracted, reference points (RP). Top left (A): Computed without any RP. Top right (B): Computed using 2 RP. Bottom left (C): Computed using 3 RP. Bottom right (D): Computed using 5 RP.

small, so much so that one can consider that by registering these 5 points, we capture a so-called "master curve" : most of the information about the curves is contained in those 5 points.

More quantitatively : without any reference point, inferred bid functions would lie in the interval shown in graph A of figure 2.2.7. With two reference points (namely the minimum an the maximum quantity), the uncertainty is reduced as shown by the smaller error interval in graph B. Graph C adds a third point (the inflection point) and Graph D adds another two points (the two points of maximum curvatures). Figure 2.2.7 shows clearly that with an increasing number of reference points, we obtain a more precise information about the original bid function. We quantify the informational gain by measuring the pink shaded area in each graph A to D. The result is shown in figure 2.2.8 and reveals decreasing marginal information for each additional point. By selecting $K = 5$ points, we are able to reduce the shaded area by a factor of about 50 when compared to figure A (see figure 2.2.8). We see this insight as support for using $K = 5$ points for further work.

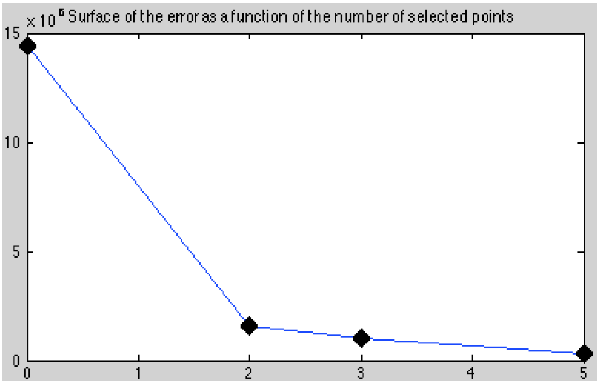


Figure 2.2.8: Proxy for degrees of freedom on master curve

Note: The graph plots a proxy for the number of degrees of freedom for the inference of the original bid function on the number of reference points. Specifically, it plots the size of the pink shaded area in figure 2.2.7 against the number of points.

While the graphs in figure 2.2.7 are displayed on inverted axes and rescaled units, we show the final master curve and uncertainty interval on the original axes and units in figure 2.2.9.

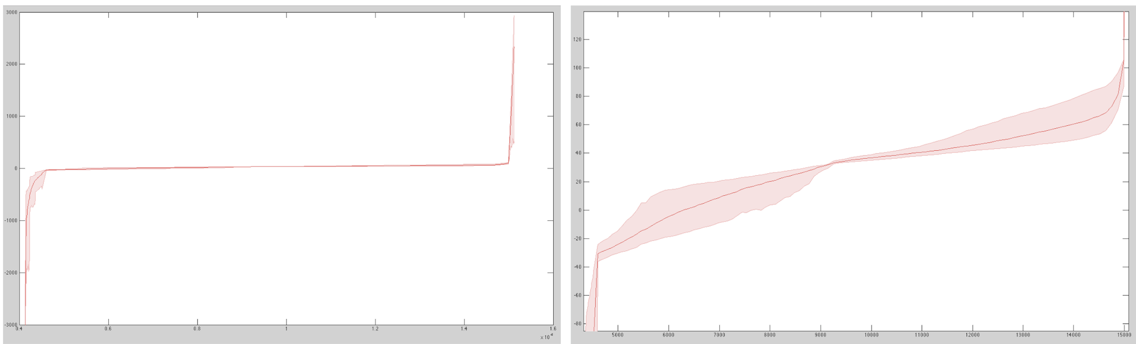


Figure 2.2.9: Overall (left) and zoomed (right) Mastercurve with confidence interval

Note: Master curve in the quantity - price dimension.

2.2.5 Discussion

In this section, we have developed an alternative technique to run a cross-section reduced form model on data generated by a market that keeps track of the full aggregate demand

and/or supply functions. While we apply it to aggregate demand functions, the methodology is fit for the anaylsis of aggregate supply functions and individual bid functions of either market side.

The methodology is inspired by the techniques used in the literature on Treasury auctions, but has been set up from scratch to allow treatment of more heterogenous data. Furthermore, the hard assumption of an underlying logistic function is relaxed and our non-parametric point selection avoids the storing of bid function information in the form of estimated function parameters, which are difficult to interpret.

Smoothing of the original bid functions is a component in both the traditional logistic function approach and our comparable point selection methodology. The smoothing enables the user to abstract of small bid function particularities and imprecision, e.g. steps in the function. However, in the traditional approach, the reduction of plus 1000 bid points into very few parameters resulted in the blurring of "local" bid function information from all parts of the function at once. Our non-parametric approach allows specifically to control the extent to which one smoothes the underlying data through the amount of registration points considered.

The results of the comparable point selection are encouraging. We show that each type of point is distinctly chosen and that patterns of the original bid functions do not influence the quality of derivative information extracted at the selected points. We acknowledge the existence of bidding frictions in the original data and highlight this observation for further work.

2.3 Methodology to aggregate geographically dispersed information on a national level

The theoretical results of chapter 1 indicate that a key ingredient in explaining the dynamics of the bids submitted by suppliers on the electricity market is the uncertainty about demand shocks.

Energy demand addressed to the electricity markets depends on temperature (through the heating of buildings), on wind speed (through the production of wind turbines which reduces the net demand) and on luminosity (through the production of solar panels which reduces the net demand). However, these three weather variables vary in space, whereas the market is at the national level. We introduce here the methodology with which we estimate the associated uncertainty.

We have two types of meteorological data: observations and forecasts. We use both types of data to estimate the underlying uncertainty, the methodologies for each differ slightly.

2.3.1 Dealing with meteorological data

Interpolation methodology on weather observations

Observations are obtained from MétéoFrance for three parameters of particular interest: temperature, wind speed and light intensity. These observations take the form of tables of hourly observations for a given set of weather stations. Each parameter is observed on a different set of stations.

Due to their hourly nature, the analysis of the electricity market's sensitivity to weather requires a very high number of observations. Therefore we select between one and two stations per Département[7], a French administrative unit of roughly 6000 km^2, i.e. of a typical lengthscale of about 75 km. We have 161 stations for temperature, 113 stations for wind speed and 106 for light intensity, as shown in Fig 2.3.1.

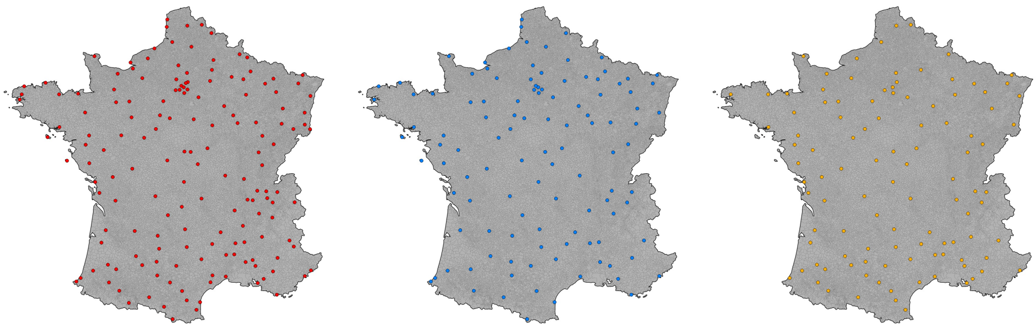


Figure 2.3.1: Stations for which we have hourly data. Left : temperature, center : wind speed, right : light intensity.

For each hour, we select the corresponding observations and interpolate them in order to reconstruct the weather on the entire french territory. An interpolation consists on inferring the value of a variable at query points using a reference data set of known values. One very important underlying assumption of interpolation methods is that of the continuity of the process underlying the data generation. The easiest interpolation method is the linear interpolation: think about a dataset of hourly observations with one missing value; to reconstruct the missing value, take the average of the value of the preceding and following hour. There are numerous methods of interpolation, even more so when the data is spatial in nature, all revolving around two main steps. First, given a query point at which one would like to infer the value of the variable, there needs to be a selection rule to know which of the points from the reference data set should be used (in our example the preceding and following values). Second, once these points are selected, one needs a weighting function to know their relative importance in order to obtain the interpolated value (in our example it is a simple averaging, that is weights of 0.5).

We use the natural neighbor interpolation method, well known for its good balance

[7]There are 95 Département in France

between speed and accuracy. In short, in this case, the first step makes use of the Voronoi tessellation algorithm[8], one is able to define the natural neighbors of a point for which one seeks an interpolated value. These natural neighbors are used in the second step, which performs the actual interpolation as a weighted average of the values of these natural neighbors using a ratio of surfaces as weights (see Fig 2.3.2 for more details).

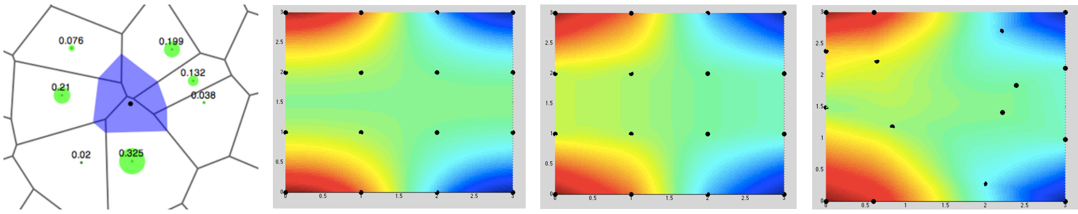


Figure 2.3.2: Left: Voronoi's algorithm is applied once on the reference points highlighted in green to obtain the white surfaces, and a second time on the same points to which is added the query point in the center to obtain the new blue cell. The green circles, which represent the interpolating weights, are generated using the ratio of the shaded area to that of the cell area of the surrounding points. Center left: example of a reference surface (color mapped) to be reconstructed through a natural neighbor interpolation. Center right: interpolated surface with a reference set of 16 evenly organized points, represented in black. Right: interpolated surface with a reference set of 16 unevenly organized points, represented in black. From 16 points one is able to reconstruct the color mapped surfaces which are approximations of the reference one, represented in the center left image.

Image transformation to recover weather forecasts

Forecasts are obtained from the Global Forecast System (GFS), and come in the form of colormaps, as shown in Fig 2.3.3. We are going to illustrate our methodology on temperature data, but the same exact approach is performed on wind speed data. The general idea is that the pointwise precision is low ($2°C$ per color) but that the overall map contains quite a lot of information through the topology of the colored regions. We describe below how to extract this information.

First: image cleaning To extract the relevant data we first clean the color map from its irrelevant information, namely the temperature in numbers and the administra-

[8]The Voronoi tessellation algorithm takes a collection of points $\{p_k\}$ in the plane, and then partitions the plane as regions "belonging" to each point, called cells. A Voronoi cell associated with a given point p_k is defined as the collection of every point in the plane whose distance to p_k is less than or equal to its distance to any other p_{-k}. Each such cell is a convex polygon.

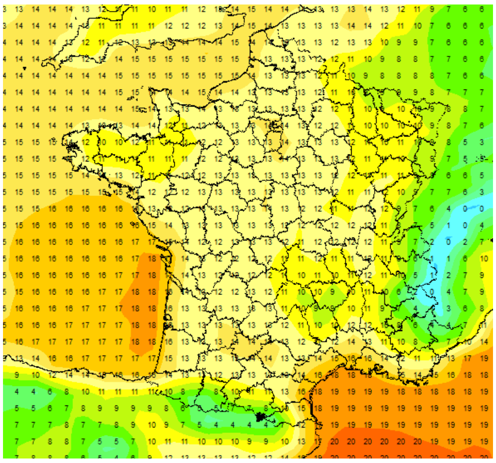


Figure 2.3.3: Temperature forecast from a simulation run by the GFS at 6 a.m. on the 3rd of november 2011, for a forecast at 22 p.m.

tive borders. Note that this step introduces a small amount of high spatial frequency noise, see Fig 2.3.4 left and center left.

Second: removal of redundant information A lot of information is lost from the actual GFS simulations by using a color map representation, as temperature is described as a discontinuous variable: each color has a precision of $2°C$. In order to correct for this, we leverage the fact that all the information contained in this color map, that is the color at each pixel, is actually contained in a smaller set of points. Consider the value at the boundaries between different color regions: by knowing that the interior of a constant color region has a constant value, one is able to represent all the information contained in the original image by keeping only track of the values at the boundaries. To recognise those boundaries we perform image analysis, more precisely we use edge recognition methods based on finding high gradient regions, thus obtaining Fig 2.3.4 center right.

Third: surface fitting Once we represent the information in this denser form we can perform the last step, which consists in fitting a surface to our newly defined dataset, i.e. the temperature values at the boundaries, which take the form of (x, y, T) triplets. We could perform an interpolation, but these methods are not well suited to reference sets having so much structure. Here, data points lie on curves representing iso-temperatures, so that along such a curve there is a lot of data points, whereas the information is very sparse along the direction of the gradient. In addition the first step introduced some spatial noise which we want to correct to some extent : we allow our fitted surface to take different values than our data points, so as to smoothen out this noise. We define the rigidity of our fitted surface, i.e. a penalty associated to fast changes in the value of the surface, and therefore reduce the importance of the high frequency noise introduced in the first step. The end result is presented in Fig 2.3.4 right.

It is key to understand that this image is displayed using a colormap close to the one in the original picture to facilitate comparison but that its underlying data is continuous whereas the original image describes temperature by bins of $2°C$. It can therefore be used to query the value at any given point, and these values will change continuously in space instead of discrete jumps in the raw format.

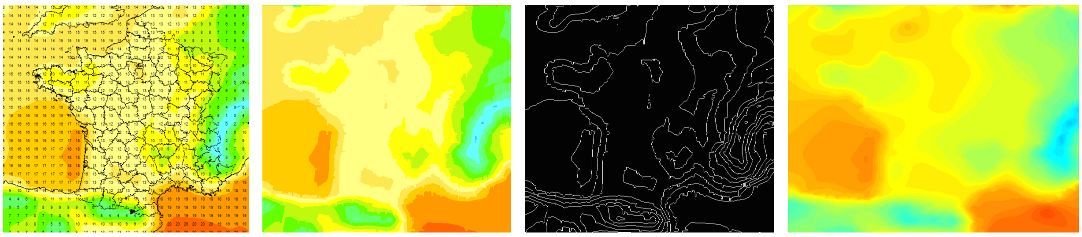


Figure 2.3.4: Left: reference image. Center left: borders and numbers are removed. Center right: edge recognition. Right: final fitted surface.

Autocorrelation lengthscale

We use this dataset to build measures of the weather uncertainty. To do so we measure the auto-correlation lengthscale of our three weather variables of interest : temperature, wind speed and light intensity. This lengthscale measures how much are the weather variables correlated spatially. Our hypothesis is that the auto-correlation lengthscale is inversely proportional to uncertainty about the variable we are interested in. When it is small, the variable is less spatially correlated, which we interpret as being more difficult to forecast. Conversely, when the auto-correlation lengthscale is large, the variable is very correlated spatially, that is that the informational content of one datapoint is higher for the prospect of using it for the evaluation of a national effect.

Take two points on a plane and a spatially correlated bounded variable. If those points are infinitely distant, the value of the variable at these points should be uncorrelated. That is that the absolute difference between the variable taken at those two points should have a given average value. Conversely, two points infinitely close should have the same value, i.e. a zero absolute difference between the variable taken at those two points. The question is how fast is the transition between those two limit cases. First, we define the average absolute difference between two points when distant of a given value. Second we extract a typical lengthscale.

To define the average absolute difference between two points when distant of a given value, we consider at a given point in time every possible pair of points in our dataset. For a given pair we compute its distance and its absolute difference in value (in black in Fig.2.3.5). For 100 datapoints we obtain 4950 pairs. We then use a kernel smoother in order to obtain the average non parametric autocorrelation function (in blue in Fig.2.3.5).

To recover a typical lengthscale we make the parametric assumption that the autocorrelation is exponential in nature. We fit an exponential function through our smoothed data (in red in Fig.2.3.5), and recover the exponential decay parameter as our lengthscale (in green in Fig.2.3.5). We perform this operation for every hour in our dataset and every weather variable. The results are timeseries for the characteristic lengthscale of the weather parameters.

2.3.2 Aggregation of local information

Wind1DA *Wind speed (average speed in km/h):* Wind speeds influence the productivity of wind turbines, which are a source of unreliable electricity generation. In general, renewable technologies benefit from a feed-in guarantee by the state. That is,

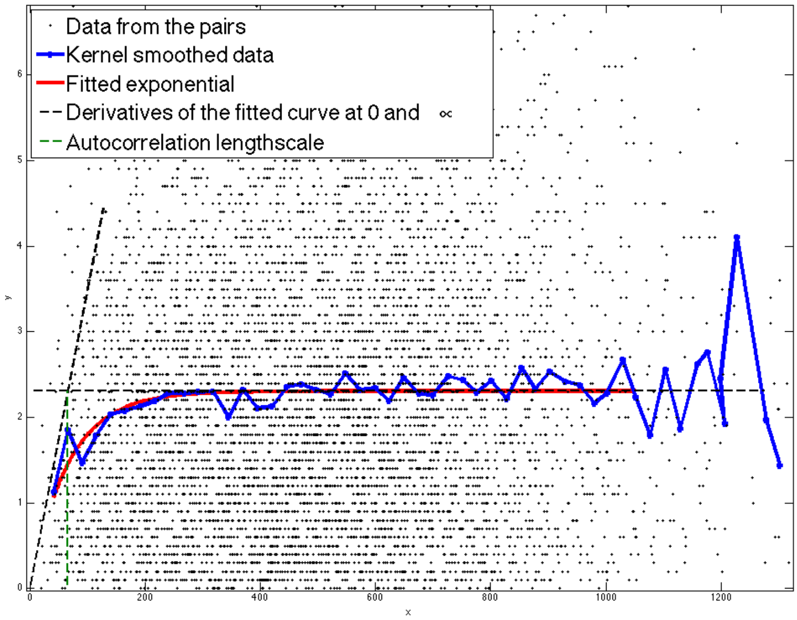


Figure 2.3.5: Auto-correlation lengthscale computation. In black are the points obtained from all the pairs from our original data, that is absolute wind speed differences as a function of the distance between the two points. In blue is the kernel smoothed function from those points. In red is the exponential fit. In black are the derivatives of the fit at 0 and ∞. In green is the recovered auto-correlation lengthscale. The unit for the lengthscale is in km.

regardless of the trading outcome on all markets, renewable energies will be the first to be fed into the power grid at a guaranteed price.

Consequently, the electricity production of renewable technologies represents a production shock for all actors on the market. The production shock means that the demand to be served by traditional electricity producing firms is reduced by the amount that is serviced by the electricity gained from renewable sources.

In the case of wind turbines, the average speed of the wind per hour allows to proxy for the size of the production shock due to the electricity generation from wind energy.

We use hourly windspeed forecast in the form of color maps from the Global Forecast System (GFS), giving the speed by bin of 5 km/h at 10m above ground, and the location and production capacity of the wind turbines present on the french territory, given by the SOeS (service d'observations et d'études statistiques - observations and study department) a department of the french environment ministry.

We consider that all turbines in France are of the same type, that is that they have the same response curve and height.

A typical response curve is represented in Fig. 2.3.6. It has three main characteritics : the wind speed at which the turbine starts to produce electricity, called the cut-in speed, the speed at which the turbine reaches its rated output, called the rated ouput speed, and the speed at which the turbine has to stop to avoid damage, called the cut-out speed. We use data publicly available[9] to obtain a rough estimate of the french average wind turbine characteristics. We use a cut-in speed of 2.5 m/s, a rated output speed of 14 m/s, and reduce arbitrarily the cut-out speed from an estimate of 24 m/s to 20 m/s to account for the fact that a turbine is shut down not when the average speed is too high but when the maximal speed becomes dangerous for the turbine.

Wind speed also increases with height, and turbines are typically between 60 and 80m high. We therefore apply a multiplier to the reconstructed wind speed at 10m.

We seek to reconstruct the french wind energy production from meteorological data. The two adjusted values, the cut-out speed and the speed multiplier, are adjusted by hand to obtain reasonnable fits. The reason for this is that the reconstruction of wind speed and aggregate production is computationnally intensive, therefore we cannot perform a full blown estimation. We choose these values with a precision of roughly 10% with repect to their admissible range of values.

We obtain a reconstruction of wind production from day-ahead wind speed forecasts that we compare to actual observed production and to day-ahead wind production fore-

[9]http://www.thewindpower.net

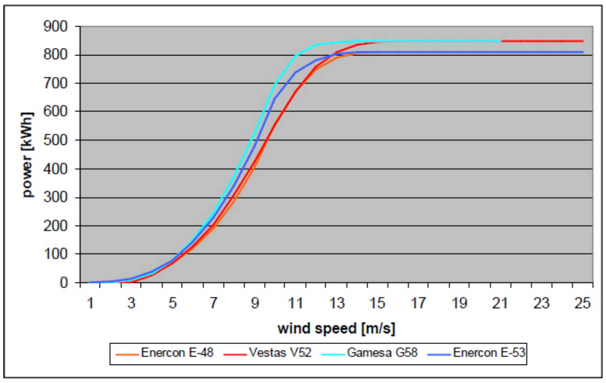


Figure 2.3.6: Typical response curves of different wind turbines

cast computed by RTE, the french grid operator as shown in Fig.2.3.7. We stress here that our aim is two-fold: to link wind turbines' production to weather data and to use forecast data as the market actors only possess this information when bidding. We do not aim at producing better forecasts than the grid operator, the figure is only displayed to show that our methodology produces reasonable estimates (we obtain a correlation coefficient between our forecast and the observation of 0.85 where the grid operator obtains 0.97).

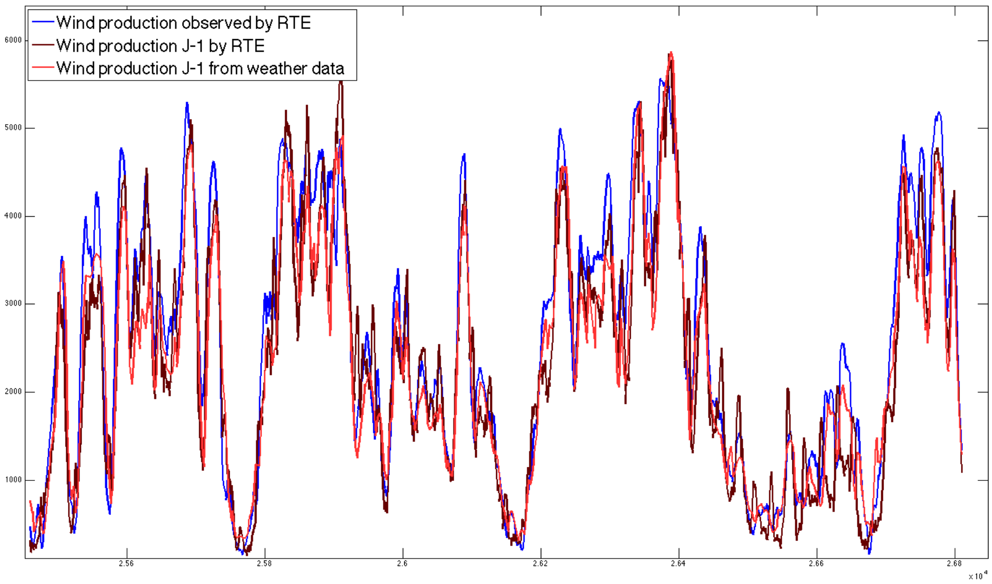


Figure 2.3.7: All curves are hourly production data. The origin of the hours is the first of January 2011, and the production is in MWh. In blue: the observed wind production. In dark red: the day-ahead predictions from the grid operator. In light red: the day-ahead predictions from weather data.

Tempeff15 We focus on the effect of temperature on the demand of electricity first. In France, a high percentage of the population heats their housing with electricity, therefore cold waves have a high impact on electricity consumption: 2300MW of additional power consumption for every drop of $1^{o}C$ below $15^{o}C$, as shown in Fig.2.3.8 sourced from [RTE, 2014], the French grid operator.

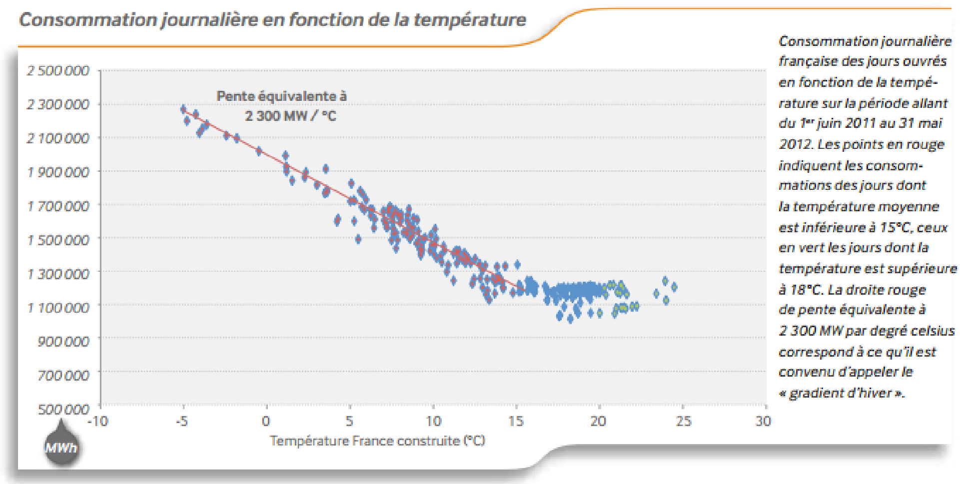


Figure 2.3.8: Daily electricity consumption in France as a function of the temperature

We apply this information to our observed meteorological data in order to build an effective temperature for France aimed at capturing its effect on consumption. To do so, we reconstruct temperature data for every french *commune*, the smallest administrative unit in France (there are around 36000 of those). We consider population as being a good proxy for potential heat consumption, therefore we apply it as a weight to the *commune* temperature. Lastly, we consider that temperatures saturate at 15°C. This allows us to build an effective temperature taking into account where the population is located and the nonlinearity of heat start up which in turn allows us to account at the country-level for the local impact of temperature on the electricity consumption.

Tempeff We also build an effective temperature that does not account for the nonlinearity at 15°C following the same methodolgy otherwise as a control.

Roll_Temp$H$ Variable capturing seasonal trends by using the rolling average temperature on effective temperature (Tempeff15) over the last H hours, i.e. the last $H/24$ days.

Solar Light intensity (in $W.m^{-2}$) impacts the electricity market through multiple channels. The most obvious one is the associated electric production from photovoltaic panels. But there is another channel through which lighting can be seen as impacting electricity consumption : more sunlight decreases artificial light usage. In France, annually, the electric consumption that can be attributed to lighting represents roughly 50 TWh where solar production is roughly 4 TWh[10].

We have photovoltaic production data, which in itself is a blackbox. As we aim to link meteorological data to consumption we first want to validate the quality of our meteorological data. To do so we reconstruct the photovoltaic production from weather data. We know what are the hourly luminosity conditions on the french territory but also where is installed the photovoltaic production capacity. The SOeS (statistical observation and study department), a branch of government, publishes each year a file containing the installed capacity of renewable energy sources per communes, a french administrative unit with a typical size of roughly 3 km.

We use observed luminosity data from MétéoFrance, as there is no hourly forecast of luminosity, and assume a sigmoid response from photovoltaic panels to light intensity with a saturation towards high light intensity, that is approximately a linear response up to a certain threshold. The results are shown in Fig.2.3.9.

We observe that solar production is much more regular than wind production, therefore it is not possible to build a proxy for lighting consumption that would allow us to decorrelate the effects from production and lighting. We therefore stick to this proxy to

[10]These estimates are computed by the authors based on numbers coming from [Bertoldi and Atanasiu, 2007], INSEE and EDF

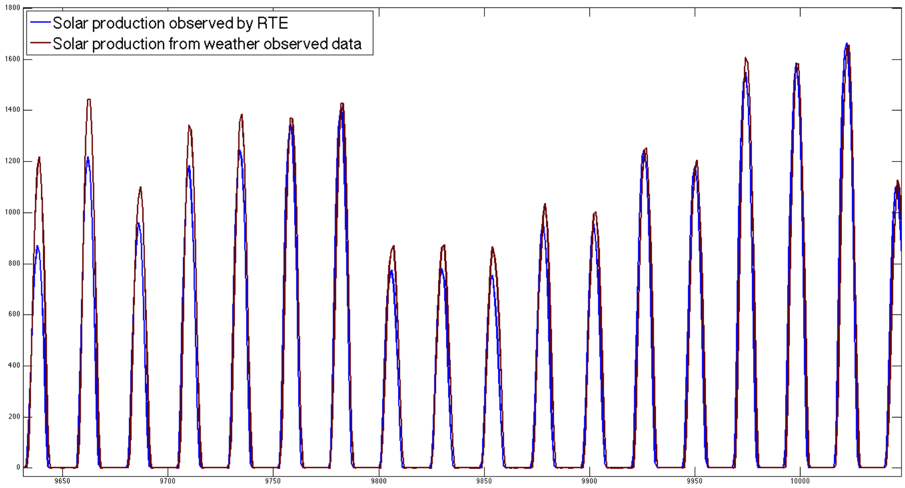


Figure 2.3.9: Hourly solar production in MWh. The time origin is the first January 2011. In blue: observed production by RTE. In dark red: reconstructed production from observed weather data.

capture the net effect of both channels.

SolarRest Solar represents estimates of solar production. Therefore, it is highly collinear to the daily suncycle variable since solar production is light dependent. SolarRest is the residual from a regression of Solar on suncycle and captures the unexplained part of solar production on top of pure light intensity considerations. Table 2.1 gives the results of the regression.

	(1)	
	Solar	SE
suncycle	1,500***	3.903
Constant	0.876**	0.383
Observations	150,959	
R^2	0.702	

*** p<0.01, ** p<0.05, * p<0.1

Table 2.1: Regression of Solar on suncycle

RteBlackBox RTE, the French grid operator gives day ahead predictions of the total hourly consumption, which are available at the time of bidding. This variable is called PrevConsoH.

We do not have access to the exact definition of the index and it is thus a black box. However, it is available to the firms at the time of bidding and we want to include it in the demand estimations.

At the same time, it is evident that the Index uses much of the information that we explicitly control for in the regressions, therefore collinearity is an issue. In order to have correct coefficient estimates, we adopt an instrumental variable approach by regressing the RTE prediction on our exogenous factors, extracting the residuals and only including the unexplainable component of the RTE prediction in the demand estimation in the form of a separate variable called RteBlackBox.

Formally, RteBlackBox is equal to the predicted residuals (u) of the following regression, where X stands for the vector of explanatory variables: Tempeff15, Roll_Temp24 , Roll_Temp240, suncycle, morning, deltasun and EWH.

$$\text{PrevConsoH} = a + bX + u \tag{2.3.1}$$

In table 2.2 we give the output of regression 2.3.1 in column 1, which is strong support that our prepared data for exogenous variables is of very high quality. We highlight the significance of all explanatory variables at the 1% level and the R^2 statistic of 85.3%.

We highlight that the comparison of columns 1 and 2 gives very strong support to our adjusted measure of effective temperature (Tempeff15 instead of Tempeff), which takes into account the demand behaviour as a function of the temperature. Temperatures above 15^{o}C are considered not to impact demand behaviour [RTE, 2014].

	(1) PrevConsoH	(2) PrevConsoH
Tempeff15	-682.6***	
Roll_Temp24	-802.0***	
Roll_Temp240	-1,175***	
SolarRest	-0.860***	-0.345***
suncycle	7,849***	7,418***
morning	-4,759***	-4,398***
deltasun	10,108***	9,010***
EWH	-1,245***	-1,254***
Tempeff		-301.4***
Roll_avgT24		-687.3***
Roll_avgT240		-918.2***
Constant	77,701***	76,651***
Observations	146,909	146,909
R^2	0.853	0.816

*** p<0.01, ** p<0.05, * p<0.1

Table 2.2: "Black box" regression on RTE predicted consumption
Note: The dependent variable PrevConsoH is the day ahead prediction by RTE of the total consumption in France.

2.4 Conclusion

In this methodological chapter we present the different methods that we developed to study in the next chapter the impact that uncertainty about demand shocks can have on suppliers' bids.

We want to be able to describe how bids change shape as a function of a number of regressors. To do so we apply functional data analysis to the bids, and argue that the landmark registration technique allows us to compare important features across bids.

Finally, as we are interested in the impact of uncertainty about demand shocks, we note that weather is an important source of uncertainty, and introduce a number of metrics, based on the intrinsic structure of weather data or on its relationship to the processes at play when considering consumption or production of electricity.

In the next chapter we will therefore be able to focus on the econometric analysis of our data.

Appendix

Appendix 2.A Technical details

2.A.1 Using the kernel density estimation (KDE) in our setting

In order to estimate the first and second derivatives of the bid functions, we use a kernel density estimation. The estimator is essentially a smooth version of a histogram and counts the number of points in moving intervals (called a window) of predefined width along a dimension of the data. In our case, it counts bid points per price interval. In addition, the KDE assigns a weight to each observation based on the distance from the observation to the center of the window. The weighing function is called the kernel.

The observed bid functions are each a multitude of price-quantity combinations. However, a kernel density estimation on the observed points of the bid function would be useless since the number of points per price interval does not vary much with the slope of the curve.

We are able to transpose the observed bid function to one that suits our needs. This is done by adding linearly interpolated points at the unit cent level (corresponding to the minimum bidding unit). The kernel density estimation is then able to estimate the slope of the function by simply counting the points in an interval since the number of points per price interval of constant width varies proportionally with the slope of the function over that interval.

Hard choices in the code of the KDE

A few specific choices have been made in the code and are detailed here.

Kernel choice: First, we use the default Epanechnikov kernel for simplicity. It is generally considered that the kernel choice has significantly less impact than the choice of the bandwidth. The use of the kernel is to weigh more the observations close to the centre of the moving window. The performance of a kernel is judged on the trade-off between variance and bias. The used Epanechnikov kernel is optimally efficient. However, even simplistic kernel functions, such as the rectangular, have a relative efficiency of 93%. Thus, kernel choice is not important and other factors may influence the decision, such as computational effort [Salgado-Ugarte et al., 1994, Silverman, 1986].

Bandwidth choice: Second, we hard code the bandwidth selection for computational reasons. The bandwidth of the kernel (and thus the width of the price interval over which points are counted) is determined on the basis of a trade-off between smoothing the original bid function and mixing up information of different parts of the bid function. By smoothing the original bid function, we obtain estimates of the information that our KDE measures (i.e. points in the interval and thereby the slope) that are less sensitive to local specificities of the bid functions. The larger the selected bandwidth, the larger the interval over which points are counted and the stronger the smoothing of the estimates. However, as the width of the interval increases, we mix up more information of a selected point of interest with the information of its neighbouring points. Therefore, in setting the bandwidth we aim to achieve smoothed estimates with a reasonable compromise between respecting local curve information, while not being fragile to steps in the bid function.

For estimates of the first order derivative, these considerations are minor and we could use the default bandwidth, optimal for a Gaussian distribution, to extract the point of maximum slope from the distribution. However, one reason we slightly increase it is to ensure that the distribution of the first derivatives is uni-modal[11]. Furthermore, the selection of the bandwidth in the first stage density estimation impacts both the precision and speed of the second stage estimation. A better smoothing in the first stage gives a large advantage in the second stage estimation[12], thus we have a further incentive to

[11]Uni-modal at the point of inflection in the price-quantity dimension. The smoothing ensures that the selected point is not mistaken due to steps in the bid function that have a very large slope locally, but which is not representative of the neighbouring portion of the bid function.

[12]The gain in computation in the second stage arises from the fact that a stronger smoothing in the first stage produces a more homogenous dataset for the second stage estimation. By more homogenous,

increase the bandwidth.

For the second derivative the trade-off is more critical: We want to obtain a reasonably broad smoothing to obtain a meaningful selection of points that is not driven by random noise. On the other hand, a large bandwidth reduces the importance of local information of a part of the curve as a consequence of which, selected points (points $k = 2$ and $k = 4$) are pushed towards the point of inflection ($k = 3$). This is due to the maximum point of the first derivative gaining more weight in the second derivative's estimation. The fact that first derivative estimates are already smoothed rather strongly, we can choose a narrow bandwidth in the second stage KDE.

In the end, we select a rather broad bandwidth of 45 units in the first estimation. This gains robustness of the point selection mechanism against noise in the data and estimation speed in the second stage. The bandwidth in the second stage is set more narrowly at a level of 2 units to keep as much information as possible from the first stage estimation and allow sufficient variation to select the k points.

To support our choice, we illustrate the impact of different bandwidths on the first and second stage estimation in figures 2.A.1 and 2.A.2. Our choice is based on an adequate point selection and the fastest runtime.

In these graphs, the top row shows the first stage KDE, over the whole function on the left and zoomed on the right. The large bandwidth in figure 2.A.1 shows the impact of smoothing on the estimates of the first derivative as compared to figure 2.A.2. The second row in both graphs shows the second stage KDE in two versions: Using a wide kernel bandwidth on the left and a tight bandwidth on the right. Again, we disclose the result as seen over the whole function (left) and zoomed on the central price range (right).

The third row details the original demand function with the final point selection given

we mean that fewer monotone regions of the graph of first derivatives must be interpolated at the unit cent level to ensure that our algorithm works correctly.

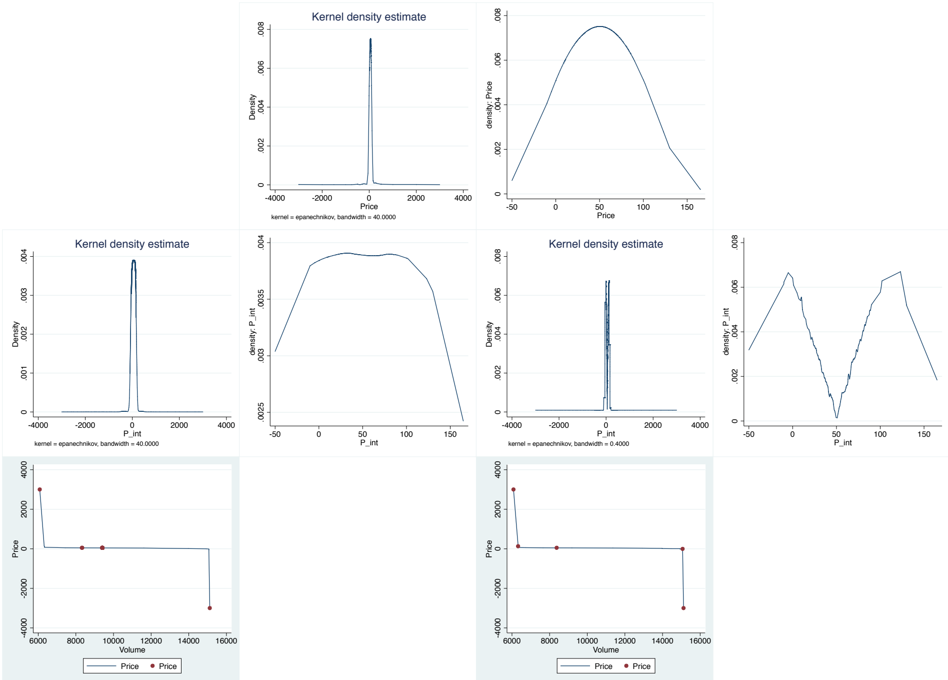


Figure 2.A.1: Comparison of bandwidths: Large bandwidth in first stage

Note: Large bandwidth in first stage (top row), large bandwidth in second stage (second row left), small bandwidth in second stage (second row right), Resulting selection of points for large bandwidth in stage one and two (bottow row left, A) and selection of points for large bandwidth in stage one and small bandwidth in stage two (bottom row right, B).

the bandwidth selection as given by the two rows above. Regardless of the first stage bandwidth, we see that a large bandwidth in the second stage KDE easily distorts the point selection. Selected points of type $k = 2, 4$ are either two centred or too wide as a result of the second derivatives being smoothed excessively and not precisely representing the local specificities of the curve.

The right hand side of both figures show that a tighter bandwidth on the KDE can easily mistake large slope changes due to steps in the bid functions as the appropriate points of maximum curvature of the full bid function and thereby make an error. Therefore, we apply a sensitive second stage KDE on rather smooth estimates of the first derivatives, which yields an adequate point selection in our setting (figure 2.A.1B).

The bandwidth selection received much attention in this work in order to obtain a reasonable selection of points based on local information of the curves, while achieving a

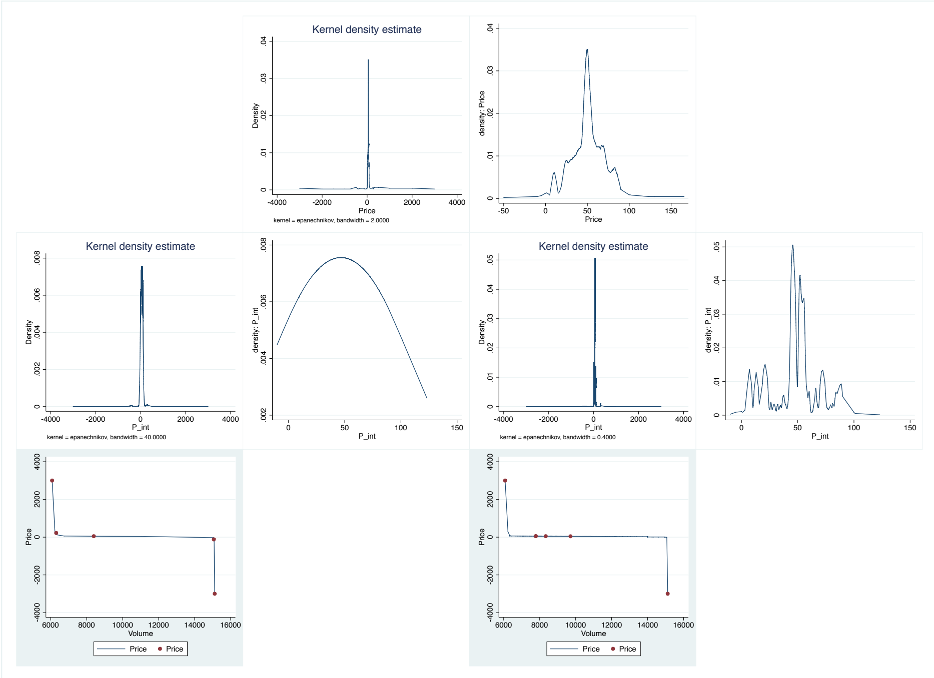


Figure 2.A.2: Comparison of bandwidths: Small bandwidth in first stage

Note: Small bandwidth in first stage (top row), large bandwidth in second stage (second row left), small bandwidth in second stage (second row right), Resulting selection of points for large bandwidth in stage one and small bandwidth in stage two (bottom row left, C) and selection of points for small bandwidth in stage one and two (bottow row right, D).

satisfying robustness to noise in the bid function. We are aware that this subjective setting of the bandwidth is not without consequence for our work. However for computational reasons[13], we do not run a full robustness test on this choice ex-post.

2.A.2 Outlier detection and removal

In some rare cases, our point selection mechanism does not work. This is the case when curves have very small number of points at a kink and it is thus very difficult to detect their curvature.

As a result, the selected points are then quasi in-differentiable from the next selected point type, i.e. a point of type $k = 2$ is almost identical to the selected point $k = 3$. The code is unable to select the right points due to a data lack on the original curve (second derivative on a constant slope up to POI is zero).

[13]The point selection algorithm ran for more than two weeks in the current setting.

We screen for adjacent points that display quasi no variation in volumes. Figure 2.A.3 shows a histogram of volumes differences over 2 selected points (from $k = 2$ to $k = 4$) and reveals a positive mass point at zero, indicating outliers that do not display any volume variation between points of the same bid function. We use the histogram to identify and drop those outliers from our dataset.

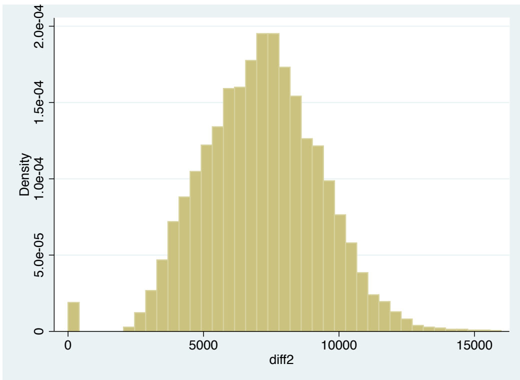


Figure 2.A.3: Histogram of volume variation between points

Note: The histogram shows the volume difference between points $k = 2$ *and* $k = 4$ *of the same bid functions.*

References

[APX, 2017] APX (2017). Type of orders on the epex day-ahead auction, uk. https://www.apxgroup.com/trading-clearing/uk-half-hour-day-ahead-1530-auction/.

[Arima, 2012] Arima, K. (2012). Dispatch modelling: Quantifying long term benefits via high resolution analysis. Technical report, Wärtsilä Power Plants.

[Berg et al., 1999] Berg, S. A., Boukai, B., and Landsberger, M. (1999). Bid functions for treasury securities; across countries comparison. Technical report, Citeseer.

[Bertoldi and Atanasiu, 2007] Bertoldi, P. and Atanasiu, B. (2007). Electricity consumption and efficiency trends in the enlarged european union, status report. Technical report, Institute for Environment and Sustainability, European Commission.

[Boiteux, 1960] Boiteux, M. (1960). Peak-load pricing. *The Journal of Business*, 33(2):157–179.

[Bolle, 1992] Bolle, F. (1992). Supply function equilibria and the danger of tacit collusion: the case of spot markets for electricity. *Energy economics*, 14(2):94–102.

[Boukai and Landsberger, 1998] Boukai, B. and Landsberger, M. (1998). Market bid functions for treasury securities as logistic growth curves. Technical report, Working Paper, Indiana University–Purdue University, Indianapolis, Indiana.

[Boyle, 2007] Boyle, G. (2007). *Renewable Electricity and the Grid, the Challenge of Variability*. Routledge.

[EPEX, 2015] EPEX (2015). Type of orders on the epex day-ahead auction, france. https://www.epexspot.com/en/product-info/auction.

[GE, 2015] GE (2015). Heavy-duty gas turbine operating and maintenance considerations. Technical report, General Electrics Energy.

[Green, 1991] Green, R. (1991). Reshaping the cegb: electricity privatization in the uk. *Utilities Policy*, 1(3):245–254.

[Green, 1999] Green, R. (1999). The electricity contract market in england and wales. *The Journal of Industrial Economics*, 47(1):107–124.

[Green and Newbery, 1992] Green, R. J. and Newbery, D. M. (1992). Competition in the british electricity spot market. *Journal of political economy*, 100(5):929–953.

[Grossman, 1981] Grossman, S. J. (1981). Nash equilibrium and the industrial organization of markets with large fixed costs. *Econometrica: Journal of the Econometric Society*, pages 1149–1172.

[Hertzler, 2003] Hertzler, G. (2003). A stochastic differential equation for modeling the "classical" probability distributions. In *2003 Conference (47th), February*, pages 12–14.

[Holmberg et al., 2008] Holmberg, P., Newbery, D. M., and Ralph, D. (2008). Supply function equilibria: Step functions and continuous representations.

[Klemperer and Meyer, 1989] Klemperer, P. and Meyer, M. (1989). Supply function equilibria in oligopoly under uncertainty. *Econometrica*.

[Newbery, 1998] Newbery, D. M. (1998). Competition, contracts, and entry in the electricity spot market. *The RAND Journal of Economics*, pages 726–749.

[Özcan, 2004] Özcan, R. (2004). The logistic function approach to discriminatory and uniform price treasury auctions.

[Préget and Waelbroeck, 2005] Préget, R. and Waelbroeck, P. (2005). Treasury bill auction procedures: Empirical perspectives from french market bid functions. *Journal of International Money and Finance*, 24(7):1054–1072.

[RTE, 2014] RTE (2014). Website of rte, the french grid operator.

[Salgado-Ugarte et al., 1994] Salgado-Ugarte, I. H., Shimizu, M., and Taniuchi, T. (1994). Exploring the shape of univariate data using kernel density estimators. *Stata Technical Bulletin*, 3(16).

[Seierstad and Sydsaeter, 1987] Seierstad, A. and Sydsaeter, K. (1987). *Optimal Control Theory with Economics Applications*. North-Holland.

[Silverman and Ramsay, 2005] Silverman, B. and Ramsay, J. (2005). *Functional Data Analysis*. Springer.

[Silverman, 1986] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

[von der Fehr and Harbord, 1993] von der Fehr, N.-H. M. and Harbord, D. (1993). Spot market competition in the uk electricity industry. *The Economic Journal*, 103(418):531–546.

[Wölfing, 2013] Wölfing, N. (2013). *Interacting markets in electricity wholesale : forward and spot, and the impact of emissions trading*. PhD thesis. Thèse de doctorat dirigée par Schubert, Katheline et Fleckinger, Pierre Sciences economiques Paris 1 2013.

List of Figures

1.1 a b 29

1.2 This graph plots $S^*(p)$ for different values of the ramping cost parameter, and compares them to the set of equilibria obtained in KM's framework. Four optimal supply schedules are plotted. The black curve (full line) corresponds to the case where $\gamma \to 0$. As before, as γ increases the optimal schedules get steeper and steeper until in the limit of $\gamma \to \infty$, the optimal schedule attains a vertical slope. In addition, we show the set of available equilibria in KM's model in light green, and the extremal demand schedules in dashed black. . 32

1.3 This graph plots an envelope of constant average value but varying width $\omega(t)$. By comparing regions of increasing or decreasing width, respectively the left or right side of a lobe, one sees that the informativeness of the envelope is being respectively reduced or increased with respect to a situation where the width would be kept constant. The change in informativeness is represented by the area between the (full) green arrows (observed level of informativeness) compared to the area between the (dashed) red arrows (level of informativeness had the width been constant). In addition, by comparing the left lobe to the right one, it is possible to see why the relative variation of the width, and not the absolute variation of the width, matters. For a larger width (right lobe) and the same rate of change in the width, there is less change in informativeness than for a smaller width (left lobe), i.e. the same rate of change matters less for the right lobe than for the left lobe. 43

2.2.1 Example of an assymetric aggregate supply function. The x axis is the quantity in MWh, the y axis if the price in [Pleaseinsertintopreamble]. In red is the actual aggregate function, in green is an estimated logistic function showcasing the large discrepancies that can arise with this parametric approach. The blue point is the market outcome. 61

2.2.2 Comparison of two aggregate demand functions for the same hour . . . 64

2.2.3 Steps of the point selection process 65

2.2.4 Heat map on selected, comparable demand and supply points 67

2.2.5 Distribution of selected demand (left) and supply (right) points 68

2.2.6 Histogram of slopes per point type 69

2.2.7 Error bars as a function of the number of extracted points 71

2.2.8 Proxy for degrees of freedom on master curve 72

2.2.9 Overall (left) and zoomed (right) Mastercurve with confidence interval . 72

2.3.1 Stations for which we have hourly data. Left : temperature, center : wind speed, right : light intensity. 75

2.3.2 Left: Voronoi's algorithm is applied once on the reference points highlighted in green to obtain the white surfaces, and a second time on the same points to which is added the query point in the center to obtain the new blue cell. The green circles, which represent the interpolating weights, are generated using the ratio of the shaded area to that of the cell area of the surrounding points. Center left: example of a reference surface (color mapped) to be reconstructed through a natural neighbor interpolation. Center right: interpolated surface with a reference set of 16 evenly organized points, represented in black. Right: interpolated surface with a reference set of 16 unevenly organized points, represented in black. From 16 points one is able to reconstruct the color mapped surfaces which are approximations of the reference one, represented in the center left image. 76
2.3.3 Temperature forecast from a simulation run by the GFS at 6 a.m. on the 3rd of november 2011, for a forecast at 22 p.m. 77
2.3.4 Left: reference image. Center left: borders and numbers are removed. Center right: edge recognition. Right: final fitted surface. 78
2.3.5 Auto-correlation lengthscale computation. In black are the points obtained from all the pairs from our original data, that is absolute wind speed differences as a function of the distance between the two points. In blue is the kernel smoothed function from those points. In red is the exponential fit. In black are the derivatives of the fit at 0 and ∞. In green is the recovered auto-correlation lengthscale. The unit for the lengthscale is in km. 80
2.3.6 Typical response curves of different wind turbines 82
2.3.7 All curves are hourly production data. The origin of the hours is the first of January 2011, and the production is in MWh. In blue: the observed wind production. In dark red: the day-ahead predictions from the grid operator. In light red: the day-ahead predictions from weather data. 82
2.3.8 Daily electricity consumption in France as a function of the temperature 83
2.3.9 Hourly solar production in MWh. The time origin is the first January 2011. In blue: observed production by RTE. In dark red: reconstructed production from observed weather data. 85
2.A.1 Comparison of bandwidths: Large bandwidth in first stage 92
2.A.2 Comparison of bandwidths: Small bandwidth in first stage 93
2.A.3 Histogram of volume variation between points 94

Table of Contents

Remerciements—Acknowledgements **i**

General introduction **1**

Summary **5**

Chapter 1: Dynamics of the Electricity Day-Ahead Market : Supply Function Equilibria and Ramping Costs **9**

1.1 Introduction 11

1.1.1 Litterature review 11

1.1.2 The day-ahead markets 14

1.2 Heuristic Description of the Model 16

1.3 Stochastic Dynamics 20

1.3.1 The stochastic process 21

1.3.2 The ramping costs 22

1.3.3 Discussion of the approximations 25

1.3.4 The maximisation program 25

1.4 The Monopoly 26

1.4.1 Results 27

1.5 The Symmetric Oligopoly 28

1.5.1 Discussion 33

1.6 Dynamic behavior of the bids 34

1.6.1 The setup 36

1.6.2 Results 39

1.6.3 Discussion 41

1.7 Limits 43

1.8 Discussion and Concluding Remarks 45

Appendix **47**

Appendix 1.A Proof of equation 1.11 47

Appendix 1.B Proof of Proposition 1.4.1 48

Appendix 1.C Proof of Proposition 1.5.1 50

Appendix 1.D Proof of proposition 1.5.2 53

Appendix 1.E Proof of proposition 1.6.1 54

Appendix 1.F Proof of proposition 1.6.2 55

Chapter 2: Methodological tools for non parametric functional data evaluation and weather data usage **57**

2.1 Introduction 59

2.2 Point selection on functional data : a non parametric approach 59

2.2.1 Information about our data 62

2.2.2 Point selection algorithm 63
2.2.3 Results of the point selection methodology 66
2.2.4 Observations of bidding frictions 67
2.2.5 Discussion 72
2.3 Methodology to aggregate geographically dispersed information on a national level 74
2.3.1 Dealing with meteorological data 74
2.3.2 Aggregation of local information 79
2.4 Conclusion 87

Appendix **89**
Appendix 2.A Technical details 89
2.A.1 Using the kernel density estimation (KDE) in our setting 89
2.A.2 Outlier detection and removal 93

References **95**

List of Figures **97**