# Firm Behaviour under Uncertainty and Legal Challenges

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2015

vorgelegt von

Henri de Belsunce

| | |
|---|---|
| Referent: | Prof. Dr. Klaus M. Schmidt |
| Korreferent: | Prof. Dietmar Harhoff, PhD. |
| Promotionsabschlussberatung: | 4. November 2015 |

Datum der mündlichen Prüfung: 27. Oktober 2015

Namen der Berichterstatter: Klaus M. Schmidt, Dietmar Harhoff, Gregor Zöttl

# Danksagung

Mein Dank geht zuerst an Klaus Schmidt für die Betreuung und Unterstützung bei dieser Dissertation. Die fachlichen Diskussionen und Kommentare haben maßgeblich zum Erfolg dieser Dissertation beigetragen. Ebenso möchte ich mich bei Dietmar Harhoff für die Betreuung des empirischen Teils dieser Arbeit bedanken, sowie bei Gregor Zöttl für die Bereitschaft als Drittgutachter zu fungieren.

Die größte Freude an der Forschung habe ich in der Zusammenarbeit mit meinen Koautor Alexis Bergès entwickelt. Der rege Austausch und die Komplementarität unserer Forschung haben im dritten Kapitel dieser Dissertation gemündet und ich freue mich auf weitere Zusammenarbeit.

Ich habe stark von den stimulierenden Forschungsumfeldern am Max Planck Institut für Innovation und Wettbewerb sowie auch am Lehrstuhl für Wirtschaftstheorie der Ludwig-Maximilians-Universität profitiert. Die vielen fruchtbaren Diskussionen und Anregungen, sowie das angenehme Arbeitsklima habe ich sehr genossen.

Ich bin für die finanzielle Unterstützung über das Stipendium des Max Planck Instituts und das Abschlussstipendium der Ludwig-Maximilians-Universität sehr dankbar.

Der wichtigste Dank gilt meiner Familie und Freunden, die immer für mich da sind und mir echten Rückhalt im Leben geben. Ihnen widme ich diese Dissertation.

# Contents

# List of Figures

# List of Tables

# Preface

This doctoral dissertation presents applied research at the level of the individual firm. It is composed of two main parts, which relate to different fields of research. Chapters 1 and 2 use economic theory for an analysis of the law on the market for intellectual property. Chapter 3 uses econometric techniques for an investigation of supplier behaviour on the electricity market.

The central theme common to all chapters is the understanding of how firms adjust their strategic behaviour in response to changes in their environment. In game theoretic parlance, how firms react to changes in the rules of the game. These rules can be artificially imposed on a game, e.g. laws imposed by a regulating authority, or may arise naturally, e.g. physical laws or technological constraints in production. We will analyse specific instances of both types.

Understanding the effects of these rules on firm behaviour is important for sophisticated policy making. The impact of a new law or policy crucially depends on how existing and future economic actors react to the change. Since in real-life it is not possible to perfectly predict the outcome of a specific policy, both the ex-ante and ex-post evaluations of these policies are important to ensure success. The importance of evaluating policies is obvious by the cost of implementing these measures to society. In Germany, for example, the expenditure for research and development made up 2.85% of gross domestic product in 2013[1]. When evaluating policies, it is important to pay attention to the market specificities, because the impact that rules have on firms is very much context dependent. The structure of the market also affects the economic tools available to us for their analysis. While some markets yield detailed and precise data that allow us to disentangle the causal effects of policies, others do not. We therefore focus our analysis of the two types of rules on different markets and employ both theoretical and empirical tools.

[1]Source: Commission of Experts for Research & Innovation, (EFI) (2015) - EFI annual report.

In the first half of this dissertation, we look at the impact of artificial rules in the context of the market for cumulative innovations. The intellectual property market provides an interesting environment to investigate the effects of some artificial rules for two reasons: First, the state's interference with the natural market for innovation by granting patents is rather invasive[2] and, second, many rules employed to incentivise innovation are heavily contested[3], e.g. the US treble damages doctrine (details below).

Furthermore, the importance of innovation for economic growth is widely recognised[4] and many government policies aim at encouraging innovative activities[5]. Patent protection is one means of encouraging innovation by granting exclusion rights to innovators. Based on the premise that stronger protection rights foster more innovation, many governments have strengthened intellectual property rights over time[6].

In chapter 1, we look at the strategic patenting incentives of incumbent firms and how these incentives are affected when a patent strengthening policy is imposed by regulators. Specifically, we analyse the effect of the US treble damages doctrine for wilful patent infringements on firms in the market. This doctrine allows the award of three times the damages to a patentee in case of a patent infringement, when the infringement is judged to have been conducted on purpose. The possibility to receive treble damages strengthens the position of patent holders, but is strongly criticised for discouraging other firms from studying competitor patents and inhibiting the diffusion of patented knowledge[7].

We show that the excessive patenting of low quality ideas can be a way to reduce market entry for incumbents that are committed to litigating entrants. In order to do so, we set up a theoretical model in which the entrant can decide to use the study of prior art in order to reduce the probability of losing in litigation upon market entry. Then, when the studying costs depend on the size of the incumbent's portfolio, the incumbent can use the inflating of the size of his patent portfolio as a "raising rival's cost" strategy. This result could contribute to explaining the surge of low value patent filings observed at the US patent offices in recent decades[8].

---

[2]See Lemley and Tangri (2003), Kitch (1977).
[3]See Means (2013), Schmidt (2010), Lemley (2008).
[4]See Menell and Scotchmer (2007) for a brief survey.
[5]See Commission of Experts for Research & Innovation, (EFI) (2015) for numerous examples.
[6]See Takenaka (2000) and Jaffe (2000) for the evolution in Japan and the US, respectively.
[7]See FTC (2003).
[8]See Moore (2005).

Furthermore, we show the utility of the treble damages doctrine in reducing the incentives of incumbents to patent excessively. This result arises because treble damages reduces the profitability of studying prior art by making studying errors more expensive. As a result, entrants invest less in the study of prior art. As the study of prior art decreases, the raising rival's cost impact that the incumbent can induce using an excess patent is reduced. While the reduction of excess patenting incentives is beneficial to society since it avoids the wasteful spending of efforts on patenting worthless ideas as well as reduces the strain on the patent office, the treble damages doctrine itself represents a burden on the entrant. We thus highlight a trade-off on the pros (reduces strain on the patent office) and cons (reduces information-sharing nature of the patent system) for the patent strengthening treble damages doctrine and give policy recommendations depending on the public goal. Thereby, we contribute to the active discussion of the treble damages doctrine in the United States.

However, the view that stronger exclusion rights are the best means to achieve the goal of more innovation has been challenged by developments in the recent past. Faster product life and innovation cycles have emphasized the impediment that stronger patent protection may bear for cumulative innovators and the diffusion of knowledge. The Information and Communication Technology (ICT) sector is a strong example. The recent growth of this sector has been accompanied by a high rate of patent disputes and litigation activity. This has stimulated much research by academics and policy changes to deal with the challenges imposed by the new innovation environment[9].

In the ICT sector especially, technologies often rely on intellectual property contained in many patents and held by many parties. Furthermore, the need for interoperability is particularly pronounced in this sector and has been one of the driving forces behind the creation of standards in that industry. When patents become essential for the implementation of a standard, they gain a lot of value due to the hold-up potential that standard-essential patents (SEPs) have. Standard setting organisations (SSOs) and regulators alike have increasingly become aware of the hold-up potential in composite technologies and have adopted voluntary rules or general regulation, respectively, in order to restrain SEP holders' ability to abuse their market power. A commonly imposed commitment is the FRAND commitment for SSOs and patent pools. This commitment demands that the licensing of SEPs occurs at fair, reasonable and non-discriminatory rates and aims

[9] See Lerner and Tirole (2013) and Regibeau and Rockett (2011)

to curb the bargaining power of SEP holders in licensing negotiations. While FRAND terms represent a serious commitment enforced by the courts[10], there exists no generally accepted definition of FRAND terms nor ex-post tests. Consequently, the strength of a FRAND commitment depends crucially on the wording of a FRAND commitment and how the commitment is embedded with other clauses in the agreements. In particular, other clauses may provide exceptions to the rule or soften conditions.

In chapter 2, we look at a specific contractual clause, called partial termination, from the patent pooling agreement of the MPEG-2 patent pool and investigate its effects in light of the recent developments of US antitrust law with respect to patent holders. The partial termination clause gives pool members a bargaining advantage in the negotiations for non-essential patents owned by licensees of the patent pool. This clause merits attention, because the MPEG-2 pool is the only patent pool to include such a clause in its licensing agreement and some puzzling observations from the literature suggest that the FRAND commitment of the MPEG-2 pool is weaker than that of other pools[11].

We use a simple theoretical model to illustrate the competitive and licensing relationship of a patent pool member and a licensee of that patent pool. This setting allows us to compare the incentives of a licensee to invest in cumulative innovation under the two regimes, where the patent pool member has access to the partial termination clause and where it does not. We show that when the threat of partial termination is credible, the pool licensor is able to capture some of the rents from cumulative innovations made by the licensee of the pool. Also, we show that the credibility of the partial termination threat depends on the strength of the essential patent right granted by the courts[12].

Our analysis is able to shed light on the previously mentioned puzzles from the literature. First, we are able to contribute a possible explanation for the drop in innovation rates of firms that research in the technological proximity of the MPEG-2 pool. This result follows from the capture of innovation rents by pool members. Second, we provide a possible explanation for why vertically integrated firms are overrepresented (under-represented) among the MPEG-2 pool members (outsiders) as compared to other patent pools. This structural effect arises from the fact that the partial termination clause increases the value of joining the pool for vertically integrated firms only, the joining incentives of pure upstream firms are unaffected.

[10]See section 2 of chapter 2 for details and references to case law.
[11]See Vakili (2012) and Layne-Farrar and Lerner (2011).
[12]In legal terms, whether a property rule or liability rule applies to holders of essential patents.

Furthermore, we show that the current organisational rules of the MPEG-2 pool exacerbate the threat of partial termination, because these rules does not actively restrain firms from over-declaring essential patents. Our analysis thus also contributes to the debate on whether SSOs and patent pools should take active responsibility in the determination of the essentiality and validity of SEPs.

In the second half of this dissertation, which is joint work with Alexis Bergès[13], we study a natural rule of the game, namely uncertainty. Theory predicts that uncertainty is costly to firms when they face dynamic costs[14]. Dynamic costs refer to the costs incurred by producers when the volume of production varies. These costs are inherent to some production technologies[15]. When faced with uncertainty, we expect that suppliers with dynamic costs smooth production volume over time in order to minimise these dynamic costs. In this third chapter, we find support for this prediction.

The question of dynamic costs on the electricity market is an important and topical one. In order to avoid costly power outages, the generation must closely match consumption at all times. Thus, given constant consumption, the outage of a generation plant must rapidly be compensated for by other plants. The quick increase in production for the other plants is expensive, if not impossible due to capacity constraints. For this reason, the intermittency of the generation from renewable energies poses major challenges to the electricity system as a whole. The topic was brought to the attention of the general public by the media recently. In Germany alone, the solar eclipse (occurring around midday of March, 20$^{\text{th}}$ 2015) induced a variation of solar generated electricity with a gradient of up to -272MW/min for the ramp-down and up to +348MW/min for the ramp-up[16]. The volumes of the total ramps have been compared by companies and media to the quantities produced by up to 10 (for ramp-down) and 20 (for ramp-up) average sized nuclear power plants. These ramps occurred within the space of a few hours and electricity producers announced to have prepared to face this challenge together for more than a year in advance. While we do not focus on this specific event, the example helps to emphasize the importance of better understanding firm behaviour in the presence of dynamic costs.

[13]Doctoral candidate at the Paris School of Economics, France.
[14]See Bergès and Martimort (2014).
[15]Most notably for traditional production techniques on the electricity market(gas, coal, nuclear).
[16]See Weniger et al. (2014) and www.opower.com.

Our analysis is based on an empirical methodology and focuses on the French electricity market. This market is a divisible goods auction and particularly suited for our analysis, because it provides a rare level of data insight: we can observe the full aggregate supply and demand functions for every hourly auction. From the demand functions, we can capture uncertainty about the realisation of market demand. Furthermore, uncertainty on the electricity market also arises from the fact that renewable production (e.g. wind and solar) is a local and dispersed production, but feeds into a national market with a single price. When meteorological conditions change, the geographic production profile also changes. This further complicates the predictability of renewables generation and contributes to the uncertainty that electricity producers face when playing on the electricity market[17]. With the growth of installed capacity for generation from renewable energies (which has strongly been supported by government policies in the recent past), this source of uncertainty has gained importance in recent years. We study how both sources of uncertainty affect the strategies of suppliers facing dynamic costs.

We show that firms on the French electricity market take uncertainty from meteorological forecasts as well as uncertainty from demand realisation into account and interpret the result as a sign that expected dynamic costs matter on this market. More precisely, we show that the electricity suppliers react to an increased level of uncertainty by bidding more volume elastically in order to minimise dynamic costs, which increase with the uncertainty. While our findings hold for the central part of the bid functions, they disappear towards the extremities of the bid functions where capacity constraint considerations seem to dominate. Our results also indicate that not only supplier bidding is affected by uncertainty, but that the level of uncertainty also impacts bidding from the demand side of the market.

The results are obtained using a reduced form model. However, in order to make most of the observed data structure, we adjust our empirical methodology. Our work is non-standard with respect to two methodological building blocks. First, we develop a methodology to concentrate the full market data from the French electricity market, that is, we reduce the hourly bids in the form of full demand and supply functions into a fixed number of points per function[18], where each point is representative of a part of the original function. This methodology allows to run a reduced form model on the selected points while circumventing traditional problems of endogeneity that occur when working with

[17]See Meibom et al. (2009).

[18]Using landmark registration techniques (Silverman and Ramsay, 2005).

equilibrium outcomes. Second, we develop a technique to aggregate the local impact of weather at the national level. This technique is based on a new bottom-up approach and allows to describe exogenous weather variables at a sufficiently aggregated level to include them in the reduced form analysis.

We contribute to the literature in a number of aspects. First, our analysis is able to find statistically significant support for the importance of dynamic costs. Our approach to separate out the uncertainty from market demand expectations and predictability of renewables generation is novel. The proxies used to capture the level of both types of uncertainty are also new. Second, we employ a number of data-intensive adjustments to leverage our dataset. For example, we control for both daytime and longer seasonality using continuous variables rather than dummies. Thereby, we are able to increase the sample size for each of our regressions, avoid the "black-box" interpretation of dummies and improve the precision of our estimates. Third, the functional focus of our empirical strategy allows to analyse a much fuller transcript of the firms' strategies. The functional approach allows to investigate the players behaviour both in the region where the equilibrium is likely to occur as well as in regions that rarely have an impact on the equilibrium outcome. Our technique relies on a non-parametric selection of representative points. Compared to previous, parametric work using functional data, we are able to overcome some underlying structural restrictions, e.g. the symmetry of the logistic function approaches[19].

With this dissertation, we hope to foster the discussion on how rules are set and managed. In particular, we want to stress the importance of both the ex-ante and ex-post evaluation of policies. A rigorous evaluation of policies will allow us to better understand firms' behaviour in changing environments. Only when policy makers learn from past policies, we can improve policy making in the future.

Both halves of the dissertation are fully independent. Furthermore, all chapters themselves are self-contained (with their own introductions and appendices) and may be read independently. The bibliography is joint for all chapters at the end of the dissertation.

[19] See Préget and Waelbroeck (2005), Özcan (2004), Belsunce (2011).

# Chapter 1

# Do More Patents Mean Less Entry?

## Patenting strategies in cumulative innovation under the threat of litigation

### 1.1 Introduction

Based on the premise that stronger patent rights foster more innovation, the United States have implemented patent strengthening policies since the 1980s. At the same time, we have witnessed an unprecedented explosion[1] of the size of patent portfolios that firms demand through patent filings or acquire via financial transactions.

However, many scholars were unable to establish a direct link between the stronger patent rights and the patent surge (Jaffe (2000), Qian (2007), Lerner (2009)). Using a survey, Cohen et al. (2000) find that the effectiveness of patents as a means to protect innovations has not increased in the eyes of the innovating firms following the pro-patent policy changes. Kortum and Lerner (1999) promote the view that the patent surge is driven by factors outside of the patent system. Their work suggests that improvements in research productivity have driven the surge in patent filings, an argument that rests on the assumption that firms have been able to create more (valuable) innovations for which the said firms demanded patent protection. We are also interested in the link between

[1]Most notably in recent years: US patent grants have increased by 58% over the 5-year period from 2008 to 2013, the ratio of grants per applications was 47.8% in 2012 (Data from US PTO Patent Monitor as of June 2014).

stronger patent rights and the patent surge, but focus on another aspect, namely the observation that numerous granted patents are commercially and even legally worthless[2]. We ask in this chapter, if the excessive patenting of low quality ideas may have contributed to the patent surge and if a particular patent strengthening policy, the treble damages doctrine (details later), could have influenced such a strategy?

The following observations motivate our analysis. First, the lack of standards at the patent office has been lamented by many scholars, e.g. Gallini (2002), Shapiro (2001), Bessen (2003). Also, the US Supreme Court has acknowledged the "notorious difference between the standards applied by the Patent Office and by the courts" (Meurer, 1989). As a result, a high number of patent disputes end up in court and almost half of the litigated patents are held invalid when contested in court (Allison and Lemley, 1998). Effectively, the true opposition of patents occurs in court and firms innovate under the threat of litigation. Second, when litigation occurs, the opposing parties contest the validity of a patent in light of the existing knowledge, i.e. prior art. Prior art is often contained in multiple patents and claims and thus in practice, the litigation process opposes patent portfolios. The court rule in favour of or against an incremental innovation depends on the value-added of the new innovation relative to the value of the prior art, i.e. on both the quality of the new patent as well as the strength of the patents composing the prior art (Llobet, 2003). Third, although a patent is publicly available, its patent strength is private information. Thus, an information asymmetry arises between the owner of a patent portfolio and an observer. The larger a portfolio, the stronger the information asymmetry concerning that portfolio will be. Only a costly[3] effort of searching and studying existing patents gives an indication of the strength of prior art.

These observations suggest the argument that firms use patents as strategic tools in competition. We set up a simple and tractable model that uses the above observations as its main ingredients. Using this model, we show that the excessive patenting of low quality ideas by incumbent firms can be rationalised as a means to reduce market entry. This result could contribute to explaining the surge of low value patent filings observed in the US.

[2]Referring to patents that have no impact on profits and which would be invalidated with certainty if contested in court. For a detailed account on worthless patents see Moore (2005).

[3]Studying of a patent is costly in effort for any reader. Furthermore, "not being engaged in the development process, a competitor is likely to know less about the content of the prior art (Meurer, 1989)". Also, the mere search for prior art may be costly, since the disclosed prior art in a patent is often insufficient. For critiques on the low performance of disclosure in patents, see Merges (1999a) and Kitch (1977).

Our result relies on the study of an entry game with two firms. An entrant aims to enter a market that is protected by a patent by designing around the existing patent, while an incumbent aims to deter entry by litigating all entrants. Market entry in this setting triggers infringement litigation in which the courts decide whether an infringement has occurred or not based on the relative strengths of the firms' patents. Given that patent strengths are private information, the outcome of the litigation is uncertain. In order to reduce this uncertainty, the entrant may invest in the study of competitor patents (called screening henceforth) before entering the market. Because the screening costs of the entrant depend on the size of the incumbent's portfolio, the incumbent can reduce screening of prior art by the entrant by inflating the size of his patent portfolio, i.e. excessive patenting. This strategy works by the mechanism of a *raising rivals' cost* argument. When the entrant switches away from screening before entering, he either stays out of the market straight away or enters without screening depending on the quality of his innovation (weaker innovators stay out, better innovators enter directly). It is the former switch that is in expectation profitable for the incumbent, who can save litigation costs and avoid competition. Inducing this switch motivates an excess patenting strategy in the model. On the other hand, the latter switch is costly for the incumbent since litigation costs have to be paid more often. We characterise the trade-off for the incumbent to pursue a strategy of excessive patenting.

We use this setting to investigate how the treble damages doctrine affects this strategic use of patents. The treble damages doctrine allows a US court to treble the damages to be paid to the defendant (284 United States Code 35, 2006) in case of a wilful patent infringement. This question is an important one since a famous concern about the treble damages doctrine is that it discourages firms from reading competitor patents (FTC, 2003; Lemley, 2008; Lemley and Tangri, 2003). When firms stop reading existing patents, the information-sharing nature of the patent system is undermined and therefore costly to society[4].

We show that the treble damages doctrine reduces the incentives of an incumbent to invest in an excess patenting strategy. The result is driven by the fact that when screening is imperfect, both false positive (enter and lose in litigation) and false negative errors (forego profitable entry) occur for the entrant with positive probability in equilibrium.

[4] "Failure to read competitors' patents can jeopardize plans for a noninfringing business or research strategy, encourage wasteful duplication of effort, delay follow-on innovation that could derive from patent disclosures, and discourage the development of competition" (FTC, 2003).

The treble damages doctrine multiplies the cost of false positive screening errors only and, thereby, represents an additional cost to screening. Given that weaker entrants are more prone to this type of error, the increased screening cost disproportionately affects weaker entrants. Thus, under treble damages, the previously described strategy switches that an excess patent can induce are affected: The profitable switch away from market entry on weak entrants is reduced more than the costly switch towards market entry for stronger entrants. Hence, the effectiveness of using an excess patenting strategy is reduced.

We conclude that treble damages have the positive effect of reducing the excess patenting incentives of incumbents. This is notable if policy makers are concerned about the strain on the patent office due to the flood of patent filings. However, the reduction in excess patenting incentives comes at the expense of entrants reading fewer competitor patents. This undermines the effectiveness of the patent system in sharing knowledge. Consequently, positive externalities for society arising from the information-spreading nature of patents are reduced under a treble damages regime. With these insights, we contribute to the legal debate on treble damages in the US.

This chapter draws on multiple strands of both the legal and economic literature. It provides a model on the economic implications of the treble damages doctrine in order to support a highly active legal debate. The model uses elements from the economic literature on innovation under the threat of litigation, searching and reading prior art, as well as entry deterrence.

A few papers are particularly notable for our modelling of litigation. In order to investigate the effect of treble damages, our model must yield that infringement findings occur in equilibrium. This can be achieved either using probabilistic patent strengths, e.g. Ménière and Parlane (2008)[5], or using imperfect information concerning these strengths, e.g. Llobet (2003). We opt for the latter approach and adopt Llobet's market structure and litigation procedure. However, we do not adopt imperfect courts that only observe noisy signals on the patent strengths and assume no settlement possibility. Llobet uses his setting to derive optimal licensing rates for the patent of an incumbent facing entry.

[5]They assume (exogenous) endowments of patent portfolios, which yield probabilistic success rates if asserted against an opponent's portfolio, in order to better understand the threat of hold-up litigation. They derive the optimal level of infringement fees based on the trade-off between positive (counter over-investment incentives from an R&D race) and negative (deter from investment) effects of infringement litigation on innovation.

We adjust the setting to focus on how the institutional framework of the patent system affects firms' patenting strategies.

A second relevant strand of the literature is the one looking at when firms invest in the reading of patents. Atal and Bar (2010) analyse the search and study incentives of innovators vis à vis the patent office. Langinier and Marcoul (2007) focus on using the effort of patent examiners as a policy tool to improve prior art disclosure in patent filings. Caillaud and Duchêne (2011) study the overload of the patent office while focussing on examination fees and toughness as policy tools. While these papers focus on the prior-art search for patent approval, our work sets itself apart by looking at firms' strategic study of rival patents in a context of market entry. Our focus on the reading of competitor patents in light of the treble damages doctrine is novel and contributes to a vibrant debate.

Moreover, the classical I.O. literature (initiated by Salop and Scheffman (1983)) on entry deterrence by means of raising rivals' cost arguments deserves a brief mention. The present model puts the traditional mechanisms into the current institutional context of the patent system and investigates the response of players using such strategies to treble damages as a policy tool. This chapter can hence be seen as a reinterpretation of the basic ideas from a raising rivals' cost consideration à la Gilbert and Newbery (1982), where this paper uses a different set-up to focus on the private nature of information on patent strengths. In contrast to their paper, we can thus rationalise the excessive patenting of worthless ideas by incumbents and see how this strategy performs in a context where courts award treble damages.

The paper proceeds as follows. In section 1.2 the model set-up is proposed. Section 1.3 presents the results for the baseline model. In section 1.4, we apply the model to shed light on the effect of the treble damages doctrine on the behaviour of both incumbent and entering firms. Section 1.5 looks at extensions and robustness of the model, section 1.6 discusses the results. Section 1.7 concludes.

## 1.2 Model set-up

The model focuses on a two-player setting with an incumbent ($I$) and an entrant ($E$). The model relies on two institutional aspects of the patenting system that were introduced in the introduction.

First, patent strength[6] is private information, but observable at a cost. This means that a competitor must spend a costly effort $s > 0$ to study an existing patent in order to learn its strength. However, the number of patents ($N$) that a competitor holds is observed for free.

Second, patents are granted by the patent office without proper scrutiny. In the model, the argument is taken at the extreme and any idea (even worthless ones) will be granted a patent. The true opposition of patents occurs in court. Thus, cumulative innovation occurs under the threat of litigation.

### Storyline

$I$ earns profits from the sale of a product that generates an exogenous consumer valuation of $v_I$. $I$'s market (or production technique) is protected by a patent portfolio of breadth $b$. We assume that $I$ is committed to litigating any entrant[7].

$E$ has an idea to enter the market by engineering around[8] existing IP. $E$'s entry generates a small exogenous increase in the consumer valuation $v_E$. $E$ gets a patent of quality $q_E$ for its alternative production technology[9].

Given $I$'s commitment to litigate, $E$'s entry triggers litigation on infringement of $I$'s patents. The court decides if an infringement has occurred and whether damages have to be paid. Market entry is only profitable for $E$ when it wins in the legal process, i.e. when the design-around does not infringe the existing patents of the incumbent.

## 1.2.1 Litigation

When $I$ litigates, it files an infringement suit against $E$. The entrant's defence is based on the validity of its design-around patent. For $E$'s patent to be upheld valid by the courts, it may not infringe on the claims of $I$'s patent. We thus have an opposition of the patent strengths in court, where the entrant's quality $q_E$ must be sufficiently large the overcome

[6]We use the terms patent "strength", patent "quality" and patent "breadth" synonymously.

[7]This is the case when an incumbent has a reputation to pursue imitators in court to fight entry. Such a strategy has been used for example by Apple under Steve Jobs against Android phones.

[8]The invention of an alternative idea that produces the same output as a given patented idea, while not infringing the original patent's claims.

[9]The insights of the paper remain the same if the entrant does not obtain a patent, but merely produces the output using an alternative production technology.

the patent breadth $b$ of the incumbent's patent portfolio[10]. Given that both $I$'s and $E$'s patents describe different ideas to produce the same output, we assume that both $b$ and $q_E$ are drawn from the same interval $[0, \bar{q}]$. More on this in subsection 1.2.2.

During the legal process, the court observes the true qualities of the patents in question and decides in favour of the stronger patent. This is based on the idea that for an infringement case, a better engineering around idea has a higher chance of surviving the legal enquiry. In the same logic, a stronger incumbent patent (e.g. broader and more general claims) will reduce the survival chances of the entrant's patent in court. We assume that courts are perfect and use a deterministic litigation technology. The probability of the entrant's patent not being invalidated is given by a threshold decision rule:

$$m = \mathbb{P}\{q_E > b\} = \begin{cases} 1 & \text{if } q_E > b \\ 0 & \text{if } q_E \leq b \end{cases}$$

where $b$ is the strength of the incumbent's existing patent and $q_E$ is the strength of the entrant's patent. We have $\frac{\partial m}{\partial q_E} \geq 0$ and $\frac{\partial m}{\partial b} \leq 0$.

The court has a re-distributional character. Under litigation, the total surplus generated by the producers is dispatched. During the litigation, the market structure generates payoffs as described below in the subsection 1.2.3. In case of infringement, the entrant has to compensate $I$ for the lost profits ($v_I$). This corresponds to single damages. The variation to treble damages is introduced and analysed in the application of the model in section 1.4.

Furthermore, going to court is inefficient as litigation costs $K > 0$ arise and will be borne by the loser of the case. This corresponds to the legal doctrine on the burden of litigation costs as practised in most European countries and does not match the analysis of treble damages which is a US concept. However, it simplifies the analysis and is thus used in this initial model. Robustness of our results to this specification is shown in section 1.5.

### 1.2.2 Innovation

The general technology for innovating is as follows. The idea for an innovation is obtained for free, but in order to realise the advantages of an innovation, the innovator must

[10] If we think of $b$ as the "radius" of ideas around $I$'s technology that are protected by $I$'s patent portfolio, then $E$'s patent must be sufficiently "far" away from that technology, where the distance between the design around and the original idea is measured by the patent quality $q_E$.

develop the idea into a patent and implement it at a cost $c_i$, $i \in \{I, E\}$. The quality of an innovation $q_i$ is drawn i.i.d. from a uniform distribution U(.) on support $[0, \bar{q}]$: $q_i \sim$U$[0, \bar{q}]$. The distribution and its support are common knowledge.

*For the entrant*, this means that it learns the idea for a design-around and its quality $q_E \sim$U$[0, \bar{q}]$ for free. Implementing the design-around costs $c_E > 0$. $E$ knows $q_E$ since it knows the existing technology (which it is trying to imitate) and the dimension in which it is engineering around. However, $E$ does not know the breadth $b$ of the patents which $I$ holds to protect its IP.

*For the incumbent*, this means the following: Its existing patent portfolio of size $N$ is protecting its intellectual property with a breadth of $b$. Implementation costs for this portfolio are sunk. We make two simplifying assumptions. (i) Although $I$ knows the individual qualities $q_I$ of the patents that compose its portfolio, it does not know the breadth of the overall protection $b$ which it actually enjoys[11], i.e. $I$ does not know its own "type". (ii) We assume that $b$ (the aggregate patent breadth of $I$) is drawn from the same distribution as $q_E$ (the quality of the entrant's patent): $b \sim$U$[0, \bar{q}]$. The first assumption allows us to abstract from signalling aspects in the model. The second assumption enables us to restrict attention to linear payoff functions[12].

*The excess patent for the incumbent* is a particular case: $I$ gets an idea to patent a commercially and legally worthless idea ($q_I = 0$) at an implementation cost $c_I > 0$. Given that it is worthless, neither the aggregate portfolio strength $b$, nor the consumer valuation $v_I$ that $I$'s product generates is affected - only the size of $I$'s patent portfolio is increased from $N = n$ to $N = n + 1$. In equilibrium, $E$ is aware that the additional patent is worthless, however $E$ cannot distinguish it from valuable patents ex-ante (e.g. the worthless patent was filed simultaneously with the valuable patents[13]).

[11] This could be motivated by a story where design-arounds can occur in many different dimensions and $I$ does not know in which dimension to measure it.

[12] It would be more realistic to set $b = \max q_n$, $\forall n = \{1, .., N\}$ and assume it is taken from a Beta distribution of the last order statistic (or impose any other dependency between $b$ and the individual $q_I$). This would, however, complicate the analysis without yielding any additional insight as long as $b$ is exogenous to the model.

[13] The excess patent can also be thought of as the costly effort by the incumbent to split a single patent claim into multiple patents of smaller complementary claims. Other strategies that fit our model setting are discussed in section 1.6

### 1.2.3 Screening

By studying ("screening") all[14] its opponent's patents, $E$ can learn $b$, the strength of the existing patent portfolio of $I$. This gives $E$ an indication of the outcome of the litigation in case entry occurs. Per patent screening costs $s > 0$ are constant. The total screening costs $S$ thus depend linearly on the size $N$ of the patent portfolio to be screened: $S_N = s * N$.

*When not screening*, the entrant forms an expectation $\mathbb{E}[b]$ based on the common knowledge about the distribution that $b$ is drawn from. The costless estimate of the incumbent's patent strength is given by $\hat{\hat{b}} = \mathbb{E}[b] = \int_0^{\bar{q}} 1d\mathrm{U}(b) = \frac{\bar{q}}{2}$.

*When $E$ screens,* it delays its entry decision and observes a signal (message) $\hat{m} : \{b\,; q_E\} \rightarrow \{1\,; 0\}$ on how the courts will decide the infringement case. We assume that the action of screening is not observable by the incumbent, but verifiable by the courts ex-post[15]. Initially, we consider perfect screening: $E$ observes $m$ without imprecision.

$$\hat{m} = m = \begin{cases} 1 & \text{if } q_E > b \\ 0 & \text{if } q_E \leq b \end{cases}$$

where $m$ is the true information on the court decision. Therefore, payment of the screening fee erodes uncertainty about $I$'s patent portfolio strength and allows $E$ to take a perfectly informed entry decision.

A technical assumption is made to obtain that screening occurs with positive probability in the baseline: $S < \bar{S} = \Delta\left(1 - \frac{\Delta}{v_I + K}\right)$, where $\Delta = v_E - c_E$ represents the net benefit from non-infringing entry for $E$. The assumption signifies that screening is not

[14] For simplicity, we assume that $E$ needs to screen the whole portfolio of $I$ to learn $b$. Our results are robust to alternative settings as long as total screening costs increase in the number of patents screened: ($S'(N) > 0$). We could thus relax this assumption as follows: The entrant screens a sub sample of $I$'s patent portfolio to learn a signal $\hat{m}$. The precision of the signal depends on the proportion of the full portfolio that is screened. A larger patent portfolio is hence more costly to screen for $E$, because either the entrant screens a larger sub sample (thus paying more for screening) or the entrant does not increase the size of the screened sub sample, but accepts a decrease in the precision of the signal $\hat{m}$ (which increases the likelihood of screening errors).

[15] The first part of the assumption is realistic as patents are public and their study occurs in the absence of the patent holder's knowledge. This assumption is necessary (at least until the decision to litigate by $I$ is taken) to exclude the signal (of whether screening has occurred or not) to $I$ in this model since it would reveal information about the quality of $E$'s patent. The second part of the assumption corresponds to the ability of courts to find proofs for the study of competitor patents ex-post. In practice this could refer to "smoking gun" proofs (e.g. explicit email correspondence or whistle-blowers) in the wilfulness investigation by courts. We need this assumption for treble damages to have deterring effect from screening in equilibrium. We give more details on why this assumption matches the beliefs of US firms in section 1.4.

prohibitively costly. It implies that avoiding infringement by screening and then deciding not to enter is profitable for $E$: $\Delta - v_I - K < -S$.

## Mechanism, Payoffs and Timing

The channel of the effects goes through the screening of the entrant. By demanding an excess patent, $I$ increases the size of its patent portfolio and thereby increases the effort demanded from $E$ to screen its portfolio, if $E$ decides to do so. Due to the additional patent, screening costs of $E$ increase by a constant factor $\tau > 0$ from $S_n = s * n = S$ to $S_{n+1} = s * (n+1) = S(1+\tau)$. In other words, $I$ can use the excess patent as a raising rival's cost strategy. We repeat for clarity, that the excess patent does not affect the success probabilities in litigation[16].

We assume Bertrand competition between the firms. As mentioned, we have that $I$ generates a consumer valuation of $v_I$. The unit consumer has utility $u(v) = v - p$. We focus on the case where $E$'s entry marginally increases the consumer valuation by $0 < v_E < \bar{v_E} = \frac{v_I + K}{2} + c_E$. The upper bound on $v_E$ ensures that, ex-post, market entry is only profitable for $E$ when its design-around technology is deemed non-infringing by the courts. Ex-ante, entry without screening is only profitable when infringement is sufficiently unlikely: $q_E > \mathbb{E}[b]$.

We simplify notation as follows: $v_I = v, \ c_I = c$. The Incumbent earns $\Pi_I^M(v) = v$ if it produces alone. If only the entrant produces, then it earns $\Pi_E^M(v, \Delta) = v + \Delta$. Under a duopoly, the profits are $\Pi_I^D(v, \Delta) = 0$ and $\Pi_E^D(v, \Delta) = \Delta$. For entry without screening, the risk of single damages payments in case of infringement leads to the following ex-ante payoffs:

$$\Pi_E^l = \Delta - (v + K)\,\mathbb{P}\{b \geq q_E\}$$
$$\Pi_I^l = v - (v + K)\,\mathbb{P}\{b < q_E\}$$

In words, $E$ earns $\Delta$ from the sale of its product when entering the market, but has to pay damages $v$ and litigation costs $K$ in case of infringement, which occurs with probability $\mathbb{P}\{b \geq q_E\}$. Similarly, $I$ earns $v$ absent market entry, but loses these profits $v$ and litigation cost $K$ in case of successful entry, which occurs with probability $\mathbb{P}\{b < q_E\}$.

The timing is as given by the tree in figure 1.1 (cubic decision nodes belong to nature):

[16] We treat commercial and legal value (affecting the litigation probabilities) of the excess patent in the extension in section 1.5

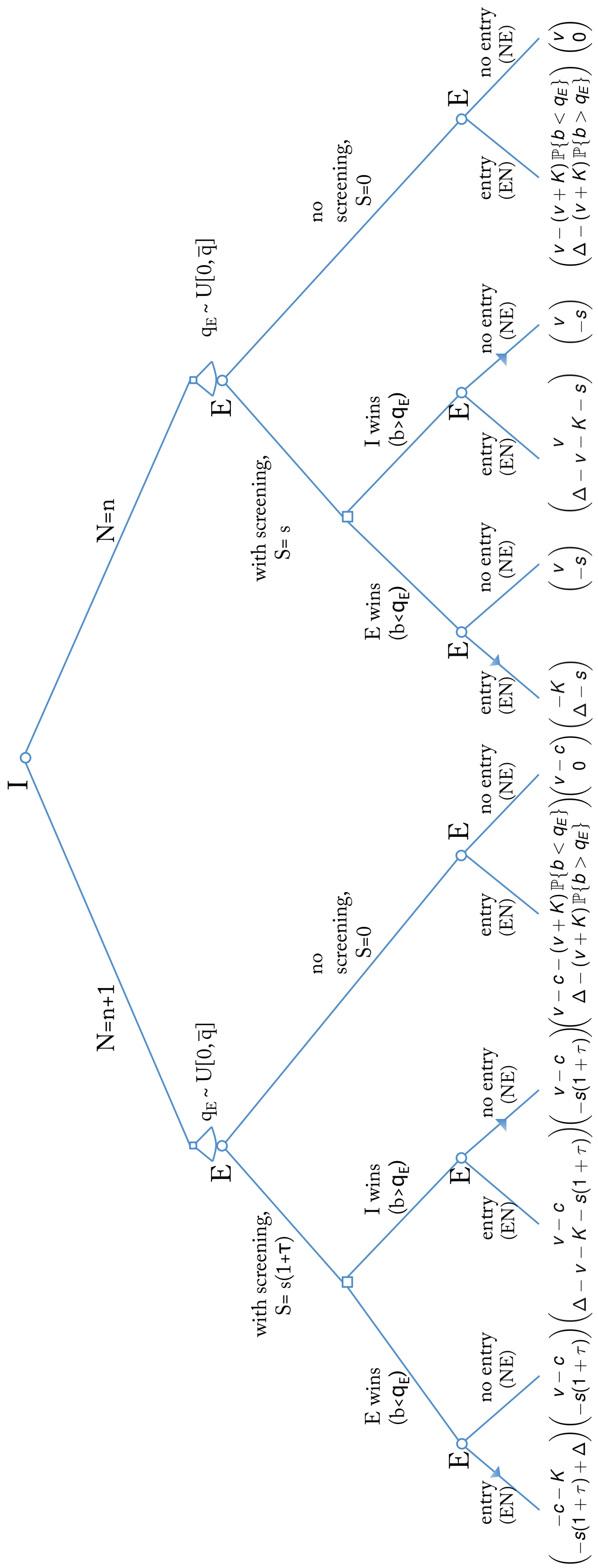



Figure 1.1: Full game tree

1. The incumbent $I$ has the option to demand the grant of a commercially and legally worthless patent. Thereby, only its portfolio size increases from $N = n$ to $N = n+1$.
2. The entrant $E$ obtains a design-around idea of quality $q_E$ for free.
3. $E$ decides whether to screen $I$'s patent portfolio to learn a signal $\hat{m}$ about the outcome of the litigation ($b \lesseqgtr q_E$).
4. $E$ decides whether to enter the market or not.
5. If market entry occurs, litigation arises. The court adjudicates whether infringement has occurred and, if so, redistributes profits.

In the game tree, the entry subgame starts at $t = 3$. We differentiate the entry subgame absent an excess patent ($N = n$, right-hand-side of the game tree) and the entry subgame given an excess patent ($N = n + 1$, left-hand-side of the game tree). In the entry subgame, $E$ has the option to choose a strategy $\sigma_E$ among the three pure strategies {(NS, NE),(NS, EN),(S, D)}, where the abbreviations stand for "Not Screen & Not Enter", "Not Screen & ENter" and "Screen & then Decide on entry" respectively.

## 1.3 Model Results

All proofs are given in the appendix, starting on page 37.

### 1.3.1 Under perfect screening

Assuming that the entrant can perfectly learn the strength of the opposing patent portfolio when screening and, hence, does not make errors when anticipating the court's rule, we obtain the following results.

**Lemma 1.3.1** (Optimal strategy for the entrant)
*In the entry subgame absent an excess patent ($N = n$), the entrant has an optimal, composite strategy ($\sigma_E^*$) conditional on the realisation of the quality of its idea ($q_E$).*

$$\sigma_E^* = \begin{cases} \textit{No Screening, No Entry (NS, NE)} & \textit{if } q_E \leq \underline{q_S} \\ \textit{Screening, then Deciding (S, D)} & \textit{if } \underline{q_S} < q_E \leq \underline{q_T} \\ \textit{No Screening, ENtering (NS, EN)} & \textit{if } \underline{q_T} < q_E \end{cases}$$

*where* $\underline{q_S} = \bar{q}\left(\frac{S}{\Delta}\right)$ *and* $\underline{q_T} = \bar{q}\left(\frac{\Delta - v - K + S}{\Delta - v - K}\right)$.

Figure 1.2 illustrates the composition of the strategy. The optimal strategy to $E$ is given by the upper envelope of the strategy payoffs. In the graph, the thresholds have an additional upper index of 0 to indicate that they result from the perfect screening specification.

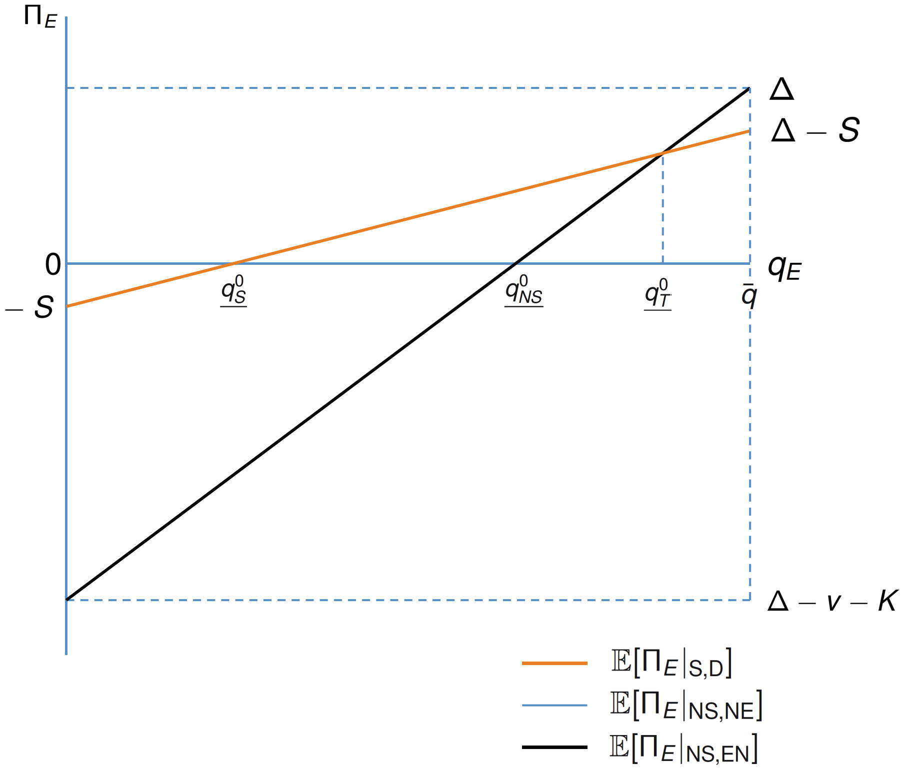



Figure 1.2: Expected payoffs to $E$ plotted against the quality of its idea $q_E$
*Note:* In order of increasing slope: The expected profit to $E$ for the strategy NS,NE is given by the blue horizontal line (coincides with the $q_E$-axis). Expected profits under S,D are given by the orange line of positive slope. Expected payoffs under NS,EN are given by the dark black diagonal of the highest slope.

For qualities less than $\underline{q_S}$, $E$ will stay out of the market without screening. For qualities above $\underline{q_T}$, $E$ will enter the market without screening. For intermediate qualities between $\underline{q_S}$ and $\underline{q_T}$, $E$ will screen $I$'s patent portfolio and enter only in case of no infringement (i.e. $q_E > b$) and stay out otherwise.

**Proposition 1.3.2** (More patents mean less screening)
*An excess patent by the incumbent increases screening costs for the entrant and reduces the range of qualities for which the entrant screens.*
*The optimal strategy for the entrant given an excess patent* $(N = n + 1)$ *is given by*

$$\sigma^*_{E'} = \begin{cases} NS,\ NE & \text{if } q_E \leq \underline{q_{S'}} \\ S,\ D & \text{if } \underline{q_{S'}} < q_E \leq \underline{q_{T'}} \\ NS,\ EN & \text{if } \underline{q_{T'}} < q_E \end{cases}$$

*where* $\underline{q_{S'}} = \bar{q}\left(\frac{S(1+\tau)}{\Delta}\right)$ *and* $\underline{q_{T'}} = \bar{q}\left(1 - \frac{S(1+\tau)}{|\Delta - v - K|}\right)$. *Note,* $\underline{q_{S'}} > \underline{q_S}$ *and* $\underline{q_{T'}} < \underline{q_T}$.

The impact of an excess patent is illustrated in figure 1.3. It becomes immediately visible that the induced increase in screening costs causes a constant downward shift of the expected payoff function under screening. Thereby, it impacts both the thresholds $\underline{q_S}$ and $\underline{q_T}$, which move closer together and reduce the range of qualities over which $E$ screens.

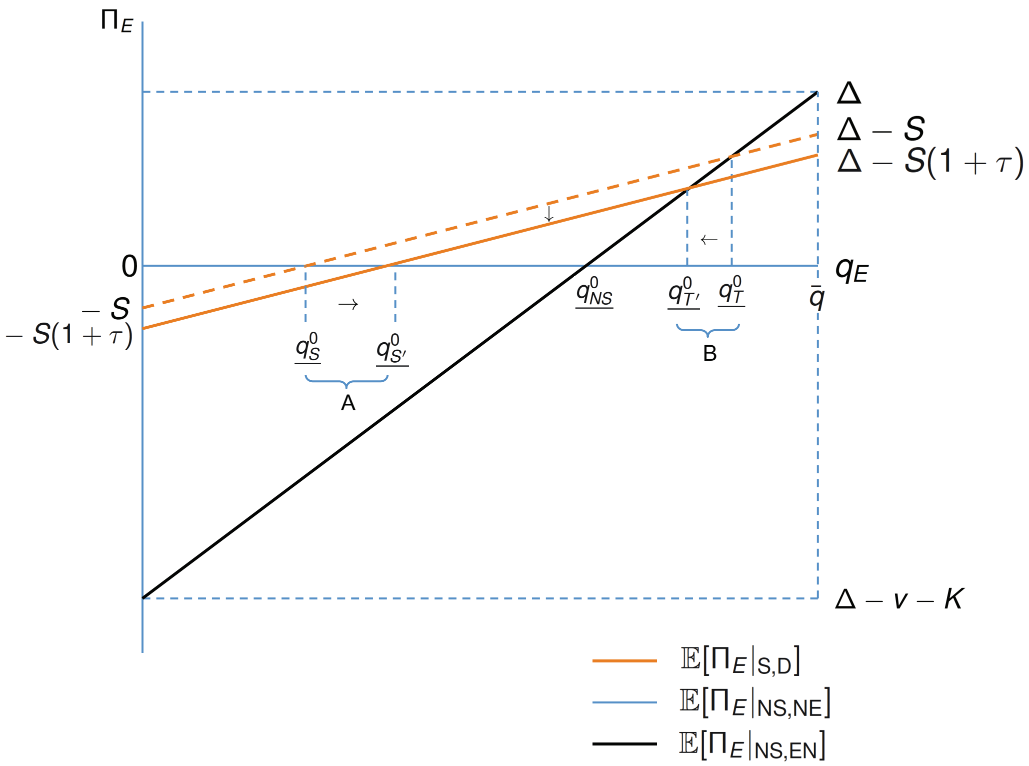



Figure 1.3: Impact of an excess patent on $E$.

*Note:* We depict the payoff function under S,D twice: The dark orange line refers to the payoffs under S,D given an excess patent. The dotted orange line is the payoff under S,D in the absence of an excess patent (identical to the dark orange line of figure 1.2). The dash on the lower index of thresholds indicates that these levels result from the situation with an excess patent.

In figure 1.3, $A$ represents the range over $q_E$ for which $E$ switches strategy from S,D to NS,NE, i.e. stays out instead of screening. $B$ represents the range where $E$ changes from S,D to NS,EN, i.e. from screening to entry without screening.

**Proposition 1.3.3** (Less screening means less entry)

*In the setting of a design-around, where the entrant's product generates only a marginal increase in consumer valuation, less screening translates into less entry:*

$$|A| > |B|$$

The graphical intuition behind proposition 1.3.3 is that A is measured directly against the $q_E$-axis ($= \mathbb{E}[\Pi_E|_{\text{NS,NE}}]$), whereas B is measured against $\mathbb{E}[\Pi_E|_{\text{NS,EN}}]$ of increasing slope. Translated onto the $q_E$-axis, B is smaller than A, because the difference in slopes between $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ and $\mathbb{E}[\Pi_E|_{\text{NS,EN}}]$ is sufficiently large when infringement is sufficiently expensive for $E$.

**Lemma 1.3.4** (Excess patenting incentives for the incumbent)
*The incentives for the incumbent to produce an excess patent for the purposes of raising rivals' costs of screening are given by:*

$$-c + \frac{S * \tau}{\Delta}\frac{v+K}{2} \geq 0$$

The condition is interpreted as follows: The cost $c$ of demanding an excess patent must be set off by the expected benefit $\frac{v+K}{2}$ that $I$ gets from reducing $E$'s screening incentives times the probability $\frac{S*\tau}{\Delta}$ that its strategy will have an impact on $E$. The first term results from avoiding litigation cost ($K$) and price competition ($v$) for $I$. The second term is the explicit expression of $\frac{A}{\bar{q}}$ and thus reflects the likelihood of meeting an entrant who will stay out due to increased screening costs. Last, we note that the European splitting for litigation costs has considerably simplified the analysis since the switch from strategy S,D to NS,EN by $E$ has no profit implication for $I$.

### 1.3.2 Under imperfect screening

By introducing screening errors, we move away from the assumption that the entrant can perfectly anticipate the court's decisions. This is more realistic since perfect screening is likely to be prohibitively costly in real life. This setting will be useful to analyse the treble damages doctrine since it yields the result that infringement occurs with positive probability in equilibrium.

The entrant now screens with an exogenous precision of $g < 1$. I.e. the signal $\hat{m}$ that $E$ observes on the patent portfolio strength of the incumbent is wrong with a probability $(1-g)$. Two types of errors thus arise for $E$. Type I errors refer to false positive signals ($\hat{m} = 1 \neq m = 0$). As a result, $E$ enters although its patent is too weak ($q_E \leq b$) and will lose in litigation. Type II errors refer to false negative signals ($\hat{m} = 0 \neq m = 1$). In this case, $E$ decides to stay out of the market even though its patent is strong enough

($q_E > b$) to be upheld valid if contested by $I$. We note that we have assumed that both types of errors occur with the same probability.

The screening precision is restrained to be sufficiently large ($g > \bar{g} = \max\{\frac{\Delta}{v+K}, \left(1 - \frac{\Delta}{v+K}\right)\}$) such that the dominant strategies in the entry subgame (conditional on the reading signal $\hat{m}$ of $I$'s patent portfolio strength) are unchanged[17]. Thus, the signal $\hat{m}$ remains informative even though errors occur. Again, we restrict attention to screening costs that are sufficiently small ($S < \bar{S}^{IS}$) in order to exclude prohibitively costly screening and to ensure that screening occurs with positive probability in equilibrium[18]. Consequently, $E$ uses screening for some positive range of values of $q_E$ and takes a misguided entry decision with the probability $(1 - g) > 0$.

**Lemma 1.3.5** (Costs of screening errors depend on their type)
*Imperfect screening introduces type I (false positive) and type II (false negative) errors in the entrant's entry decision post screening, where*

$$\mathbb{E}[\textit{Cost of a type I error}] > \mathbb{E}[\textit{Cost of a type II error}]$$

The result is driven by the assumption that the design-around only marginally improves the consumer valuation for the product. When $E$ enters while liable for infringement (type I error), it earns $\Delta$ from the sale of its product, but has to pay litigation costs ($K$) and damages ($v$). When $E$ stays out of the market when in fact no infringement would have been adjudicated by the courts, it merely foregoes profits of $\Delta$. Consequently, a type I error is more expensive than a type II error to $E$.

**Lemma 1.3.6** (Screening errors reduce the use of screening)
*In the entry subgame absent an excess patent ($N = n$), given screening errors, the entrant has an optimal, composite strategy ($\sigma_E^*$) conditional on the realisation of the quality of its idea ($q_E$).*

$$\sigma_E^{1*} = \begin{cases} NS,\ NE & \text{if } q_E \leq \underline{q}_S^1 \\ S,\ D & \text{if } \underline{q}_S^1 < q_E \leq \underline{q}_T^1 \\ NS,\ EN & \text{if } \underline{q}_T^1 < q_E \end{cases}$$

*where* $\underline{q}_S^1 = \frac{\bar{q}[S-(1-g)(\Delta-v-K)]}{g(\Delta)-(1-g)(\Delta-v-K)} > \underline{q}_S$ *and* $\underline{q}_T^1 = \frac{\bar{q}[S-(1-g)(\Delta-v-K)]}{g(\Delta)-(1-g)(\Delta-v-K)} < \underline{q}_T$. *I.e. the entrant reduces the range of qualities $q_E$ over which it screens before taking the entry decision.*

[17] For a full derivation of this assumption see the appendix on page 40.

[18] We extend the assumption used in the perfect screening setting. Precisely, $\bar{S}^{IS} = \left(\frac{\Delta-v-K}{v+K}\right)[-g(\Delta) + (1-g)(\Delta - v - K)] + (1+g)(\Delta - v - K)$. For a derivation of this assumption, see the appendix on page 42.

Screening errors impose an additional cost on the screening strategy for the entrant, similar to the excess patent analysed above. In line with the previous interpretation, a downward shift of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ occurs and the range over which the entrant screens decreases likewise.

However, an additional effect occurs due to the discrepancy between the costliness of the types of errors that can be made due to a false screening signal: Because false positive errors are more expensive than false negative errors (by lemma 1.3.5), the expected cost of screening is increased proportionally more for weaker entrants (low $q_E$). Therefore, the downward shift of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ is left-side heavy, i.e. the expected payoffs fall more for low levels of $q_E$ and thereby the slope of the payoff function under strategy S,D increases. The graph in figure 1.4 visualises this analysis.

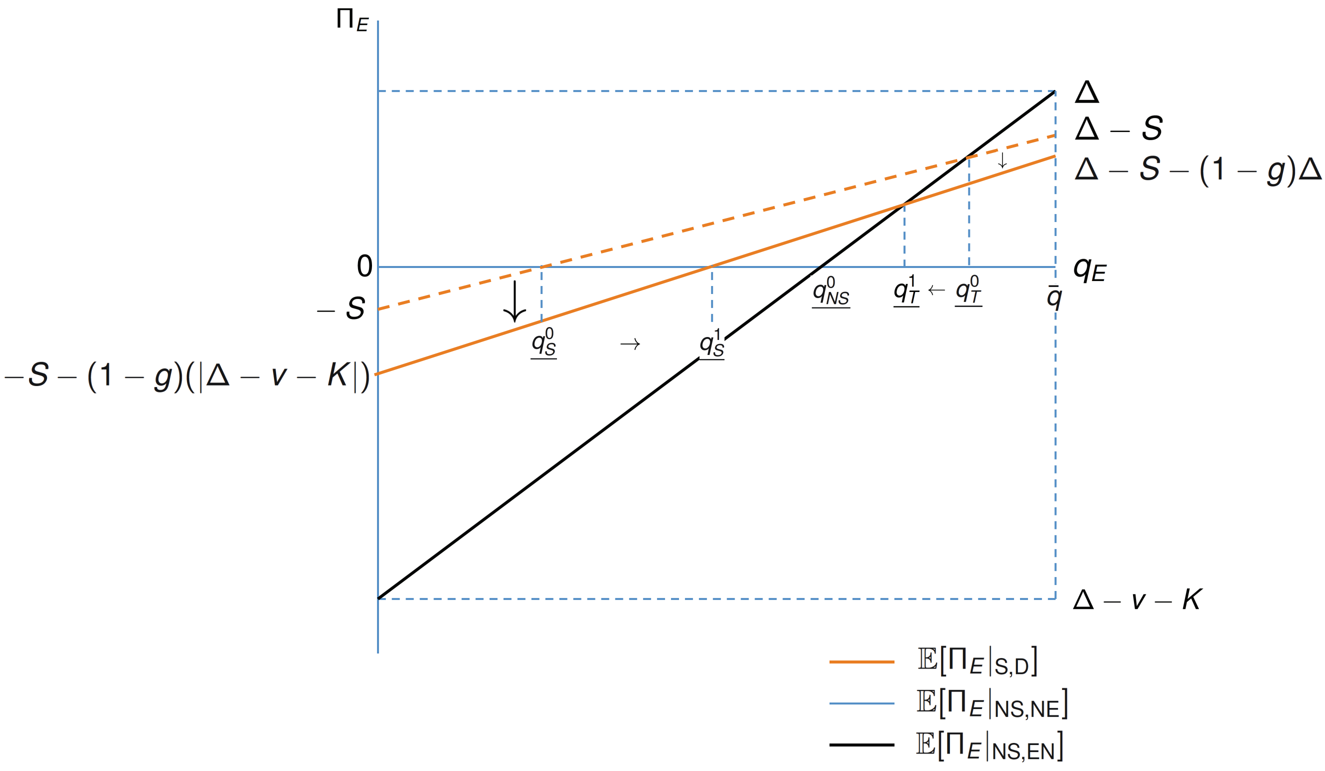



Figure 1.4: Impact of screening errors on $E$.

*Note:* The dark orange line represents payoffs under S,D given imperfect screening. The dotted orange line depicts payoffs under S,D given perfect screening (identical to the dark orange line of figure 1.2). Superscripts "1" indicate the threshold levels under imperfect screening. Superscripts "0" indicate the threshold levels under perfect screening.

**Corollary 1.3.7** (Less screening means less entry)

*Given imperfect screening, Proposition 1.3.3 holds true.*

The intuition behind corollary 1.3.7 is that excess patents and screening imperfections are substitutes since both factors represent an additional cost on screening. Proposition 1.3.3 showed that overall market entry is reduced when a constant cost is imposed on screening for entrants of all types $q_E$. The net reduction in market entry resulted from the fact that the reduction in entry of low quality entrants outweighed the entry increase from high quality entrants. It is thus not surprising that this result holds when the imposed screening costs decrease in $q_E$, i.e. affects weaker entrants more.

**Proposition 1.3.8** (Steeper slope reduces excess patenting incentives)
*Given imperfect screening, screening becomes more costly for the entrant and thus is used less by the entrant, resulting in less entry (Corollary 1.3.7). Consequently, the excess patenting strategy becomes less profitable for I, since*

$$A^1 < A$$

*where*[19] $A = \underline{q}_{S'} - \underline{q}_S$ *and* $A^1 = \underline{q}^1_{S'} - \underline{q}^1_S$ .
*The incentives to excessively patent for the incumbent are given by*

$$\left(\frac{gS\tau}{g\Delta - (1-g)(\Delta - v - K))}\right)\left[\frac{v+K}{2}\right] - c \geq 0,$$

*a condition which is more stringent than the one given in Lemma 1.3.4.*

Proposition 1.3.8 yields that the effect of an excess patent is reduced in a setting with imperfect screening as compared to a perfect screening setting. This is because the reduction in entry that the excess patent commands is reduced in a setting where the entrant makes screening errors.

The shifts (termed $A$ and $A^1$) of the intercept of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ with the $q_E$ axis determine by how much entry is reduced when the incumbent demands an excess patent (in the perfect screening and imperfect screening settings, respectively). As the slope of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ increases under imperfect screening, the downward shift of the function by a constant shift $S\tau$ results in a smaller shift of $\underline{q}_S$ to $\underline{q}_{S'}$ (given by the intersections of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ with the $q_E$ axis). Thus, the entry reduction due to the excess patent is diminished. Figure 1.5 visualises this effect[20].

[19] Recall that the upper index 1 refers to the effect due to screening errors and the dash ($'$) on the subscript refers to the effect of an excess patent.

[20] The increase in entry given by the shift of length $\text{B}^1$ of the threshold $\underline{q}^1_T$ to $\underline{q}^1_{T'}$ is neglected here in the text since under the considered European litigation cost regime, the strategy change from S,D to NS,EN has no profit implication for $I$. The results are qualitatively unaffected if a US cost regime is assumed. Robustness is shown in section 1.5.1.

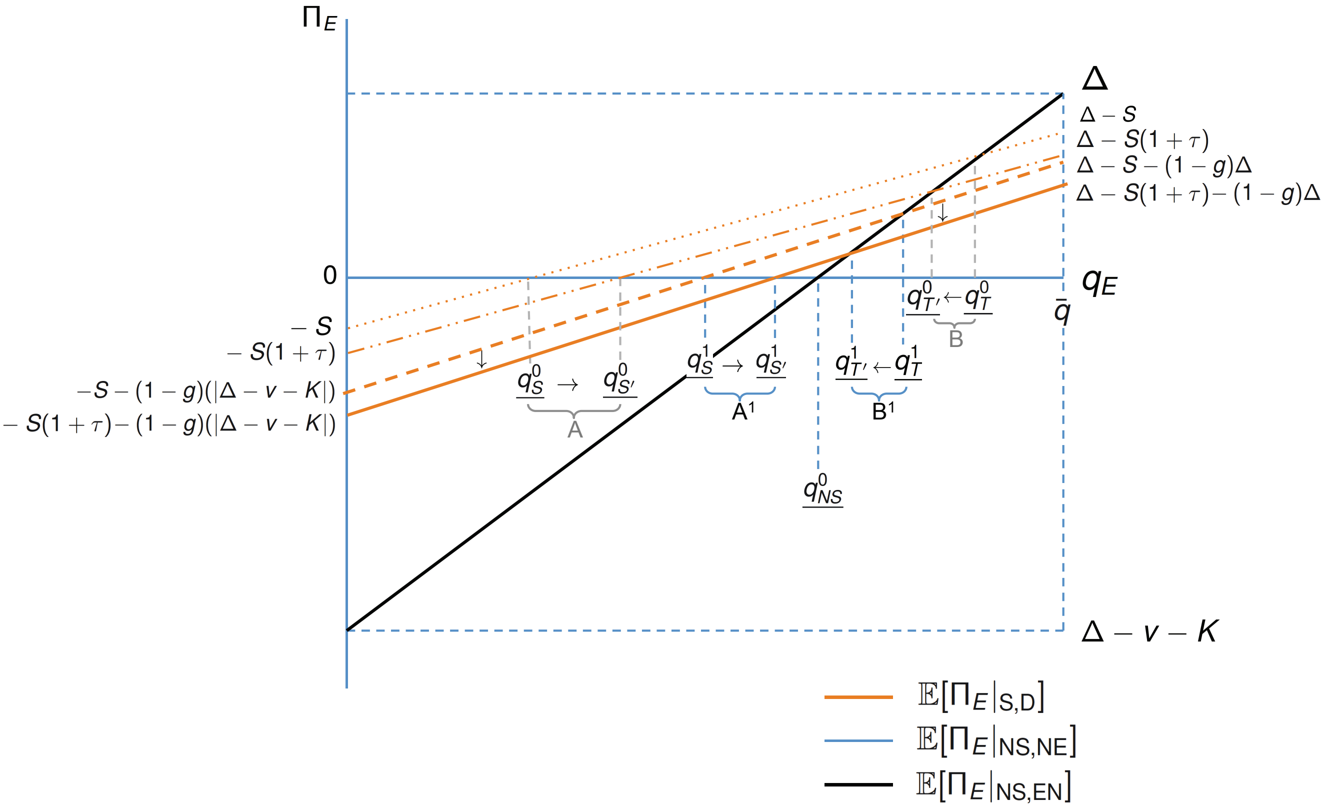



Figure 1.5: Steeper slope reduces effect of an excess patent.

*Note:* This graph adds the impact of an excess patent to the payoff functions from figure 1.4. The upper index 1 refers to the effect due to screening errors and the dash ($'$) on the subscript refers to the effect of an excess patent. From bottom to top (from solid to finely dotted), the four orange payoff functions for $E$ under S,D represent: (i) Payoff under imperfect screening given an excess patent, (ii) Payoff under imperfect screening absent an excess patent, (iii) Payoff under perfect screening given an excess patent, (iv) Payoff under perfect screening absent an excess patent.

The intuition behind this mechanism is that, when screening errors occur, the variability of the profits $\Pi_E$, that the entrant can earn, increases[21]. Given a larger range of expected profits (and consequently costs) for $E$, the same constant increase of a single cost factor loses importance in the overall entry decision. In other words, $E$ bases its entry decision not only on the screening costs, but also on the cost of (the newly introduced) screening errors. The constant change in screening costs due to an excess patent now makes up a smaller proportion of overall costs and, thus, the raising rival's cost measure of excessive patenting loses effectiveness.

[21]I.e. the range $\mathbb{E}[\Pi_E|_{\text{S,D}}]\big|_{q_E=\bar{q}} - \mathbb{E}[\Pi_E|_{\text{S,D}}]\big|_{q_E=0}$ increases.

## 1.4 Treble damages

The proposed model framework allows to analyse the effect of the treble damages (TD) doctrine. A quote[22] from the FTC (2003) motivates this application:

> [Pursuant to 35 U.S.C. §284,] a court may award up to three times the amount of damages for a defendant's wilful infringement of a patent - that is, the defendant knew about and infringed the patent without a reasonable basis for doing so[23]. Some Hearings' participants explained that they do not read their competitors' patents out of concern for such potential treble damage liability[24,25].

Furthermore, the FTC (2003) mentions that

> recent data suggests that courts enhance damages in a significant percentage of decisions that find infringement[26]. [Data] from the records of all patent cases tried from 1983 through 1999, show a finding of willfulness in 39% of the 888 decisions that found infringement and enhanced damage awards in 70% of the 219 cases in which judges considered enhancement issues.

Consequently, it attributes a "disproportionately large *in terrorem* effect" to the treble damages doctrine as testified by panelists.

### 1.4.1 Adapting the model set-up

In the model, the treble damages doctrine is translated as follows: When a firm screens and reads the wrong signal $\hat{m}$ on the outcome of the litigation process and enters, the courts will award treble damages to the incumbent as a compensation for the infringement. It is hence assumed that a screening error will be sanctioned with treble damages although entry is based on a screening error and not on a deliberate decision to enter. Courts can here not distinguish between the two types of mal-entry.

[22] Parts in square brackets [] have been sourced from a different part of the same source.

[23] [Ryco, Inc. v. Ag-Bag Corp., 857 F.2d 1418, 1428 (Fed. Cir. 1988): "The test is whether, under all the circumstances, a reasonable person would prudently conduct himself with any confidence that a court might hold the patent invalid or not infringed."]

[24] [Panelists expressed considerable dissatisfaction with a state of affairs that in effect exposes firms to greater potential damages for trying to learn if they are infringing any patents than if they keep themselves blissfully ignorant.]

[25] [Many firms discourage employees from reading patents out of fear of wilful infringement.]

[26] See Kimberly A. Moore, Judges, Juries, and Patent Cases An Empirical Peek Inside the Black Box, 11 FED. CIR. B.J. 209 (2001)

Without going into the practical, legal details on how courts know that $E$ screened[27], this application models the state of affairs of the US patent system as perceived by the firms in the market and described in the quotations from the FTC (2003). The provided quotations suggest that firms hold the belief that reading competitor patents bears the risk of treble damages adjudications and that they adjust their strategies accordingly. Seaman (2011) gives additional empirical support to the view that firms do not read out of fear for treble damages. He shows that although in theory the standard required for wilfulness findings is high[28], it seems to be much lower in practice, especially when cases are decided by a jury. We discuss his findings in more detail in section 1.6.

### 1.4.2 Model results under treble damages

**Lemma 1.4.1** *(Treble damages multiply the cost of type I errors only)*
*In a setting where screening is imperfect, the introduction of treble damages acts as a multiplier on the cost of type I errors (false positives) only. Consequently, the slope of the payoff function from screening increases as it forces a left-side-heavy translation of the expected payoff function under screening.*

Graphically, the introduction of treble damages forces an increase in slope of the screening payoff function similar to that described in figure 1.4. The difference is that the translation of the expected payoff function under treble damages shifts down the intercept of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ with the $\Pi_E$-axis, while it does not change the value of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$ at $q_E = \bar{q}$.

**Corollary 1.4.2** *(Treble damages reduce screening and thereby entry by E)*
*Treble damages reduce the range over which E screens due to the (left-side-heavy) downward shift of the screening function.*

The intuition behind corollary 1.4.2 is analogous to that of corollary 1.3.7, namely that the strategy of screening has become more expensive and is thus used less in equilibrium.

[27] In practice, courts may find proof for the study of competitor patents ex-post on the basis of *smoking gun* proof, e.g. explicit email correspondence or whistle-blowers.

[28] "Plaintiff must prove wilfulness by clear and convincing evidence. This is a higher degree of persuasion than is necessary to meet the preponderance of the evidence standard. Plaintiff proves wilful infringement if it shows that defendant (1) was aware of plaintiff's patent and (2) had no reasonable basis for reaching a good faith conclusion that its making, using or selling its device avoided infringing the patent. Plaintiff may also prove wilful infringement by proving that defendant did not exercise due care to determine whether or not it was infringing plaintiff's patent once the defendant had actual notice of plaintiff's patent. Infringement is not wilful and deliberate if defendant had a reasonable basis for believing that the patent is invalid or not infringed." Quote from the 5th Circuit Pattern Jury Instructions - Civil under Chairman Judge Martin C. Feldman, 2006.

In the present setting, the reduction of screening leads to less market entry by potential rivals.

**Corollary 1.4.3** *(Treble damages reduce excess patenting incentives for I)*
*The left-hand-side translation of the screening function increases its slope and thus diminishes the reduction of entry that an excess patent can command (intercept A on $q_E$ axis is reduced).*
*Furthermore, treble damages increase expected profits to I when E screens, thus further decreasing excess patenting incentives.*

Two effects come into play (even in the European litigation cost setting), when treble damages are introduced. This is because both shifts of the thresholds $\underline{q_S}$ and $\underline{q_T}$ have a profit implication for $I$ and are thus relevant. Both effects go in the same direction and reduce the excess patenting incentives of $I$. The first is analogous to the one analysed before, namely that the cost impact commanded by an excess patent is reduced. This effect arises from the steepened slope of $\mathbb{E}[\Pi_E|_{\text{S,D}}]$.

The second and new effect is that treble damages increase the expected profit of the incumbent when the entrant has screened. Hence, the shifting together of both the left and right hand side boundaries of the screening interval is relevant. These shifts reduce the range of qualities $q_E$ for which $E$ screens (which has now become more profitable to $I$). Therefore, the incentives for the incumbent to excessively patent are further reduced.

## 1.5 Extensions and robustness checks

### 1.5.1 US litigation cost setting

In the baseline model, an EU litigation cost splitting has been considered where the loser of a court case has to pay the total legal costs of both parties ($K^{EU}$). This simplifies considerably the terms of the calculations in the output, because then the switch by $E$ from strategy S,D to NS,EN (represented by the shift $B$ in the graphs) has no profit implication for $I$.

In the US system, each party has to bear its own litigation costs $K_i^{US}$, $\forall i = \{I, E\}$[29]. Inducing the strategy switch from S,D to NS,EN for $E$ is now costly to $I$ (litigation costs

[29]Note $K_i^{US} \neq \frac{K^{EU}}{2}$, i.e. the level of court costs is generally not the same in Europe and the US.

$K_I^{US}$ have to be paid) even when the entrant's patent is weaker than $I$'s patent ($q_E \leq b$) and $I$ wins in litigation. Under the strategy S,D, player $E$ would have stayed out of the market and saved $I$ the defence costs.

**Proposition 1.5.1** (Robustness against a US litigation cost setting)
*The model insights are robust to a US litigation cost setting. However, excess patenting incentives are lower compared to a European legal fees regime.*

The mathematical expressions change, however the qualitative insights from the model remain the same because the strategy considerations for $I$ based on reducing entry ($A$) dominate those based on fostering entry ($B$). This is the case for two reasons. First, the payoff impacts in region $A$ outweigh those in region $B$. In region $A$, $I$ benefits from foregoing litigation costs and competition ($K_I^{US} + v$) when inducing a strategy change from S,D to NS,NE by an $E$, who would have won the litigation. In $B$, $I$ only loses litigation costs ($K_I^{US}$), when $E$ changes strategy from S,D to NS,EN although $I$ wins the litigation process. Second, the region $A$ exceeds the region $B$ in size as given by proposition 1.3.3.

Although the general case in the US is that each party bears its own litigation costs, 35 U.S.C §285 allows for the recovery of reasonable attorney fees in exceptional cases. The baseline model analysed in the paper can hence be seen as the setting where recovery of attorney fees is presumed.

### 1.5.2 Valuable Patents

**Incumbent's patent value:** For simplicity, the incumbent's additional patent was assumed to be of zero commercial and legal quality. The only effect from $I$'s additional patent was an artificial increase of $I$'s portfolio size, which translated into a raise of the screening costs for the entrant. However, our model framework also allows to analyse the patenting incentives of both commercially and legally valuable patents.

**Proposition 1.5.2** (Patent value affects the incumbent's patenting constraint)
*The baseline model represents a lower-bound on the excessive patenting activity by incumbents. When additional patents bear commercial or legal added-value, the constraint to demand an additional patent for the incumbent is relaxed.*

The intuition of the results is the following. First, positive commercial value effectively reduces the implementation cost $c_I$ for the additional patent. Additional commercial value also increases the level of damages to be paid by $E$ in case of infringement. Thus, $E$'s expected profits from market entry are reduced for strategies NS,EN and S,D[30]. Therefore, the commercial value of the additional patent strengthens the entry-reducing effect from increasing the portfolio size.

Second, legal value increases the portfolio strength $b$ and thereby improves the probabilities for $I$ to win in case of litigation. This effect is entry-reducing as it makes market entry more expensive in expectation for $E$ and will induce some entrants' types to switch to less risky strategies (e.g. NS,EN to S,D in region $B$ and S,D to NS,NE in region $A$). Both types of switches thus reduce entry into $I$'s market.

Since both effects of commercial and legal value go in the same direction (namely reduce market entry and increase incentives for $I$ to demand an additional patent) and any valuable patent encompasses the increase in screening costs for the rival (because the portfolio size still increases for valuable patents), the baseline model can be viewed as a lower bound on the additional patenting activity that we observe in a market. More patents mean less entry is, hence, valid to a larger extend when additional patents by $I$ bear value. Furthermore, when additional patents have commercial or legal value, the raising rivals' cost aspect can become a by-product of the patent application.

**Entrant's patent value:** So far, the model assumes both commercial ($v_E$) and legal ($q_E$) value for the patent of the entrant. In order to obtain the results of the baseline model, the latter may take values in the interval $[0, \bar{q}]$, which corresponds to the full range.

Commercial value of the entrant's patent, however, was limited to the interval $v_E \in (0, \bar{v_E}]$ corresponding only to a marginal increase in the exogenous consumer valuation. The idea behind the technical assumption on $v_E$ is that entry should be unprofitable when infringement occurs. The sufficient condition for this is $\Delta < v + K$ rather than the one used so far: $2\Delta < v + K$, which is more stringent and results in a case restriction. When relaxing this condition to the sufficient one, the main results remain unchanged, but the intermediate steps are less clear cut[31].

**Proposition 1.5.3** (Larger commercial value of the design-around)
*When $\frac{v+K}{2} < \Delta < v+K$, the large commercial value that the entrant can generate reduces*

[30]Note that the latter is true only when screening is imperfect.
[31]In the setting of US litigation costs splitting, we cannot validate the result of proposition 1.3.3.

*its sensibility to a raising rivals' costs effort by the incumbent since the entrant relies less on screening.*

*Consequently, the model results are attenuated under larger $v_E$ and large $v_E$ may lead to the excess patenting constraint not being satisfied under a US litigation cost regime.*

It is logical that the higher the expected profit of entry is, the less important is the uncertainty about the competitor's patent portfolio strength and the less screening will be used. Therefore, we focus in our baseline model on the case when screening is an integral part of competition ($\Delta < \frac{v+K}{2}$) and we check robustness here (in proposition 1.5.3). The analysis loses its interest when $\Delta > v + K$ since the entry is profitable regardless of infringement. We neglect the last case. Overall, we take away that entrants with smaller commercial value of their innovation are more sensitive to excessive patenting by $I$.

### 1.5.3 Independent legal opinion

"Judges are more likely to find willfulness when the infringer does not present an attorney opinion as a defense" (Moore, 2004). This hints at the fact that in practice, firms can avoid findings of wilful infringement by securing a counselling letter by an independent legal counsel on the validity of the incumbent's patents. The practice in question is as follows (FTC, 2003):

> Other testimony indicated that when troublesome patents are identified, firms frequently seek to show due care and dissipate wilfulness concerns by securing opinion letters regarding invalidity or non- infringement from outside counsel[32]. Some testimony questioned the value of that practice and noted that attempts to inquire about or pierce the surface of opinion letters can raise thorny disputes over attorney-client privilege[33].

This strategy can be modelled as a method to circumvent award of treble damages, however further increases the costs of screening. I.e. it makes the constant downward shift of the screening payoff stronger and hence increases the effectiveness of an excess patent. Now, the treble damages doctrine does not have grip on wilful infringers anymore (if they

[32] [See Sung 2/8 (Patent Session) at 147 (a competent, independent legal opinion, even if incorrect, will usually help to rebut an allegation of wilful infringement).]

[33] [See, e.g, Thomas 10/25 at 155 (rather than getting quality advice from counsel... we're getting sort of pats on the back that, you might as well continue and here's your shield from triple damages), 177-78 (not suggesting that patent bar will cynically dish out any kind of opinion); Gambrell 10/25 at 169; Taylor 10/25 at 160.]

can disguise their wrongdoing by legal counsel), however it still has an externality on entrants since their screening costs have gone up in order to acquire additional legal advice. Screening and entry is thus still adversely affected by the treble damages doctrine.

Furthermore, the excess patenting strategy gains in effectiveness, since the steepening of the slope (volatility range that $E$ faces when entering) of the screening function is cancelled and replaced by an increase of the constant downward shift of the screening function due to a multiplication of the screening costs. In technical terms, $\tau$ is increased, while the slope $\frac{\partial \Pi_E|_{\mathrm{S,D}}}{\partial q_E}$ does not change.

Allowing firms to escape treble damages by obtaining legal counsel erodes the purpose of treble damages. However, Moore (2004) details some of the factors which reduce the use of legal opinion letters:

> There are two problems with disclosing attorney opinions during discovery. The first is that it gives the patentee a detailed blueprint as to likely defences early in the litigation process. The second is that the patentee can use the opinion against the defendant if at any point the defendant pursues a litigation strategy that differs from the opinion.

For the above reasons, it is not uncommon for firms to refrain from legal counselling and then the insights from the model application to treble damages hold. What is necessary for the model insights to be relevant is that there exists a positive probability of treble damages adjudication in case of infringement by the entrant.

## 1.6 Discussion of the results

The model has shown that the entrant uses screening for intermediate patent qualities. Screening reduces the costly uncertainty over the strength of the incumbent's patent portfolio. Imperfect screening makes unsurprisingly the screening incentives lower. However, the setting with imperfect screening provides the correct counterfactual to compare the setting with and without the treble damages doctrine since litigation must arise in equilibrium for treble damages to have an effect. It is clear from the model that for the setting of engineering around, the treble damages doctrine reduces the incentives of the incumbent to patent excessively. This is a good thing since it avoids society the wasteful spending of efforts on patenting worthless ideas as well as it reduces the strain on the patent office. However, the treble damages doctrine itself represents a burden on the entrant since it exacerbates the cost of type I (false positive) errors. Consequently, it can be seen as a substitute for the incumbent's excessive patenting strategy that reduces entry itself, but avoids wasteful activity by the incumbents. Policy should note these points.

The model implicitly assumes that entry post screening will result in the adjudication of treble damages. This is not perfectly accurate in real life, but the qualitative insights from the model remain valid as long as firms face a positive probability in equilibrium to be condemned for treble damages if they screen. Seaman (2011), Lemley and Tangri (2003) and Moore (2004) discuss the effects of the wilfulness doctrine pre and post the 2007 Seagate decision[34] and present the facts that treble damages have been awarded both before and after Seagate in a significant (and similar) proportion of litigation cases, with an especially high number of treble damages adjudications when the court cases were decided by a jury. This lends validity to the model as it suggests that treble damages apply in unpredictable situations[35] and supports the idea that firms shy away from activities sanctioned by treble damages where commonly cited defence tools (e.g. legal opinions on infringement or validity of prior art) have no statistically significant power in avoiding enhanced damages (Seaman, 2011).

In the same line of thought, the model can be seen as providing differentiated predictions based on the nature of an industry. For example in the pharmaceutical industry, patent claims can be very precise with the protection of a molecular formula, whereas in other industries the limits of claims are more blurry. If wilful infringement is more likely to be found in case of infringement in specific sectors, then the model results gain importance in that field.

When treble damages are granted, the court may in addition also award attorney's fees to the prevailing plaintiff (Moore, 2004). This is equivalent to passing a US court case under a European litigation costs splitting, equal to the main setting presented in this paper. This strengthens the results from the model since it increases the amount of the damages. The case without attorney's fee awards is compared in section 1.5.1, the qualitative insights of the model remain valid.

[34]by the Federal Circuit in the US which toughened the standard applied to determine a wilfulness finding, for details see *re Seagate Tech., LLC, 497 F.3d 1360, 1365*

[35]E.g. the following factors had no significant predictive ability of wilfulness findings in Seaman (2011): opinions of counsel, attempts to design around the patent, re-examination at the PTO, and bifurcation of wilfulness from liability at trial.

The model and its application have a strong validity in the context of technologies that rely on the bundling of patents from different stake holders. This fragmentation of patent rights is also known as *patent thickets*[36]. When each individual patent holder of a fragment of the composite technology owns IP protection for this fragment in the form of patents, then each of these right-holders may have an incentive to excessively patent. The accumulation of excess patents exacerbates the effect of the raising rival's cost argument as compared to the setting of a single incumbent. The insights of this model thus give an additional rationale for incumbent firms to invest in creating a dense web of overlapping patent rights à la Shapiro (2001).

In the context of multiple entrants, a reasoning à la Choi (1998) is applicable. In his paper, patent litigation serves as an information transmission mechanism (revealing the strength of an incumbent's patent portfolio) where the costs of transmission are borne by the first plaintiff. Other firms interested in the free information transmitted by the litigation process have a free-riding incentive to wait and let another entrant initiate costly litigation. The threshold for the first player to litigate is thus increased. In the current setting, it would apply by increasing the threshold to enter for the first entrant. Subsequent entrants would learn for free the quality of the incumbent's portfolio. Therefore, entry would occur after a longer time lag, but then all entrants with higher quality ideas than the incumbent would enter.

The current paper is particularly suited to explain excess patenting, however also applies to other situations, where a raising rival's cost strategy works by the mechanism of furthering information asymmetries. The basic considerations of a raising rival cost strategy are widely discussed in the economics literature, even their existence in patenting (Gilbert and Newbery, 1982). This paper sets itself apart from the previous literature in the sense that our set-up allows to investigate the impact of the treble damage doctrine and respects the private information character of patent strength. By introducing the novel screening mechanism on patent qualities, this paper is able to rationalise the excessive patenting of even commercially and legally worthless ideas. This is in contrast to Gilbert and Newbery (1982), where preemptive patenting steals the legal and commercial value of the entrant's idea in a patent race. Our setting does not need to rely on valuable preemptive patenting to reduce market entry, a result that seems to fit strongly the observed explosion of low quality patent filings at the patent office.

[36] For accounts on thickets, see Galasso and Schankerman (2010); Graevenitz et al. (2011).

Finally, it is useful to detail what is meant by "more patents" in the title of this paper when talking about reducing entry. The model insights apply to all raising rival's cost strategies which have an effect on the entrant through the channel of increasing screening costs. Next to a mere additional patent, the following strategies fit the model set-up: (i) Introducing ambiguity in patent claims (Chiang, 2012) and thereby blurring the boundaries of the claim, (ii) Citing irrelevant prior art (Means, 2013) and thereby blurring the limits of entrant's affirmative duty to read all relevant patents, (iii) Monitoring entry in order to warn and inform potential entrants of patent infringement (Crampes and Langinier, 2002) and thereby focussing their resources on the study of patents that they would otherwise not be aware of.

## 1.7 Conclusion

This is a first step looking at the incentives of firms to flood the market with patents and their interaction with treble damages as a policy tool. The predictions of this model are testable since both patent numbers and entry are easily observable. The screening decision, on which read intermediate results, is not observable, but the account by the FTC (2003) matches the results of the model.

The model is able to put the treble damages doctrine into an analysis that provides policy recommendations depending on the public goal. If a policy objective is the fostering of the mutual reading of competitor patents (e.g. if large positive externalities are to be expected from this activity), then the treble damages doctrine should be removed to re-incentivise firms to screen competitor patent portfolios. If, however, the reduction of excess patenting of worthless ideas by the industry is the goal (e.g. because the stress on the patent office shall be reduced), then the maintenance of the treble damages doctrine ensures reduced incentives for firms to so. This effect however comes at the expense of a reduced information-sharing ability of the patent system as entrants will read less.

# 1.8 Appendix: Proofs

**Proof of Lemma 1.3.1**

We solve the game by backwards induction. In the entry subgame, $E$ has a dominant strategy to play depending on the relative strength of its patent ($q_E$) against that of the incumbent ($b$).

First, look at t=4. Given $I$'s commitment to litigate, $E$ will enter iff: $\Pi_E \geq c_E$. For any given strategy, $E$'s ex-ante payoffs under litigation increase monotonely with the quality of its own patent: $\frac{\partial \Pi_E}{\partial q_E} > 0$. This relationship comes from the fact that the ex-ante probability of winning litigation $\mathbb{P}\{q_E > b\}$ increases with $q_E$. Hence, entry is defined by a threshold argument where $E$ realises all ideas above a certain threshold $\underline{q}$.

The outside option is always to "Not Screen, Not Enter", which yields payoff $\mathbb{E}_b[\Pi_E|_{\text{NS, NE}}] = 0$. Without screening, ex-ante profits to $E$ from entering are as follows:

$$\begin{aligned}
&\mathbb{E}_b[\Pi_E|_{\text{NS, EN}}] = \mathbb{E}_b[\mathbb{P}\{b < q_E\}(\Delta) + \mathbb{P}\{b \geq q_E\}(\Delta - v - K)] \\
&= \mathbb{E}_b[\Delta - (v + K)\,\mathbb{P}\{b \geq q_E\}] \;= \Delta - (v+K)\int_{q_E}^{\bar{q}} 1 d\mathrm{U}(b) \;= \Delta - (v+K)\left(\frac{\bar{q} - q_E}{\bar{q}}\right)
\end{aligned}$$

where $_{\text{NS, EN}}$ refers to the strategy of "Not Screening, then ENtering". $E$ enters the market when $\mathbb{E}_b[\Pi_E|_{\text{NS, EN}}] \geq 0$. In order words, $E$ realises all ideas of quality $q_E \geq \underline{q}_{\text{NS}}$, where $\underline{q}_{\text{NS}}$ is defined by: $\underline{q}_{\text{NS}} = \bar{q}\left(1 - \dfrac{\Delta}{v+K}\right)$.

Given screening efforts at a total cost $S$, $E$ observes the signal $\hat{m}$ and the message is precise $\hat{m} = m$. By screening, the entrant has delayed the entry decision and can now take this entry decision given perfect information about the outcome of the litigation. Ex-post screening, payoffs to $E$ are given by

$$\Pi_E|_{\text{S, D}} = \begin{cases} -S & \text{if } q_E \leq b \text{ and } E \text{ stays out} \quad \leftarrow \\ \Delta - v - K - S & \text{if } q_E \leq b \text{ and } E \text{ enters} \\ \Delta - S & \text{if } q_E > b \text{ and } E \text{ enters} \quad \leftarrow \\ -S & \text{if } q_E > b \text{ and } E \text{ stays out} \end{cases}$$

where the arrows indicate the dominant strategies to be played depending on whether the quality $q_E \gtrless b$ and where $_{\text{S, D}}$ refers to the strategy of "Screening, then Deciding". Ex-ante, before learning the true strength of $I$'s patent, $E$'s payoff is given by:

$$\mathop{\mathbb{E}}_{b}[\Pi_E|_{\text{S, D}}] = \mathop{\mathbb{E}}_{b}[\mathbb{P}\{b \geq q_E\}(-S) + \mathbb{P}\{q_E > b\}(\Delta - S)]$$
$$= \mathop{\mathbb{E}}_{b}[-S + \Delta\,\mathbb{P}\{q_E > b\}] \quad = -S + \Delta \int_0^{q_E} 1 d\mathrm{U}(b) \quad = -S + \Delta \left(\frac{q_E}{\bar{q}}\right)$$

Absent the strategy of entry without screening ($_{\text{NS,EN}}$), $E$ will thus screen and then decide whether to enter when $\mathbb{E}_b[\Pi|_{\text{S, D}}] \geq 0$. At equality, this defines the threshold $\underline{q_S}$ above which ideas will be realised: $\underline{q_S} = \bar{q}\left(\frac{S}{\Delta}\right)$.

Between the strategies $_{\text{S, D}}$ and $_{\text{NS, EN}}$, $E$ goes for the latter whenever $\mathbb{E}_b[\Pi_E|_{\text{NS, EN}}] \geq \mathbb{E}_b[\Pi_E|_{\text{S, D}}]$. This inequality defines at equality the threshold $\underline{q_T}$ above which ideas' qualities must lie for $E$ to realise them without screening:

$$\Delta - (v+K)\left(\frac{\bar{q} - q_E}{\bar{q}}\right) \geq -S + \Delta\left(\frac{q_E}{\bar{q}}\right) \qquad \Rightarrow \underline{q_T} = \bar{q}\left(\frac{\Delta - v - K + S}{\Delta - v - K}\right)$$

The thresholds are ordered in the following way: $0 < \underline{q}_S \leq \underline{q}_{NS} \leq \underline{q}_T < \bar{q}$. To show that the thresholds are in increasing order, two steps: First show that the single crossing property holds for any two of the payoff functions of $E$: $\mathbb{E}[\Pi_E|_{\text{S, D}}]$, $\mathbb{E}[\Pi_E|_{\text{NS, EN}}]$ and $\mathbb{E}[\Pi_E|_{\text{NS, NE}}]$. Second, find the condition for $\mathbb{E}[\Pi_E|_{\text{S, D}}]_{|q_E = q_{NS}} > 0$ to hold.

The derivatives with respect to the quality of the $E$'s idea are (both constant in $q_E$):

$$\frac{\partial\,\mathbb{E}[\Pi_E|_{\text{NS, EN}}]}{\partial q_E} = \frac{v+K}{\bar{q}}, \quad \frac{\partial\,\mathbb{E}[\Pi_E|_{\text{S, D}}]}{\partial q_E} = \frac{\Delta}{\bar{q}} \quad \Rightarrow \frac{\partial\,\mathbb{E}[\Pi_E|_{\text{NS, EN}}]}{\partial q_E} > \frac{\partial\,\mathbb{E}[\Pi_E|_{\text{S, D}}]}{\partial q_E}$$

because $v > \Delta$, thus the single crossing property is verified. For the ordering to be as requested above (and not the other way round), it suffices to show that $\mathbb{E}[\Pi_E|_{\text{S, D}}]_{q_E = \underline{q_T}} > 0$. Simplifying, it yields the requirement that the screening cost $S$ must be sufficiently small for screening to occur with positive probability in the baseline equilibrium. The precise condition is given by: $S < \bar{S} = \Delta(1 - \frac{\Delta}{v+K})$.

Hence in $t = 3$, $E$ will use a different strategy conditional on the quality $q_E$ of its idea.

$$\sigma_E^* = \begin{cases} \text{No Screening, No Entry} & \text{if } q_E \leq \underline{q}_S \\ \text{Screening, then Deciding} & \text{if } \underline{q}_S < q_E \leq \underline{q}_T \\ \text{No Screening, ENtering} & \text{if } \underline{q}_T < q_E \end{cases}$$

Thus, screening of competitor patents will occur only for mediocre quality ideas (not the best, nor the worst ones). □

**Proof of Proposition 1.3.2**

Given the excess patent, the ex-ante ($t = 3$) expected payoff to $E$ from the strategy (S,D)

is altered to :
$$\mathop{\mathbb{E}}_{b}[\Pi_E|_{\text{S, D}},\text{ n+1}] = -S(1+\tau) + \Delta\left(\frac{q_E}{\bar{q}}\right)$$
which corresponds to a constant, linear downward shift of the curve by $-S\tau$ as shown in figure 1.3. The payoffs of the strategies NS,NE and NS,EN are unaffected.

As becomes immediately visible from the graph, the induced increase in screening costs has an impact on both thresholds $\underline{q_S}$ and $\underline{q_T}$, which move closer together. The new thresholds (indicated by the dash on the subscript) are given by

$$\underline{q_{S'}} = \bar{q}\left(\frac{S(1+\tau)}{\Delta}\right) > \bar{q}\left(\frac{S}{\Delta}\right) = \underline{q_S}$$
$$\text{and} \quad \underline{q_{T'}} = \bar{q}\left(1 - \frac{S(1+\tau)}{|\Delta - v - K|}\right) < \bar{q}\left(1 - \frac{S}{|\Delta - v - K|}\right) = \underline{q_T}$$

The new optimal composite strategy for $E$, given an excess patent by the incumbent is:

$$\sigma^*_{E'} = \begin{cases} \text{NS, NE} & \text{if } q_E \leq \underline{q_{S'}} \\ \text{S, D} & \text{if } \underline{q_{S'}} < q_E \leq \underline{q_{T'}} \\ \text{NS, EN} & \text{if } \underline{q_{T'}} < q_E \end{cases}$$

The differences $|\underline{q_{S'}} - \underline{q_S}| = \bar{q}\left(\frac{S\tau}{\Delta}\right) = A$ and $|\underline{q_{T'}} - \underline{q_T}| = \bar{q}\left(\frac{S\tau}{|\Delta-v-K|}\right) = B$ represent by how much screening decreases given the excess patent. The assumptions $\Delta > 0$ and $|\Delta - v - K| > 0$ ensure that both $A$ and $B$ are positive, hence yielding that more patents mean less screening. □

**Proof of Proposition 1.3.3**

From Proposition 1.3.2, we have expressions for $A$ and $B$. We have $|A| > |B|$:

$$\Leftrightarrow \frac{\bar{q}S\tau}{\Delta} - \frac{\bar{q}S\tau}{|\Delta - v - K|} > 0 \Leftrightarrow \bar{q}S\tau\left[\frac{-(2\Delta - v - K)}{(\Delta)(|\Delta - v - K|)}\right] > 0 \Leftrightarrow |\Delta - v - K| > \Delta \quad (1.1)$$

which is always satisfied by assumption since we restrict our analysis to design-arounds yielding only marginal increases in consumer valuation, $v+K > 2\Delta$. Hence, less screening is synonymous with less entry. □

**Proof of Lemma 1.3.4**

Given the impact of an excess patent on $E$'s strategy, it is now possible to analyse the excess patenting incentives of the incumbent[37]. The incumbent has two strategies, namely to produce an excess patent (ExP) or not (NP), $\sigma_I = \{\text{ExP, NP}\}$. Depending on the strategy that $I$ plays, $E$ will be confronted to a patent portfolio of size $N = n + 1$ or $N = n$, respectively.

[37] This part simplifies due to the European setting for litigation costs. The consequence is that the switch by $E$ from strategy S,D to NS, EN has no profit implication for $I$.

Without the excess patent, $I$'s payoff is:

$$\mathbb{E}[\Pi_I|\sigma_E^*, n] = \begin{cases} v & \text{if } q_E \leq \underline{q_S} \\ \mathbb{P}\{b \geq q_E\}(v) - K\,\mathbb{P}\{b < q_e\} & \text{if } \underline{q_S} < q_E \leq \underline{q_T} \\ \mathbb{P}\{b \geq q_E\}(v) - K\,\mathbb{P}\{b < q_e\} & \text{if } \underline{q_T} < q_E \end{cases}$$

An excess patent affects the ranges over $q_E$ for which the $E$ plays a different strategy (according to $\sigma_E^*$). It becomes clear that under the EU litigation cost setting, only the variation of $\underline{q}_S$ by a distance A, has an impact on $I$'s profits. Therefore, it is profitable for $I$ to develop the excess patent if the following condition holds (the profitable region of $q_E$ is the following: $\underline{q}_S < q_E < \underline{q}_{S'}$) :

$$\mathop{\mathbb{E}}_{b}\mathop{\mathbb{E}}_{q_E}\left[-c + \mathbb{P}\{\underline{q}_S < q_E < \underline{q}_{S'}\}[v - v\,\mathbb{P}\{b \geq q_E\} + K\,\mathbb{P}\{b < q_E\}]\right] \geq 0$$

$$\Leftrightarrow \mathop{\mathbb{E}}_{b}\mathop{\mathbb{E}}_{q_E}\left[-c + \frac{A}{\bar{q}}(v+K)\,\mathbb{P}\{b < q_E\}\right] \geq 0$$

$$\Leftrightarrow -c + \frac{S\tau}{\Delta}\frac{(v+K)}{2} \geq 0 \tag{1.2}$$

For completeness: The full game can now be solved for the Nash equilibrium in the baseline scenario, when no screening errors are made by $E$. Assuming that the incentive constraint for excess patenting in equation 1.2 is satisfied, the NE$(\sigma_I^*, \sigma_E^*)$ of the full game under perfect screening is NE$[_{\text{ExP}}, \sigma_{E'}^*]$. If the excess patenting condition is not satisfied, then the NE of the full game is the following: NE$[_{\text{NP}}, \sigma_E^*]$. We focus on the former out of interest. □

**Proof of Lemma 1.3.5**

Note first that both type I (false positives) and type II (false negatives) errors can occur with the same probability $(1-g) = \mathbb{P}\{b \geq q_E|\hat{b} < q_E\} = \mathbb{P}\{b < q_E|\hat{b} \geq q_E\}$.

**Assumption 1.8.1** *(Condition on the likelihood of screening errors)*
*The screening errors occur with probability $(1-g)$ in both directions, i.e. for both false positive and false negative signals. The condition on the probability of "no error":*

$$g \geq \bar{g} = \max\left\{1 - \frac{\Delta}{v+K}, \frac{\Delta}{v+K}\right\}$$

*ensures that the optimal screening strategies for the entrant in the screening subgame remain unchanged, conditional on reading the signal $\hat{m}$.*

**Derivation of Assumption 1.8.1** In the screening subgame post screening, for $_{\text{EN}}$ to remain optimal when observing signal $\hat{m} = 1 = [b < q_E]$, need:

$$g(\Delta - S) + (1-g)(\Delta - v - K - S) > -S \quad \Leftrightarrow \Delta - (v+K)(1-g) > 0$$
$$\therefore g > \bar{g} = \frac{\Delta - v - K}{-v - K}$$

In the entry subgame post screening, for $_{\text{NE}}$ to remain optimal when observing signal $\hat{m} = [b > q_E]$, we need:

$$-S > g(\Delta - v - K - S) + (1-g)(\Delta - S) \quad \Leftrightarrow \Delta - g(v+K) < 0$$
$$\therefore g > \bar{g} = \frac{\Delta}{v+K}$$

Thus, screening errors must be sufficiently small for equilibrium strategies to remain unchanged. This also implies that $g > \frac{1}{2}$ and, in other words, means that the signal must be informative for screening to be profitable. End of derivation.

The imperfect screening affects the expected payoff of $E$ when it screens.

$$\begin{aligned}
\mathbb{E}[\Pi_E|_{\text{S,D}}] = & \,\mathbb{E}\big[\mathbb{P}\{b \geq q_E\}[\mathbb{P}\{b \geq q_E|\hat{b} \geq q_E\}(-S)] \\
& + [\mathbb{P}\{b \geq q_E|\hat{b} < q_E\}(\Delta - v - K - S)] \quad \leftarrow \text{Type I error} \\
& + \mathbb{P}\{b < q_E\}[\mathbb{P}\{b < q_E|\hat{b} < q_E\}(\Delta - S) \\
& + \mathbb{P}\{b < q_E|\hat{b} \geq q_E\}(-S)]\big] \quad \leftarrow \text{Type II error} \\
& = -S + (1-g)(\Delta - v - K)\left(\frac{\bar{q} - q_E}{\bar{q}}\right) + (g)(\Delta)\left(\frac{q_E}{\bar{q}}\right)
\end{aligned}$$

The costs of both types of errors are given by:

$$\begin{aligned}
C_I &= \mathbb{E}[\text{Cost of type I error}] = (1-g)(\Delta - v - K) \\
C_{II} &= \mathbb{E}[\text{Cost of type II error}] = (1-g)\Delta
\end{aligned}$$

The cost of a type I error exceeds that of a type II error if $v + K > 2\Delta$, which is satisfied by assumption: $\Rightarrow C_I > C_{II}$. Furthermore, the slope of the expected payoff function under screening increases from $\frac{\partial\,\mathbb{E}[\Pi_E|_{\text{S,D}}]}{\partial q_E} = \frac{\Delta}{\bar{q}}$ under perfect screening ($g = 1$) to

$$\frac{\partial\,\mathbb{E}[\Pi_E|_{\text{S,D}}]}{\partial q_E} = \frac{1}{\bar{q}}\left[g\Delta + (1-g)(\underbrace{|\Delta - v - K|}_{>\Delta})\right]$$

under imperfect ($g < 1$) screening. $\square$

**Proof of Lemma 1.3.6**

For $E$, only the screening payoff is affected by the screening errors. Strategies $_{\text{NS,EN}}$ and $_{\text{NS,NE}}$ remain unaffected. Consequently, the threshold $\underline{q_{NS}}$ (intersection of the latter two functions) is also unaffected. The new thresholds, that define the range (over $q_E$) for which $E$ screens, are given by $\underline{q_S^1}$ and $\underline{q_T^1}$.

$\underline{q_S^1}$ is defined by the intersection of $\mathbb{E}[\Pi_E|_{\text{NS,NE}}]$ and $\mathbb{E}[\Pi_E|_{\text{S,D}}]$:

$$-S + (1-g)(\Delta - v - K)\left(\frac{\bar{q} - \underline{q_S^1}}{\bar{q}}\right) + (g)(\Delta)\left(\frac{\underline{q_S^1}}{\bar{q}}\right) = 0$$

$$\text{Reordering yields the threshold: } \Rightarrow \underline{q_S^1} = \frac{\bar{q}\,[S - (1-g)(\Delta - v - K)]}{g(\Delta) - (1-g)(\Delta - v - K)}$$

$\underline{q_T^1}$ is defined by the intersection of $\mathbb{E}[\Pi_E|_{\text{NS,EN}}]$ and $\mathbb{E}[\Pi_E|_{\text{S,D}}]$:

$$-S + (1-g)(\Delta - v - K)\left(\frac{\bar{q} - \underline{q_T^1}}{\bar{q}}\right) + (g)(\Delta)\left(\frac{\underline{q_T^1}}{\bar{q}}\right) = \Delta - (v+K)\left(\frac{\bar{q} - \underline{q_T^1}}{\bar{q}}\right)$$

$$\text{Reordering yields the threshold: } \Rightarrow \underline{q_T^1} = \frac{\bar{q}\,[S - (1-g)(\Delta - v - K)]}{g(\Delta) - (1-g)(\Delta - v - K)}$$

Both thresholds $\underline{q_S^1}$ and $\underline{q_T^1}$ are in the same ordering as before ($0 < \underline{q_S^1} < \underline{q_{NS}} < \underline{q_T^1} < \bar{q}$), when screening occurs with positive probability in equilibrium.

**Assumption 1.8.2** *(Screening occurs with positive probability)*
*Screening occurs with positive probability in equilibrium when screening costs are sufficiently low: $S < \bar{S}$. In the case of imperfect screening, the precise condition is given by:*

$$\bar{S} = \left(\frac{\Delta - v - K}{v + K}\right)[-g(\Delta) + (1-g)(\Delta - v - K)] + (1+g)(\Delta - v - K)$$

**Derivation of Assumption 1.8.2** For screening to occur in equilibrium, $E$ must be able to generate positive profits by screening when it obtains a quality $q_E = \underline{q_{NS}}$, where $\underline{q_{NS}}$ is the indifference threshold between entering without screening and staying out: $\mathbb{E}[\Pi_E|_{\text{S,D}}]\Big|_{q_E = \underline{q_{NS}}} > 0 \Leftrightarrow -S + (1-g)(\Delta - v - K)\left(\frac{\bar{q} + \bar{q}\frac{\Delta - v - K}{v+K}}{\bar{q}}\right) + g(\Delta\left(\frac{-\bar{q}\frac{\Delta - v - K}{v+K}}{\bar{q}}\right) > 0$
Simplifying at equality yields the threshold cost of screening $\bar{S}$. End of derivation.

The optimal, composite strategy to $E$ is given by :

$$\sigma_E^{1*} = \begin{cases} \text{NS, NE} & \text{if } q_E \leq \underline{q_S^1} \\ \text{S, D} & \text{if } \underline{q_S^1} < q_E \leq \underline{q_T^1} \\ \text{NS, EN} & \text{if } \underline{q_T^1} < q_E \end{cases}$$

where the definitions of the thresholds are given above. The shifts of the thresholds are given by the following distances (Note: the upper index 1 refers to the effect due to screening errors):

$$\underline{q}_S^1 - \underline{q}_S = \frac{\bar{q}\left[S-(1-g)(\Delta-v-K)\right]}{g(\Delta)-(1-g)(\Delta-v-K)} - \left(\frac{\bar{q}S}{\Delta}\right) > 0$$
$$\underline{q}_T^1 - \underline{q}_T = \left(\frac{\bar{q}\left[S-(1-g)(\Delta-v-K)\right]}{g(\Delta)-(1-g)(\Delta-v-K)}\right) - \bar{q}\left(\frac{\Delta-v-K+S}{\Delta-v-K}\right) < 0$$

Both thresholds $\underline{q}_S$ and $\underline{q}_T$ move closer together, hence the range of qualities for which $E$ will screen is reduced. $\square$

**Proof of Corollary 1.3.7**

Threshold $\underline{q_S}$ moves to the right to $\underline{q_S^1}$, a shift of distance $\underline{q}_S^1 - \underline{q}_S$, which represents a decrease in entry since firms who screened before, now stay out of the market without screening.

Threshold $\underline{q_T}$ moves to the left to $\underline{q_T^1}$, a shift by distance $\underline{q}_T^1 - \underline{q}_T$. This shift represents an increase in the market since firms who previously screened (and possibly stayed out after reading a signal of $I$'s patent portfolio strength), now enter the market without screening.

The net effect of increased and decreased entry is an entry reduction if :

$$|\underline{q}_S^1 - \underline{q}_S| > |\underline{q}_T^1 - \underline{q}_T|$$
$$\Leftrightarrow \frac{(g-1)\bar{q}\left(-\Delta^2+\Delta(K-2S+v)+S(K+v)\right)}{(-\Delta+K+v)(\Delta+g(-2\Delta+K+v))} + \frac{\bar{q}((g-1)K+(g-1)(v-\Delta)-S)}{\Delta-2\Delta g+(g-1)(K+v)} > 0$$

which is always true[38] given the assumptions made so far on screening errors (Assumption 1.8.1) and screening costs (Assumption 1.8.2) in addition to those made in the baseline model of perfect screening. $\square$

**Proof of Proposition 1.3.8**

Visually, as before, the excess patent results in a constant downward shift of the screening payoff function. The threshold by which entry reduction is affected is given by $A^1 = \underline{q}_{S'}^1 - \underline{q}_S^1$. This change in $E$'s strategy is beneficial to $I$ in terms of profits. On the other hand, the entry increase, due to entrants who enter without screening instead of screening first (given by $B^1 = \underline{q}_T^1 - \underline{q}_{T'}^1$), is without any profit implication for $I$ since litigation costs are borne by the loser of the case.

The profits to I given $E$'s optimal strategy $\sigma_E^*$ are:

[38] Mathematica file available upon request.

$$\mathbb{E}[\Pi_I|\sigma_E^*, n] = \begin{cases} v & \text{if } q_E \leq \underline{q}_S^1 \\ \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(v)] & \text{if } \underline{q}_S^1 < q_E \leq \underline{q}_T^1 \\ \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(v)] & \text{if } \underline{q}_T^1 < q_E \end{cases}$$

The intercept on the $q_E$-axis that an excess patent generates determines how profitable such a move is. In the case of imperfect screening, the slope of the screening function increases, because of the cost discrepancy between type I and II errors. We see that:

$$\begin{aligned} A^1 = \underline{q}_{S'}^1 - \underline{q}_S^1 &= \frac{\bar{q}\,[S(1+\tau) - (1-g)(\Delta - v - K)]}{g(\Delta) - (1-g)(\Delta - v - K)} - \frac{\bar{q}\,[S - (1-g)(\Delta - v - K)]}{g(\Delta) - (1-g)(\Delta - v - K)} \\ &= \frac{\bar{q}S\tau}{g(\Delta) - (1-g)(\Delta - v - K)} \qquad < \quad \frac{\bar{q}S\tau}{\Delta} = A \end{aligned}$$

Thus the entry reduction that can be obtained by an excess patent has decreased. The full expression capturing the incentives to excessively patent for $I$ is given by

$$\mathop{\mathbb{E}}_{b}\mathop{\mathbb{E}}_{q_E}\left[-c + \mathbb{P}\{\underline{q}_S^1 < q_E < \underline{q}_{S'}^1\}[v - \mathbb{P}\{b \geq q_E\}(v) - \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(v)]]\right] \geq 0$$

$$\Leftrightarrow -c + \mathop{\mathbb{E}}_{q_E}[\mathbb{P}\{A\}][v - \frac{v}{2} - \frac{1}{2}[g(-K) + (1-g)(v)]] \geq 0 \quad \Leftrightarrow \mathop{\mathbb{E}}_{q_E}[\mathbb{P}\{A\}][g\frac{(v+K)}{2}] - c \geq 0$$

$$\Leftrightarrow \frac{gS\tau}{g(\Delta) - (1-g)(\Delta - v - K)}\left[\frac{v+K}{2}\right] - c \geq 0 \tag{1.3}$$

We verify that the incentive constraint in (1.3) is more stringent than the one of Lemma 1.3.4. □

**Proof of Lemma 1.4.1**

Introducing treble damages, the expected costs of errors are given by:

$C_I^{\mathrm{TD}} = \mathbb{E}[\text{Cost of a type I error}] = (1-g)(\Delta - 3v - K)$

$C_{II}^{\mathrm{TD}} = \mathbb{E}[\text{Cost of a type II error}] = (1-g)(\Delta) \qquad \Rightarrow C_I^{\mathrm{TD}} \gg C_{II}^{\mathrm{TD}}$

Comparing to the single damages setting, we have $C_I^{\mathrm{TD}} > C_I > C_{II} = C_{II}^{\mathrm{TD}}$. By the same argument as in the proof of Lemma 1.3.5, we have an increase of the slope $\frac{\partial \mathbb{E}[\Pi_E|\mathrm{S,D}]}{\partial q_E}$. □

**Proof of Corollary 1.4.2**

Identical argument as in the proof of Lemma 1.3.6 and Corollary 1.3.7. □

**Proof of Corollary 1.4.3**

Identical argument as in the proof of Proposition 1.3.8, where the expected costs of screening errors are as given in the proof of Lemma 1.4.1. Furthermore, treble damages have a profit implication for $I$. The expected profits to I given $E$'s optimal strategy $\sigma_E^*$ are:

$$\mathbb{E}[\Pi_I|\sigma_E^*, n] = \begin{cases} v & \text{if } q_E \leq \underline{q}_S^2 \\ \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(3v)] & \text{if } \underline{q}_S^2 < q_E \leq \underline{q}_T^2 \\ \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(v)] & \text{if } \underline{q}_T^2 < q_E \end{cases}$$

Therefore, now both of the shifts of the thresholds $A_{TD}^2 = \underline{q}_{S'}^2 - \underline{q}_S^2$ and $B_{TD}^2 = \underline{q}_T^2 - \underline{q}_{T'}^2$ affect $I$'s incentives to excessively patent.

$B_{TD}^2$ represents a shift from S,D to NS,EN, which is payoff costly to $I$.

$A_{TD}^2$ represents the shift of $E$ from S,D to NS,NE. When $v > \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(3v)]$,we have two effects that yield that the excess patent is less profitable than before. First, reduced screening by the entrant under treble damages gives $A_{TD}^2 < A^1$. The region impacted by the excess patent is reduced. Furthermore, the gain from the excess patent under treble damages to $I$ is lower under treble damages since: $v > \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(3v)] > \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(v)]$.

If $v < \mathbb{P}\{b \geq q_E\}(v) + \mathbb{P}\{b < q_E\}[g(-K) + (1-g)(3v)]$, then screening entrants are more profitable to $I$ than excluded rivals and the utility of the excess patent is none.

All effects go in the same direction and therefore, the introduction of treble damages reduces the excess patenting incentives of the incumbent. □

**Proof of proposition 1.5.1**

When $E$ has to pay legal costs under a US litigation scheme, its behaviour is unchanged from the baseline model modulo a scaling factor due to the new scheme for litigation costs. The calculations of the baseline model apply for $E$ by substituting $K = 0$ and $\Delta = v_E - c_E - K_E^{US}$.

Given US litigation costs, payoffs to $I$, in the absence of an excess patent, are given by:

$$\mathbb{E}[\Pi_I|\sigma_E^*, n] = \begin{cases} v & \text{if } q_E \leq \underline{q_S} \\ \mathbb{P}\{b \geq q_E\}(v) - K_I^{US}\,\mathbb{P}\{b < q_E\} & \text{if } \underline{q_S} < q_E \leq \underline{q_T} \\ \mathbb{P}\{b \geq q_E\}(v) - K_I^{US} & \text{if } \underline{q_T} < q_E \end{cases}$$

We thus see that the negative profit implication on $I$ due to increased entry by high quality entrants (represented by the shift $B$) is equal to $-K_I^{US}\,\mathbb{P}\{b > q_E\}$. The profit gain from inducing an entrant to stay out in region $A$ is still $(v + K_I)\,\mathbb{P}\{b < q_E\}$. We see easily $\mathbb{E}_b\,\mathbb{E}_{q_E}\left[|(v + K_I^{US})\,\mathbb{P}\{b < q_E\}|\right] > \mathbb{E}_b\,\mathbb{E}_{q_E}\left[| - K_I^{US}\,\mathbb{P}\{b > q_E\}|\right] \quad\Leftrightarrow\quad v(\frac{1}{2}) +$

$K_I^{US}(\frac{1}{2}-\frac{1}{2}) > 0 \quad$ if $v, K_I^{US} > 0$. Thus, expected profits in region $A$ exceed those in region $B$ (abstraction made of different sizes of the regions $A$ and $B$ - see next step).

To compare the overall payoff effect of an excess patent, we need to weigh the profit effects due to changes in regions $A$ and $B$ by their likelihood of occurring. We have from proposition 1.3.3 that $|A| > |B|$, i.e. the probability that the affected entrant will switch strategy to staying out exceeds the probability that $E$ will switch to entry without screening. Thus, strategy considerations based on $A$ dominate those based on $B$. Consequently, the baseline analysis focussing on the shift $A$ yields qualitatively the same insights as one which would have taken both $A$ and $B$ into account.

For completeness and as a comparison, we provide the expressions of some milestones in the calculations of the model. The condition for less screening to mean less entry is given by the following and compares to condition 1.1:

$$|A|-|B| = \frac{\bar{q}S\tau}{\Delta - K_I^{US}} - \frac{\bar{q}S\tau}{|\Delta - v - K_I^{US}|} > 0 \quad \Leftrightarrow \bar{q}S\tau\left[\frac{|v|}{(\Delta - K_I^{US})(|\Delta - v - K_I^{US}|)}\right] > 0 \tag{1.4}$$

The excess patenting incentive for $I$ under the US cost regime:

$$\begin{aligned}\mathop{\mathbb{E}}_{b}\mathop{\mathbb{E}}_{q_E}\Big[-c + \mathbb{P}\{\underline{q}_S < q_E < \underline{q}_{S'}\}[v - v\,\mathbb{P}\{b \geq q_E\} + K_I^{US}\,\mathbb{P}\{b < q_E\}] \\ + \mathbb{P}\{\underline{q}_{T'} < q_E < \underline{q}_T\}[-K_I^{US}(1 - \mathbb{P}\{b < q_E\})\Big] \geq 0 \\ \Leftrightarrow -c + \frac{S\tau}{\Delta}\left(\frac{v + K_I^{US}}{2}\right) + \frac{S\tau}{|\Delta - v - K_I^{US}|}\left(\frac{-K_I^{US}}{2}\right) \geq 0\end{aligned} \tag{1.5}$$

Inequality 1.5 can also be rewritten in the form of

$$\frac{1}{2}\left[\mathbb{P}\{A\}[v + K_I^{US}] - \mathbb{P}\{B\}[K_I^{US}]\right] - c \geq 0$$

where $\mathbb{P}\{.\}$ interprets as the probability that $q_E$ falls in the range of $A$ or $B$. Inequality 1.5 compares to inequality 1.2. To show that excess patenting incentives are lower under a US regime versus a European litigation cost regime, see the following which compares inequality 1.5 (US payoffs, left) to inequality 1.2 (EU payoffs, right):

$$\frac{1}{2}\left[\mathbb{P}\{A\}[v + K_I^{US}] - \mathbb{P}\{B\}[K_I^{US}]\right] - c \geq \frac{1}{2}\left[\mathbb{P}\{A\}[v + K^{EU}]\right] - c \quad \Leftrightarrow \frac{-K_I^{US}}{K^{EU} - K_I^{US}} \leq \frac{\mathbb{P}\{A\}}{\mathbb{P}\{B\}}$$

which is satisfied since the LHS is negative and the RHS is positive (by prop. 1.3.3). □

**Proof of proposition 1.5.2**

By introducing commercial (increase $v_I$) or legal (increase $b$) value, the excess patenting incentives are satisfied more easily.

*Commercial value (increase in $v_I$:)* When the additional patent by $I$ increases the consumer valuation $v$ by a factor $\zeta > 0$, then the patenting constraint for the additional patent (under European litigation costs) is given by:

$$\begin{aligned}\mathop{\mathbb{E}}_{b}\mathop{\mathbb{E}}_{q_E}[\underbrace{-c+\mathbb{P}\{\underline{q}_S < q_E < \underline{q}_{S'}\}[(v(1+\zeta)+K)\,\mathbb{P}\{b<q_E\}]}_{>\text{ payoff from condition (1.2)}}+\underbrace{\mathbb{P}\{q_E<\underline{q_S}\}[v\zeta]}_{>0}\\ +\underbrace{\mathbb{P}\{q_E>\underline{q_{S'}}\}[v\zeta\,\mathbb{P}\{b\geq q_E\}]}_{>0}] > 0 \quad (1.6)\end{aligned}$$

Compared to condition 1.2 ($\zeta = 0$), we see that commercial patent value introduces slack in the excess patenting condition as all additive terms are increased. We see that $I$ benefits from $\zeta > 0$ in all cases when $E$ abstains from market entry.

Furthermore, commercial value of the additional patent affects the entrant in case of damages payments. The threshold $\underline{q_S}$ increases even more given an additional patent (with now also commercial value) when screening is imperfect. This contributes in further making the first line of condition 1.6 larger than its counterpart in condition 1.2. Moreover, the threshold $\underline{q_{T'}}$ increases (compared to no commercial value). This reduces the size of $B$ due to an excess patent and is a profitable change for $I$. Hence, commercial value results in enhanced entry deterrence (in $A$) and reduced entry fostering (in $B$). Both effects relax the excess patenting constraint.

*Legal value (increase in $b$:)* When the additional patent by the incumbent bears legal value, the litigation probabilities are affected. Assume that the new expected success probability for $I$ in litigation $\mathbb{P}\{\tilde{b} \geq q_E\} = \mathbb{P}\{b \geq q_E\} + \gamma, \gamma > 0$ and $\gamma$ is small s.t. $\mathbb{P}\{\tilde{b} \geq q_E\} \in [0, 1]$. Then for $E$, the expected payoff function for NS,EN shift downwards by a constant shift of $-(v + K)\gamma$. The payoff for $E$ from S,D shifts down by constant shift $-\Delta\gamma$. The latter shift is identical to an increase in screening costs if $-\Delta\gamma = -S\tau$.

Given the assumption $2\Delta < v + K$, we have that the NS,EN curve shifts down more and threshold $\underline{q_T}$ shifts right (shift of region B), inducing some $E$ to switch from NS,EN to S,D. Furthermore, we have in region A that the downshift of $\Pi_{S,D}$ induces some $E$ to switch from S,D to NS,NE. Both effects reduce market entry and are profitable for $I$.

In addition, the portfolio size has increased and causes an increase in screening costs of $S\tau$. Thus $\Pi_{S,D}$ shifts down again with the effects as shown in proposition 1.3.2. Overall, we see that legal value has relaxed the patenting constraint for $I$, because the legal value has strengthened the entry reduction in $A$ and reduced the entry fostering effect in $B$ over a mere increase in the size of the portfolio. $\square$

**Proof of Proposition 1.5.3**

When $\frac{v+K}{2} < \Delta < v + K$, two effects occur: First, the expected payoff function of S,D has a larger slope. The slopes of the payoff functions of the outside option as well as the strategy NS,EN remain unaffected. Higher $\Delta$ allows $E$ to take more risks ex-ante ($E$ can face a stronger $I$ without screening) and will result in more entry without screening (for lower values of $q_E$).

The argument is identical to corollary 1.4.3: Due to the increased slope of the screening function, the entrant's sensitivity to an excess patent goes down, because the raising rivals' cost impact is relatively smaller. Consequently, the model results are attenuated under large $v_E$ as the incumbent faces a tighter excess patenting condition.

The second effect is that due to the decreased slope differential between the expected payoffs of the S,D and NS,EN function, we do not validate the result of proposition 1.3.3: $|A| \leq |B|$. Consequently, less screening does not mean less entry anymore. This is because the relative magnitudes of the shifts $A$ and $B$ have changed, however the directions of the shifts remain as before (i.e. thresholds move together given an excess patent). In the European litigation cost setting, the main results are unaltered since the shift $B$ is payoff irrelevant for $I$.

In the US litigation cost setting, we thus see that excess patenting constraint may not be satisfied even if it were satisfied under European litigation costs. This is because the switch of an entrant from S,D to NS,NE is still more profitable to $I$ than a switch from S,D to NS,EN, but not necessarily more likely. Hence, larger $v_E$ affects the behaviour of $E$ by influencing the size of the shifts $A$ and $B$, which determine whether excess patenting by $I$ is profitable. We see that the excess patenting constraint ($\frac{1}{2}\left[\mathbb{P}\{A\}[v + K_I] - \mathbb{P}\{B\}[K_I]\right] - c \geq 0$) may not be satisfied when $\mathbb{P}\{B\}$ sufficiently exceeds $\mathbb{P}\{A\}$. $\square$

# Chapter 2

# Partial Termination Clauses

## On hold-up potential in patent pools

## 2.1 Introduction

Technology standards often rely on the intellectual property (IP) contained in many patents owned by different firms. Those patents that are essential to comply with a standard are called standard essential patents (SEPs). By bundling SEPs into a patent pool, the complements problem[1] in licensing can be avoided (Schmidt (2014)). Furthermore, one-stop shopping reduces the transaction costs for implementers of the standard to obtain those licenses. Both effects lead to a lower cost of access to a standard's technology and are generally expected to increase innovation rates for a standard (Shapiro (2001), Lerner and Tirole (2004)).

On the other hand, the market power of SEP holders is a problem in standard setting. When a standard becomes widely used, the owners of SEPs generally enjoy a large increase in the market power of their patent, because it is pivotal for the implementation of the standard. In order to reduce the ability of SEP holders to abuse this market power, the concept of fair, reasonable and non-discriminatory (FRAND) licensing terms has been developed[2]. However, the effectiveness of FRAND critically depends on how the

[1] The complements problem arises when licensors of complementary patents do not take into account the fact that an increase in the price of the license on their own patent reduces the demand for the licenses of each of the other complementary patents as well as the final good.

[2] European Commission (2012), "Case no COMP/M.6381 - Google/Motorola Mobility" *Merger Procedure Article 6(1)(b) - Decision*

FRAND commitment is implemented in practice, i.e. how FRAND terms are embedded in the IPR policies and licensing agreements of standard setting organisations and patent pools.

Two observations suggest that the FRAND commitment of the MPEG-2 patent pool[3] is weaker than that of other pools and motivate our analysis. First, Vakili (2012) shows that the innovation rate of firms in the technological proximity of the MPEG-2 standard dropped by 35% in the wake of the formation of the MPEG-2 patent pool. Second, Layne-Farrar and Lerner (2011) show that the MPEG-2 patent pool is composed of a higher proportion of vertically integrated firms than other patent pools. They also show that proportionally fewer vertically integrated firms stay out of the MPEG-2 pool than for other patent pools. These two observations taken together could suggest that, after the set-up of the pool, vertically integrated members of the MPEG-2 patent pool were able to capture more of the rents of outside innovators than in other patent pools. We wish to understand if this conjecture is true and if it can be explained by the policy of the MPEG-2 patent pool.

The policy of the MPEG-2 patent pool differs slightly from the policies of other patent pools. In particular, MPEG-2 is the only patent pool to feature a contractual clause, called partial termination, in its licensing agreement[4]. This clause gives pool members a bargaining advantage in the negotiations for non-essential patents owned by licensees of the patent pool. Specifically, partial termination[5] allows a licensor (owning a standard essential patent) of a patent pool to opt out of the pool with respect to a single licensee of the pool when negotiating a licence on a related patent[6] owned by that licensee. Without the excluded patent, the pool licence does not allow the pool licensee to implement the

[3]The MPEG-2 patent pool is "a group of parties that created a one-stop-shopping clearinghouse by pooling their patents. These patents are needed by entities looking to manufacture electronic equipment that stores or transmits compressed video data. A jointly owned licensing administrator (known as MPEG LA) licenses these patents in a single package that enables manufacturers to meet the standard known as MPEG-2 video compression technology. The technology standard eliminates redundant information, such as images that are all the same color or figures that do not change from one moment to the next, reducing the amount of data, storage and transmission space required to reproduce video sequences" (Quote taken from Department of Justice (1997), slight adaptations made for brevity). See definition 5 for details.

[4]Patent pools subsequently reviewed by the Department of Justice (e.g. the Toshiba DVD, UHF RFID and 3G pools) did not feature the partial termination clause. Also, Glader (2006) confirms this (p. 160).

[5]The partial termination clause may be invoked under two conditions (further specified in definition 6): (i) legal action must prevail between the licensor and the licensee and (ii) the proposed royalty rate on the related patent may not exceed the licensor's royalty for the essential patent. We give and discuss the full legal definition of partial termination in section 2.2.

[6]A related patent is by definition not essential for the implementation of the standard. Details p. 53.

standard without infringement. A separate licence for the excluded essential patent is necessary and the pool licensee (who owns a related patent) must negotiate bilaterally with the licensor of the excluded patent as if the pool did not exist. In the bilateral negotiations, the pool member is not bound by the FRAND commitment, which the pool member is subjected to under the pool licensing agreement. By recovering the discrimination power over its essential patent under partial termination, the essential patent holder has a stronger bargaining power to negotiate a low royalty on the related patent.

The question whether the partial termination clause compromises the effectiveness of FRAND in preventing essential patent holders from abusing their market power is an important question since the Department of Justice (1997) reviewed the partial termination clause in its review letter of the MPEG-2 patent pool and waived the clause since "on balance, [the clause appears] unlikely to be anticompetitive".

This work shows that the partial termination clause introduced a hold-up threat in the patent pool before the 2006 eBay decision. Before *eBay*, courts in the USA employed a property rule for essential patent holders, which backed up the patent entitlement by exclusion rights (i.e. injunctive relief). In a FRAND-committed patent pool like the MPEG-2 pool, however, the SEP holder gives up its discrimination power over the patent. By using partial termination to circumvent the FRAND commitment of the patent pool in certain situations, the SEP holders were able to leverage the importance of their SEPs in negotiations for licences to related patents of licensees of the standard. Since *eBay*, the threat of partial termination has been significantly reduced. This is due to the fact that the advantage, which partial termination bore for essential patent holders pre *eBay* (namely the stalling of the FRAND commitment), got lost with the adoption of a liability rule for essential patent holders by the courts. The liability rule limits the entitlement under a patent to a right to reasonable damages only and de facto imposes a FRAND commitment on all essential patent holders, even if the patent is not included in a patent pool. We note that the threat of partial termination is not completely eroded since permanent injunctions are not forbidden per se. However, the conditions required to obtain an injunction are much harder to satisfy.

Our analysis of partial termination is able to shed light on a few empirical results that have been established in the literature. First, this paper gives a possible explanation for why vertically integrated firms are overrepresented (under-represented) among the

MPEG-2 pool members (outsiders) as compared to pools that do not feature this clause (Layne-Farrar and Lerner, 2011). This structural effect arises from the fact that the partial termination clause increases the value of joining the pool for vertically integrated firms, while the pool joining incentives of pure upstream firms remain unaffected by the partial termination clause. Second, our model suggests an explanation for the drop in innovation rates of firms that research in the technological proximity of the MPEG-2 pool (Vakili, 2012). This result arises, because partial termination can hold-up the innovator of a cumulative innovation ex-post when that innovator is a licensee of the pool. The expected profits from investing in cumulative innovation are thus decreased.

In the extension of the model, we specifically look at the conditions under which partial termination may be invoked. We highlight that the partial termination threat is exacerbated by the number of declared SEPs and the rules used by the MPEG-2 pool to identify and check the essentiality and validity of SEPs. The analysis gives support to the view that the standard setting organisations should take active responsibility in the determination of the essentiality and validity of SEPs. This would significantly reduce the hold-up threat from partial termination, while ensuring the benefits from partial termination.

This chapter is closely related to the literature on the licensing of patents. Kamien and Tauman (1986), Kamien et al. (1992) and Wang (1998) investigate the performance of different types of licensing fees (e.g. auction vs. fixed vs. royalty) for innovators of cost-reducing innovations. More specifically on licensing in the context of a standard or patent pool, Kim (2004) and Schmidt (2014) investigate how the licensing of essential patents is affected by horizontal or vertical integration. They show that when the upstream market consists of perfect complements, vertically integrated firms have an incentive to discriminate against downstream rivals. The situation where an essential patent holder exploits the dependency of a manufacturer on standard essential patents, especially after the latter incurred standard-specific investments, is well documented under the term of patent hold-up (Shapiro (2001, 2010), Bessen (2004), Farrell et al. (2007)). The effects of the hold-up problem on innovation as well as possible remedies are investigated by Ganglmair et al. (2012) and Lerner and Tirole (2004, 2013). Also from a legal perspective, the literature on hold-up is extensive (Geradin (2009), Cotter (2009) and many more) and concludes that when the used market power is derived from the inclusion of the patent into a standard (rather than from its intrinsic value), the hold-up strategy is considered to be an abuse of a dominant position (Schmidt, 2010). We contribute a simple model to reveal the effects of the partial termination clause and give a possible explanation for the empirical puzzles by Layne-Farrar and Lerner (2011) and Vakili (2012).

The paper proceeds as follows. Section 2.2 details the institutional context of the analysis. We set up the model and give the results in section 2.3. Section 2.4 extends the model to generate a few predictions. We discuss the results and their limitations in section 2.5 and conclude in section 2.6.

## 2.2 Institutional context

We define key terms and the context of our analysis. First, we define two types of patents since regulation relating to each type differs. Second, we describe the regulation that exists and provide a brief account of the recent policy shift by the US Department of Justice on essential patent holders. Third, we present the partial termination clause from the licensing agreement of the MPEG-2 patent pool, which we want to study.

### 2.2.1 Types of patents

Many industry or technology standards rely on patent protected technologies. We distinguish two types of patents.

**Definition 1 (Essential Patent)** *"An essential patent[7] shall mean a Patent which is necessarily infringed in connection with the use or implementation of the technology standard."*

**Definition 2 (Related Patent)** *"A related patent[8] shall mean any Patent which is not an Essential Patent but which has one or more claims directed to an apparatus or a method that may be used in the implementation of a product designed in whole or in part to exploit the technology standard."*

The distinction between the two types of patents lies in their importance to comply with a standard. While the related patent covers inventions that are optional to the compliance with a given standard (e.g. small improvements and add-on features), the inventions contained in an essential patent are absolutely necessary to implement and comply with a fully functioning standard. For clarity, we emphasize that the difference does not lie

[7]Source: MPEG-2 PPL (2009), The Patent Portfolio License, §1.13, Adaptations made for brevity.
[8]Source: MPEG-2 PPL (2009), The Patent Portfolio License, §1.17, Adaptations made for brevity.

in the inventive step, i.e. the intrinsic quality, of the patents. The essential part comes merely from the proximity of the invention to the standard.

Consequently, the market power that an IP holder can derive from his patent differs for both types of patents. When the standard becomes widely adopted, the essential patent holder enjoys market power that is not due to the quality of his invention, but driven by the success of the standard. As mentioned in the introduction, the concept of FRAND was developed in order to limit the ability to abuse the market power granted by essential patents.

FRAND stands for fair, reasonable and non-discriminatory. The commitment applies to the licensing terms of essential patents. In our context, we interpret the FRAND commitment as the legal obligation to propose a single price (= non-discriminatory) for the licence on a patent to all interested parties (= fair), where the price must be lower than a reference price set by the courts (= reasonable). We note the soft nature of the definition and acknowledge some critiques on FRAND[9].

### 2.2.2 Court rules

The market power considerations of SEP holders have driven the recent policy change in the US. In litigation cases until 2006, the courts awarded the prevailing patent owner a permanent injunction absent ”exceptional circumstances” (Cotter, 2013). This meant that “a patent owner who prevailed in an infringement suit was effectively entitled to a permanent injunction as a matter of right” (Torrance and Tomlinson, 2011).

In *eBay v. Mercexchange (2006)*[10], the US Supreme Court acknowledged that the Federal Circuit “erred in its categorical grant of (injunctive) relief”. The Supreme Court decided that a traditional four-factor test should be applied when considering to award permanent

[9] Layne-Farrar et al. (2007) propose some critiques: Goldstein and Kearsey (2004) identify the lack of a universally accepted operational definition as one of the “core problems in the licensing of rights” to essential patents. Legal uncertainty arises from the lack of a generally accepted test for FRAND commitments (Swanson and Baumol, 2005). Rapp and Stiroh (2002) discuss the imprecision of a FRAND definition in determining the size of royalties since the reasonability criterion can mean very different things to the intellectual property owners and the licensees. Furthermore, Cotter (2013) shows that the magnitude of FRAND royalties is very much context dependent.

[10] “When (..) the threat of an injunction is employed simply for undue leverage in negotiations, legal damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest.” *eBay Inc. v. MERCEXCHANGE, LL*, 547 U.S. 388, 126 S. Ct. 1837, 164 L. Ed. 2d 641 (2006)

injunctive relief to a defendant under the Patent Act[11]. Since the *eBay* case, courts do not seem to grant permanent injunctions any more against infringers of standard essential patents subject to a FRAND commitment (Cotter, 2013). Two major cases are notable[12]: *Microsoft v. Motorola*[13] and *Apple v. Motorola*[14]. In these cases, the courts first confirmed the validity of Motorola's FRAND commitment. Second, the four factor test was not deemed to be satisfied and motions for an injunction were rejected. Third, the rulings emphasized the importance of the courts for the determination of the level of FRAND royalties[15].

The policy shift induced by *eBay* is a move from a property rule to a liability rule with respect to holders of essential patents. Note that this rule shift only applies to essential patents and not to related patents.

**Definition 3 (Property rule)** *Under a property rule*[16]*, an entitlement is backed up by a right to injunctive relief.*

**Definition 4 (Liability rule)** *Under a liability rule*[17]*, an entitlement is backed up by a right to reasonable damages only.*

A liability rule differs from the property rule in that it does not grant the defendant a right to exclude others from using its patent henceforth, but merely awards reasonable (FRAND) damages for the future use of the intellectual property by the infringer. Common to both rules is the award of reasonable damages for past use of the patent. Thus, in practice under a liability rule, the defendant is free to breach the entitlement and pay

[11]Specifically, the test requires a plaintiff to demonstrate: (1) it has suffered an irreparable injury; (2) remedies available at law, such as monetary damages, are inadequate to compensate for that injury; (3) considering the balance of hardships between the plaintiff and defendant, a remedy in equity is warranted; and (4) the public interest would not be disserved by a permanent injunction.

[12]The US evolution is congruent with European law. For details, see Cotter (2013) and *UK High Court Denies a Patent Injunction Against Nokia in Light of a FRAND Commitment*, Foss Patent Blog (30.03.2012).

[13]*Microsoft Corp. v. Motorola, Inc.*, No. C10 - 1823 JLR, WL 5993202 (W.D. Wash. Nov. 30, 2012)

[14]*Apple, Inc. v. Motorola, Inc.*, 869 F. Supp. 2d 901 (N.D. Ill. 2012).

[15] Judge Crabb states in the decision of a related ruling of the latter case (*Apple, Inc. v. Motorola Mobility, Inc.*, 2012 WL 5416941 (W.D. Wis. Oct. 29, 2012)): "It makes sense to allow [...] for the court to determine license terms, if necessary. In fact, in situations such as this in which the parties cannot agree on the terms of a fair, reasonable and non-discriminatory license, the court may be the only forum to determine license terms."

[16]Patent law interpretation after Cotter (2013) based on Calabresi and Melamed (1972).

[17]Patent law interpretation after Cotter (2013). Torrance and Tomlinson (2011) give further details: "the owner of an entitlement is legally powerless to keep it exclusively for herself. The owner must accept compensation at the objectively determined value by some neutral third-party authority."

damages. The right holder is not able to invoke an exclusionary right by means of a permanent injunction.

### 2.2.3 Patent pool (MPEG-2)

**Definition 5 (Patent Pool)** *A patent pool*[18] *is an aggregation of patent rights for the purpose of joint package licensing. The portfolio combines essential patents*[19] *, which are complementary*[20] *and interdependent*[21] *for implementing the standard. Licensing of the bundled patents underlies a commitment to license on FRAND terms.*

For the sake of pool stability, many pooling agreements oblige licensors and licensees to grant back existing and future essential patents. These standard grantback clauses require all pool members (licensors & licensees) to include all essential patents in the pool, even if they have been developed after the pool has been set-up[22]. Grantback removes the hold-up potential by licensees owning essential patents. This is generally considered pro-competitive and welfare enhancing[23].

By definition, a related patent is not subject to the grantback clause[24]. However, the partial termination clause extends this obligation to related patents of licensees of the pool in some situations. We want to study this clause and formally define it here. In the following, the *licensing administrator* refers to the managing entity (MPEG LA) of the patent pool, *the licensors* refers to the owners of essential patents included in the pool and *the licensees* refers to firms paying for a pool licence.

[18] We lean on Department of Justice (1997) for the definition of a patent pool.

[19] For the MPEG-2 pool: Absent disputes, the licensors merely need to reach consensus that a patent should be included in the pool. In case of disputes, the evaluation is left to an independent expert.

[20] There is no technical alternative to any of the portfolio patents within the standard.

[21] Each portfolio patent is useful for standard compliant products only in conjunction with the others.

[22] For convenience, this formulation is imprecise. However, the main idea is preserved. Formally, all pool licensors and licensees can get access to essential patents of other licensors or licensees at a cost that is equivalent or less to what they could have negotiated absent the patent pool, be it by including the new essential patent into the pool (MEPG-2 §2.3 AAL) or by denying the owner of the invention access to the fully functioning pool bundle until an agreement has been reached (MPEG-2 §2.8 AAL).

[23] A grantback clause is used by 71% of the pools not litigated in the sample of Lerner, Strowjas & Tirole (2007, p.619). They take non-litigated pools as a proxy for pools of complements based on the assumption that pools of substitute patents are more likely to be litigated (since they can avoid price competition).

[24] "No matter how attractive the licensee's patented implementation of the [...] standard may be, by definition the Related Patent will not be essential to compliance with the standard. And, not being essential, the patent is not subject to [...] grantback (Department of Justice, 1997).

**Definition 6 (Partial Termination)** [25] *The Licensing Administrator, upon the instruction of a Licensor, shall terminate Licensee's sublicense under any Pool Patent(s) licensed or sublicensed to the Licensing Administrator by such Licensor in the event that*

> ***Condition 1 (Legal Action)*** *the Licensee has brought a lawsuit or other proceeding for infringement of a Pool Related Patent(s) against such Licensor, and*
>
> ***Condition 2 (Non-FRAND Terms)*** *Licensee has refused to grant the Licensor a license on fair and reasonable terms and conditions under the Pool Related Patent(s) upon which the lawsuit or other proceeding is based. [...] The Licensor's per patent share of royalties payable shall be presumed to be a fair and reasonable royalty rate for Licensee's patent(s)*[26].

In simple terms the clause does the following: Assume that we have a patent pool, which licenses out a portfolio of patents needed for a standard. Consider two firms, where firm $A$ is a licensee of this pool and firm $B$ a licensor of the pool (i.e. owns an essential patent included in the pool portfolio license). Licensee $A$ owns a related patent to the standard in question, which makes the object of a lawsuit between licensee $A$ and a licensor $B$. Furthermore, $A$ does not agree to license the related patent to $B$ at a royalty that is less than or equal to the royalty that $B$ earns with its essential patent through the pool license[27]. Given these two conditions, $B$ can invoke partial termination and stall the portfolio license for $A$ only (other licensees of the pool are unaffected). Thereby, $B$ makes $A$ unable to comply with the standard. If $A$ wants to comply with the standard, it must obtain a separate licence for the excluded patent from $B$.

In summary, the partial termination clause allows individual pool members to opt out of the pool with respect to a single licensee (Merges, 1999b)[28].

[25] Emphasis and formatting added. Slight adaptations made for brevity. Definition taken from MPEG-2 PPL (2009): §6.3. The PT clause applies both to related as well as essential patents of licensees. For essential patents, however, the PT clause has no additional grip on top of MPEG-2 PPL (2009) §7.3 (grantback). Consequently, we suppress the references to essential patents in definition 6.

[26] Note that "Licensors who are Licensees are exempted from this presumption as a result of, among other things, the costs incurred in connection with this licensing program".

[27] Recall that the quality of the patents is not specified in this argument and irrelevant for the definition of an essential or related patent. Therefore, there is no indication that this upper bound is appropriate.

[28] Merges (1999b) confirms that "the purpose of this provision [...] is to provide bargaining leverage to any MPEG-2 (pool-)member that is negotiating for a license to a complementary patent held by an MPEG-2 (pool-)licensee."

Whilst the above definition applies partial termination to situations with existing licensees of the pool, an alternative version of the clause applies to prospective licensees of the pool. See appendix 2.7.1 for details. We note that partial termination has been invoked only once against a prospective licensee of the MPEG-2 patent pool.

## 2.3 Model

We are interested in the innovation incentives of a firm that is dependent on a licence for an essential patent from another firm. Specifically, we investigate how these incentives are affected by changes in the legal environment, i.e. by the rules of the game. The model is based on Wang (1998). We first outline the status quo, we then detail the cumulative innovation game.

### 2.3.1 Set-up

**Market structure** Our industry is composed of two markets, an upstream and a downstream market. The upstream market relates to intellectual property in the form of patents, which provide know-how for the production of goods on the downstream market. We focus on a setting with two firms, both of which are downstream manufacturers, but only one of which owns a standard essential patent at the start of the game. The firm owning the patent is thus vertically integrated (called the *upstream incumbent* $I$) and stands for a simplified version of a patent pool. The other firm is not yet vertically integrated (called the *upstream entrant* $E$).

**Downstream market** Consider the downstream market first without any upstream IP licensing considerations. We assume a simple linear Cournot model. Suppliers $i = \{I, E\}$ produce homogenous goods and face a linear and downwards sloping demand function: $P = a - Q$, $Q = \sum_i q_i$. Fixed costs are assumed to be sunk. Manufacturers incur only constant marginal costs $c > 0$ in the production of the final good. Demand is large enough for a market to exist: $a > c > 0$. Both manufacturers maximise profits $\Pi_i = (a - Q - c)q_i$, $\forall i$. From standard existence theorems, we know that a unique, symmetric Nash equilibrium exists on the downstream market with $q_i = \frac{a-c}{3}$ and $\Pi_i = \left(\frac{a-c}{3}\right)^2$ $\forall i$.

**Upstream market** As mentioned, downstream production uses know-how contained in patents, which are licensed on the upstream market. We assume that all patents are valid with certainty and that an infringement of a patent is never profitable for the infringer. We assume that the patents are licensed using linear royalty rates as is the case for the MPEG-2 patent pool (Kim, 2004)[29]. Royalties paid by firm $i$ ($r_i$) thus add to the marginal costs of downstream production. Total downstream marginal costs are given by $C_i'(r) = c + r_i$.

**Licensing of the essential patent** We assume that the patent owned by the incumbent $I$ is an essential patent. A license for the essential patent is required to produce goods on the downstream market. $I$ is bound by a FRAND commitment for this essential patent and licenses it out at a fee $l \geq 0$. The licence fee $l$ is assumed to be exogenous and cannot be adapted in the game in response to, for example, competition[30]. The reasonability criterion of FRAND means that $l \leq \bar{l}$, where $\bar{l}$ is the threshold level that courts understand as reasonable. To simplify the model, we set the exogenous $l = \bar{l} = 0$ [31]. The fairness criterion also means that $I$ is not able to exclude $E$ from a licence[32].

Thus in the status quo, both firms have access to the essential patent for free. The downstream equilibrium remains unchanged. Both firms behave like pure downstream manufacturers since there are no profits to be made upstream.

### 2.3.2 Cumulative innovation game (without partial termination)

In the context of the status quo duopoly, we are interested in the incentives of $E$ to invest in cumulative innovations. The (until now) pure downstream manufacturer $E$ can invest in a cost-reducing innovation, which is not essential, but related to the standard. If the

[29]Generally, linear royalties are abundantly used in practice (Layne-Farrar and Lerner, 2011), one of the reasons for this choice is their risk sharing property (Schmidt, 2014).

[30]This represents the case, when multiple licensees of the essential technology make a royalty adjustment unprofitable (e.g. a FRAND commitment would require to change the royalty rates for all licensees simultaneously). This is also representative of a patent pool, where royalty determination is delegated to the pool managing entity and revised rarely. The MPEG-2 pool in particular is known for not adjusting its royalty over time, even as patents expire.

[31]Thereby, we abstract of the double mark-up problem for the entrant and focus our analysis on the boundary case in which $I$ only cares about the downstream market. We address the implication of this assumption in more detail in section 2.5.2.

[32]The non-discrimination criterion requires the same price to apply for all licensees of the pool bundle. This is irrelevant in a two-player setting.

investment is made, $E$ obtains a patent (called the related patent) for its invention, which can be licensed to $I$.

**Cumulative innovation** When $E$ decides to develop a cost-reducing innovation, it pays investment costs $K \geq 0$ and receives an innovation, which yields an expected constant marginal cost reduction of $\delta(K) \geq 0$. When investing, $E$ is rational, risk-neutral and only cares about the expected value of the innovation, not about the uncertainty on the outcome of the innovation. We assume decreasing marginal returns of investment[33]: $\delta'(K) > 0$ and $\delta''(K) < 0$. The expected cost reduction $\delta(K)$ is bounded by $\delta(0) = 0$ and $\lim_{K\to\infty} \delta(K) = (a-c)$. We also impose $\lim_{K\to 0} \delta'(K) = \infty$ and $\lim_{K\to\infty} \delta'(K) = 0$.

The innovation is not essential to produce downstream, but optional. We therefore limit the size of the cost-reduction due to the related patent to $\delta < a - c$. This restriction excludes drastic innovations that would lead to the exclusion of the incumbent downstream[34] (Arrow, 1962) and is motivated by the idea that the legal threshold for distinguishing a related patent from an essential patent would likely be crossed by a drastic innovation. A drastic patent excludes all firms without access to its IP from the downstream market and is thus de-facto essential. Given that essential patents do not underlie the same regulation as related patents, we do not include them in the model.

**Licensing of the related patent** The new related patent can be licensed to $I$. $E$ decides on and proposes a take-it-or-leave-it offer $\hat{r} \geq 0$ for a licence on its new related patent. Note that $E$ is not bound by a FRAND commitment for its related patent[35]. $I$ can decide whether to accept the offer or reject it.

If $I$ accepts the offer, it pays the royalty $\hat{r}$ and both firms enjoy the marginal cost reduction of size $\delta$ in downstream production. The payoff to the firms in the Cournot stage is then $\Pi_I^L(\delta-\hat{r},\delta) = (a-Q^L-c+\delta-\hat{r})q_I^L$ and $\Pi_E^L(\delta,\delta-\hat{r}) = (a-Q^L-c+\delta)q_E^L+\hat{r}q_I^L$. In terms of notation, we abbreviate equilibrium quantities to firm $i = \{I, E\}$ by $q_i^S$ and payoffs by $\Pi_i^S(\Delta C_i', \Delta C_j'),\ i \neq j$ . The superscript $S = \{L, N\}$ stands for whether licensing occurs ($L$) or not ($N$). The first argument ($\Delta C_i'$) stands for the cost-reduction enjoyed by firm $i = \{I, E\}$ and the second argument ($\Delta C_j'$) stands for the cost reduction enjoyed by the other firm. $\Pi_i(0, 0)$ refers to firm $i$'s status quo profits.

[33]These assumptions are in line with the endogenous investment model of Baye and Hoppe (2003).
[34]This occurs when the monopoly price under the new technology does not exceed the competitive price under the old technology $p_m \leq p_c \Leftrightarrow \delta \geq a - c$.
[35]This corresponds to the situation of a licensee, who is not a licensor of essential IP, of a patent pool.

If $I$ rejects the offer, it cannot use the cost-reducing innovation for downstream production. The firms then maximise $\Pi_I^N(0,\delta) = (a - Q^N - c)q_I^N$ and $\Pi_E^N(\delta,0) = (a - Q^N - c + \delta)q_E^N$ on the downstream market.

**Timing and structure** The game develops as follows:

- $t = 1$ $E$ decides on the investment level $\hat{K}$
- $t = 2$ $E$ makes a take-it-or-leave-it royalty offer $\hat{r}$ to $I$
- $t = 3$ $I$ decides whether to accept the offer or not
- $t = 4$ Cournot downstream competition ensues

We derive the equilibrium of the cumulative innovation game without partial termination in proposition 2.3.1.

**Proposition 2.3.1** *There exists a unique subgame perfect equilibrium in the cumulative innovation game without partial termination, where the entrant invests $\hat{K} > 0$, obtains a cumulative innovation of size $\delta = \delta(\hat{K})$ and proposes a royalty rate $\hat{r} = \delta$ to the incumbent, who accepts the offer. Both firms produce positive quantities on the downstream market.*

$$q_I^L = \frac{a-c-\delta}{3},\ q_E^L = \frac{a-c+2\delta}{3} \text{ and } Q^L = \frac{2(a-c)+\delta}{3} \tag{2.1}$$

*Equilibrium profits to the firms are*

$$\Pi_I^L(0,\delta) = \Big(\frac{a-c-\delta}{3}\Big)^2 \quad \text{and} \quad \Pi_E^L(\delta,0) = \Big(\frac{a-c+\delta}{3}\Big)^2 + \frac{5}{9}\delta(a-c) \tag{2.2}$$

*The level of investment in innovation ($\hat{K}$) is implicitly defined by*

$$\Big[\frac{7}{9}(a-c) + \frac{2}{9}\delta(\hat{K})\Big]\frac{\partial\delta}{\partial\hat{K}} = 1 \tag{2.3}$$

**Proof** See the appendix. □

In equilibrium, we obtain that $E$ invests a positive amount in cumulative innovation and licensing of the related patent occurs ex-post. The intuition for the result is the following. For licensing to occur, both firms must earn at least as much as given no licensing. In $t = 3$, $I$ thus accepts any royalty $0 \leq r \leq \delta$. $E$'s profits are increasing in $r$ and $E$ will propose the highest royalty that will be accepted by $I$, i.e. $\hat{r} = \delta$. In equilibrium, $I$ is indifferent between accepting and rejecting, while $E$ is strictly better off than under no licensing. There exists no unilateral profitable deviation for any player. $E$ is also strictly better off than in the absence of innovating and thus invests a positive amount $\hat{K}$ in innovation. The level of investment in innovation is implicitly characterised by the first order condition in (2.3).

### 2.3.3 Partial termination (under a property rule)

We use the model to analyse the effect of the partial termination (PT) clause. Partial termination is formally defined on page 57. When PT is invoked, the licensor is able to exclude its essential patent from the pool bundle with respect to a single licensee. In order to comply with the standard, the licensee must obtain a separate licence for the excluded patent. Continued production downstream without the excluded patent would constitute an infringement. Therefore, the use of PT is equivalent to returning to the bilateral negotiations between firms in the absence of the pool and the pool's commitments (such as FRAND). Outside of the pool, we assume that the owner of the essential patent enjoys a property right over its patent and is able to exclude competitors from using it.

In the model, when $I$ invokes PT, it is able to cut the license of $E$ to the essential patent and thereby, $E$ is excluded from the downstream market. Thus using PT, $I$ becomes a monopolist. Even when $E$ is excluded downstream, licensing of the related patent may occur. Profits to the firms are $\Pi_E^{\text{L, PT}} = \hat{r}^{\text{PT}} q_I^{\text{PT}}$ and $\Pi_I^{\text{L, PT}} = (a - q_I^{\text{PT}} - c + \delta - \hat{r}^{\text{PT}}) q_I^{\text{PT}}$ when licensing occurs and $\Pi_E^{\text{N, PT}} = 0$ and $\Pi_I^{\text{N, PT}} = (a - q_I^{\text{PT}} - c) q_I^{\text{PT}}$ given no licensing.

However, PT can only be invoked if two conditions are satisfied. Condition 1 (Legal Action) is assumed to be exogenously given[36] when the Incumbent rejects the proposed royalty $\hat{r}$. Condition 2 (Non-FRAND Terms) is satisfied, when the proposed royalty is not FRAND. The PT clause specifies that for condition 2 to be satisfied, the proposed royalty on the related patent ($\hat{r}$) must exceed the per patent royalty received by the owner of the essential patent ($l = 0$). Therefore in our setting, $I$ may invoke PT iff $\hat{r} > l = 0$.

**Proposition 2.3.2** *There exists a unique subgame perfect equilibrium in the cumulative innovation game with partial termination (under a property rule for essential patents),*

[36] In particular, we note that condition 1 for PT (originating from section 2.10 of the Agreement Among Licensors) is satisfied when the essential patent owner (licensor of the pool, here called $I$) initiates an infringement suit against a prospective licensee ($E$) on the patent to be excluded. Legal action must thus not be initiated by the $E$. See appendix 2.7.1 for details.

*where the entrant invests* $0 < \hat{K}_{PT} < \hat{K}$ *and obtains a cumulative innovation of size* $\delta^{PT} = \delta(\hat{K}_{PT}) < \delta$.

*When* $0 < \delta(\hat{K}_{PT}) \leq \frac{a-c}{2}$, *the proposed royalty is* $\hat{r}^{PT} = 0 < \hat{r}$ *and the partial termination clause cannot be invoked in equilibrium. A duopoly remains on the downstream market, where equilibrium quantities and profits are*

$$q_I^{PT} = q_E^{PT} = \frac{a-c+\delta^{PT}}{3} \quad and \quad Q^{PT} = \frac{2(a-c+\delta^{PT})}{3} \tag{2.4}$$

$$\Pi_I^L(\delta^{PT}, \delta^{PT}) = (\frac{a-c+\delta^{PT}}{3})^2 > \Pi_I^L(0,\delta) \quad and \quad \Pi_E^L(\delta^{PT}, \delta^{PT}) = (\frac{a-c+\delta^{PT}}{3})^2 < \Pi_E^L(\delta, 0) \tag{2.5}$$

*When* $\frac{a-c}{2} < \delta(\hat{K}_{PT}) < (a-c)$, *the proposed royalty is* $\hat{r}^{PT} = \delta^{PT}$ *(same linear price as* $\hat{r}$*) and the partial termination clause is invoked in equilibrium. A monopoly results on the downstream market, where equilibrium quantities and profits are*

$$q_I^{PT} = \frac{a-c}{2},\ q_E^{PT} = 0 \quad and \quad Q^{PT} = \frac{a-c}{2} \tag{2.6}$$

$$\Pi_I^{L,PT} = (\frac{a-c}{2})^2 > \Pi_I^L(0,\delta) \quad and \quad \Pi_E^{L,PT} = (\frac{a-c}{2})\delta^{PT} < \Pi_E^L(\delta, 0) \tag{2.7}$$

*The level of investment in innovation* $\hat{K}_{PT}$ *is implicitly defined by*

$$\begin{cases} \left[\frac{2}{9}(a-c+\delta(\hat{K}_{PT})\right]\frac{\partial\delta}{\partial\hat{K}_{PT}} = 1 & if\ \ 0 < \delta(\hat{K}_{PT}) \leq \frac{a-c}{2} \\ \left(\frac{a-c}{2}\right)\frac{\partial\delta}{\partial\hat{K}_{PT}} = 1 & if\ \ \frac{a-c}{2} < \delta(\hat{K}_{PT}) < (a-c) \end{cases} \tag{2.8}$$

**Proof** See the appendix. □

The intuition of the results is the following. Under a property rule for essential patents outside of the pool, the threat of partial termination is credible, because invoking PT is always profitable for $I$ and always costly to $E$. Two cases arise since $E$'s optimal behaviour depends on the expected value of the cumulative innovation. In both cases (detailed below), we obtain the result that expected profits from innovating go down for $E$ and hence, the optimal level of investment in innovation for $E$ is decreased. This is because, by using PT or the credible threat of PT, $I$ can capture some of the rents from innovating.

When the expected cost-reduction from the related patent is small ($0 < \delta^{\mathrm{PT}} \leq \frac{a-c}{2}$), then the attainable revenue from licensing the related patent is not sufficient to compensate $E$ for the loss of its downstream profits. $E$ actually finds it more profitable to share the related patent for free rather than be excluded from the downstream market. Consequently, $E$ proposes the highest possible royalty, which does not allow $I$ to invoke PT: In our setting this is the zero royalty. Since the proposed royalty is lower than the optimal royalty in the absence of PT, $E$ forgoes profits and its incentives to invest in innovation are reduced.

When the expected cost-reduction is larger ($\frac{a-c}{2} < \delta^{\mathrm{PT}} < (a-c)$), the best reaction by $E$ to the partial termination threat is accepting exclusion from the downstream market and maximising revenue from the licensing of the related patent. Hence, $\hat{r}^{PT} = \delta$. While in this case the proposed (proportional) royalty rate is unchanged compared to proposition 2.3.1, $E$'s total profits and thus innovation incentives are hurt since $E$ is excluded downstream. In both cases, the equilibrium levels of investment in innovation are lower than in the absence of a PT threat. The exact levels are characterised by the first order conditions in (2.8).

We note that even at $\hat{r} = 0$, $I$ would like to invoke PT, but cannot do so because condition 2 is not satisfied. This point is important and shows that the limiting conditions for the partial termination clause create a safe harbour for $E$ to license his innovation. However, the above situation exemplifies that the value of the safe harbour depends strongly on the definition of FRAND as applied to the related patent. We return to this point later.

### 2.3.4 Partial termination (under a liability rule)

When a liability rule prevails for essential patent holders outside of a patent pool, the threat of PT is not credible any more. The difference between a liability and property rule is formally defined on page 55. Recall that this is the policy shift that we have observed in the US in the wake of the eBay decision in 2006. Where as under a property rule an essential patent owner was able to exclude competitors, this is not possible under a liability rule any more. Instead, the essential patent holder must accept compensation at reasonable rates.

In the model, when $I$ invokes PT and excludes its patent from the pool bundle, it is still subject to a liability rule and must grant a license for the essential patent to $E$ on reasonable terms. Reasonable terms for the essential patent were set equal to a zero by assumption ($l = 0$). This means that $E$ continues to get the license for the essential patent for free and cannot be excluded from the downstream market. In the negotiations for the royalty on the related patent, $I$ can hence only revert to not accepting the license offer. Bargaining based on $I$'s power to cut the licence on the essential patent is not possible any more. The game is thus identical to the baseline scenario without PT.

**Proposition 2.3.3** *When a liability rule applies to essential patent holders, the partial termination has no effect and the equilibrium is identical to that of proposition 2.3.1*

**Proof** In the text above. $\square$

We have now seen that when PT is a credible threat, $E$'s incentives to innovate as well as competition on the downstream market are affected. In a linear Cournot model, the welfare implications for consumers are directly related to the effect on the total downstream quantity, where an increase in total quantity translates into an increase in consumer surplus.

**Corollary 2.3.4** *If we assume that the cost reductions of the cumulative innovation ($\delta$ and $\delta^{PT}$) are given exogenously, then the total downstream quantity under the threat of partial termination*

*(i) is lower when* $\delta^{PT} > \frac{a-c}{2}$

*(ii-a) is lower when* $0 < \delta^{PT} \leq \frac{a-c}{2}$ *and* $\delta^{PT} < \frac{\delta}{2}$

*(ii-b) is higher when* $0 < \delta^{PT} \leq \frac{a-c}{2}$ *and* $\delta^{PT} > \frac{\delta}{2}$

*compared to the total downstream quantity in the absence of a partial termination threat.*

**Proof** Immediate by comparing total volumes in equations 2.1, 2.4 and 2.6. $\square$

We compare the situation with and without PT. When $E$ innovates in the absence of PT, then $E$ benefits from reduced marginal production costs, while $I$ does not benefit from any cost reduction. The result is that total downstream volume is increased over the situation where no cost-reducing innovation is realised. The magnitude of the increase in total downstream volume depends on the size of the cost reduction.

When PT is a credible threat and the expected cost-reduction is larger ($\frac{a-c}{2} < \delta^{\mathrm{PT}} < (a-c)$), then in equilibrium, the PT clause will be invoked by $I$. Now, $I$ benefits from the decreased competition as $E$ is excluded from the downstream market, but $I$ does not benefit from a marginal cost reduction. The downstream quantity resulting from a monopoly served by $I$ is always lower than that of the duopoly absent a threat of partial termination.

When PT is a credible threat and the size of the cost reduction is small ($0 < \delta^{\mathrm{PT}} \leq \frac{a-c}{2}$), then $E$ gives out the cost reduction for free (PT cannot be invoked) and stays in the downstream market. Consequently, both $E$ and $I$ benefit from the cost reduction - an effect that results in an increase of the total downstream quantity. However, the fact that $E$ does not earn profits from the licensing of the innovation decreases its investment incentives and the expected level of the realised innovation falls - an effect that decreases the total downstream quantity. The net effect of the PT threat on the downstream volume is thus ambiguous and depends on the specification of the innovation investment function. In our setting, total volume will decrease (increase) if the PT threat reduces the expected level of the cost reduction by more (less) than half its level in the absence of a PT threat. We look at the results for a specific investment function in section 2.4.1.

## 2.4 Extensions

### 2.4.1 Comparative statics

To generate specific welfare implications, we specify a functional form for our innovation investment function. The cost reduction of the cumulative innovation ($\delta$) depends on $E$'s investment ($K$): $\delta(K) = A\sqrt{K}$, where the parameter $A > 0$ is a multiplier constant that controls the concavity of the investment function. To align the functional form with the conditions used in the model so far, we impose decreasing marginal returns (by the use of the square root) and focus on non-drastic innovations[37]: $A < \sqrt{2}$.

**Proposition 2.4.1** *When $\delta(K) = A\sqrt{K}$, then propositions 2.3.1 and 2.3.2 hold for $0 < A < \sqrt{2}$.*

*Respectively, without and with a credible threat of partial termination, the equilibrium levels of investment and the realised cost-reductions are given by*

$$\hat{K} = \big[A\frac{\frac{7}{2}(a-c)}{9-A^2}\big]^2, \quad \delta = A^2\big[\frac{\frac{7}{2}(a-c)}{9-A^2}\big] \quad < a-c \tag{2.9}$$

$$\textit{and} \quad \hat{K}_{PT} = \big[A\frac{(a-c)}{9-A^2}\big]^2, \quad \delta_{PT} = A^2\big[\frac{a-c}{9-A^2}\big] \quad < \frac{\delta}{2} \quad < \frac{a-c}{2} \tag{2.10}$$

[37]The condition is derived in the proof of proposition 2.4.1. It ensures that $0 < \delta < a - c$.

*The total downstream quantity is reduced under a credible partial termination threat ($Q^L > Q^{PT}$) for the whole range $0 < A < \sqrt{2}$.*

**Proof** See the appendix. □

We have previously seen that, in equilibrium, the investment in cumulative innovation and the resulting cost-reduction are reduced when the threat of partial termination is credible. Proposition 2.4.1 shows that these results hold for a specific set of innovation investment functions. Furthermore, it shows for a credible threat of PT that case (i) of corollary 2.3.4 will not occur in equilibrium. This means that the realised innovation by $E$ is always of a sufficiently low quality ($0 < \delta^{PT} < \frac{a-c}{2}$), such that $E$'s best response to the partial termination threat is the free-sharing of the innovation with $I$. By sharing the innovation for free, $E$ avoids the expulsion from the downstream market due to an exercise of PT.

Moreover, proposition 2.4.1 shows that case (ii-b) of corollary 2.3.4 will not occur in equilibrium either. This means that we observe a reduction in the quality of $E$'s innovation due to the PT threat that is more than half its value in the absence of a PT threat ($\delta^{PT} < \frac{\delta}{2}$). As a result, the total quantity on the downstream market is always reduced. In our linear Cournot model, this downstream volume reduction leads to an increase in the price on the downstream market and a reduction of the consumer surplus on the market. In addition, we have from proposition 2.3.2 that under PT, consumers expect less innovation related to the standard. Both effects harm consumers.

We note that these results are largely driven by two assumptions: (i) the focus on non-drastic innovations only and (ii) the set of concave investment functions considered. The first assumption (detailed on page 60) ensures that the partial termination clause may be invoked in principle and restricts the parameter range of $A$ considered. The second is a technical restriction. By focussing on variations of $A$, we investigate the effects of applying a constant scaling factor on the concavity of the investment function for all levels of $K$. For this set of functions and the considered range of $A$, we always observe that investment incentives under PT are reduced and that total downstream volume is decreased[38]. While we take away that the results of propositions 2.4.1 are likely to

[38] Another set of concave investment functions, e.g. $\delta(K) = K^{\alpha},\ \alpha < 1$, could be considered by varying $\alpha$. This would non-linearly scale the concavity of the investment function, i.e. change the concavity of the function differently at differently levels of $K$. However, no closed-form solution is available for this specification and, even with numerical methods, we could not prove that no counterexamples exist.

occur, we do not exclude the possibility that the obtained results may fail for alternative specifications of the innovation investment function. We keep this caveat in mind for the next section.

### 2.4.2 Pool joining incentives

We investigate how the incentives of firms to join the patent pool are affected by the existence of a PT clause in the pool's policy (corresponding to $t = 0$ of our game). For this analysis, we briefly clarify two aspects of our understanding upfront. First, we understand that PT may be invoked by a pool member only if the PT clause features in the pool's licensing agreement. Second, firms not member of the patent pool are not bound by the pool's general FRAND commitment and are thus not restricted from using their essential patents as leverage in bargaining situations (i.e. they may cut the licences for their essential patent in a manner similar to that of PT)[39].

**Proposition 2.4.2** *The partial termination clause (under a property rule for essential patents outside of the pool) increases the value of joining the patent pool for vertically integrated firms, but not for pure upstream firms.*

**Proof** :
*For pure upstream firms:* Proposition 2.4.1 gives that a credible PT threat results in a decrease of the total downstream quantity. This effect hurts upstream firms since fewer unit licences for the essential patents are taken out. Consequently, a pure upstream firm will never invoke PT. Furthermore, the loss of upstream revenue due the existence of a PT threat negatively affects licensors of essential patents in and outside of the pool equally. Therefore, the inclusion of PT in the pool licence does not affect pure upstream firms' incentives to join or leave the patent pool.

*For vertically integrated firms:* We have just seen that the PT threat hurts licensors of essential patents, i.e. also the upstream part of a vertically integrated firm. Earlier, proposition 2.3.2 showed that the PT threat benefits the downstream division of vertically integrated pool members and, moreover, that the downstream gain outweighed the upstream cost[40]. Therefore, the PT clause has a net positive value for a vertically integrated firm, who will use PT when possible. Invoking PT is possible as a pool outsider

[39]Abstraction made of any additional commitments, such as individual FRAND commitments, which are often demanded for the inclusion of patents into a standard.

[40]The constraint given by equation 2.20 on page 79 must be satisfied. It states that $I$ must find it profitable to invoke PT (otherwise the threat would not be credible). We show that this is the case for any non-drastic innovation.

or as a pool member when PT features in the pool's licensing agreement. Consequently by joining a patent pool that does not allow PT, the vertically integrated firm loses in terms of its bargaining power vis-à-vis cumulative innovators. Hence compared to a pool without PT, a pool with PT should attract more vertically integrated firms. □

We note that proposition 2.4.2 holds even when the PT threat results in an increase of the total downstream volume[41]. We also note that the argument is robust to the situation, where a standard requires a FRAND commitment by all owners of essential IP, regardless whether they join the patent pool or not[42]. Finally, we note that this analysis does not provide information on which way the causality runs: Whether vertically integrated firms influenced the inclusion of a partial termination clause in the pool agreements or whether the existence of the PT clause drew in proportionally more vertically integrated firms[43].

## 2.5 Discussion

### 2.5.1 Findings

The context of the analysis is the following. In 1997, the MPEG-2 pool was inaugurated and the Department of Justice waived the partial termination clause in the pool licensing agreements. Outside of the pool, essential patent owners enjoyed a property right on their inventions. In 2006, with the *eBay* decision, the Department of Justice changed its

[41] In this case, it is evident to see that the joining incentives for a vertically integrated firm to join the patent pool are increased even more by the PT clause since the firm now benefits both up and downstream from invoking PT. For the pure upstream firms, the argument is as follows: First, note that PT is only useful for upstream firms, when there exists a downstream quantity effect, i.e. when a vertically integrated pool member uses it to increase total downstream volume or the pure upstream firm itself becomes a producer. Thus, consider a patent pool with at least one vertically integrated firm. Proposition 2.3.2 showed that incentives to invoke PT are always higher for vertically integrated firms than for pure upstream firms. Hence, ceteris paribus, the pure upstream firm will never invoke PT (even when profitable for the upstream firm), since the threshold for invoking PT is satisfied earlier for vertically integrated firms who will invoke PT. Furthermore, pure upstream firms benefit equally in and outside of the patent pool. Therefore, we conclude that the inclusion of PT in the pool licence does not affect pure upstream firms' incentives to join or leave the patent pool. This is because a pure upstream firm will not invoke PT itself and the cost or benefit of the PT threat is the same in or outside of the patent pool.

[42] This is often the case for standard setting organisations, e.g. ETSI and IEEE (since 02/2015). In this case, the pool joining incentives of a vertically integrated firm are increased even more by the PT clause, since the value of staying outside of the pool is reduced.

[43] The fact that for most standards (incl. MPEG, but also IEEE and ETSI), the policy is fixed before the standard is set up would suggest that the latter relationship is more probable. However, we do not have information on the type of stakeholders who took part in the decision process of the MPEG policy.

rule on standard essential patent holders and imposed a general liability rule on them for the licensing of essential patents.

The empirical puzzles from the literature, which were introduced in the introduction, rely on data from the time-period before the *eBay* decision. Our model provides a possible explanation for these.

Vakili (2012) empirically looks at the impact of the MPEG-2 pool formation on the innovation activities of firms that are in the technological proximity of the pool. He finds that the innovation rate of technologically proximate firms drops by 35% as a result of the MPEG-2 pool formation. Proposition 2.3.2 shows that the threat of partial termination adversely affects innovation incentives of the pool licensees and hence can contribute to explaining this phenomenon. The effect in our model goes through a reduction of the marginal return on investment in innovation.

Layne-Farrar & Lerner (2011) empirically investigate the joining incentives of firms with regards to patent pools and the effects of the different types of rent sharing rules. They investigate nine different patent pools, only one of which has recourse to PT, namely the MPEG-2 pool. They find that in the MPEG-2 pool, 89% of the firms are vertically integrated, where as only 25% of the outsiders are vertically integrated. The average over all pools, including the MPEG-2 pool, is 83% of firms within the pool and 32% of firms outside of the pool are vertically integrated[44]. Proposition 2.4.2 shows that the partial termination clause increases pool joining incentives for vertically integrated firms only. This result could help explaining the above average participation of vertically integrated firms in the MPEG-2 pool.

Proposition 2.3.3, however, shows that the threat of PT is not credible any more since the *eBay* decision. We therefore predict that these observations should disappear in data after 2006. Unfortunately, we do not have any data for this latest period. It would be interesting to track these observations in an event study of the policy changes following the *eBay* decision.

[44] Removing the MPEG-2 pool from the overall average would only exacerbate the difference.

### 2.5.2 Limitations and robustness of the model

We address an important simplification of our model set-up. We have assumed a zero license fee on the essential patent ($l = 0$). This simplification allows us to ignore situations when *not* invoking partial termination is profitable for the incumbent.

Vertical integration of the incumbent, i.e. the incumbent does not need to pay for a license to its own essential patent, erodes the double mark-up problem for the incumbent himself. By assuming that $l = 0$, we also abstract of the double mark-up problem for the entrant. At the same time, this assumption forces the incumbent to rely exclusively on the downstream market since there are no profits to be made upstream. Thereby, we look at the boundary case, where downstream considerations are the sole focus of the incumbent's profit maximisation programme and focus on the case where $I$'s incentives to invoke partial termination are highest.

We relax this assumption and allow a positive license fee ($l > 0$) to be charged by the incumbent in the game. This royalty remains exogenous and cannot be adapted in response to play in the game. When $l > 0$, $I$ now earns positive upstream profits and $E$ needs to pay for its use of the pool bundle. The payoff functions are adjusted accordingly. For example in the status quo (before $E$'s innovation), the objective functions are $\Pi_I(0, -l) = (a - Q - c)q_I + lq_E$ and $\Pi_E(-l, 0) = (a - Q - c - l)q_E$ and similarly for the other model situations[45].

**Proposition 2.5.1 (Robustness against $l > 0$)** *Given a positive licence fee ($l$) on the essential patent, the threat of partial termination remains credible and reduces the innovation incentives of the entrant, when the following conditions hold:*

*(i) the licence fee on the essential patent is not too high ($0 < l \leq \min[\frac{a-c}{2}, 2(a-c-\delta)]$),*

*(ii) the cumulative innovation is of sufficient quality ($l < \delta$) and*

*(iii) the marginal cost distortion due to licensing rates is not too large ($\delta \leq \frac{5(a-c)}{2} - 5l$).*

*Furthermore, the strength of the partial termination threat ($\hat{r} - \hat{r}^{PT}$) is reduced when $l > 0$ compared to the situation where $l = 0$.*

**Proof** See the appendix. □

[45]Given licensing, the payoff functions are: $\Pi_I^L(\delta - \hat{r}, \delta - l) = (a - Q^L - c + \delta - \hat{r})q_I^L + lq_E^L$ and $\Pi_E^L(\delta - l, \delta - \hat{r}) = (a - Q^L - c + \delta - l)q_E^L + \hat{r}q_I^L$. In the absence of licensing the payoffs are $\Pi_I^N(0, \delta - l) = (a - Q^N - c + \delta - \hat{r})q_I^N + lq_E^N$ and $\Pi_E^N(\delta - l, 0) = (a - Q^N - c + \delta - l)q_E^N + \hat{r}q_I^N$.

When the above conditions are given, the insights from our model hold in a setting, where the pool license is costly. The intuition for the result is the following. Condition (i) guarantees a downstream duopoly situation before and after the development of the related patent by E, when there is no threat of PT. This is the required setting for our analysis of PT. When the entrant innovates, the better the entrant's innovation (in terms of the cost reduction $\delta$), the more the entrant's downstream quantity in the Cournot game will be increased (cannibalising some of the incumbent's sales). The incumbent feels this in two ways. First, downstream profits decrease due to a lower volume being supplied by the incumbent himself. Second, upstream profits increase whenever $l > 0$, since the entrant will pay more unit licences for the use of the essential patent.

When the threat of PT is credible, the effects are again twofold and in the opposite direction of those just described. The effect of PT is pro-downstream and anti-upstream profits for the incumbent. Hence, it is clear that cases may exist, when the incumbent has no incentive to use partial termination (unlike in the baseline model where PT was always profitable). For example, this is the case when the profit margin on the license for the essential patent is sufficiently large compared to the downstream margin of the incumbent. Then, instead of invoking PT, allowing the entrant to become a monopolist would generate a sufficient increase in the number of paid licences for the essential patent such that the incumbent finds exit of the downstream market profitable. Therefore, the PT threat is only credible if the downstream benefits outweigh the upstream costs of invoking PT for the incumbent (point *(iii)* of proposition 2.5.1).

Furthermore, we note that the ability of the incumbent to invoke PT has been reduced compared to the baseline model with respect to condition 2 of the PT clause. Condition 2 stipulates that the incumbent's unit licensing revenue ($l$) should be used as a benchmark to judge the reasonability of the royalty proposal ($\hat{r}$) by the entrant. Only when the royalty offer on the related patent exceeds the unit licensing revenue of an essential patent ($\hat{r} > l$), the PT clause may be invoked. Therefore given the optimal pricing of the related patent by the entrant ($\hat{r} = \delta$), the incumbent is only able to invoke PT on related patents of sufficient quality (point *(ii)* of proposition 2.5.1).

Overall, we see that the results of our model are robust to the introduction of a positive licence rate on the essential patent (assuming $l$ is not too large and different from $\delta$). Both conditions (ii) and (iii) given in proposition 2.5.1 indicate that the utility of partial termination decreases as the size of the pool licence increases. In addition, we see that the strength ($\hat{r} - \hat{r}^{PT}$) of the partial termination clause has decreased due to the relaxed requirement of condition 2 (FRAND criterion applied to the related patent). This aspect highlights the importance of the definition of FRAND reciprocity as applied to related patents, because it restricts the incumbent in his use of PT.

### 2.5.3 Restrictions on the use of partial termination

We have already mentioned that conditions 1 and 2 of the PT clause create a safe haven for the innovator. With respect to condition 2, the size of the haven depends crucially on the definition of reciprocal FRAND for the related patent. For the MPEG-2 pool, the FRAND limit for the size of the royalty of the related patent is defined equal to the incumbent's per patent revenue from licensing the pool bundle. This aspect merits attention. Given a constant price for the pool licence, we have that the larger the patent pool, the more important is the threat of PT. Applied to the MPEG-2 patent pool, this argument is interesting, because since its inauguration, the number of declared standard essential patents has increased significantly, while its royalty rate has not been adjusted. Currently, absent disputes, member firms only need to mutually agree on which patents are to be included in the portfolio licence. This leaves scope for collusion, one possible side-effect of which would be a strengthening of the threat of partial termination. This result supports the view that standard setting organisations and patent pools should take up responsibility in checking the essentiality and validity of notified patents.

With respect to condition 1, the Department of Justice (1997) restricted the use of the PT clause in its review letter in that the essential patent holder may not force legal action by deliberate infringement of the related patent. Our model does not address this aspect by assuming that the condition is given exogenously. From a legal perspective, the question of the validity of the PT clause is controversial and feeds into a larger debate. For a brief account of the legal questions with respect to condition 1, see appendix 2.7.2.

Lastly, our model supports the Department of Justice's opinion that licensors are themselves restrained in their use of partial termination. Proposition 2.5.1 indicates that licensors' dependency on upstream profits reduces their incentives to invoke partial termination. Furthermore, we note that incentives to abstain from using partial termination in order to promote the development of the standard may exist, e.g. if a reputation of not using partial termination generates an exogenous demand increase for the standard. These aspects are not accounted for in our model.

## 2.6 Conclusion

When a partial termination clause is available to members of a patent pool, we show that two effects arise provided that a property rule applies to owners of essential patents outside of the pool: (i) incentives to invest in cumulative innovation related to the standard decrease for licensees of the pool and (ii) more vertically integrated firms find it profitable to join the pool as compared to a patent pool that does not allow such a clause. These predictions match some observations from the literature. This paper also shows that the threat of partial termination is no longer credible when a liability rule applies to standard essential patent holders (i.e. since the *eBay* decision in 2006).

Finally, we note that this work contributes to two current legal debates. First, this work contributes to the debate on reciprocal FRAND conditions. We highlight that the level of the reasonability criterion, which applies to cumulative innovations, can be used as a policy tool to trade-off investment incentives of cumulative innovators and the ability of pool licensors to capture some of the cumulative innovators' profits. This work is thus in line with many others (e.g. Layne-Farrar et al. (2007), Cotter (2013), etc.) arguing for more legal certainty on the implications of FRAND.

Second, this work also contributes to the larger debate on whether patent pools (or standard-setting organisations) should have a liability in determining the validity and essentiality of patents included in the patent pool bundle (or standard). The case of the partial termination clause shows that, in the absence of objective, independent controls for the validity and essentiality of patents included in a pool, other clauses in the licensing agreements may have soft constraints. In our example, the collective ability of pool members to overdeclare essential patents reduces the effectiveness of a constraint (i.e. condition 2 of partial termination), because the constraint itself is based on the number of declared essential patents.

# 2.7 Appendix

## 2.7.1 Technical details on partial termination

Partial termination[46] strictly speaking refers to section 6.3 of the MPEG-2 Patent Portfolio License (PPL) between the Licensing Administrator (MPEG LA) and a Licensee. It is as given in definition 6 in the paper and concerns both essential and related MPEG-2 patent(s). Being part of the PPL, Section 6.3 can be invoked only against an *actual* (not a prospective) Licensee. It is rooted in Section 2.8 of the MPEG-2 Agreement Among Licensors (AAL) which provides for the right of a Licensor to instruct the Licensing Administrator regarding the termination provided for in Section 6.3 of the PPL.

Furthermore, any Licensee is free to add MPEG-2 essential patents to the Portfolio that it or an affiliate may own on the same terms and conditions as all other Licensors (Section 7.4, "grantback"). If a Licensee chooses not to do so, however, it must agree to license such patents to any Licensor or Licensee on fair and reasonable terms (one example of which is the Licensors' per patent share of royalties payable under the License). Note that by definition, section 7.3 does not apply to related patents.

A second contractual tool with a very similar idea to PT is an agreement only among licensors, not with licensees, contained in sections 2.9 and 2.10 of the AAL. Section 2.9 allows a Licensor to instruct the Licensing Administrator to exclude any of its MPEG-2 Patent Portfolio Patent(s) from an MPEG-2 Patent Portfolio License to a *prospective* pool licensee that has brought a claim(s) for patent infringement of an MPEG-2 Essential Patent or an MPEG-2 Related Patent in a lawsuit or other proceeding against the Licensors, where the Licensor has decided to bring a claim(s) for infringement of the patent(s) to be excluded in the same or another lawsuit of other proceeding against the *prospective* pool licensee. Section 2.10 allows a Licensor to instruct the Licensing Administrator to exclude any of its MPEG-2 Patent Portfolio Patent(s) from an MPEG-2 Patent Portfolio License to a *prospective* sublicensee against whom the Licensor has brought a claim for infringement of the patents to be excluded. Unlike Section 6.3 of the PPL which can be invoked only against a Licensee, Sections 2.9 and 2.10 of the AAL would be exercised against a *prospective* Licensee.

[46] Many thanks go to Larry Horn and the MPEG-LA team for providing information (and some concise paraphrases used in the appendix) of the AAL and PPL contracts as well as making the originals available.

Section 2.10 of the AAL, unlike section 6.3 of the PPL, has been exercised in practice once.

### 2.7.2 Related legal question on the validity of PT

We indicate that from a legal perspective, the question of the validity of the PT clause is closely related to that of a standard essential patent holder seeking injunctive relief albeit a FRAND commitment against a willing licensee.

Recently, the Federal Trade Commission (FTC) emphasized its critical viewpoint on injunctive relief demands for FRAND-encumbered standard essential patents: "There is increasing judicial recognition, coinciding with the view of the Commission, of the tension between offering a FRAND commitment and seeking injunctive relief. Patent holders that seek injunctive relief against willing licensees of their FRAND-encumbered SEPs should understand that in appropriate cases the Commission can and will challenge this conduct as an unfair method of competition under Section 5 of the FTC Act. Importantly, stopping this conduct using a stand-alone Section 5 unfair methods of competition claim, rather than one based on the Sherman Act, minimizes the possibility of follow-on treble damages claims."[47].

For the partial termination clause, the competitive concerns are very similar. In our setting, the question is whether a licensee who sues for infringement of a related patent to the standard (where the related patent is not subject to a FRAND commitment), is considered a "willing" licensee according to the definition of the FTC. If so, the antitrust concerns for the PT clause should be immediate.

### 2.7.3 Proofs

**Proof of proposition 2.3.1**

We solve the game by backwards induction. In stage $t = 4$, Cournot competition arises with the proposed royalty rate on the related patent considered fixed.

*No licensing (upper index N):* When $I$ rejects the proposed royalty, the equilibrium quantities are (given by maximising $\Pi_i^N(.,.)$ with respect to $q_i$ $\forall i = \{I, E\}$): $q_I^N = \frac{a-c-\delta}{3}$, $q_E^N = \frac{a-c+2\delta}{3}$ and $Q^N = \frac{2(a-c)+\delta}{3}$. Equilibrium payoffs to the firms are

[47]Source: Statement of the Federal Trade Commission *in the matter of Robert Bosch GmbH* FTC File Number 121-0081 (Nov. 26, 2012).

$$\Pi_I^N(0,\delta) = (\frac{a-c-\delta}{3})^2 \text{ and } \Pi_E^N(\delta,0) = (\frac{a-c+2\delta}{3})^2 \tag{2.11}$$

*With licensing (upper index L):* When I accepts the proposed royalty $\hat{r}$ for a licence on the related patent, equilibrium quantities in the Cournot game are (given by maximising $\Pi_i^L(.,.)$ with respect to $q_i \forall i$): $q_I^{\mathrm{L}} = \frac{a-c-2\hat{r}+\delta}{3}$, $q_E^{\mathrm{L}} = \frac{a-c+\hat{r}+\delta}{3}$ and $Q^{\mathrm{L}} = \frac{2(a-c+\delta)-\hat{r}}{3}$. Corresponding equilibrium profits are:

$$\Pi_I^L(\delta-\hat{r},\delta) = (\frac{a-c+\delta}{3})^2 - \frac{4}{9}\hat{r}(a-c+\delta-\hat{r}) \tag{2.12}$$

$$\Pi_E^L(\delta,\delta-\hat{r}) = (\frac{a-c+\delta}{3})^2 + \frac{5}{9}\hat{r}(a-c+\delta-\hat{r}) \tag{2.13}$$

In $t=3$, $E$ will thus choose its royalty according to the following maximisation program in equation 2.14. Constraint 2.15 refers to the fact that $I$ must be better off under licensing than in the absence of licensing. Constraint 2.16 means that $E$ must earn more given licensing than in the absence of licensing.

$$\hat{r} = \arg\max_r \Pi_E^L(\delta,\delta-r) \tag{2.14}$$

$$\text{s.t.} \quad \Pi_I^L(\delta-r,\delta) \geq \Pi_I^N(0,\delta) \tag{2.15}$$

$$\Pi_E^L(\delta,\delta-r) \geq \Pi_E^N(\delta,0) \tag{2.16}$$

First, we find the maximum of equation 2.14 at $r = \frac{a-c+\delta}{2}$ and see that $\frac{\partial \Pi_E^N(\delta,\delta-r)}{\partial r} > 0$, if $r < \frac{a-c+\delta}{2}$. Constraint 2.15 gives that $I$ will accept $r$ iff the following holds:

$$(\frac{a-c+\delta}{3})^2 - \frac{4r}{9}(a-c+\delta) + \frac{4r^2}{9} \geq (\frac{a-c-\delta}{3})^2 \tag{2.17}$$

which is satisfied for $r \leq \delta$. This is smaller than the optimal royalty for $E$ since $\delta < \frac{a-c+\delta}{2}$, if $\delta < a-c$. Constraint 2.16 is satisfied at $r=\delta$ for all $0 < \delta < a-c$, since

$$(\frac{a-c+\delta}{3})^2 + \frac{5}{9}\delta(a-c) \geq (\frac{a-c+2\delta}{3})^2 \tag{2.18}$$

is true. Thus in equilibrium there exists a unique licensing equilibrium for any non-drastic innovation $0 < \delta < a-c$, where $E$ proposes a royalty rate $\hat{r} = \delta$, which will be accepted by $I$. Payoffs to each firm are $\Pi_I^L(\delta-\hat{r},\delta) = (\frac{a-c+\delta}{3})^2 - \frac{4}{9}\delta(a-c) = \Pi_I^N(0,\delta)$, and $\Pi_E^L(\delta,\delta-\hat{r}) = (\frac{a-c+\delta}{3})^2 + \frac{5}{9}\delta(a-c) > \Pi_E^N(\delta,0)$.

*The innovation incentives* to $E$ are given by $\Phi_E(K) = \Pi_E^L(\delta, \delta - \hat{r}) - \Pi_E^L(0,0) - K$. The optimal investment level $\hat{K}$ is given by $\hat{K} = \arg\max_K \ \Phi_E(K) \quad \text{if } \Phi_E(\hat{K}) > 0$ and $\hat{K} = 0$ otherwise.

Precisely, $\Phi_E(K) = (\frac{a-c+\delta(K)}{3})^2 + \frac{5}{9}\delta(K)(a-c) - \left(\frac{a-c}{3}\right)^2 - K$ and we always have an interior solution for $\hat{K}$ given our assumptions on the investment, i.e. $\Phi_E(K) > 0$ for $K > 0$. At the optimum, the following first order condition holds when $E$ invests a positive amount $\hat{K}$ in innovation :

$$\Big[\frac{7}{9}(a-c) + \frac{2}{9}\delta(\hat{K})\Big]\frac{\partial \delta}{\partial \hat{K}} = 1 \tag{2.19}$$

□

**Proof of proposition 2.3.2**

*Outside option:* It is immediate to see that, even if excluded downstream, $E$ wants to license out its related patent for any $\hat{r}^{\mathrm{PT}} > 0$ since $\hat{r}^{\mathrm{PT}} q_I^{\mathrm{PT}} > 0$. $I$ accepts any licence (even non-FRAND) iff $(a - q_I^{\mathrm{PT}} - c + \delta - \hat{r}^{\mathrm{PT}})q_I^{\mathrm{PT}} > (a - q_I^{\mathrm{PT}} - c)q_I^{\mathrm{PT}}$, which is satisfied for all $\hat{r}^{\mathrm{PT}} \leq \delta$. Following the same argument as for propostion 2.3.1, $E$ proposes $\hat{r}^{\mathrm{PT}} = \delta$ in equilibrium, which will be accepted by $I$. If $E$ was to be excluded, the monopolist $I$ would produce $q_I^{\mathrm{PT}} = \frac{a-c}{2}$ downstream. Thus, if $E$ is excluded in equilibrium, payoffs to the firms are (we drop the upper index L): $\Pi_E^{\mathrm{PT}} = \delta(\frac{a-c}{2})$ and $\Pi_I^{\mathrm{PT}} = (\frac{a-c}{2})^2$.

We see that $E$'s outside option payoff has decreased under PT since for $0 < \delta < a - c$: $\Pi_E^N(\delta, 0) - \Pi_E^{\mathrm{PT}} = (\frac{a-c+2\delta}{3})^2 - \delta(\frac{a-c}{2}) = \frac{1}{18}[(a-c)(a-c-\delta) + (a-c)^2] + \frac{4}{9}\delta^2 > 0$. Therefore, constraint 2.16 of $E$'s maximisation programme (defining the lower bound of $\hat{r}$) is relaxed.

We see that $I$ prefers to invoke PT rather then not get a licence for the related patent since $\Pi_I^{\mathrm{PT}} > \Pi_I^N(0, \delta), \quad \text{if } \delta < a - c$. Furthermore, $I$ is always better off using PT rather than licensing (accepting any positive or zero royalty fee), because the monopoly is more profitable. This is shown by the following: $\frac{\partial \Pi_I^L(\delta - r, \delta)}{\partial r} < 0$ for $0 < \delta < a - c$ and $\Pi_I^{\mathrm{PT}} - \Pi_I^L(\delta - r, \delta)$

$$= (\frac{a-c}{2})^2 - \big[(\frac{a-c+\delta}{3})^2 - \frac{4}{9}r(a-c+\delta-r)\big] > 0, \quad \text{if } r \geq 0 \text{ and } \delta < a - c \tag{2.20}$$

Thus constraint 2.15 of $E$'s maximisation programme (defining the upper bound of $\hat{r}$) is never satisfied and $I$ invokes PT whenever possible ($r > 0$). The new proposed royalty rate under PT ($\hat{r}^{\mathrm{PT}}$) is hence defined by

$$\hat{r}^{\mathrm{PT}} = \arg \max_r \ \Pi_E^L(\delta, \delta - r) \tag{2.21}$$

$$\text{s.t.} \quad \Pi_I^L(\delta - r, \delta) \geq \begin{cases} \Pi_I^{\text{N, PT}} & \text{if } r > \delta \\ \Pi_I^{\text{L, PT}} & \text{if } \delta \geq r > 0 \\ \Pi_I^N(0, \delta) & \text{if } r = 0 \end{cases} \tag{2.22}$$

$$\text{s.t.} \quad \Pi_E^L(\delta, \delta - r) \geq \begin{cases} \Pi_E^{\text{PT}} = 0 & \text{if } r > \delta \\ \Pi_E^{\text{PT}} = r(\frac{a-c}{2}) & \text{if } \delta \geq r > 0 \\ \Pi_E^{\text{PT}} = 0 & \text{if } r = 0 \end{cases} \tag{2.23}$$

From equation 2.20, we know that $I$ will always invoke PT when possible ($r > 0$). $E$ thus decides whether it proposes $r = 0$ and stay in the downstream market or whether it proposes $r = \delta$ and is excluded from the downstream market. E proposes $r = 0$ iff

$$\Pi_E^L(\delta, \delta) = (\frac{a - c + \delta}{3})^2 > \delta(\frac{a - c}{2}) = \Pi_E^{\text{PT}} \tag{2.24}$$

which is satisfied iff $\delta < (\frac{a-c}{2})$. We thus obtain two cases in equilibrium: For all $0 < \delta < (\frac{a-c}{2})$, $E$ prefers to license out its patent for free ($\therefore \hat{r}^{PT} = 0$) in order to prevent invoking of PT (and to stay in the downstream market). For all $(a - c) > \delta > (\frac{a-c}{2})$, $E$ proposes the optimal royalty $\hat{r}^{PT} = \delta$ and becomes a pure upstream licensor. Quantities and payoffs to firms in the two cases are

$$q_i^{PT} = \frac{a - c + \delta}{3} \text{ and } \Pi_i^{PT} = (\frac{a - c + \delta}{3})^2 \quad \forall i = \{I, E\} \qquad \text{if } 0 < \delta \leq \frac{a - c}{2} \tag{2.25}$$

$$q_I^{PT} = \frac{a - c}{2},\ q_E^{PT} = 0 \text{ and } \Pi_I^{PT} = (\frac{a - c}{2})^2,\ \Pi_E^{PT} = \delta(\frac{a - c}{2}) \qquad \text{if } \frac{a - c}{2} < \delta < (a - c) \tag{2.26}$$

*The innovation incentives* to $E$ are given by

$$\Phi_E^{\text{PT}}(K) = \Pi_E^{PT} - \left(\frac{a - c}{3}\right)^2 - K = \begin{cases} (\frac{a-c+\delta}{3})^2 - \left(\frac{a-c}{3}\right)^2 - K & \text{if } 0 < \delta \leq \frac{a-c}{2} \\ \delta(\frac{a-c}{2}) - \left(\frac{a-c}{3}\right)^2 - K & \text{if } \frac{a-c}{2} < \delta < \frac{a-c}{1} \end{cases} \tag{2.27}$$

At the optimum, the following first order conditions define the investment level $\hat{K}_{\text{PT}}$:

$$\left[\frac{2}{9}(a - c + \delta(\hat{K}_{\text{PT}})\right] \frac{\partial \delta}{\partial \hat{K}_{\text{PT}}} = 1 \qquad \text{if } 0 < \delta(\hat{K}_{\text{PT}}) \leq \frac{a - c}{2} \tag{2.28}$$

$$\left(\frac{a - c}{2}\right) \frac{\partial \delta}{\partial \hat{K}_{\text{PT}}} = 1 \qquad \text{if } \frac{a - c}{2} < \delta(\hat{K}_{\text{PT}}) < (a - c) \tag{2.29}$$

The following reasoning applies to both cases: Both left-hand-side terms of the FOC are positive. Compared to the situation without PT, the following changes occur: Given constant $\delta$, the first term falls, because: $\frac{2}{9}(a-c+\delta) < \frac{7}{9}(a-c)+\frac{2}{9}\delta$ and $\frac{a-c}{2} < \frac{7}{9}(a-c)+\frac{2}{9}\delta$. The constant on the right hand side is unaffected. Therefore, the second term $\delta'(K)$ must rise in order to balance the equation. Given that $\delta''(K) < 0$, we have that $\hat{K}_{\mathrm{PT}} < \hat{K}$. $\square$

**Proof of propostion 2.4.1**

We impose $\delta(K) = A\sqrt{K}$. This implies $\delta'(K) = \frac{A}{2\sqrt{K}}$.

*Absent a threat of PT (use subscript* 0 *for this case:)* Using the FOC in equation 2.19, we obtain the optimal investment level and resulting quality of the innovation as, respectively: $\hat{K}_0 = (A\frac{\frac{7}{2}(a-c)}{9-A^2})^2$ and $\delta_0 = (A^2)(\frac{\frac{7}{2}(a-c)}{9-A^2})$.

*Conditions on A:* The conditions for positive and non-drastic cumulative innovations only are given by (i) $0 < \delta_0$ and (ii) $\delta_0 < a-c$. In terms of $A$, we require for (i): $A < 3$ and for (ii): $A < \sqrt{2}$. The latter is more stringent. Proposition 2.3.1 thus holds for $0 < A < \sqrt{2}$.

*Given a credible PT threat:* Here, we have to differentiate for "low" and "high" realised qualities of the cumulative innovation.

(Use subscript 1 for this case:) When the realised quality is low ($0 < \delta_{PT} \leq \frac{a-c}{2}$), the FOC in (2.28) defines the optimal investment and realised quality: $\hat{K}_1 = (A\frac{(a-c)}{9-A^2})^2$ and $\delta_1 = (A^2)(\frac{a-c}{9-A^2})$. The required conditions are

For $\delta_1 > 0$, need $A < 3$

For $\delta_1 < \delta_0$, need $A > 0$

For $\delta_1 < \frac{a-c}{2}$, need $A < \sqrt{3}$.

(Use subscript 2 for this case:) When the realised quality is high ($\frac{a-c}{2} < \delta_{PT} < a-c$), the FOC in (2.29) defines the optimal investment and realised quality: $\hat{K}_2 = \frac{A^2}{1}(a-c)^2$ and $\delta_2 = \frac{A^2}{4}(a-c)$. The required conditions are:

For $\delta_2 > 0$, need $A < 0$

For $\delta_2 < \delta_0$, need $A < 3$

For $\delta_2 > \frac{a-c}{2}$, need $A > \sqrt{2}$.

We see that this is outside the allowable range $0 < A < \sqrt{2}$. Hence, $E$ will never choose $K$, such that $\delta_{PT} > \frac{a-c}{2}$. The equilibrium levels are thus $K_1$ and $\delta_1$ and from above we know that innovation incentives for $E$ have decreased. Therefore, we confirm that proposition 2.3.2 holds for $0 < A < \sqrt{2}$.

*Net effect on total downstream volume* Total downstream volume decreases when $Q^L > Q^{PT} \Leftrightarrow \frac{2(a-c)+\delta_0}{3} > \frac{2(a-c+\delta_1)}{3} \Leftrightarrow \delta_0 > 2\delta_1 \Leftrightarrow (A^2)(\frac{\frac{7}{2}(a-c)}{9-A^2}) > 2(A^2)(\frac{a-c}{9-A^2}) \Leftrightarrow \frac{7}{2} > 2$, which is true for all $0 < A < 3$. $\square$

**Proof of proposition 2.5.1**

To show that the effects of partial termination are robust to $l > 0$, we need to show robustness of the following: First, the threat of PT must be credible. $I$ must find it profitable to invoke PT. Second, for PT to reduce the innovation incentives of $E$, the threat of PT must be costly to $E$. We first give the explicit expressions of payoffs of the firms in different scenarios:

*In the status quo,* equilibrium quantities and profits are given by (superscripts $^S$): $q_I^S = (\frac{a-c+l}{3})$, $q_E^S = (\frac{a-c-2l}{3})$, $\Pi_I^S(0,-l) = (\frac{a-c+l}{3})^2 + l(\frac{a-c-2l}{3})$, $\Pi_E^S(-l,0) = (\frac{a-c-2l}{3})^2$ and both firms produce positive quantities downstream iff: $l < (\frac{a-c}{2})$.

*Under no licensing of the related patent,* the equilibrium quantities and profits are (superscripts $^N$): $q_I^N = (\frac{a-c+l-\delta}{3})$, $q_E^N = (\frac{a-c-2l+2\delta}{3})$, $\Pi_I^N(0,\delta-l) = (\frac{a-c+l-\delta}{3})^2 + l(\frac{a-c-2l+2\delta}{3})$, $\Pi_E^N(\delta-l,0) = (\frac{a-c-2l+2\delta}{3})^2$. We maintain a duopoly iff: $\delta < a-c$ (as before) and $l < \frac{a-c}{2}$.

*Under licensing of the related patent* at $\hat{r}$, equilibrium quantities and payoffs are given by: $q_I^L = (\frac{a-c+l-2\hat{r}+\delta}{3})$, $q_E^L = (\frac{a-c+\delta-2l+\hat{r}}{3})$, $\Pi_I^L(\delta-\hat{r},\delta-l) = (\frac{a-c+l-2\hat{r}+\delta}{3})^2 + l * (\frac{a-c+\delta-2l+\hat{r}}{3})$, $\Pi_E^L(\delta-l,\delta-\hat{r}) = (\frac{a-c+\delta-2l+\hat{r}}{3})^2 + \hat{r} * (\frac{a-c+l-2\hat{r}+\delta}{3})$.

The optimal royalty charged by $E$ for the cumulative innovation is unchanged at $\hat{r} = \delta$. To see that, see the adjusted maximisation programme of $E$'s royalty setting in $t = 3$ in equation 2.30. $I$ must be better off under licensing than in the absence of licensing (Constraint 2.31). $E$ must earn more given licensing than in the absence of licensing (Constraint 2.32).

$$\hat{r} = \arg\max_r \Pi_E^L(\delta-l,\delta-r) \tag{2.30}$$

$$\text{s.t.} \quad \Pi_I^L(\delta-r,\delta-l) \geq \Pi_I^N(0,\delta-l) \tag{2.31}$$

$$\Pi_E^L(\delta-l,\delta-r) \geq \Pi_E^N(\delta-l,0) \tag{2.32}$$

The optimal royalty rate in 2.30 is $r_{opt} = (\frac{a-c+\delta}{2}) - \frac{l}{10}$. We confirm that $\frac{\partial \Pi_E^L(\delta-l,\delta-r)}{\partial r} > 0 \quad$ if $r < r_{opt}$ and that the second derivative is negative. $I$ accepts $r$ iff

$$(\frac{a-c+l-2r+\delta}{3})^2 + l(\frac{a-c+\delta-2l+r}{3}) \geq (\frac{a-c+l-\delta}{3})^2 + l(\frac{a-c-2l+2\delta}{3})$$

which (Constraint 2.31) is satisfied for $r \leq \delta$. Furthermore, for $I$ to accept, we need $I$ to stay downstream post innovation and not change its status to an upstream licensor:

$$\Pi_I^N(0,\delta-l) \geq l * q_E^{M,N} \;\Leftrightarrow\; (\frac{a-c+l-\delta}{3})^2 + l(\frac{a-c-2l+2\delta}{3}) \geq l(\frac{a-c+\delta-l}{2}) \tag{2.33}$$

where $q_i^M(\delta)$ is the downstream monopoly quantity supplied by a firm $i$ when it has access to the related patent. Condition 2.33 is satisfied for $0 < l \leq 2(a-c-\delta)$. This condition guarantees that $\delta < r_{opt}$ since: $\delta \leq (\frac{a-c+\delta}{2}) - \frac{l}{10} \Leftrightarrow \delta < (a-c) - \frac{l}{5} \Leftrightarrow l < 5(a-c-\delta)$. Constraint 2.32 gives that $E$ accepts to licence iff $(\frac{a-c+\delta-2l+\hat{r}}{3})^2 + \hat{r} * (\frac{a-c+l-2\hat{r}+\delta}{3}) \geq (\frac{a-c-2l+2\delta}{3})^2$

$$\Leftrightarrow \hat{r} \geq \underbrace{\frac{a-c+\delta}{2} - \frac{l}{10}}_{=r_{opt}} - \frac{1}{10}\sqrt{\underbrace{((5(a-c)-l)^2 + 5\delta[2(a-c)+7(2l-\delta)])}_{>0}} \tag{2.34}$$

which is satisfied at $r = \delta$. Thus, absent PT, $\hat{r} = \delta$.

We now see under which conditions the effects of PT hold. For that we need the conditions outlined above: (i) a profitable threat of PT for $I$ and (ii) that the threat is costly to $E$.

<u>*Case (i)*</u> First, note that we have introduced the restriction necessary to exclude the case where $I$ finds it profitable to become a pure upstream licensor, i.e. $I$ does not want to quit the downstream market when conditions 2.33 is satisfied.

Second, $I$ finds invoking PT profitable iff: $\Pi_I^{PT} > \max[\Pi_I^N(0,\delta), \Pi_I^L(\delta-\hat{r},\delta-l)]$. The incentive to invoke PT for $I$ is given at $\hat{r} = \delta$, when $\Pi_I^{\text{PT}} \geq \Pi_I^L(\delta-r,\delta-l)$:

$$\Leftrightarrow (\frac{a-c}{2})^2 \geq (\frac{a-c+l-\delta}{3})^2 + l * (\frac{a-c+2\delta-2l}{3}) \tag{2.35}$$

which is satisfied for $l < \delta < \frac{5(a-c)}{2} - 5l$. The same condition guarantees that we have that $\Pi_I^{\text{PT}} > \Pi_I^N \;\; \Leftrightarrow (\frac{a-c}{2})^2 \geq (\frac{a-c+l-\delta}{3})^2 + l(\frac{a-c-2l+2\delta}{3})$. Thus, when $l < \delta < \frac{5(a-c)}{2} - 5l$, $I$ will invoke PT and the threat is credible.

<u>*Case (ii)*</u> The threat of PT must be costly to $E$ for it to reduce innovation incentives. Thus, we need: $\Pi_E^{PT} < \Pi_E^L(\delta-l,\delta-\hat{r})$, where $\Pi_E^{PT} = \max[\Pi_E^L(\delta-l,\delta-\hat{r}^{PT})\ , r_{UP}q_I^M]$. $q_I^M$ is the monopoly quantity of $I$ given $E$'s exclusion from the downstream market and $r_{UP}$ is the optimal price charged by $E$ if it is a pure upstream firm.

Focus on $\Pi_E^{PT} = r_{UP}q_I^M$ first: In this case, $E$ accepts exclusion from downstream and licenses at the rate maximising upstream profits only. It is trivial to see that $r_{UP} = \delta$ and $q_I^M = \frac{a-c}{2}$. Thus we need: $\Pi_E^L(\delta-l,\delta-\hat{r} > r_{UP}q_I^M \quad \Leftrightarrow \quad (\frac{a-c+\delta-2l+\hat{r}}{3})^2 + \hat{r}*(\frac{a-c+l-2\hat{r}+\delta}{3}) > \delta(\frac{a-c}{2})$, which is satisfied for : $l < \frac{a-c}{2}$.

Now focus on $\Pi_E^{PT} = \Pi_E^L(\delta - l, \delta - \hat{r}^{PT})$: Given that PT tightens constraint 2.31, i.e. $I$'s outside option increases. Therefore, if $E$ wants to keep licensing the cumulative innovation, it must reduce its royalty offer $\hat{r}^{PT} < \hat{r}$. We have seen before from our maximisation in 2.30, that $\frac{\partial \Pi_E^L(\delta - l, \delta - r)}{\partial r} > 0 \quad \text{if } r < r_{opt}$. Therefore, when $E$ reduces it royalty offer in order to maintain a licensing relation, it incurs a profit loss: $\therefore \Pi_E^L(\delta - l, \delta - \hat{r}^{PT}) < \Pi_E^L(\delta - l, \delta - \hat{r})$.

Hence, we see that the credible PT threat reduces the expected payoff to $E$ when innovating and thereby $E$'s incentives to invest in innovation. $\square$

# Chapter 3

# Investigating the Impact of Uncertainty on Firms with Dynamic Costs

## A Case Study on the French Electricity Market

## 3.1 Introduction

There exists a consensus that dynamic costs, also referred to as ramping or adjustment costs, are important on the electricity market[1]. These are the costs incurred by a producer when production varies. The importance of uncertainty for the expectation of dynamic costs is shown in Bergès and Martimort (2014). Uncertainty itself on the electricity market has been studied by Wolak (2007). We focus on two sources of uncertainty for traditional electricity suppliers, namely uncertainty about the realisation of the market demand and uncertainty from the inherently unpredictable meteorological situation (which affects renewables generation). We propose a methodology to measure this uncertainty and its impact on firm strategies on the electricity market.

Electricity as a market is very important in and of itself ($2 trillion in worldwide sales in 2010). It is also a crucial input for many industries; power outages induce very large costs to society (LaCommare and Eto (2004), Reichl et al. (2013)). The electricity market is,

[1] Anderson and Xu (2005), Hobbs (2001), Hortacsu and Puller (2008), Reguant (2014), Sewalt and De Jong (2003).

however, quite different from the markets for other commodities in a few respects. First, electricity cannot be efficiently stored. As a consequence, electricity markets are high frequency (prices can update down to 15-min intervals) and firm strategies are purer as they are free of stock management considerations.

Second and in addition to non-storability, a generation surplus cannot be disposed of freely[2]. Thus, generation of electricity must always be matched with consumption in real time (modulo a small tolerance). This represents a hard constraint on the market[3] and forces suppliers to be reactive. However, this reactivity is costly as plant operators incur dynamic costs when adjusting production and the larger the adjustment made, the larger the cost. Hence, suppliers face a trade-off between cheap generation of electricity and costly reactivity to the demand realisation. Indeed, no single generation technology exists that satisfies both cheap generation and sufficient reactivity to allow production fluctuations at a reasonable price. Existing generation techniques are either cheap and unresponsive, e.g. nuclear plants, or expensive and flexible, e.g. gas turbines.

Interestingly, we also observe negative prices. In France for example, during the weekend of the 15$^{\text{th}}$ June 2013, the price per MWh dropped to $-200$€. This contrasts to the yearly average of approx. 45€/MWh and is generally understood as a sign that subsidising consumption temporarily is cheaper for a supplier than shutting down a plant (EPEX, 2014)[4]. The increase of the share of renewable generation in the energy mix contributes to the occurrence of negative prices on the market. The intermittency of renewables causes large residual demand shocks (EPEX, 2014). The unreliability of renewable generation also means that more flexible plants (i.e. plants with lower dynamic costs) are required to provide rapid responses to fluctuations in production from renewables (REN21, 2013).

Furthermore, uncertainty arises from the fact that renewable production is a local and dispersed production, but feeds into a national market with a single price. When meteorological conditions change, the geographic production profile also changes. This further complicates the predictability of renewables generation and contributes to the uncer-

[2]The common assumption of free disposal as made in standard microeconomics is violated.

[3]Mismatches between consumption and generation ultimately result in power outages.

[4]"Negative prices are a price signal on the power wholesale market that occurs when a high inflexible power generation meets low demand. Inflexible power sources cannot be shut down and restarted in a quick and cost-efficient manner. Renewables do count in, as they are dependent from external factors (wind, sun)."

tainty that electricity producers face when playing on the electricity market (Meibom et al., 2009).

This paper explores the effect that the absolute level of uncertainty about residual demand has on players' strategies on the electricity market. In the light of the existence of dynamic costs, which are inherent to the production technologies, uncertainty is costly to suppliers (Bergès and Martimort, 2014). Thus when faced with uncertainty, we expect that electricity producers smooth production volume over time in order to minimise dynamic costs. In a single market interaction with a symmetric oligopoly and linear demand functions this translates to playing a steeper supply function when uncertainty is high. The detailed intuition behind the predictions tested is given in section 3.1.2.

We show that uncertainty does impact supplier strategies. However, this prediction and result only apply locally to the central, flat and linear part of the supply bid function. Towards the high and low volume extremities of the bid functions when capacity constraints start to matter, bid functions become vertical and the effect of uncertainty vanishes. Furthermore, we observe results that indicate that demand-side bidding is also impacted by uncertainty.

We focus on the French one-day ahead market, EPEX Spot. This market is a divisible goods auction and particularly suited for our analysis as we observe data on the full aggregate bid functions for both supply and demand. We introduce the market's auction format and rules in section 3.2. The dataset and its sources are presented in section 3.3.

For our investigation on the effect of uncertainty on bidding behaviour, we develop an identification methodology, which relies on the non-parametric, comparable point selection technique presented in section 3.4.1.

We present and interpret the results in section 3.5. Finally, we discuss some overarching points in section 3.6 and conclude in section 3.7.

### 3.1.1 Literature review and contribution

There exists a literature on supply function equilibria initiated by Klemperer and Meyer (1989). In traditional models, firms choose between quantities (Cournot) or prices (Bertrand) as their strategic quantities. In the intermediate case, firms choose a relationship between quantities and prices, namely a supply function. This is the focus of the supply function equilibrium models. A key ingredient of these models is uncertainty.

Supply function equilibrium models are very relevant for the analysis of electricity markets, since many electricity market designs allow firms to submit a price-volume function rather than a specific price or quantity. Green and Newbery (1992), Newbery (1998) and Bolle (1992) have used these models to analyse competition on the electricity markets. These papers have contributed to a broader investigation of the competition on the electricity markets, which has also been looked at from empirical perspectives (Wolfram, 1998; Borenstein et al., 2002). While those initial papers have focussed on the supply function equilibria of the market, they have abstracted from some technological specificities for the sake for simplification.

One such aspect that we are interested in and that has been the subject of research in recent years is the importance of dynamic costs for electricity production. Bergès and Martimort (2014) extend Klemperer and Meyer (1989) to derive predictions on firms facing dynamic costs in a supply function oligopoly under uncertainty. They find that when varying production is costly, suppliers take these costs into consideration by submitting steeper functions when facing more uncertainty, in order to limit the range of variation in production. Reguant (2014) develops a model and an empirical strategy to measure dynamic costs on the Spanish one-day-ahead electricity market. She finds that "complex bids", which allow firms to minimise dynamic costs by linking production in one time period to production in a subsequent time period, reduce the volatility and the level of prices on the market. Her work is also unique in terms of data availability. By using individual bid functions she is able to produce estimates of start-up and ramping costs per production technology. In order to quantify dynamic costs on the Australian electricity market, Wolak (2007) derives a methodology to recover estimates of the parameters of parametric cost functions at the level of the production unit. His identification is based on the assumption that each profit maximising supplier knows the distribution of shocks on the demand function when playing on the market. Uncertainty is thus an explicit ingredient of his paper and he captures two sources of uncertainty in a single index: (i) the uncertainty from not knowing the aggregate supply function served by all other suppliers and (ii) the uncertainty about the realisation of the market demand. The recovered cost functions quantify the cost of varying output. Forward contracts are useful to avoid output variations. By comparing the observed level of forward contracting (assumed to be the profit maximising choice for production variation) with the theoretical minimum cost production pattern, he does not find support for ramping costs.

We contribute to this literature by providing an empirical analysis of the French electricity market. Specifically, we look at the impact of uncertainty on supplier strategies and take this as evidence that dynamic costs matter. Our approach to separate out the uncertainty from market demand expectations and predictability of renewables generation is novel. Both proxies for uncertainty used are new, uncertainty from market demand is inferred from the prediction errors that firms make in a demand estimation and uncertainty from renewable production is computed in a bottom-up approach from local weather forecasts. Instead of opting for a time series regression, we understand all hourly auctions as a cross-sectional dataset and control for the time of the day by using continuous transition variables for daytime periods. Similarly, we control for seasonality using continuous variables rather than dummies. Thereby, we are able to leverage our dataset and increase the sample size for each of our regressions and improve the precision of our estimates.

Furthermore, our work contributes to the empirical literature testing strategic behaviour of market participants. Generally, these studies focus on point-wise analyses for reasons of data availability. Not only does this cause endogeneity problems when the data used is equilibrium data, but also the analysis is restricted to an understanding of the usually observed outcomes of the market. In our setting, we benefit from an interesting dataset in which we observe full aggregate bid functions of players. The functions describe the players' behaviour both in the region where the equilibrium is likely to occur as well as in regions that rarely have an impact on the equilibrium outcome. As such, they provide a much fuller description of the firms' strategies. The additional information contained in the full aggregate bid functions has been used extensively in theoretical work (notably in the supply function equilibria literature mentioned above). However, few papers exploit these full bid functions empirically. For the government bond market, Préget and Waelbroeck (2005) and Özcan (2004) use a parametric approach to this functional data for a description of the variation of bid functions with respect to exogenous factors and an investigation of the revenue superiority of the uniform or discriminatory multi-unit auction mechanism, respectively. While this parametric approach has worked well for the bond market, Belsunce (2011) shows that the technique is ill-suited for an analysis of the electricity market due to the strong heterogeneity of observed bid functions. Using a different approach, which relies on conditioning the analysis on different demand levels, Wolfram (1999) leaves the analysis of equilibrium data to investigate the duopoly power of firms on the UK day-ahead electricity spot market. She uses information from the whole aggregate supply function to investigate the impact of price caps for electricity producers. She shows that the introduction of price caps resulted in a counter-clockwise rotation of the aggregate supply function. She relates these results to produce a lower bound on the extent to which firms can increase their prices above marginal costs when regulatory pressure makes it advantageous to do so. Thereby, she contributes empirical evidence for the distorting effects of price caps.

Our work adds to this empirical literature using the information contained in the full bid functions by developing a non-parametric approach which allows to condition our analyses on multiple, representative points of the bid functions. The statistical ingredients rely on Silverman and Ramsay (2005) and are detailed in section 3.4.1. Thereby we are able to leverage our dataset, increase the sample size in individual regressions as well as obtain a fuller picture of the effects of exogenous variables on the behaviour of electricity producing firms. We emphasize that our approach allows to overcome structural restrictions underlying previous parametric approaches, e.g. the symmetry of the logistic function used in Belsunce (2011).

### 3.1.2 Theoretical prediction

We test the impact of uncertainty on supplier strategies by testing the prediction that suppliers bid steeper supply bid functions when faced with a larger uncertainty concerning the outcome of the (residual) demand realisation.

In a discontinuous setting, where the supplier produces volume $Q_H$ of electricity in hour $H$, we assume that he faces a cost function $C_i(.)$ for each production plant $i$. This cost function depends on both marginal costs of production as well as the dynamic costs for changing production rapidly: $C_i\big((Q_H), (Q_H - Q_{H-1})^2\big)$. The larger the variation in production between hours, the larger the dynamic costs. Even when the expected residual demand is constant, there are still fluctuations in the production due to possible shocks to the residual demand. The larger the shocks, the larger the change in production and thus the larger the dynamic costs. Consequently, increased uncertainty (as represented by shocks on the demand function) translates into increased expected dynamic costs. We assume that the profit maximising supplier knows the distribution of shocks on the demand function when choosing his supply function. In order to minimise these costs, the producer can choose a steeper supply function when uncertainty is high. We want to test this prediction.

We illustrate the intuition behind this prediction using a stylised case in figure 3.1. The graphs depict a situation in which a single, risk-neutral supplier bids a supply function to supply electricity in the hours 9 and 10 of the next day. For both hours, the supplier faces a constant expected residual demand function represented by $E(D)$. In a static optimisation problem, the supplier would bid a supply function $S_0$ in both auctions.

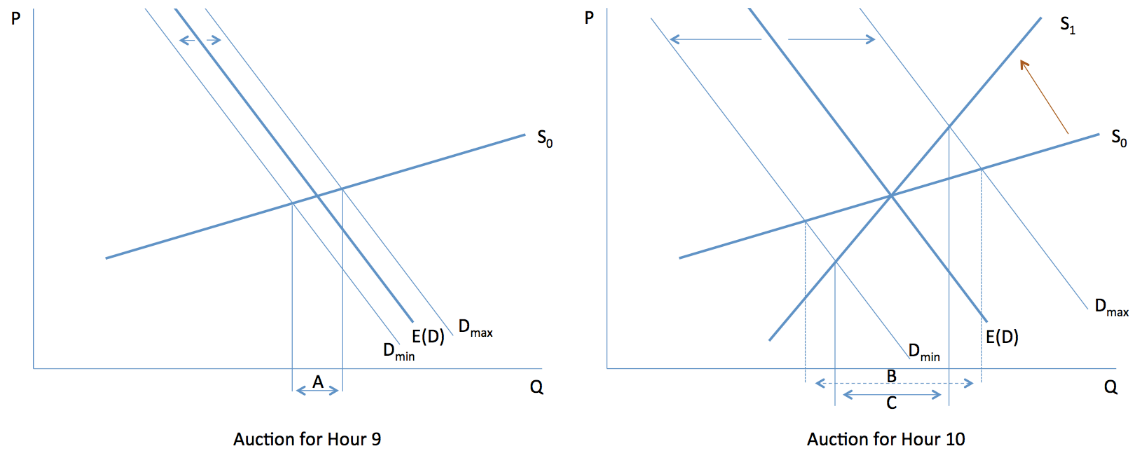



Figure 3.1: Illustrating the effect of increased uncertainty.

The uncertainty in the market is represented by the width of the envelope of shocks that affect the residual demand function (represented by the arrows on $E(D)$). Thus, in each hour, the residual demand fluctuates between $D_{min}$ and $D_{max}$, where the range between the extremal demands may vary from one hour to the next.

Before submitting a supply function to the market, the supplier estimates the distribution of probabilities of demand shocks that he will face. In hour 9, the supplier is able to rather precisely predict the realisation of the demand function in the auction, i.e. it realises within a tight confidence interval. In hour 10, however, uncertainty in predicting the outcome of the demand realisation has grown strongly as represented by the much wider confidence interval on the demand realisation.

Given a fixed supply bid function $S_0$, the possible range of quantities to be produced by the supplier when going from hour 9 to hour 10 has increased due to the increase in the size of the uncertainty (interval on the Q-axis has grown from length $A$ for hour 9 to the dotted length $B$ in hour 10).

Now, we assume that the supplier faces dynamic costs, i.e. it is costly for production to vary on top of any traditional marginal cost consideration and the larger the variation, the larger the cost. Then in the case of a fixed supply bid function ($S_0$ in both auctions), an increase in uncertainty implies an increase in expected dynamic costs.

The supplier's reaction to increased uncertainty is therefore to bid a steeper supply function $S_1$ in order to trade-off static optimality and dynamic effects. As a consequence, the range of volumes produced in equilibrium is reduced (the firm produces in the range $C$ instead of $B$). When seen over time, these considerations lead to a smoother production as compared to a constant supply curve: demand shocks are absorbed through a higher price volatility and a lower production volatility.

If cautious behaviour under high uncertainty is true for all firms on the market and each firm has the same expectation of the probability distribution of the uncertainty, then the reaction of bidding a more price inelastic supply function to increased uncertainty should be observable on the aggregated supply function.

We emphasize that this prediction relies on linear demand and supply functions and does not incorporate capacity constraint considerations (both upper and lower bounds on the production volume of plants), which are also important on the market. Furthermore, we have outlined our prediction using a discrete time-setting. The continuous version of this analysis on dynamic costs is explored in detail by Bergès and Martimort (2014).

The present paper tests this mechanism empirically and understands an increase in the slope of aggregate supply bid functions due to an increased level of uncertainty as evidence that firms minimise dynamic costs across auctions.

## 3.2 The EPEX spot market

### 3.2.1 General background

The EPEX Spot market is an auction market, which allows firms to trade electricity 12-36h ahead of delivery. It covers France, Germany with Austria and Switzerland. The volume traded on Epex Spot represents 12%, 40% and 30% of the total electricity consumption in these countries respectively in 2013 (EPEX, 2014).

The EPEX Spot market has considerably gained in importance over time and the daily trading volume has almost quadrupled since 2005, whereas the total electricity consumption has essentially remained constant. The graph in figure 3.2 shows these trends very clearly. Furthermore, it shows the significant volatility of the market trading volume (as indicated by the width of the grey-shaded confidence interval).

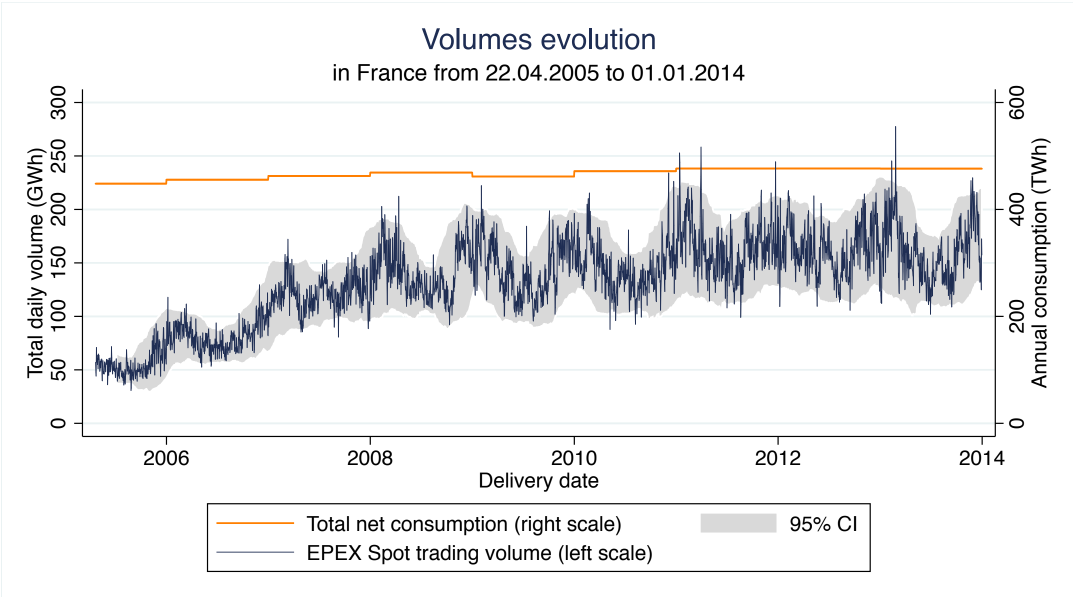



Figure 3.2: Traded volume plotted against total annual consumption
Note: Total consumption is netted of the electricity withdrawal at the level of the production unit. The 95% confidence interval is based on a 150-days moving window and assumes that volumes are normally distributed in the time window. GWh and TWh stand for giga and terawatt hours, respectively.

On the EPEX Spot market, the participants submit supply or demand bid functions to be able to meet their next day's supply commitment. This market is important, because it allows the firms to adjust their portfolio to the upcoming demand. The market matches business to business trades, where producers (the suppliers and transmission system operators) and industrial consumers may participate.

The EPEX Spot market settles in a three-pronged market that firms use to achieve their desired power position: The long-term bilateral contracting market, the day-ahead market and the intra-day market. Energy cannot be stored, thus an precise power position must be achieved at each point in time. Firms thus face a trade-off between cheap up-front sourcing and costly uncertainty. The closer the market gets to the delivery of its power, the less uncertainty does the firm face in determining its power requirements (pushing firms to wait until the last minute to fill their energy position). However, the imperfect flexibility of the electricity production landscape cannot satisfy the whole demand short-term at a reasonable price, hence firms must anticipate their requirements in order to obtain cheaper power. Consequently, these three markets complement each other to allow firms to gather a power position at a reasonable price.

### 3.2.2 Auction rules and mechanism

The EPEX Spot auction occurs daily, all year-round, and proceeds as follows: the order book closes every day at noon for contracts of the following day, results are published two hours later. Bids may be submitted 24/7 from 45 days prior until the closing of the books.

Tradable contracts exist for each hour of the day and firms submit an individual bid function for each of these hours, i.e. a separate, simultaneous auction is run for all hours of the following day and trading is specific for each of these hourly tranches.

The bid submission must be a supply function (or a demand function depending on the position of the firm) with at least 2 and at most 256 price/quantity combinations for single contract orders. The final bid function, thus, consists of the explicitly submitted points and all linearly interpolated points between them. The bid curves must be monotonically increasing for a supply function and vice versa for a demand function. Orders are transmitted via an online IT-platform and a redundant confirmation process aims to avoid erroneous bids. Bids are anonymous and the final electricity distribution is done via the French distribution network controlled by RTE EDF Transport SA.

Prices are specified in €/MWh with two decimal digits and must range from -3000€/MWh to +3000€/MWh. Quantities are specified in whole MWh. In addition to single contract orders for an individual hour, bidders may submit block orders. These are combined single contract orders with a minimum of two consecutive hours. The vital difference with multiple single contract orders is the "All-or-None" condition, namely that the executions of the individual contract orders forming the block are dependent on one another. That is for a block order covering hours 17 to 20, the quantity demanded for the hour 17 is only awarded if the corresponding quantity is also awarded for the hours 18, 19 and 20. Each registered bidder account is limited to a maximum of 40 block orders per delivery day, each of which is limited in volume to 400 MWh (approx. equal to 0.25% of the total daily volume traded on EPEX Spot).

The price-quantity determining mechanism is a uniform price, multi-unit auction mechanism: the summed demand and supply curves are computed and the intersection of these gives the equilibrium price and quantity pair. The market clearing mechanism takes into account single and block orders simultaneously and hence solves the corresponding programme by an algorithm of full enumeration of possible solutions, where each partial solution is verified to provide real, compatible prices. The mechanism works under a time limit. In the case of a curtailment, i.e. a disequilibrium with disproportionate prices due to unmatched supply and demand or an abnormal price for a specific hourly contract, the system proceeds to a second price fixing.

Of particular interest is the clear distribution of information. Ex-ante bidding, firms in the market know the identities of the rival bidders they face (but neither their individual bid functions nor their results in past auctions), the history of aggregated equilibrium prices and quantities up to that day, their clients' past demand realisations and their individual long term contracting position. Upon the clearing of the market, the aggregated supply and demand bid functions, equilibrium quantity and the equilibrium price become common knowledge. Each bidding account is informed of the contracts it has been awarded, i.e. the individual quantities to be sold and bought through the system.

## 3.3 Our data explained

### Auction market data

We have data from the French EPEX Spot market for the period 01.01.2011 to 30.06.2013. This is the latest period, where no significant changes in the auction rules have occurred and where data for all variables can be observed.

We observe the full aggregate bid functions for the day-ahead auctions of each hourly contract for both supply and demand. We understand the dataset as a cross-section rather than a time-series[5] and focus on weekday trading contracts only. This sums up to about 31 500 observations[6]. A single aggregate bid function is the sum of the individual bid functions, which are not available. We also observe the equilibrium price and quantity for each auction.

Moreover, we observe the block bidding results at the equilibrium solution only. We ignore the blocked aspects and treat subsequent auctions as independent from one another.

[5]This is supported by the graph in figure 3.2, which shows a flat total consumption and average trading volume on EPEX Spot since 01.01.2011.

[6]31 500 observations $\approx$ 2.5 years of hourly ($*365*24$) demand and supply ($*2$) functions for weekday trading ($*5/7$).

The two graphs in figure 3.3 show the aggregate supply and demand bid functions for the same hour of the same day. For a glimpse at the variation of bid functions over time, see figure 3.4. The table 3.1 sheds some light on the raw data. For further details as well as the plotted distribution of realised market equilibria, refer to appendix 3.10.2.

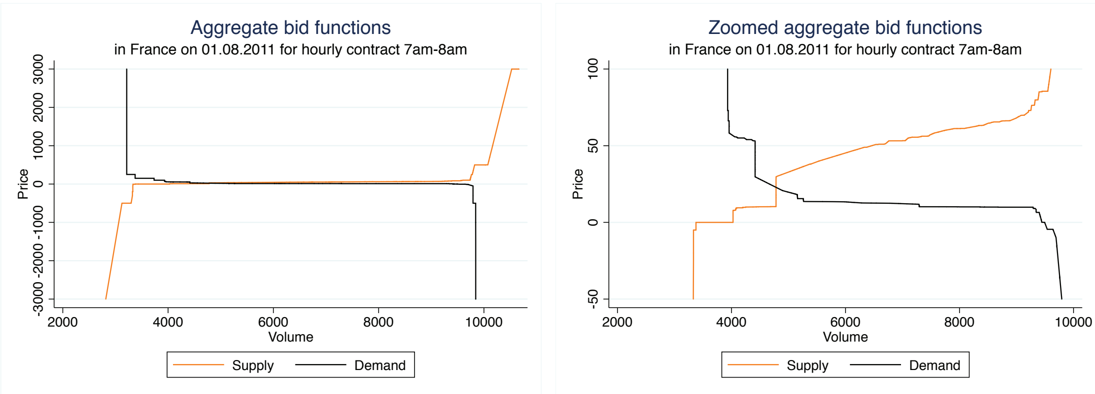



Figure 3.3: Example aggregate demand and supply bid functions
Note: The right-hand-side graph is a zoom of the left graph on for the price range −50€ /MWh to +100€/MWh.

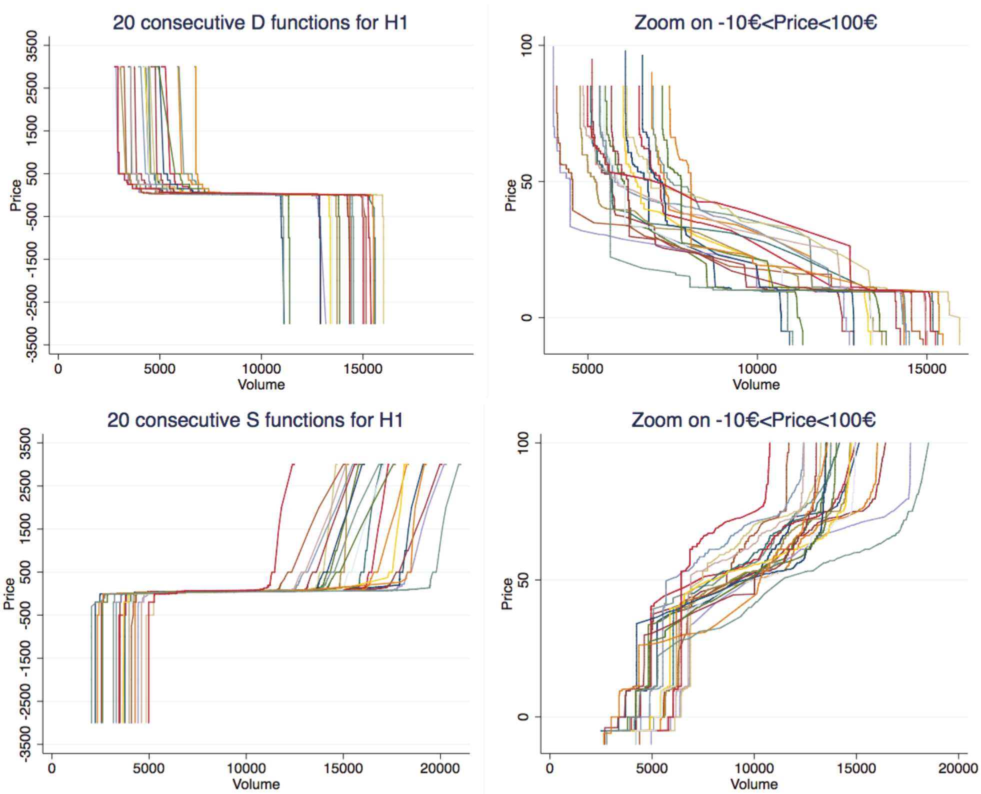



Figure 3.4: Aggregate bid functions for 20 consecutive days
Note: The graph shows 20 consecutive aggregate demand and supply functions for the contracts on hour 1 (between 12am and 1am) for the time period 11/12/2011 to 31/12/2011. The graph on the right is a zoom on the price elastic region of the curves on the left.

| | Mean | Median | Std. Dev | Min | Max |
|---|---|---|---|---|---|
| Total daily volume | 161,912 | 159,313 | 25,059 | 99,054 | 277,531 |
| Average realised daily price[7] | 46.6 | 48.3 | 17.2 | -39.0 | 381.2 |
| Minimum demanded agg. volume[8] | 5,030 | 4,968 | 1,467 | 914 | 11301 |
| Maximum demanded agg. volume | 13,327 | 13,222 | 2,212 | 4,990 | 23,254 |
| Minimum supplied agg. volume | 3,721 | 3,526 | 1,344 | 618 | 10594 |
| Maximum supplied agg. volume | 14,390 | 14,142 | 3,051 | 6,580 | 35,356 |
| Bid points per demand function | 543 | 531 | 163 | 115 | 1,253 |
| Bid points per supply function | 640 | 632 | 143 | 184 | 1,283 |
| Bidders per auction[9] | - | - | - | 1 | 101 |

Table 3.1: Some descriptive statistics

Finally, we reuse the data output from section 3.4.1. Specifically, we reuse the specific points extracted from the aggregate demand and supply bid functions, which are comparable across auctions. Why these points are useful for our analysis is explained in the methodology (section 3.4).

## Exogenous factors

Regarding weather statistics, we have hourly previsions for temperature, wind and cloudiness from the GFS (Global Forecast System) as well as hourly observations for these quantities and luminosity from MétéoFrance. The previsions from the GFS are in the form of weather maps that are outputted from simulations that run one-day ahead at 6 am. This is the weather information that market participants have access to when bidding

[7]Average price is volume weighted over the 24 hourly contracts of the delivery day.

[8]Minimum and maximum volumes for both demand and supply refer to the aggregate volume bid on the market for a single hour contract at the extremal prices of +3000€/MWh or −3000€/MWh.

[9]Due to the anonymity of the auction procedure, it is unknown which bidders submitted bids. Consequently, it cannot be deduced how many bid steps a typical bidder submits. Number of registered bidders for the French EPEX Spot market as of 01.10.2014.

on EPEX Spot[10]. The weather observations are in the form of tables for specific weather stations (between 100 and 200 depending on the specific parameter of interest).

Moreover, we have the location of the total installed capacity per generation type (i.e. wind turbines, solar panels, etc.) at the level of the postcode, that is roughly a 3km precision. We obtain this data from the SOeS, a branch of the French government producing data on environmental issues at large.

Population data and data on the level of the domestic production from the manufacturing industry is obtained in monthly steps from the French National Institute of Statistics and Economic Studies (INSEE). From the same source, we obtain the spot prices for petrol and natural gas as well as the import prices at the border for coal, which we use as a proxy for the domestic prices. Prices for the European CO2 emission certificates are taken from the Portuguese secondary market (SENDECO$_2$) for European Unit Allowances (EUA)[11].

As a very coarse proxy for generation from hydro power plants, we have the total weekly stock of water in domestic dams (in the form of the summed height of all dam water levels in France) from RTE the grid operator.

## 3.4 Methodology

We want to identify the impact that the level of uncertainty has on the price elasticity of the aggregate supply function. In data terms, this means that we aim to regress the slope of (aggregate) supply bid functions on a proxy corresponding to the uncertainty that existed at the time of bidding. The uncertainty may come from two different sources: (i) uncertainty about the realisation of market demand and (ii) uncertainty on meteorological forecasts (which affect the generation from renewables). Both types of uncertainty affect the residual demand curve faced by each supplier[12].

This regression is able to explain how supply firms adjust their bidding strategies to the expectation of demand shocks that they face. Statistical significance of the level of uncertainty on the slope of the supply function would be evidence that firms take the strategic

[10]The next weather simulation run takes place at 12 noon, and is therefore not being used by the bidders on the EPEX day-ahead market, as the deadline for submitting bids is precisely 12 noon.

[11]Each unit EUA permit allows one tonne of CO2 emissions.

[12]Renewable generation benefits from a feed-in guarantee on the market and thus reduces the residual demand for all traditional electricity producers.

considerations of dynamic costs into account. We emphasize again that our theoretical prediction does not take capacity constraints into account. These capacity considerations drive the vertical parts of the bid functions. Effectively, capacity constraints are a type of infinite dynamic costs which arise towards the extremities of the bid functions. We therefore focus on the central part of the bid functions.

We lay out the methodology top-down. First, in section 3.4.2 we show the final regression of interest. Sections 3.4.3 and 3.4.4 then detail the theory and empirics underlying the variables that feed into the final regression.

Some of the information used in our analysis is drawn from the bid functions of the EPEX Spot market. As introduced in section 3.3, we observe the full aggregate bid functions for both supply and demand, the shape of which (and thus the information that we aim to extract from them, e.g. their slope) varies differently at different points (recall the graphs in figure 3.4). We therefore adopt a functional data analysis approach, which allows us to condition our analysis on specific points $k = 1, .., K$ of the functions. These points must be comparable across auctions in order to derive insights.

The methodology to select comparable points across auctions is presented in section 3.4.1 and discussed in more detail in appendix 3.8. This appendix also evaluates the results when applying the technique to our data from the Epex Spot market. We use the obtained points in the subsequent stages of our work. Figure 3.5 shows the selected points on an exemplary demand and supply curve and labels points according to their type $k = 1, .., 5$.

The different types of points selected represent different parts of the aggregate bid functions and capture different information. The most important point is $k = 3$, which corresponds to the point of inflection were a smooth underlying logistic function assumed. This point is most relevant for equilibrium determination[13]. We expect to measure our theoretical prediction at this point, as compared to points towards the extremities of the bid function which are affected by capacity constraints.

The points $k = 2, 4$ are the points of maximum curvature and represent the transition points between the central (very price elastic) region and the outer (very price inelastic) regions of the bid function. The analysis of the slope at points $k = 2, 4$ does not make much sense since they are defined as the points of maximum curvature, i.e. where the

[13]See figure 3.29 on page 160 for a glimpse at the distribution of equilibrium outcomes.

slope changes most. Last, we have the points $k = 1, 5$ which are imposed by the auction rules and are the endpoints of the bid functions. We suspect these points to be strongly influenced by capacity considerations and thus do not focus on them.

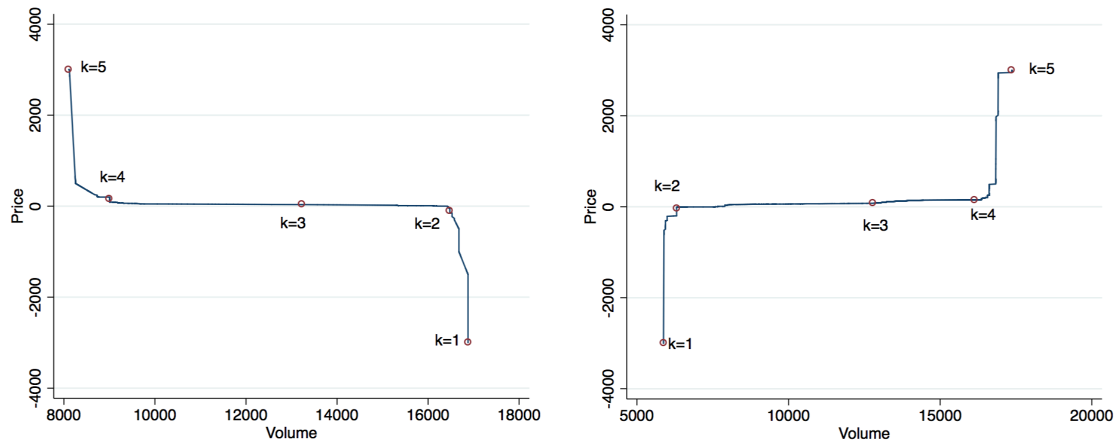



Figure 3.5: Selected points on original bid functions

*Note:* The demand function left, the supply function right, the graph superposes and names the points selected according to the methodology of section 3.4.1.

### 3.4.1 Point selection

We develop a methodology to analyse data of a specific format. The focus of this subsection lies on the methodological details. The evaluation of the performance of our technique is detailed in appendix 3.8. The aim of this methodology is to extract points of interests from functional data. The economic interpretation is secondary in this subsection.

#### Motivation

Reduced form models often rely on exploiting market outcomes, i.e. equilibrium prices and quantities, for their analysis in order to identify the determinants of firm behaviour and test predictions of the theory. On a few markets, we observe sufficient information to get around the problem of using endogenous equilibrium data. For example on the government bond markets, we observe both the full aggregate demand and supply functions. 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. Préget and Waelbroeck (2005) do this to investigate the determinants of demand bid functions in French government bond auctions. They rely on the propositions first put forward by 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. This relation can be exploited in a cross-section (cross-auction) reduced form analysis. Doing so, Préget and Waelbroeck (2005) show that cross-auction variation of the demand functions arises from differing auction covariates. Özcan (2004) applies the methodology 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.

This method has worked remarkably well in the context of treasury auctions[14]. In Belsunce (2011), the same logistic function approach is applied to data from the French day-ahead electricity market. Although the auction mechanism is identical to that of the Treasury market and data availability is comparable, the author shows that the logistic function approach does not suit the context of the electricity market due to the strong heterogeneity of the bid functions.

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 the supply side of the market (unlike the market for government bonds, where the supply is monopolistically determined by the Treasury itself). Furthermore, supplier bidding is strongly influenced by the underlying (step-function-like) marginal cost of the production

[14]As an example, Préget and Waelbroeck (2005) results provide a forecasting tool of remarkable fit for upcoming treasury auctions. Their correlation coefficient between the observed and estimated stop out rates is 0.99997. This forecasting tool is still in use by the French Treasury (Source: Personal discussions with one of the authors, June 2014).

technology - in particular towards the extremities of the bid functions[15]. 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 utility 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 of different parts of the bid function is not mixed. 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.

In terms of the literature related to our technique, we are only aware of Wölfing (2013). He uses a functional approach to measure the cost impact of $CO_2$ emission prices on supply functions on the electricity market. While his results are mixed, his work is notable since it is the first effort to implement these techniques. Our approach is more refined with respect to the relative (non-parametric) measures that we use to identify our landmarks as opposed to his approach of selecting landmarks based on absolute (hard-coded) criteria.

### Purpose

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

$$S_i' = \alpha + \boldsymbol{\beta X_i} + \epsilon$$

[15]Low volume bids are strongly impacted by marginal costs of base load production technologies (e.g. nuclear), while high volume bids are driven by peak-load production technologies (e.g. gas). The latter especially motivates bids closer to a linear function and not an S-shaped form.

where $S_i'$ is the steepness of the bid function in auction $i$, $\boldsymbol{X_i}$ the stacked vector of exogenous variables (not specified further here), $\alpha$ the regression constant, $\boldsymbol{\beta}$ the stacked vector of regression coefficients and $\epsilon$ the error term.

This is a cross-sectional regression, where the information $S'$ is drawn from the observed bid functions. These bid functions have the specificities detailed in section 3.3. The information ($S'$) that we want to measure varies along the bid function (see figure 3.3) and furthermore, the information also varies differently for bid functions of different auctions (see figure 3.4). In order to derive insights with our regression, we must make sure that the information included in the regression is comparable across auctions. We therefore adopt a functional data analysis approach, which allows us to condition our analysis on specific points $k = 1, .., K$ of the functions.

In terms of comparability, we require that a chosen point $k$ from a supply function must be comparable to the $k^{\text{th}}$ point from the supply functions of another auction. The same goes for chosen points of the demand functions. 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 3.6). One could compare the k$^{th}$

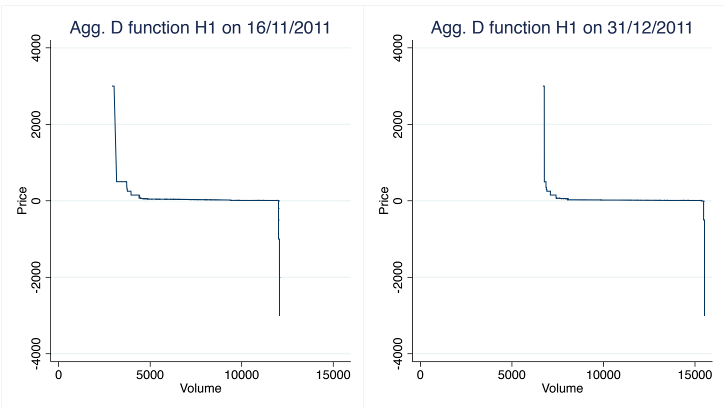



Figure 3.6: Comparison of two aggregate demand functions for the same hour

point of each function to one another. Unfortunately the number of points varies from one auction to another, so this approach would be meaningless. Instead 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.

To recognise these features, we perform two successive kernel density analyses[16]. For details on the bandwidth and kernel selection as well as algorithm specificities, see appendix 3.8.5. 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 3.7.

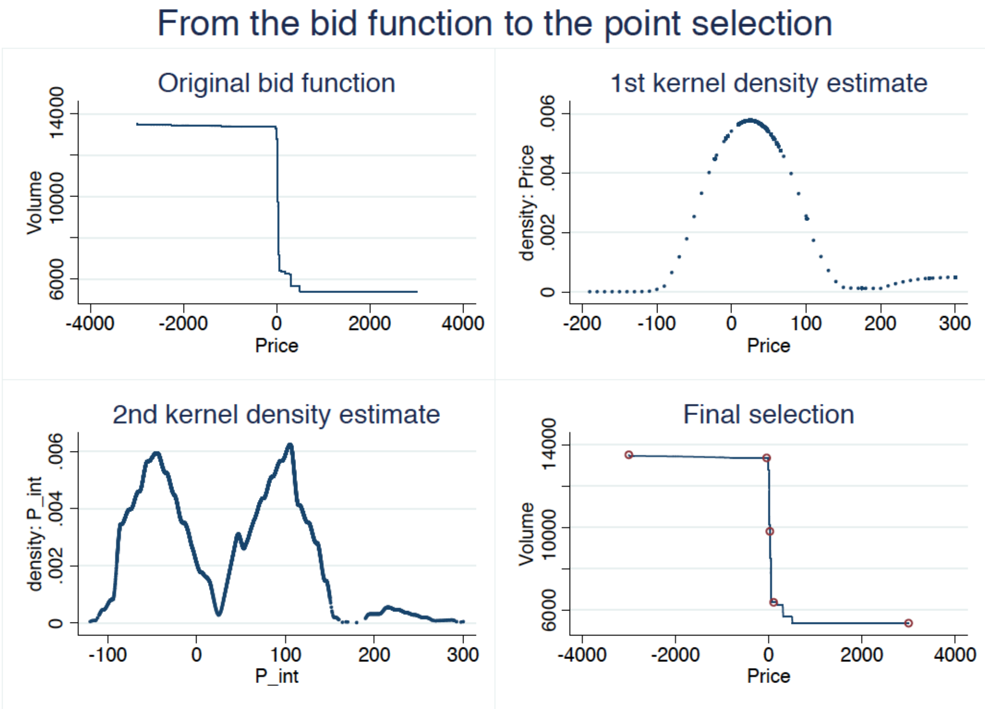



Figure 3.7: Steps of the point selection process
Top left (A): The full original aggregate demand bid function for hour 8 on 15.01.2011 in the Price- Volume dimension. Top right (B): Kernel density estimates of the first derivative, zoomed on the relevant price range. Bottom left (C): Zoomed kernel density estimates of the second derivative. Bottom right (D): The full original bid function with the $K = 5$ selected points.

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).

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

We can develop this method further and define intermediary points[17] that can again be compared to one another. This method allows to define as many points as needed, for computational reasons we limit ourselves to $K = 5$ points[18].

Graph D of figure 3.7 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 the point of inflection were a smooth functional form imposed (determined by the maximum of the first derivative, ($k = 3$)) and the points separating 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 point of inflection).

We described the technique here for the case of a demand function. The information measured at these points (e.g. price, volume or slope) can thereby be compared across demand bid functions of different auctions. The method is used analogously for selecting comparable points on the supply function.

As the focus of this paper is not on this methodology but on what it allows us to study, we describe the results of this specific methodology in appendix 3.8 (starting page 135). In the appendix, we also provide a discussion of the robustness of these results.

### 3.4.2 Regression methodology: identification

At each of these comparable points, we want to identify the effect of uncertainty on the slope of the supply function. Defining $S'_{i,k}$ the slope of the supply function of auction $i$ at point $k$ in the quantity (X-axis) - price (Y-axis) dimension, $\boldsymbol{X^S}$ being the vector of exogenous variables, $\text{PLU}^D_{i,k}$ being the proxy for the level of demand uncertainty, $\text{PLU}^R_i$ being the proxy for the level of uncertainty from renewables, $\alpha$ being the regression

[17] 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.

[18] 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.

constant and $\epsilon$ being the error term, we estimate the following:

$$S'_{i,k} = \alpha^S_k + \beta^S_k \text{PLU}^D_{i,k} + \gamma^S_k \text{PLU}^R_i + \boldsymbol{\delta^S_k X^S_i} + \epsilon^S_{i,k} \tag{3.1}$$

We are interested in the sign and magnitude of the coefficients $\beta^S$ and $\gamma^S$, which identify the effects of the PLUs ($\text{PLU}^D$ and $\text{PLU}^R$, respectively) on the shape of the supply bid function. From the predictions outlined in section 3.1.2, we expect a positive coefficient when uncertainty levels increase[19].

### 3.4.3 Left-hand-side variables

We extract the slope of the aggregate supply function at any given point $k$ from a kernel density estimation with a bandwidth of 45 units[20].

Effectively, this is a smoothed version of the slope. Thereby, we aim to make our slope estimates robust to steps in the bid function[21]. Steps in the bid functions mostly arise towards the extremities of the bid functions and could arise from marginal cost bidding. Working with smoothed slopes is in line with previous work à la Préget and Waelbroeck (2005) and Özcan (2004), who also apply reduced form models to aggregate bid function data.

### 3.4.4 Right-hand-side variables

We are regressing an ex-post measure of the auction market (realised slope of the supply bid function) on ex-ante information that bidders have at the time of bidding, i.e. which is available at midday of the day ahead of delivery. We thus keep a strict separation of the ex-post and ex-ante information to the left and right hand side of equation 3.1, respectively. This separation allows us to circumvent endogeneity problems and validates the use of simple OLS regressions.

[19] Specifically, we want $\beta^S$ to be positive, $\gamma^S_1$ positive and $\gamma^S_2$ negative. For details on $\gamma^S$, see section 3.4.4.2.

[20] The slope is a by-product of the point selection mechanism and the bandwidth selection for the smoothing thus follows the same considerations as for the latter. The details of this choice are specified in appendix 3.8.5.

[21] In our data, we observe that bid functions are effectively step functions. On EPEX spot 256 price-quantity combinations are allowed per bidder. When additional bid points are costly, then stepwise bidding behaviour may be very different from a setting where continuous functions can be bid (Kastl, 2011). Due to the fact that, on average, we do not observe that firms use up all available price-quantity combinations , the cost argument of an additional bid point seems weak. Hence, by smoothing the slope we approximate the unconstrained, continuous bid function.

For this reason, we construct our PLUs on the basis of predicted uncertainty. However, for data availability reasons we cannot exclude endogeneity problems completely. For details, see the discussion in section 3.6.2.

In this subsection, we first outline how we generate the proxies for the level of market demand uncertainty (PLU$^D$) in section 3.4.4.1. Second, we construct the proxies for the level of uncertainty from renewables energies (PLU$^R$) in section 3.4.4.2. Third, we detail how the vector of exogenous variables ($\boldsymbol{X}$) is constructed in subsection 3.4.4.3.

#### 3.4.4.1 Generating proxies for uncertainty from market demand (PLU$^D$)

We construct a proxy for the level of the demand uncertainty (PLU$^D$) by using the residuals from a demand estimation on exogenous parameters as a measure of the uncertainty that bidders face in an auction. Specifically, our PLU$^D$ is the expected squared level of the prediction errors that firms expect to make when anticipating the demand level of the day ahead. Under conditions of normality, ex-post prediction errors give consistent estimates of the uncertainty at the time of bidding.

The uncertainty proxy is obtained as detailed next in a three-step procedure. In the first step, we explain what kind of uncertainty our PLU$^D$ refers to. The second step details the conceptual details of constructing the PLU$^D$. The third step computes the PLU$^D$.

**In the first step,** we reduce the demand functions to a fixed number $K$ of comparable points across auctions by using the non-parametric point selection technique outlined in section 3.4.1. Each k$^{th}$ point is defined by a price and a quantity, which we regress independently on the exogenous variables.

Let us call $P_{i,k}^D$ and $Q_{i,k}^D$ the price and quantity of point $k$ of the realised demand function in auction $i$, $\boldsymbol{X_i^D}$ the vector of exogenous variables relevant for the demand estimation.

$$P_{i,k}^D = \alpha_k^{D,P} + \boldsymbol{\beta_k^{D,P}} \boldsymbol{X_i^D} + \epsilon_{i,k}^{D,P} \tag{3.2}$$

$$Q_{i,k}^D = \alpha_k^{D,Q} + \boldsymbol{\beta_k^{D,Q}} \boldsymbol{X_i^D} + \epsilon_{i,k}^{D,Q} \tag{3.3}$$

In regressions 3.2 and 3.3, firms try to anticipate the realisation of the demand using the exogenous information available. We consider that the producers are able to do such an analysis at the time of bidding.

The prediction errors $\epsilon_{i,k}^{D,J}$, $J = \{Q, P\}$ are a consequence of the stochastic nature of the demand and hence an manifestation of the uncertainty. We consider that more uncertainty will lead to larger prediction errors being made in equilibrium and adopt the square of the residuals $\left(\epsilon_{i,k}^{D,J}\right)^2$ as our measure for the realised level of demand uncertainty.

**In the second step,** we recover the residuals from the demand estimation in regressions 3.2 and 3.3 and test for heteroskedasticity using White (1980), which is clearly confirmed (see tables 3.3 and 3.4).

Heteroskedasticity means here that the variation of error terms varies conditional on the levels of the exogenous factors: $E(\epsilon_i^2|\boldsymbol{X_i}) = g(\boldsymbol{X_i})$. However, they are still orthogonal: $E(\epsilon_i|\boldsymbol{X_i}) = 0$, thus ensuring that the prediction is unbiased, but not "best" in the sense of the best linear unbiased estimator (BLUE). Thus, heteroskedasticity results in inefficient regressions where the estimator is not minimum variance. Since we do not interpret regressions 3.2 and 3.3 for causality, but only for predictive purposes, we stick to the unbiased OLS.

The heteroskedasticity regression is given for $J = \{P, Q\}$ by

$$\left(\,\epsilon_{i,k}^{D,J}\,\right)^2 = \alpha_k^{U,J} + \boldsymbol{\beta_k^{U,J}}\boldsymbol{X_i^D} + \epsilon_{i,k}^{U,J} \tag{3.4}$$

**In the third step,** we compute the predicted $\text{PLU}_{i,k}^D$ that firms use when bidding in the auction as:

$$\underbrace{\widehat{\left(\,\epsilon_{i,k}^{D,J}\,\right)^2}}_{\widehat{\text{PLU}}_{i,k}^D} = \alpha_k^{U,J} + \boldsymbol{\beta_k^{U,J}}\boldsymbol{X_i^D} \tag{3.5}$$

The idea is that by experience, firms in the market know that their predictions are more or less accurate depending on the environmental conditions (in the sense of realisations of exogenous factors). In other words, firms can use the realisations of $\boldsymbol{X^D}$ to infer the accuracy of their demand predictions. Technically speaking, they can use the heteroskedastic nature of the residuals to forecast the level of uncertainty that they face.

The $\text{PLU}^D$ subs into regression 3.1. For simplicity, we do not include the uncertainty proxies $\text{PLU}_{i,k}^D$ measured at all $K = 5$ points in regression 3.1 simultaneously, but only a single $\text{PLU}_{i,k}^D$ at a time. Therefore in the final regression 3.1, we regress the slope at a point of the supply function on the $\text{PLU}_{\cdot,\cdot}^D$ estimated at the corresponding point on the demand function. The pairing is done in the quantity dimension. This means that the slope of the supply function at point $k = 3$ is regressed on the uncertainty measured at point $k = 3$ of the demand function (recall the labelling of the points as given in figure 3.5)[22]. We indicate this quantity paring in the index $k^{-1}$ of the PLU:

$$\mathrm{PLU}^D_{i,k} = \widehat{\mathrm{PLU}}^D_{i,k^{-1}} \tag{3.6}$$

An increase in $\mathrm{PLU}^D_i$ corresponds to an increase in the uncertainty about the market demand realisation. We thus expect $\beta^S$ to be positive in regression 3.1.

#### 3.4.4.2 Generating proxy for uncertainty from renewable energies ($\mathbf{PLU}^R$)

We have already referred to the statement that the intermittency of renewables causes large residual demand shocks (EPEX, 2014). Suppliers are thus wary of the expected production of renewables generation.

Given that renewable generation is an exogenous source of supply, it affects the residual demand curve for each supplier, but does not enter the $\mathrm{PLU}^D$, which captures the uncertainty on market demand only.

In predicting the generation from renewables, we assume that suppliers are able to infer renewables generation from meteorological forecasts[23]. When forecasting the residual demand shocks due to generation from renewables, we consider that suppliers have an idea of the precision of their estimate based on the "look" of the meteorological forecasts that they have. By look, we mean the geographical heterogeneity or homogeneity of the forecasts. Depending on the disparity of local weather forecasts, inference of the national level of renewables generation is more or less difficult. The geographical disparity of the forecasts is captured by the characteristic lengthscale of autocorrelation of weather forecasts, which feeds into our proxy for the level of uncertainty from renewables production ($\mathrm{PLU}^R$).

Intuitively, the characteristic lengthscale of autocorrelation represents the distance required between two geographical points on a map of weather forecasts to observe a decorrelation of half of its maximum value. For example on the wind speeds prediction, a characteristic length of 80 km means that if we observe two very distant points to have a difference in wind speeds of, on average, 50km/h (this being the maximum

[22] This also means that we would pair point $k = 2$ from the supply function with the point $k = 4$ of the demand function.

[23] We specify the technique in appendix 3.9.3 and use it to construct our controls in section 3.4.4.3.

difference), then we will observe, on average, wind speed differences of 25km/h for points distant from each other by 80km.

We compute this characteristic lengthscale ($L$) as described in appendix 3.9. Our PLU$^R$ is defined as the two proxies

$$\mathrm{PLU}^R_{1,m} = \frac{1}{L_m}, \qquad \text{where } m = \{\text{Wind, Solar, Temperature}\} \tag{3.7}$$

$$\text{and} \quad \mathrm{PLU}^R_{2,m} = \Big(\frac{1}{L_m}\Big)^2 \tag{3.8}$$

Generally, we expect firms to face less uncertainty in predicting weather conditions when the lengthscale of autocorrelation $L$ is longer since the overall weather conditions will be more homogenous. A longer length $L$ (less uncertainty), will yield a smaller PLU$^R$ and we expect a flattening of the supply curve. I.e. we expect a positive coefficient $\gamma_1^S$ on the PLU$^R_{1,m}$ variables in the final slope regression.

However, we also expect the effect of $L$ on the slope to be attenuated, if not counter-balanced, by the squared term[24]. This means that for very short or long $L$, we expect an additional effect, which signifies reduced uncertainty. In the former case of short $L$, we observe a high amount of noise in weather predictions. According to the law of large numbers, these errors should cancel out and the uncertainty decrease. In the latter case of long $L$, we expect that exponentially less uncertainty results from very homogenous weather conditions. We thus expect a negative coefficient $\gamma_2^S$ on the squared PLU$^R$ term in the final slope regression (equation 3.1).

#### 3.4.4.3 Controls

This section details the exogenous variables, which we use for our study. The stacked vector of exogenous variables is not identical for the supply and demand regressions of equations 3.1 and 3.2.

The vector $\boldsymbol{X^D}$ for the demand equation includes the variables: Tempeff15, Roll_Temp24, Roll_Temp240, suncycle, morning, deltasun, EWH, SolarRest, RteBlackBox.

For the supply regression we include in $\boldsymbol{X^S}$ the following variables[25]: Coal, Brent, Gas, IT2, EUA, Wind1DA, Hydro.

[24] We expect the effects of $L$ on the slope to be of the shape of a Laffer curve.

[25] We do not include the variables used for the demand estimation as they indirectly feed into the final regression via the PLU$^D$.

Table 3.2 gives a brief overview of the controls used. Details on the computation of some variables are given in the appendix (see links in table). The last column indicates the frequency with which we observe the variable in question.

| Name | Explanation | Unit | Frequency |
|---|---|---|---|
| Wind1DA | The day-ahead predicted electricity volume generated from wind turbines. Details on p. 151. | MWh | Hourly |
| Solar | The electricity volume generated from photovoltaic sources. Details p. 153 | MWh | Hourly |
| Tempeff15 | Effective predicted temperature in France (with a cutoff point at 15°C to reflect demand patterns), aggregated on a national level. Details on p. 155. | °C | Hourly |
| Roll_Temp24 | Mean of *Tempeff15* over the last 24 consecutive hours. | °C | Hourly |
| Roll_Temp240 | Mean of *Tempeff15* over the last 240 consecutive hours. | °C | Hourly |
| suncycle | Luminosity as a percentage of maximum luminosity of the day. *Midday* defined as *suncycle*=1. Details on p. 156. | % | Hourly |
| morning | Indicator variable for hours before *Midday*. | $\{0, 1\}$ | Hourly |
| deltasun | Absolute value of the change in *suncycle*. Details on p. 156. | $[0, 1]$ | Hourly |
| EWH | Indicator variable for hours between 10pm and 4am. | $\{0, 1\}$ | Hourly |
| SolarRest | The unexplained component of photovoltaic generation. Specifically, the residuals from a regression of *Solar* on *suncycle*. Details on p. 156. | MWh | Hourly |

*Continued on next page...*

*... table 3.2 continued*

| Name | Explanation | Unit | Frequency |
|---|---|---|---|
| RteBlackBox | The unexplained component of the day ahead prediction of total consumption in France issued by the grid operator (RTE). Specifically, the residuals from a consumption estimation. Details on p. 157. | MWh | Hourly |
| Coal | Average coal import prices at the French border. | €/ton | Monthly |
| Brent | Average of spot prices for crude oil on the London based stock exchange. | $/bl | Monthly |
| Gas | Average of closing prices for natural gas at 1 month on the London market (NBP). | £/Therm | Monthly |
| IT2 | Interaction term: *Gas* weighted by an hourly index for the demand level. Details on p. 158. | £/Therm | Hourly |
| EUA | Price of $CO^2$ emissions. | €/ton | Daily |
| Hydro | Sum of dam level heights on a national level. | % | Weekly |

Table 3.2: Overview of exogenous variables.

The rationale for the included variables is the following: First, Wind1DA and Solar control for the expected level of renewables generation[26] on the day ahead market. These are computed using a novel bottom-up methodology described in the appendix 3.9.1. Second, Tempeff15 controls for the demand patterns as a function of the temperature[27]. Tempeff15 includes a cut-off at 15$^o$C in order to take into account the demand pattern as a function of temperature according to RTE (2014). Table 3.13 on page 158 reveals the improved fit over a simple temperature variable that does not respect the demand cut-off (Tempeff). Third, Roll_Temp24 and Roll_Temp240 capture the demand seasonality

[26]For data availability reasons, Solar is computed on realised luminosity values rather than forecasts of luminosity.

[27]Note that electric heating is widely spread in France. It is used in 32% of principal residences (INSEE, RP2011 exploitation principale).

via the temperature. The former gives the daily average temperature, while the latter captures the average temperature over the last 10 days. The demand cut-off at 15$^o$C for Tempeff15 is respected for these means. Including these as seasonality controls allows to get away from using dummy variables for the seasonality. In short, avoiding dummies yields more transparency of the results as we do not have the problem of interpreting the dummies, which are often black boxes[28]. Figure 3.8 illustrates the complementarity of the three continuous seasonality controls. Fourth, we use the four variables suncycle,

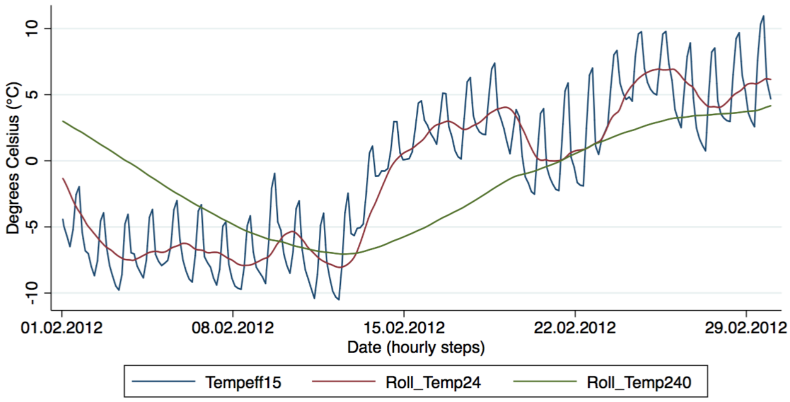



Figure 3.8: Temperature based seasonality controls

*Note:* The graph shows the evolution of the temperature based controls for seasonality for the month of February 2012. The graph shows the lagged nature of the rolling average temperature controls.

morning, deltasun and EWH collectively to continuously control for the time of the day. The reasoning is again the ability to get away from using dummies and being able to interpret the results. Figure 3.9 shows how the controls describe the daily patterns continuously. Fifth, SolarRest and RteBlackBox are the residual information gained from the variables Solar and the day ahead consumption prediction of RTE (PrevConsoH) over other variables[29] included in $\boldsymbol{X^D}$. Sixth, Coal, Brent, Gas, IT2 and EUA are rough proxies for the input prices for electricity suppliers. Hydro is used as a crude proxy for dam operator's ability to generate short term electricity using hydro reserves.

We briefly emphasize that novel methodologies have been used to compute all variables derived from weather forecasts or observations. When tracing back the shape of aggregate bid functions on exogenous factors in the second stage estimation, we use aggregated

[28]See section 3.6.1 for a full discussion on the advantage of avoiding dummies.

[29]E.g. Solar is strongly correlated with suncycle, thus SolarRest is the residual from a regression of the former on the latter. RteBlackBox is computed as the residuals from regressing PrevConsoH on Tempeff15, Roll_Temp24, Roll_Temp240, suncycle, morning, deltasun and EWH. See appendix 3.9.5 and 3.9.5 for details.

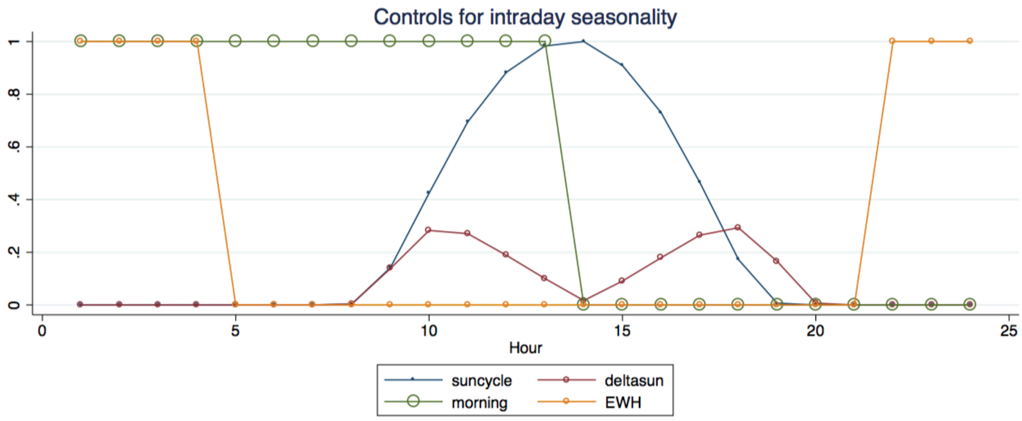



Figure 3.9: Continuous controls for daily patterns
*Note:* With the exception of EWH, all intraday seasonality controls (suncycle, morning, deltasun) are determined endogenously by the prevalent luminosity as captured by Solar.

statistics (at the national level) for the exogenous variables. We thus use an aggregation methodology to summarise local information (collected at the level of the individual postcodes) in order to generate an aggregate statistic at the national level. The general methodology for the aggregation is explained using the example of Solar and as follows: We observe the value of a weather parameter (e.g. luminosity) every hour at known weather stations in France (for details see appendix 3.9.1 and 3.9.2). We apply an interpolation technique (for details see appendix 3.9.3) in order to obtain parameter values for all possible geographic locations in France. At any local point, we can thus infer the electricity volume generated by using the information of the locally installed capacity (of solar panels) and the renewable energy available (i.e. sunlight inferred from luminosity). We then take the sum of all solar generated electricity per hour in France and use this as our aggregate statistic at the national level in our regression analyses. We used forecast data wherever possible in order to approximate the level of information that bidders have at the time of bidding and circumvent endogeneity problems. For cases where forecast data was not available, e.g. Solar, realised weather data was used. We emphasize that both Tempeff15 and Wind1DA are computed from forecast data.

### 3.4.5 Extensions and robustness checks

In order to test the robustness of our results and circumvent some drawbacks of the baseline model, we use a few alternative specifications of our empirical model.

#### 3.4.5.1 Bootstrapping standard errors

The set-up of our empirical analysis relies on stochastic variables, e.g. PLU$^D$, which are computed in the first stage of our identification. The assumption made for an OLS regression of normally distributed residuals is a very strong one (particularly with the forecast variable) and one which can flaw the precision of estimates in the second stage regression. We therefore bootstrap the standard errors of the final regression by using random sampling with replacement at each stage of the analysis, i.e. for both the PLU computation and the final slope regression with 300 repetitions.

Bootstrapping allows us to non-parametrically approximate the distribution of the forecast PLUs and thus enables us to correct the standard errors of our coefficient estimates.

#### 3.4.5.2 Kernel based uncertainty forecasts (PLU$^D_{\boldsymbol{X}_i}$)

The PLU$^D$ computed as described in section 3.4.4.1 is noisy since we assume a linear forecast model to be valid for any combination of realisations of exogenous parameters, i.e. the same model applies winter and summer, day and night. While the results are as desired for the baseline PLU$^D$, a bootstrapping of the standard errors indicates that the first stage forecast is too imprecise for effects of a satisfactory significance level[30].

We therefore develop an extension of the uncertainty prediction model in which we use the idea of demand forecasts (equation 3.5) only locally, i.e. for a limited range of variation in the exogenous parameters. In other words, we estimate the PLU$^D$ corresponding to an auction only in the neighbourhood of this auction, i.e. over all auctions that occurred in similar conditions. By conditions, we mean realisations of exogenous parameters and the neighbourhood refers to the concept of measuring the similarity of these realisations by means of a range. The next steps explain how this is done formally.

The methodology is analogous to the computation of the baseline PLU$^D$. However, we now consider that firms predict the level of the uncertainty by comparing it with the level of uncertainty in past[31] auctions of similar exogenous conditions. Thus, the suppliers

[30]Table 3.5 shows that the corrected standard error using a bootstrapping technique (column 2) is considerably larger than what an OLS regression suggests under the assumption of normally distributed PLUs (column 1). The coefficients of the PLUs change from a 1% significance level (col.1) to insignificant (col. 2) . Furthermore, the lack of significance of the PLUs in column 2 indicates that the PLU measures could include too much noise, which is likely in our simple linear baseline forecast.

[31]For data availability reasons, we pool all (past and future) auctions for the computation of this PLU. This introduces some endogeneity. For a discussion of this choice, please see section 3.6.2.

forecast the precision (squared residuals) of their demand estimation as before, but only on a subsample of the data. The subsample is defined as all observations which lie within a distance $b_{X_e}$ of the observation of interest with respect to each control variable $X_e$, $\forall e = \{1, .., E\}$. Effectively, this is a multi-variate kernel regression and subsequent forecast with a rectangular (also called *boxcar*) weighting function. Observations within the kernel window are given equal weight, while observations outside the kernel window are given zero weight. We set the bandwidth $b_{X_e}$ with respect to each variable equal to $\frac{1}{3}$ of the range of that variable[32].

Call $\Omega$ the set of all auctions $i$. Let $\boldsymbol{X_i}$ be the realisation of the stacked vector of exogenous variables ($X_e$) for an auction $i$. The simple weight function $W_{i,j} : \{\boldsymbol{X_i}, \boldsymbol{X_j}\} \rightarrow \{1, 0\}$ is an indicator function from the set of the $\boldsymbol{X}$ for two observations $i, j \in \Omega$, which indicates if the auctions $i$ and $j$ are considered similar for our purposes based on the similarity of their realisations of exogenous variables. Specifically,

$$W_{i,j} = \prod_e W_{i,j}(X_e) \ , \qquad \text{where } W_{i,j}(X_e) = \begin{cases} 1, & \text{if } |(X_e)_i - (X_e)_j| \leq b_{X_e} \\ 0, & \text{otherwise.} \end{cases} \tag{3.9}$$

This weight function defines a subsample of comparable auctions for each auction $i$: $\Omega_i \subset \Omega, \quad \Omega_i = \{j : j \in \Omega, \ W_{i,j} = 1\}$, where the size and exact composition of that subsample depend on the auction's specific $\boldsymbol{X_i}$. In the following, the index $\omega$ is taken from these $\Omega_i$ and we iterate the regressions over all subsamples. The individual subsample based regressions are then for the demand estimation

$$P^D_{\omega,k} = \alpha^{D,P}_{k,\Omega_i} + \boldsymbol{\beta}^{D,P}_{k,\Omega_i} \boldsymbol{X}_\omega + \epsilon^{D,P}_{\omega,k} \tag{3.10}$$
$$Q^D_{\omega,k} = \alpha^{D,Q}_{k,\Omega_i} + \boldsymbol{\beta}^{D,Q}_{k,\Omega_i} \boldsymbol{X}_\omega + \epsilon^{D,Q}_{\omega,k} \tag{3.11}$$

where the estimated coefficients are specific to each subsample, which is indicated using the additional index ${}_{\Omega_i}$. Similarly, the local uncertainty regression is given $\forall J = \{P, Q\}$ and $\forall \omega \in \Omega_i$ by

$$\left(\epsilon^{D,J}_{\omega,k}\right)^2 = \alpha^{U,J}_{k,\Omega_i} + \boldsymbol{\beta}^{U,J}_{k,\Omega_i} \boldsymbol{X}_\omega + \epsilon^{U,J}_{\omega,k} \tag{3.12}$$

[32]See appendix 3.10.1 for details. Column 2 of table 3.14 indicates the choice of $b_{X_e}$ for each exogenous variable considered.

Finally, we compute the PLU for each auction $i$ based on the subsample-estimated coefficients and the specific realisation of exogenous variables $\boldsymbol{X_i}$:

$$\underbrace{\widehat{\left(\epsilon_{k,\boldsymbol{X_i}}^{D,J}\right)^2}}_{\widehat{\text{PLU}}_{k,\boldsymbol{X_i}}^{D}} = \alpha_{k,\Omega_i}^{U,J} + \boldsymbol{\beta}_{k,\Omega_i}^{U,J}\boldsymbol{X_i} \tag{3.13}$$

When firms infer the upcoming uncertainty by looking at the uncertainty in past auctions, the precision of their estimate depends on the number of comparable auctions available, i.e the sample size $\Omega_i$. Given that the sample size varies greatly across auctions, we use a sample-size-weighted OLS regression in the final estimation of equation 3.1. Finally, we bootstrap the standard errors on the kernel-based PLUs using 50 repetitions[33].

#### 3.4.5.3 Locality of prediction

To support our local prediction for the central part of the bid function, we also disclose the results for the points $k = 1, 2, 4, 5$ in on page 126. We expect the effect of uncertainty to be insignificant for all points not on the centre portion of the supply bid function.

As discussed in section 3.1.2, the theoretical prediction (of an increased slope of the supply function when uncertainty is high) is a local prediction. We have two reasons for the locality of our prediction: (i) it is based on linear functions and (ii) we ignore capacity constraints.

The first point implies that we can only test this prediction on bid functions that are (at least roughly) linear. However, the overall bid functions on EPEX Spot that we observe have the characteristic S-shape that we have introduced in section 3.3. By focussing on the central part of the bid functions, we have a (sub-)bid function which exhibits a linear shape over our window of interest, i.e. the volumes interval that occurs normally in equilibrium. The analysis of this part of the bid function is important as it is the most relevant in equilibrium. The extremities of the bid functions rarely determine the equilibrium outcome - they only have an impact when there is a major discrepancy between demand and supply. The central portion of the bid function is represented by the point $k = 3$, which we focus on for our main results.

[33]For computational reasons, we only bootstrap the kernel based PLUs for the point of inflection ($k = 3$). We choose only 50 repetitions for the same reason. Given the size of our dataset, we consider it acceptable. The general criterion for convergence is that each observation is selected at least once in the bootstrapping exercise.

The second point refers to the fact that our prediction abstracts of firms' considerations regarding capacity constraints. These considerations drive the vertical extremities of the bid functions and (by virtue of representing an infinite dynamic cost) erode dynamic cost considerations due to uncertainty. We therefore see a lack of significant results on the PLUs for the points $k = 1, 2, 4, 5$ as support for our local prediction.

## 3.5 Results

We first present the results for the demand estimation in both the Price and Volume dimension since this step is identical for all PLU specifications. We then present the results of the final regression in the baseline and alternative specifications.

### 3.5.1 Demand estimation

Table 3.3 gives the results for the demand estimation on volumes (equation 3.3). Table 3.4 shows the results for the demand estimation on prices (equation 3.2).

These tables are interesting for two reasons. First, they provide the basis for our computation of the PLU$^D$. Second and the reason why we disclose them in such detail, they are already a result in themselves.

It is comforting to see that all variables used are significant and, more importantly, of the expected sign. Thus, these results provide support for our specification of the demand estimation. For the interpretation here, we focus on the effects at the point of inflection[34] ($k = 3$).

First, looking at the volume effects of the exogenous variables: All variables included in the regression are highly significant at the 1% level. All temperature statistics (Tempeff15, Roll_temp24, Roll_temp240) bear coefficients with a negative sign and confirm that electricity demand falls with increasing ambient temperature. All daytime controls show up the expected sign as well: suncycle and deltasun have positive coefficients. This is sensible as electricity demand is higher during the day than at night (proxied for by suncycle) and rush or activity hours (proxied for by deltasun) in the morning and evening are also characterised by increased demand. The variables morning and EWH have coefficients

[34] As mentioned, the point of inflection is the centre point of the bid curves and the most relevant for equilibrium determination.

of a plausible negative sign. The morning as controlled for by our indicator variable[35] is shorter than the afternoon and evening together, thus total electricity consumption is lower as well. EWH stands for the deep night between 10pm and 4am and thus also corresponds to low demand periods. SolarRest controls for selfgeneration to cover own consumption and has a plausible negative coefficient. RteBlackBox on the other hand has a positive coefficient and confirms that actual demand is higher when the grid operator expects it to be the case.

The analysis of the price effects of these controls on demand functions is in line with the analysis of volume effects. This is coherent since for a linear downwards sloping demand curve, a left shift (volume decrease) is synonymous for a downwards shift (price decrease) of the curve. We consider that at the point $k = 3$, the demand functions are locally linear. We note the only exception for the coefficient of SolarRest which has a positive price effect, while a negative volume effect[36].

Second, these tables already give a descriptive analysis of the effects of exogenous variables on the shape of the demand bid function: We now compare all coefficients for a specific variable on the $K = 5$ different points on the demand function (we read the table horizontally and compare sign changes across columns). In table 3.3, we observe for each row at most a single sign change across the coefficients for the different points. Furthermore (and with few exceptions), the magnitudes of the coefficients generally increase or decrease monotonically along a row. This is very convincing as it suggests that exogenous variables have a monotone effect on the shape of the bid function. We thus only observe one-directional shifts (e.g. a unilateral left shift) or two-directional shifts (extension or contraction) in the volume dimension induced by the variation in exogenous variables. While the unilateral effects are explained analogously to our point specific interpretation on the point $k = 3$ above, we do not have a story to tell about two-directional effects. Tempeff15 results in a contraction of the bid function in terms of volumes (right shifts on low volume points, $k = 5, 4$ and left shifts on high volume points $k = 3, 2, 1$). Roll_Temp24 has the opposite effect and results in a volume extension of the

[35] The morning is defined as the hours before midday, which occurs when luminosity is at its daily maximum.

[36] We emphasize in the construction of our variable (appendix 3.9.4) that it is not possible to build a proxy for lighting consumption that would allow us to decorrelate the effects from photovoltaic production and lighting consumption. We therefore stick to the SolarRest proxy, which aims to capture the effect of Solar which is not captured by suncycle.

curve. Roll_Temp240 induces a pure left shift of the whole function[37]. For the intraday seasonality controls, the results are very clear. While suncycle results in an extension of the demand function[38], all other intraday controls (morning, deltasun, EWH) have unilateral effects. When the indicators morning and EWH are positive, we observe volume decreases at all points and thus a left shift of the function. Higher values of deltasun induces volume increases at all points of the bid function. Finally, we have SolarRest which induces an expansion of the curve and RteBlackBox which has a unilateral right shifting effect on the aggregate demand bid function.

The price variation of the demand bid function yields interesting results, too. Given that the prices of points $k = 1, 5$ are fixed, we only observe effects for the interior points. We thus focus on the effects on the points $k = 4, 3, 2$ only (called the "central demand function" here). Again, we only observe at most a single sign change across columns for any exogenous variable. Both Tempeff15 and Roll_Temp240 lead to an extension of the central demand function (we are now looking at vertical variation of the bid function as shown in fig. 3.5), while Roll_Temp24 causes a unilateral downwards shift. For intraday seasonality controls, we see that suncycle and deltasun have a contracting effect on the central demand function and morning a unilaterally negative effect. EWH leads to an expansion of the central demand function. SolarRest and RteBlackBox indicate an extension of the central demand function in the price dimension.

Overall, we take away a solid $R^2$ with coefficients of the correct sign. We furthermore have disclosed the White statistic which unanimously confirms heteroskedasticity in these regressions. The significance levels have been measured using robust standard errors (Huber-White). We point to the fact that the explanatory power of our demand estimations is highest for the point of inflection, in line with our expectations. Points of maximum curvature $k = 2, 4$ reveal lower $R^2$ statistics. This is likely due to the underlying data patterns that arise from bidding frictions, e.g. focal price points. For these points, it is thus not surprising that we do not observe convincing demand estimates - we note in particular the lack of explanatory power for the demand estimation in the price dimension for points of type $k = 4$.

[37]Excluding interaction effects, we note that the net effect of a simultaneous $1^o$C increase for all three temperature variables results in a net left shift of the function. In the price dimension (table 3.4) we observe a net downwards shift. Both effects suggest that electricity demand decreases with the prevailing temperature.

[38]Combined with the observed price effects from table 3.4, this suggests that demand is more price elastic during the day.

| | $k=5$ Volume | $k=4$ Volume | $k=3$ Volume | $k=2$ Volume | $k=1$ Volume |
|---|---|---|---|---|---|
| Tempeff15 | 50.72*** | 38.58*** | -130.3*** | -189.3*** | -204.0*** |
| | (9.942) | (10.13) | (10.94) | (13.32) | (13.20) |
| Roll_Temp24 | -63.57*** | -67.13*** | -48.87*** | 19.76 | 34.16** |
| | (11.78) | (12.06) | (13.14) | (15.83) | (15.76) |
| Roll_Temp240 | -60.15*** | -68.38*** | -78.49*** | -78.44*** | -87.38*** |
| | (6.655) | (6.867) | (7.450) | (10.05) | (10.00) |
| suncycle | -894.0*** | -652.1*** | 508.2*** | 1,351*** | 1,400*** |
| | (44.27) | (45.50) | (48.52) | (56.36) | (55.73) |
| morning | -101.2*** | -220.3*** | -814.8*** | -872.2*** | -885.8*** |
| | (27.52) | (28.33) | (30.44) | (37.71) | (37.28) |
| deltasun | 2,659*** | 2,850*** | 3,201*** | 1,721*** | 1,821*** |
| | (153.5) | (158.5) | (166.1) | (197.8) | (196.5) |
| EWH | -803.1*** | -833.1*** | -782.7*** | -354.7*** | -322.8*** |
| | (30.74) | (31.91) | (33.15) | (42.09) | (41.78) |
| SolarRest | -0.595*** | -0.363*** | -0.145*** | -0.0137 | 0.246*** |
| | (0.0282) | (0.0305) | (0.0342) | (0.0418) | (0.0407) |
| RteBlackBox | -0.00259 | 0.0127*** | 0.105*** | 0.107*** | 0.0979*** |
| | (0.00235) | (0.00243) | (0.00255) | (0.00316) | (0.00317) |
| Constant | 6,054*** | 7,086*** | 11,446*** | 15,215*** | 15,502*** |
| | (33.71) | (35.04) | (37.15) | (48.68) | (48.27) |
| Observations | 14,691 | 14,691 | 14,691 | 14,690 | 14,691 |
| $R^2$ | 0.201 | 0.219 | 0.478 | 0.344 | 0.346 |
| White | 548.6 | 524.9 | 407.9 | 961.8 | 944.8 |

Robust standard errors in parentheses

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

Table 3.3: Estimation results for demand volumes

*Note*: The estimated constants of this table or the left graph of fig. 3.5 indicate to which portion of the demand function the types of points $k = 1, .., 5$ refer.

| | $k=5$ Price | $k=4$ Price | $k=3$ Price | $k=2$ Price | $k=1$ Price |
|---|---|---|---|---|---|
| Tempeff15 | 0 | 4.675*** | -0.969*** | -1.308*** | 0 |
| | (0) | (1.523) | (0.0599) | (0.0980) | (0) |
| Roll_Temp24 | 0 | -10.07*** | -0.124* | -0.0470 | 0 |
| | (0) | (2.233) | (0.0713) | (0.116) | (0) |
| Roll_Temp240 | 0 | 4.250*** | -0.0901** | -0.353*** | 0 |
| | (0) | (1.147) | (0.0404) | (0.0607) | (0) |
| suncycle | 0 | -10.98** | 6.870*** | 11.60*** | 0 |
| | (0) | (5.020) | (0.258) | (0.445) | (0) |
| morning | 0 | -0.226 | -5.748*** | -9.009*** | 0 |
| | (0) | (4.133) | (0.173) | (0.285) | (0) |
| deltasun | 0 | -16.54 | 10.60*** | 18.72*** | 0 |
| | (0) | (19.16) | (0.881) | (1.497) | (0) |
| EWH | 0 | 5.136 | -1.756*** | -3.014*** | 0 |
| | (0) | (4.448) | (0.192) | (0.302) | (0) |
| SolarRest | 0 | 0.000532 | 0.00192*** | 0.00253*** | 0 |
| | (0) | (0.00307) | (0.000193) | (0.000326) | (0) |
| RteBlackBox | 0 | 9.91e-05 | 0.000906*** | 0.00147*** | 0 |
| | (0) | (0.000301) | (1.47e-05) | (2.26e-05) | (0) |
| Constant | 3,000 | 131.3*** | 39.45*** | -39.43*** | -3,000 |
| | (0) | (4.210) | (0.217) | (0.319) | (0) |
| Observations | 14,691 | 14,691 | 14,691 | 14,690 | 14,691 |
| $R^2$ | | 0.005 | 0.463 | 0.420 | |
| White | | 138.2 | 640.9 | 761.2 | |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.4: Estimation results for demand prices
*Note*: The estimated constants of this table or the left graph of fig. 3.5 indicate to which portion of the demand function the types of points $k = 1, .., 5$ refer.

### 3.5.2 Final regression

In the final regression, we investigate the effect of uncertainty on the aggregate supply bid function. We lay the focus on the point of inflection ($k = 3$) for a detailed interpretation of our results. We choose the point $k = 3$, because this type of point represents the region of the supply function that is most relevant for equilibrium determination. As mentioned, our prediction is only applicable to this centre portion of the supply function. We disclose the results for all other points $k \neq 3$ as a robustness check.

Each result table has four (three[39]) columns to show the results for different estimators and two specifications of the $\text{PLU}^D$. All other variables remain unchanged across the columns. In the tables, column 1 refers to the baseline specification of the $\text{PLU}^{D,J}$, where standard errors are calculated using the Huber-White sandwich estimator. Column 2 reports the results for the baseline model using bootstrapped standard errors with 300 repetitions. Column 3 reports the results for the regression on the kernel based $\text{PLU}^{D,J}_{\tilde{X}}$, using the sample size of each kernel as weights in the regression. Column 4 reports the results of the kernel based model using bootstrapped standard errors using 50 repetitions[40].

*Regarding notation:* In the results tables, PLUvRvar'm' stands for $\text{PLU}^R_{1,m}$ with 'm' being replaced by the initial of the variable in question (W, S and T, respectively). $\text{PLU}^R_{2,m}$ is indicated by the extension "sq". PLUvDvar'J' stands for $\text{PLU}^{D,J}$ with $J = \{P, Q\}$ representing the dimension in which the demand uncertainty is measured. The kernel based $\text{PLU}^D_{\tilde{X}}$ are given by PLUvDvarK'J' in the tables. To facilitate the reading of the tables, we adopt this notation for the discussion of the results.

**For the point of inflection ($k = 3$),** the results are shown in table 3.5. Regarding uncertainty from renewables production, only that of wind has a significant and robust impact. PLUvRvarW has a positive effect (significant at the 1% level) on the slope in all specifications. PLUvRvarWsq has a negative effect on the slope in all specifications, however this second effect is not robust to bootstrapping the standard errors. The signs of the estimated coefficients are in line with our expectations. To show this, we recall

[39]For computational reasons, we do not run the bootstrapping of the kernel based $\text{PLU}^D$ for the points $k \neq 3$, thus we only have three columns for these tables.

[40]Coefficients vary slightly ($< \pm 20\%$, no sign change), because the bootstrapping loop includes the kernel-based prediction of the uncertainty and thus varies the kernel sample sizes, which are used as weights in the final regression. Furthermore, the estimator has probably not yet fully converged with 50 repetitions, however for computational reasons we stick to this choice.

that both versions of the PLUvRvarW are based on the inverse of the characteristic lengthscale $L_W$ of autocorrelation of the wind speed measurements. Thus, when $L_W$ increases (it represents a decrease in the uncertainty since wind speeds are homogenous over longer distances), the PLU decreases (corresponding to a decrease in uncertainty). While an increase in the PLUvRvarW leads to an increase in the slope of the supply function, the effect is attenuated by the squared term PLUvRvarWsq for very small and large $L_W$[41]. The estimated coefficient for the latter is negative and suggests that for very short $L_W$ (i.e. very heterogenous wind speeds over the country), prediction errors cancel out. For very long $L_W$ (i.e. very homogenous wind speed profile), the marginal impact of $L_W$ on the level of uncertainty decreases.

With respect to the uncertainty from temperature forecasts, the results are always of the anticipated sign, but insignificant in most specifications. Note that we expect the main impact of temperature uncertainty to go via the demand response, which we account for in our proxy for the uncertainty from demand realisation (PLUvD). Uncertainty from Solar production is attributed no effect. This is not surprising as generation from solar is only a fraction of that generated from wind power and thus negligible. Furthermore, we are unable to disentangle the effect of solar generation from the reduced demand effect from high luminosity (which results in low demand for lighting). We do not find evidence for a direct response from suppliers to uncertainty in solar predictions.

Uncertainty from the realisation of market demand has a negative and significant effect when proxied for by price-based PLUvDvarP (see table 3.5) as opposed to a positive and significant effect when proxied for by a volume-based PLUvDvarQ (see table 3.5). The positive effect on PLUvDvarQ is in line with our prediction made in section 3.1.2. This results supports the theory that firms take uncertainty when bidding into account and consequently adjust their bidding strategy in order to minimise dynamic costs. However, our theory produces a prediction for volume based uncertainty only. We include the uncertainty proxy for price PLUvDvarP as a control and its effect seems rather robust. The effects of PLUvD in either the price or volume dimension are robust to the exclusion

[41]By looking at the variation of our data, we see that the negative effect of the PLUvRvarWsq term merely attenuates, rather than overrides, the positive effect of the PLUvRvarW term on the slope since in our dataset we very rarely observe PLUvRvarW values sufficiently large to exceed the maximum of the Laffer curve of the impact on the slope.

of the other[42]. We do not have a story to explain the opposing signs for the coefficients of the two proxies.

Furthermore, table 3.5 gives support to our extension using kernel based PLUvDs. Column 2 shows that the effects of the baseline PLUvD are not significant when bootstrapped. Our alternative is to use a more elaborate uncertainty prediction model. These kernel based PLUvD are more sophisticated in two respects: (i) the forecasting model is only applied locally, that is auctions are only compared to similar auctions and (ii) the obtained forecast is weighted by the sample size used for its prediction. Thereby, we control for the confidence of the firms in making those predictions. The results of the weighted regression are given in column 3. The results using the more elaborate prediction model are in line with those from the baseline regression, while being more accurate as indicated by the improved explanatory power of our model (we see a 16.5% increase of the $R^2$ from columns 1-2 to columns 3-4). Finally, the results of our kernel based model are more precise as indicated by the higher significance level for the PLUvDvarKP and PLUvDvarKQ, which are now also robust to a bootstrap (column 4).

We explicitly include the controls for the levels of the input prices of electricity producers ($\boldsymbol{X^S}$). We do not interpret these coefficients since there are no ex-ante expectations of their levels to affect the slope of the supply bid function. We briefly mention that intraday seasonality controls as well as other demand related variables are not included in this regression to avoid multicollinearity problems with the PLUvD, which are themselves computed as a linear combination of the demand control variables ($\boldsymbol{X^D}$).

Overall, we take away a goodness of fit of $\geq 20\%$ for our empirical model as well as the robust positive coefficients for both the demand based uncertainty proxy (PLUvDvarQ) and the weather based uncertainty proxies (PLUvRvarW and PLUvRvarWsq). We note the puzzling result for the PLUvDvarP, which we discuss in section 3.6.3.

[42]Thereby controlling for collinearity (correlation coefficient of 0.62) does not affect the results as otherwise OLS remains unbiased, standard errors are larger. Full results available from the authors.

For k=3 (Point of inflection)

| | (1) fxInvertQP | (2) fxInvertQP | (3) fxInvertQP | (4) fxInvertQP |
|---|---|---|---|---|
| PLUvRvarT | 0.000882 | 0.000882 | 0.00374** | 0.00508 |
| | (0.00152) | (0.00415) | (0.00155) | (0.00354) |
| PLUvRvarTsq | -0.000529 | -0.000529 | -0.00161*** | -0.00215 |
| | (0.000584) | (0.168) | (0.000603) | (0.183) |
| PLUvRvarW | 0.00790*** | 0.00790*** | 0.00647*** | 0.00574*** |
| | (0.00123) | (0.00257) | (0.00121) | (0.00207) |
| PLUvRvarWsq | -0.00235*** | -0.00235 | -0.00192*** | -0.00170 |
| | (0.000373) | (0.0644) | (0.000370) | (0.0479) |
| PLUvRvarS | -5.20e-10 | -5.20e-10 | -2.28e-09 | -2.23e-09 |
| | (2.68e-09) | (3.58e-08) | (3.16e-09) | (3.69e-08) |
| PLUvRvarSsq | 0 | 0 | 0 | 0 |
| | (0) | (0) | (0) | (0) |
| Coal | 6.90e-06*** | 6.90e-06*** | 5.18e-06*** | 6.29e-06*** |
| | (4.35e-07) | (4.64e-07) | (4.39e-07) | (6.87e-07) |
| Brent | -2.36e-05*** | -2.36e-05*** | -1.18e-05*** | -1.40e-05*** |
| | (1.51e-06) | (1.96e-06) | (1.53e-06) | (2.01e-06) |
| Gas | -2.82e-07 | -2.82e-07 | 1.37e-05*** | 1.36e-05*** |
| | (1.89e-06) | (9.41e-06) | (1.67e-06) | (2.46e-06) |
| IT2 | -2.71e-05*** | -2.71e-05 | -1.73e-05*** | -1.99e-05*** |
| | (2.17e-06) | (1.80e-05) | (1.34e-06) | (1.69e-06) |
| EUA | 7.20e-05*** | 7.20e-05*** | 2.62e-05*** | 2.71e-05*** |
| | (2.31e-06) | (4.49e-06) | (3.34e-06) | (6.84e-06) |
| Wind1DA | 1.04e-07*** | 1.04e-07*** | 1.18e-07*** | 1.25e-07*** |
| | (6.45e-09) | (1.03e-08) | (6.51e-09) | (7.63e-09) |
| Hydro | -7.55e-06*** | -7.55e-06*** | -4.08e-06*** | -5.88e-06*** |
| | (8.33e-07) | (2.24e-06) | (8.61e-07) | (1.11e-06) |
| PLUvDvarP | -0.000219*** | -0.000219 | | |
| | (4.57e-05) | (0.000203) | | |
| PLUvDvarQ | 0.000567*** | 0.000567 | | |
| | (9.44e-05) | (0.000585) | | |
| PLUvDvarKP | | | -0.000600*** | -0.000462*** |
| | | | (2.69e-05) | (4.24e-05) |
| PLUvDvarKQ | | | 0.000151*** | 0.000170** |
| | | | (3.39e-05) | (6.80e-05) |
| Constant | 0.00651*** | 0.00651*** | 0.00513*** | 0.00538*** |
| | (0.000208) | (0.000789) | (0.000195) | (0.000257) |
| Observations | 11,702 | 11,702 | 11,702 | 11,702 |
| $R^2$ | 0.200 | 0.200 | 0.233 | 0.234 |

Standard errors in parentheses. | p-values: *** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Regressions of slope on $\text{PLU}^R$ and $\text{PLU}^D$ at $k = 3$

*Note*: Standard errors are reported in parenthesis. Column 1 refers to the baseline specification. Column 2 reports bootstrapped results for the baseline model. Column 3 reports the results for the (weighted) regression on the kernel based $\text{PLU}^D_{\hat{X}}$. Column 4 reports bootstrapped results of the model in column 3.

**For the other points ($k = 1, 2, 4, 5$),** the results are given in tables 3.6, 3.7. Table 3.6 compares the bootstrapped results of the baseline specification of the PLUvD across all points $k = 1 - 5$. Table 3.7 compares the weighted regression for the kernel based PLUvD across all points. Recall our labelling for the points on the supply curve refers to low volume points for $k = 1$ and high volume points for $k = 5$. We comment on the effects over all points collectively in order to give an overview of the full bid function behaviour.

The results give support to our restriction that the prediction is local on the point $k = 3$. First of all, the explanatory power is highest for the point $k = 3$. Second, we observe coefficients of the anticipated sign for all[43] uncertainty proxies at the point $k = 3$, but we do not observe this for the points $k = 1, 2, 4, 5$.

However, we do observe a rather robust effect in both tables 3.6 and 3.7 for the uncertainty from market demand. Market uncertainty measured in volumes has a robust positive effect[44], while it has a a negative effect when measured in prices. While we note the robustness of the result for market uncertainty, we are cautious to rely much on the points $k \neq 3$ for the reasons outlined in section 3.4. Specifically, we referred to the definition of the points $k = 2, 4$ (based on maximum curvature) which make an analysis of the slope at these points questionable. We also referred to the importance of capacity constraints for the analysis of high and low volume points ($k = 1, 5$).

On the remaining controls, we do not observe a clear pattern on the effects at the different points. We run the analysis without these controls and note that the signs of all significant variables remain unchanged[45].

While we look at multiple points on the whole bid function (similar to a functional analysis), we do not learn about the functional behaviour of the bid function precisely in the central region of the bid function, where our prediction applies. There we unfortunately only have a single point to analyse. We highlight the importance of more reference points on the centre part of the bid function (close to $k = 3$) for further work in the discussion in section 3.6.1.

[43] We note the mentioned exception of PLUvDvarP, which we address in section 3.6.3.
[44] We note the only exception for $k = 1$ in the bootstrapped baseline regression model.
[45] Results available from the authors.

Comparison of column 2 for points $k = 1 - 5$

| | (k=1) fxInvertQP | (k=2) fxInvertQP | (k=3) fxInvertQP | (k=4) fxInvertQP | (k=5) fxInvertQP |
|---|---|---|---|---|---|
| PLUvRvarT | -4.14e-05 | -0.00252 | 0.000882 | -0.00442 | -0.000252 |
| PLUvRvarTsq | 1.56e-05 | 0.00106 | -0.000529 | 0.00149 | 9.10e-05 |
| PLUvRvarW | -6.04e-06 | -0.00549** | 0.00790*** | -0.000137 | -0.000555** |
| PLUvRvarWsq | 1.71e-06 | 0.00158 | -0.00235 | 0.000173 | 0.000169 |
| PLUvRvarS | 0 | -6.82e-10 | -5.20e-10 | 2.59e-09 | -4.17e-10 |
| PLUvRvarSsq | -0 | 0 | 0 | -0 | 0 |
| Coal | -8.54e-09*** | 2.36e-06*** | 6.90e-06*** | 2.22e-07 | -8.70e-07*** |
| Brent | 8.64e-08*** | -1.86e-05*** | -2.36e-05*** | -7.46e-06*** | 1.72e-06*** |
| Gas | -6.20e-08*** | -8.94e-06*** | -2.82e-07 | 9.04e-06*** | 4.53e-06*** |
| IT2 | 4.95e-08*** | 1.98e-05*** | -2.71e-05 | -1.96e-05*** | 2.23e-06*** |
| EUA | -3.14e-08*** | 8.69e-05*** | 7.20e-05*** | 4.71e-05*** | 2.89e-06*** |
| Wind1DA | -3.38e-10*** | 6.13e-09 | 1.04e-07*** | 1.64e-08** | -5.41e-10 |
| Hydro | 4.69e-08*** | -5.82e-06*** | -7.55e-06*** | -8.73e-06*** | 1.78e-06*** |
| PLUvDvarP | | -4.81e-05*** | -0.000219 | -0.000212*** | |
| PLUvDvarQ | -3.87e-06*** | 0.000442*** | 0.000567 | 0.000110** | 4.29e-05*** |
| Constant | 2.11e-06** | 0.00319*** | 0.00651*** | 0.00370*** | -0.000494*** |
| Observations | 11,702 | 11,702 | 11,702 | 11,701 | 11,702 |
| $R^2$ | 0.152 | 0.158 | 0.200 | 0.086 | 0.128 |

Standard errors available upon request | p-values: *** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Regressions of slope on $\text{PLU}^R$ and $\text{PLU}^D$ at all points

*Note*: Comparison of regressions using baseline PLUvD with bootstrapped SE.

Comparison of column 3 for points $k = 1 - 5$

| | (k=1) fxInvertQP | (k=2) fxInvertQP | (k=3) fxInvertQP | (k=4) fxInvertQP | (k=5) fxInvertQP |
|---|---|---|---|---|---|
| PLUvRvarT | -4.35e-05*** | 0.00231 | 0.00374** | 0.000559 | 0.000734*** |
| PLUvRvarTsq | 1.65e-05*** | -0.000735 | -0.00161*** | -0.000368 | -0.000280*** |
| PLUvRvarW | -7.84e-07 | -0.00527*** | 0.00647*** | -0.00205 | -0.000545*** |
| PLUvRvarWsq | 1.25e-07 | 0.00152*** | -0.00192*** | 0.000739* | 0.000163*** |
| PLUvRvarS | 0 | -1.09e-09 | -2.28e-09 | 2.40e-09 | -3.07e-10 |
| PLUvRvarSsq | -0 | 0 | 0 | -0 | 0 |
| Coal | -1.23e-08*** | 2.20e-06*** | 5.18e-06*** | 1.48e-06*** | -4.90e-07*** |
| Brent | 6.62e-08*** | -1.59e-05*** | -1.18e-05*** | -1.30e-05*** | 4.90e-07** |
| Gas | -6.50e-08*** | -2.68e-06 | 1.37e-05*** | 2.04e-05*** | 2.96e-06*** |
| IT2 | 3.02e-08*** | 1.12e-05*** | -1.73e-05*** | -2.61e-05*** | 2.47e-06*** |
| EUA | -6.75e-08*** | 8.19e-05*** | 2.62e-05*** | 3.19e-05*** | 8.35e-06*** |
| Wind1DA | -4.91e-10*** | 3.77e-08*** | 1.18e-07*** | 1.50e-08** | 3.49e-09*** |
| Hydro | 4.60e-08*** | -1.04e-05*** | -4.08e-06*** | -1.33e-05*** | 1.19e-06*** |
| PLUvDvarKP | | -4.95e-05*** | -0.000600*** | -0.000163*** | |
| PLUvDvarKQ | 2.26e-08 | 6.56e-05** | 0.000151*** | 4.08e-05 | 5.56e-05*** |
| Constant | 2.17e-06*** | 0.00341*** | 0.00513*** | 0.00406*** | -0.000351*** |
| Observations | 11,702 | 11,702 | 11,702 | 11,701 | 11,702 |
| $R^2$ | 0.107 | 0.149 | 0.233 | 0.117 | 0.131 |

Standard errors available upon request. | p-values: *** p<0.01, ** p<0.05, * p<0.1

Table 3.7: Regressions of slope on $\text{PLU}^R$ and $\text{PLU}^D$ at all points

*Note*: Comparison of regressions using kernel based PLUvD using weighted regressions.

## 3.6 Discussion

In this section we reflect on the results and use the opportunity to address a few issues, drawbacks as well as qualities of the research conducted. We first review the design of the empirical strategy and lend particular focus to how we deal with the issue of endogeneity. We then discuss the finding of the paper.

### 3.6.1 Internal and external validity

We believe that the work is credible due to many aspects of the research design. First, our set-up is based on rather intuitive relations which we test exclusively using simple OLS regressions. These regressions are econometrically unbiased given the data impurities that we observe. To guarantee precision of our estimates, we use bootstrapping techniques.

Second, considerable effort has gone into the treatment of the information that goes into the right hand side of our regressions. We do not only refer to the final PLUs used, but also point at the precise use of our controls. See for example the treatment of the variable RteBlackBox (details see page 157), which proxies for the information contained in the day ahead demand estimates (PrevConsoH) given out by the grid operator RTE. In order to extract the marginal information of the PrevConsoH estimate, which is not explained by other controls variables that we include in our analysis, we compute the residuals from a regression of PrevConsoH on our other controls, e.g. daytime controls such as suncycle. These residuals (called RteBlackBox) enable us to achieve a more sophisticated understanding of our regression output[46].

We also emphasize the aspect that we understand our dataset as a cross-sectional dataset rather than a time-series. While we do segment our dataset into weekday and weekend days and only run our analysis on the former, there is no reason why demand on a Tuesday afternoon should not be comparable to demand on a Thursday afternoon. We therefore ignore weekday dummies to increase our sample size. Furthermore, we avoid the use of dummy variables to control for the hour of the contracts in our regressions in order to further increase the sample size. However, we cannot compare electricity consumption between 4am and 4pm within a day. Neither can we compare two 4pm hours of a day in winter and another in the summer. Using dummies would first restrict our sample size, plus make our interpretation more difficult since the dummy variable aggregates the

[46] See for example the regression output of the demand estimation in tables 3.3 and 3.4.

effect over all conditions that change between samples. We use a bottom up approach that allows us to circumvent the sample size restriction and interpretation difficulties from daytime or seasonality dummies. Instead, we use continuous variables to control for the daytime and season by means of short and longer term temperature averages or other weather characteristics such as luminosity, which generates controls like deltasun[47].

Finally, we point at the empirical framework that allows us to run reduced form regressions on multiple regions of bid functions to better understand functional responses of those bids to variation in exogenous factors. We use 5 points for our analysis and refer to section 3.4.1 for the full details on this choice and the evaluation of the point selection. With hindsight, we feel that an additional two points would have been useful to better understand functional behaviour of the part of the bid functions, which is more relevant in equilibrium, i.e. on the centre part[48]. We note the computational demands of more points.

The methodology developed for our exercise on data from the French electricity market has applications in other domains. This is valid for the non-parametric point selection mechanism (section 3.4.1), the mechanism to aggregate local geographic data to a national level (appendix 3.9) as well as the identification strategy based on purely ex-ante data. In particular, we note the possibility to run reduced form estimation strategies for the analysis of markets which make access to functional data available. This includes all markets which use a multi-unit, uniform (or discriminatory) auction mechanism.

### 3.6.2 Endogeneity

The set-up of this work is specifically aimed at circumventing problems of endogeneity. For that sake, we keep a strict separation of ex-post and ex-ante information to the left and right hand sides, respectively, of any regression (which we achieve except in two instances that are described below).

To achieve this separation of ex-ante and ex-post information, both newly developed methodologies are highly useful. The point selection methodology from section 3.4.1 allows us to extract proxies for the level of uncertainty about the realisation of market demand, which are unaffected by the equilibrium interaction with the market supply. The weather data treatment methodology from appendix 3.9 enables us to base our proxies

[47]See section 3.4.4.3 for full details on our set of control variables for both demand and supply.

[48]For that we would recommend the points representing half of the maximum curvature between the current points $k = 2, 4$ and $k = 3$.

for the level of uncertainty from renewables on measures of the expected homogeneity of weather forecasts. Both methodologies allow us to recover ex-ante information on the prevailing uncertainty that firms have at their hands at the time of bidding. The information contained in all other controls used is also available at the time of bidding.

However for data availability reasons, we are not able keep this strict separation at all times in practice and revert to using ex-post data to compute some variables that should ideally be computed on ex-post information only. This is the case twice in this work: (i) we use observed weather data to compute the variable Solar[49] and (ii) we use the pooled data over all auctions for the demand estimation and subsequent uncertainty forecast of equations 3.2 - 3.5.

In both cases, we do not believe that this choice compromises our results. For the case of Solar, we use realised luminosity instead of forecast data. This is as if weather forecasts were perfectly accurate. Given that solar production only accounts for a small fraction of total electricity generation and that we extract the very informative component of the Solar variable by using the variable suncycle (which is arguably well predictable), we do not see the use of ex-post data as problematic.

For the case of the $\text{PLU}^D$ computation, we run the demand estimation pooled over all observed auctions (i.e. past and future) and say that firms have this level of information when bidding in each auction of our sample. We do so, because we do not have the necessary data before 01.01.2011 and thus cannot calibrate our forecasting model on a "learning" dataset. Instead, we assume that demand patterns conditional on the explanatory variables have remained constant over our 2.5 years time period of analysis. The estimation based on pooled data then yields, on average, the same insights as an analysis conducted purely on past data.

We could test robustness of our pooled approach by investigating the effect of a restriction on using only past data in the demand estimation. A learning effect could arise from more precise estimations of demand functions. However, due to the long experience of most firms on the market in reality, this learning effect would be artificial and not represent a real insight. We therefore accept the possibility of a (small) endogeneity concern in this paper and further work could fully circumvent this issue by extending the database appropriately.

[49] Contrary to Wind1DA and Tempeff15, which we are able to compute purely on forecast data.

### 3.6.3 Findings

This chapter investigates whether uncertainty affects supplier bidding as predicted by the theory. We find that uncertainty from weather forecasts indeed affects the suppliers' bid function as expected. The central part of the aggregate supply function steepens when the level of uncertainty increases. We take this as evidence that firms take dynamic cost considerations into account and adjust their behaviour when facing increased expected dynamic costs.

We also find significant results for the effect of the level of uncertainty about the realisation of market demand on the suppliers' behaviour. However, we observe a strong discrepancy between the effect of uncertainty as measured on price volatility and the effect of uncertainty as measured on volume volatility. While the former is attributed a negative effect, the latter is attributed a positive effect on the slope of the aggregate supply function. The differing opposing results are robust in all specifications and seem to be of too much importance to be neglected.

The two proxies in question ($\text{PLU}^{D,P}$ and $\text{PLU}^{D,Q}$) are two variables designed to measure the same information, namely the prediction error of the demand function. As such, they are identical with respect to the set-up, computation as well as point at which they are extracted. They only differ with respect to the dimension in which the variation of the demand function is measured, the former in the price dimension and the latter in the volume dimension.

A theory using linear functions would predict that these measures of the shifts of the demand line are identical and interchangeable (modulo a translation by the slope). Also our data, i.e. the observed bid functions, suggests that, at least locally at the point $k = 3$, the bid functions are linear[50]. Furthermore, our demand estimation models for both price and volume variation[51] indicate that the prediction model used works well in both dimensions. In particular at $k = 3$, significance and equal signs on coefficients for all terms included as well as similar explanatory power[52] in both regressions confirm the similar nature of the two proxies.

Our recovered $\text{PLU}^{D,P}$ and $\text{PLU}^{D,Q}$ are, as expected, collinear[53]. While OLS remains unbiased in the presence of collinearity between two regressors, its precision is reduced.

[50]Recall the graph in figure 3.4.

[51]Precisely look at columns 3 of tables 3.3 and 3.4.

[52]$R^2$ of 0.463 for the price and 0.478 for the volume regression.

[53]Not perfectly, but with a correlation coefficient of 0,62

We correct for the collinearity by dropping one proxy or the other, but the individual results remain unchanged - the coefficients of the two proxies keep opposite signs.

Assuming that our empirical strategy is valid to test the relationship of interest, a possible reason for our intriguing observation could be that the slope of the demand function, which relates $\text{PLU}^{D,Q}$ and $\text{PLU}^{D,P}$, is endogenous on the uncertainty. Uncertainty could not only unilaterally shift the demand function in one dimension (either P or Q), but also affect the shape and thus the slope of the curve. This effect is not accounted for in our research design and could drive the opposing results for both proxies. Also, our theoretical model gives no prediction on this effect. Thus, the possibility of a demand function that reacts to uncertainty hints at the fact that we need new theories to explain both demand and supplier bidding behaviour on the electricity market. This calls for new theoretical models to better explain the shape of aggregate bid functions, which are S-shaped overall. Special attention in these models should be placed on the effect of uncertainty and its importance for bidders via the link of dynamic costs.

Finally, we give attention to the functional data analysis. For the demand estimation, we have disclosed a descriptive analysis of the full bid function. Results are coherent with a single sign change per "row" in tables 3.3 and 3.4. We also highlight specifically the strengthening or weakening effects of exogenous variables on different points in these tables. On the slope estimation in tables 3.6 and 3.7, we show that our prediction and results are local for the central part of the supply function. When we leave this central region, the effects of uncertainty on the suppliers' behaviour disappear for uncertainty from meteorological forecasts and weaken for uncertainty from market demand.

We conclude that the impacts of variations in exogenous factors on the shape of the bid functions are not uniform. Non-linear effects are neither predicted by our linear theory nor have been shown in previous studies apart from Wolfram (1999) and Wölfing (2013). Our results hint at more intricate mechanisms which drive the shape of these bid functions.

## 3.7 Conclusion

This chapter applies two novel methodologies, the point selection technique (section 3.4.1) and the aggregation of geographic information methodology (appendix 3.9.1), for an analysis of the electricity market. We observe that bidders take uncertainty from meteorological forecasts, which affects renewables generation, as well as uncertainty from demand realisation into account. The results indicate that electricity suppliers react to an increased level of uncertainty by bidding more volume elastically (steeper supply functions in the dimension Q (x-axis) - P (y-axis)) in order to minimise expected dynamic costs, which increase with the uncertainty. This is particularly true for the central part of the supply bid function, where capacity constraints are negligible. The results also indicate that not only supplier bidding is affected by uncertainty, but that the level of uncertainty also impacts bidding from the demand side of the market.

We identify three aspects for further empirical work. (i) More points on the central region of the bid functions would allow to have a functional analysis of the effects of uncertainty on the supply function. This analysis could reveal the interval of points on the supply function for which we see significant effects of uncertainty on bidding. (ii) Future work should focus on investigating the endogeneity of the demand function and address the differing results for market uncertainty as measured in prices or quantities. (iii) We believe that frictions in the bidding, in particular the focal price points discovered using our point selection methodology (details appendix 3.8), deserve further attention.

Concurrently, the results also call for more advanced theoretical work on the shape of bid functions of players, in particular to explain non-linear shapes. This is also suggested by our functional analysis of the demand functions which hints at non-unilateral effects of exogenous variables on the shape of the functions. The economic insight hidden in full bid functions is vast and a better understanding of these could be applied to address important welfare questions[54].

[54] One such application, which the authors currently focus on is the question of the optimal choice of the geographic installation of renewable electricity generation units (solar panels and wind turbines) with respect to minimising the intermittency of renewables generation. A clear understanding of the effects of uncertainty on the market is vital to close the analysis on organisational questions of the market. This is outside of the focus of this chapter.

## 3.8 Appendix: Results of the point selection

### 3.8.1 Precision of point selection

We have selected $K = 5$ types of comparable points for each of the 31500 demand and supply functions. This section details the results of the point selection methodology and presents evidence why the point selection algorithm has worked reliably.

The graphs in figure 3.10 show the local density of selected points in the quantity - price 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.

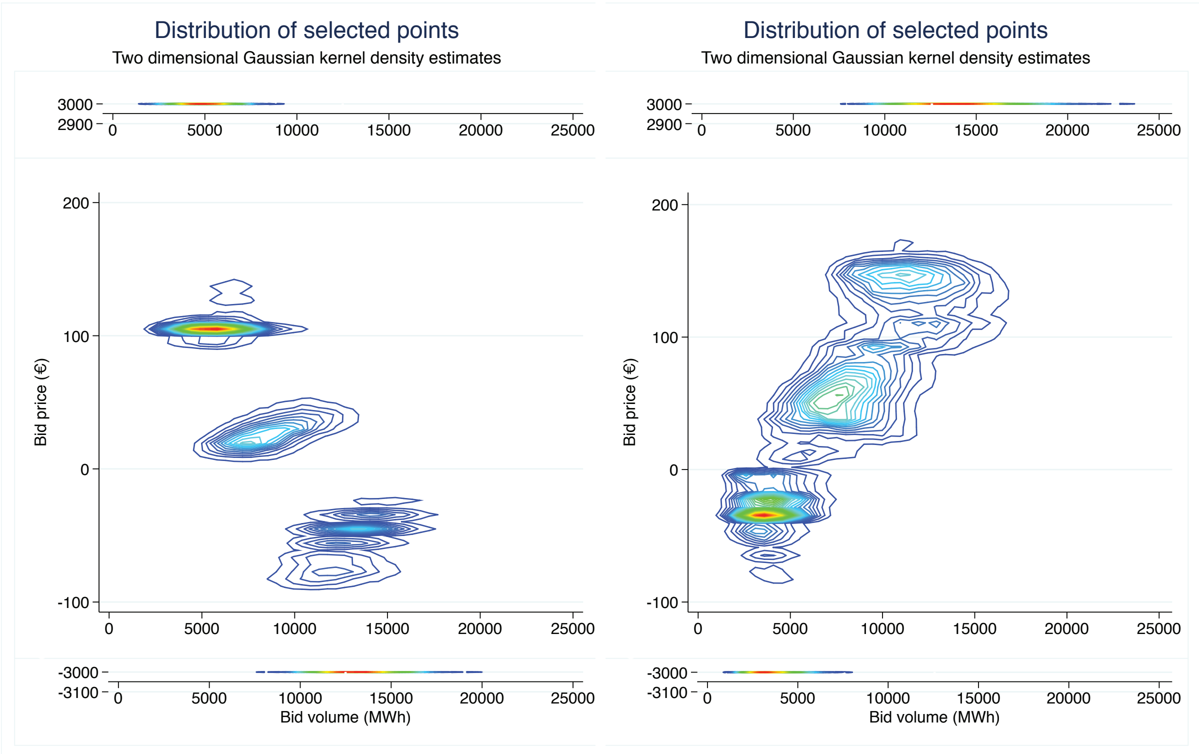



Figure 3.10: Heat map on selected, comparable demand and supply points
*Note:* Please note the discontinuity in the scale of the y-axis. The three separate graphs are arranged to be understood as a single one. The warmer the colours of the heat map, the higher the frequency of selected price-quantity pairs. The colour legend is omitted for brevity, density changes between contours are of the order of $10^{-4}$.

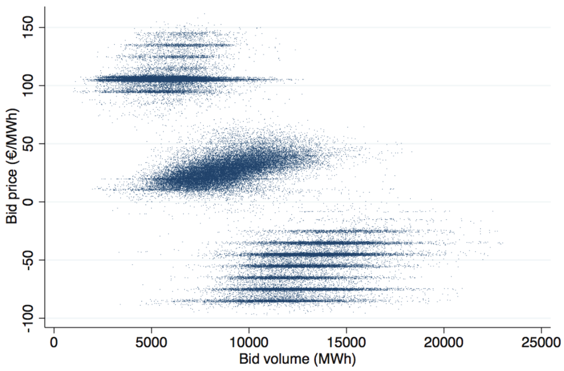
In figure 3.10, selected points of type $k = 1$ manifest at the bottom of the graph with prices fixed at $-3000$€/MWh. Similarly, $k = 5$ points appear at the top of the graph with prices fixed at $+3000$€/MWh. The three distinct groups of data points refer to points of type $k = 4$, $k = 3$ and $k = 2$, respectively, when reading the zoomed, center part of the graph from top to bottom.

In appendix 3.8.7, tables 3.8 and 3.9 allow to match data frequencies in the left graph of figure 3.10 with their types. Tables 3.10 and 3.11 relate to the data of the right graph in figure 3.10.

We note that the point selection for the demand curves has produced groups of points that are more distinct (and thus more robustly attributed to a certain type $k$) then for the supply function. While the smooth logistic function approach was unable to cope with the variations in the data from the electricity market (Belsunce, 2011), our more flexible non-parametric approach is more robust. 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[55]. Overall, this is strong evidence that the algorithm is able to distinctly differentiate between points of different types.

### 3.8.2 Observations of bidding frictions

Distinct point selection is further supported by the evidence in figure 3.11. These graphs are analogous to those in figure 3.10 and show the distribution in the quantity - price space of the selected points separately for the demand and supply function. Distinct clouds are an indication that selected points are different across types $k$.



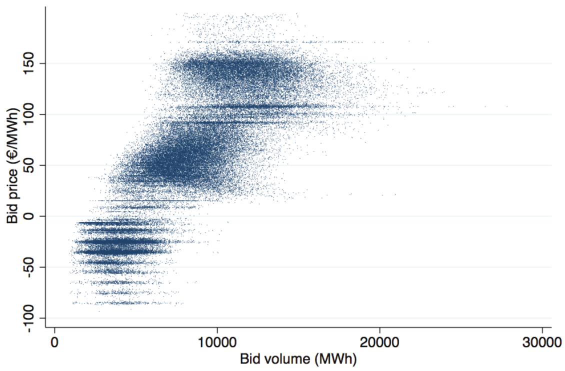




Figure 3.11: Distribution of selected demand (left) and supply (right) points

However, a feature of the graphs is striking: patterns (horizontal lines) seem to exist for the selected points of type[56] $k = 2$ 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

[55] On very rare occasions, our algorithm was unable to distinctly select between neighbouring point types, because the original bid function was linear for a large part. E.g. point $k = 2$ cannot not be identified if the bid function is linear between points $k = 1$ and $k = 3$. We give details on the outlier removal step in appendix 3.8.6.

[56] 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.

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[57], we acknowledge the existence of bid point patterns in the values (i.e. prices and quantities) of selected points. 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.

We are interested in $S'$, the slope at each selected point - an information measured at the selected point. Therefore, we investigate whether the values of the first derivative at the point $k = 3$ display a pattern. Figure 3.12 shows the histogram of the slope of the supply function for the point $k = 3$. No pattern in the values of the derivatives is apparent.

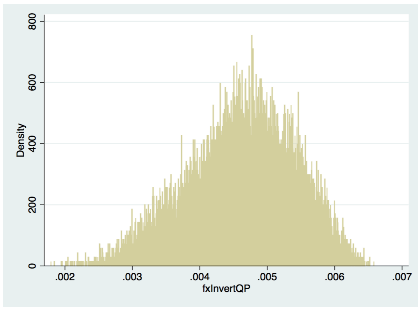



Figure 3.12: Histogram of the slope of the supply curve at $k = 3$

[57]We do not investigate the origins of bidding frictions in this chapter, which focuses 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.

### 3.8.3 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. From 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, we want to measure how well our information (as captured by the selected points) captures the variation of the original bid functions. This allows to quantify the marginal gain of information for an additional selected point.

For this investigation, we first recover the master curve (the mean expectation of a demand curve) and its confidence interval[58] for $K = 0$ to $K = 5$ points. Then, we look at the decrease in uncertainty achieved by including an additional point, obtained using our technique. Figure 3.13 shows the master curves (red line) and the expected error (pink shaded interval above and below the master curve) as a function of the number of reference points[59].

Without any reference point, the uncertainty on the inferred bid function would lie in the interval shown in graph A of figure 3.13. With two reference points (namely the minimum and the maximum quantity), the uncertainty is reduced as shown by the smaller error interval in graph B. Graph C adds a third point (the point of inflection) and Graph D adds another two points (the two points of maximum curvatures). Figure 3.13 shows clearly that with an increasing number of reference points, we obtain a more precise information about the original bid function. We quantify the gain in precision by measuring the surface of the pink shaded area in each of the graphs A to D. The result is shown in figure 3.14 and reveals decreasing marginal information for each additional point. By selecting $K = 5$ points, we are able to reduce the uncertainty about the original curve by

[58] To compensate for asymmetric variation above or below the master curve, we do not use the standard deviation to compute the confidence interval. Instead our upper (or lower) bounds are given by the mean of all curves below (or above) the expected master curve respectively.

[59] The master curve in A is obtained by rescaling all demand functions by their mean value. The master curves in B - D are obtained by rescaling the reference points, such that they coincide with corresponding point on the master curve in A plus rescaling all points between the reference points by a vector obtained as a linear combination of the displacement vectors of the closest reference points.

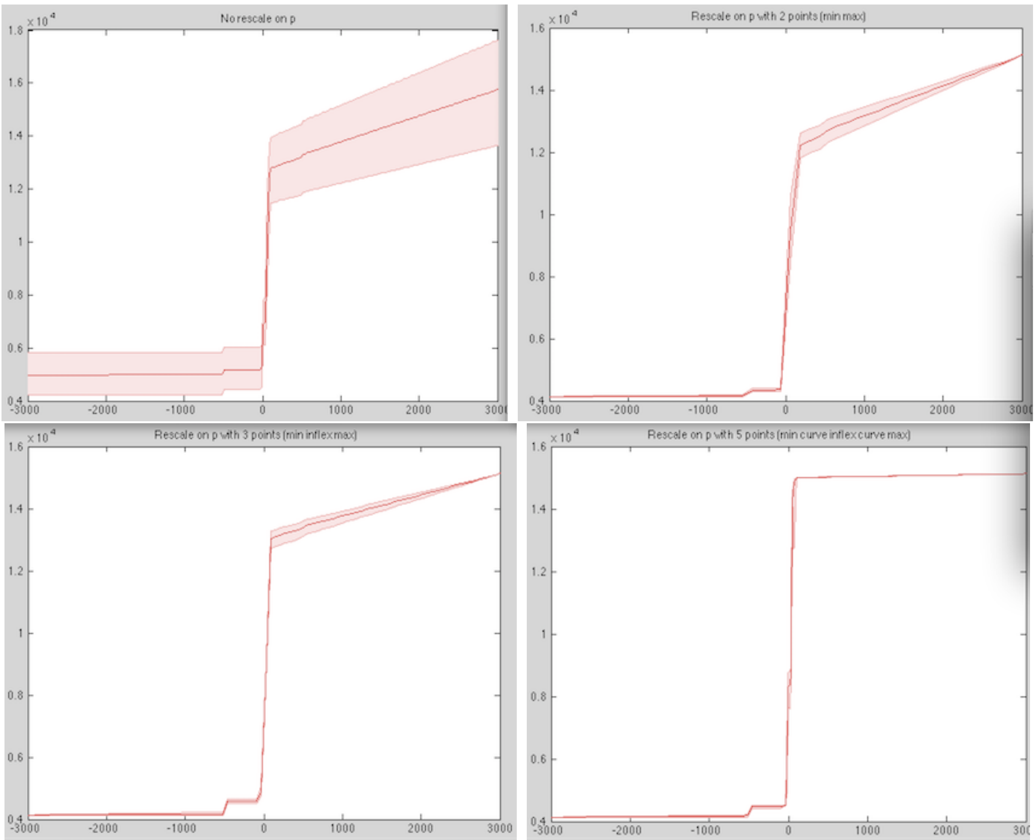



Figure 3.13: Error bars as a function of the number of extracted points
Note: In all graphs: Price on the x-axis, Volume on the y-axis. The graphs represent the master curve with the error interval for inferring the original bid function, conditional on the number of extracted, reference points (RP). 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.

a factor of about 50 (see figure 3.14). In other words, we capture between 95% and 98% of the variability of the supply and demand functions. We see this insight as support for using $K = 5$ points for further work.

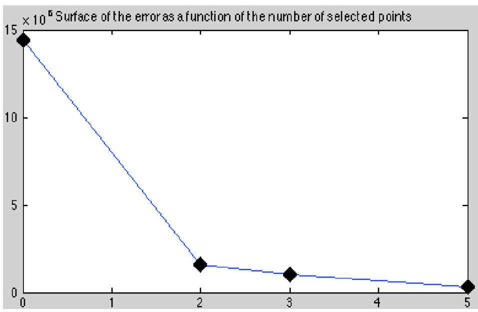



Figure 3.14: Surface of the error function as a function of reference points
Note: The graph plots the size (y-axis) of the pink shaded error area in figure 3.13 against the number of reference points (x-axis).

While the graphs in figure 3.13 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 3.15.

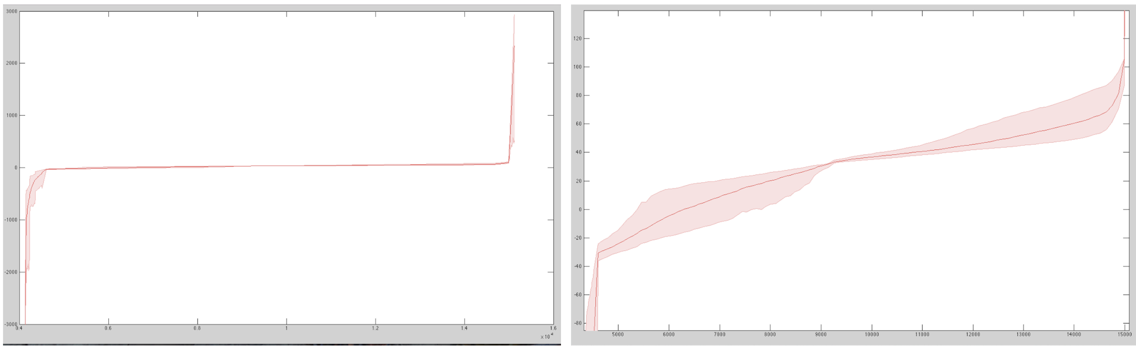


Figure 3.15: Overall and zoomed mastercurve with confidence intervals
Note: Overall (left) and zoomed (right) master curve in the quantity (x-axis) - price (y-axis) dimension.

### 3.8.4 Discussion and conclusion of point selection

In this short appendix, 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 in this paper we apply it to aggregate demand functions, the methodology is fit for the analysis 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 mixing up of “local” bid function information from all parts of the function at once. Our non-parametric approach allows specifically to control the smoothing parameter and thus enables the researcher to choose the smoothness of a bid function when extracting the points of interest. In any case, the smoothing range of the new technique is merely a fraction of that of the traditional approach and, hence, we do not mix up information of the different parts of the bid function.

The results of the comparable point selection are encouraging. We show that each type of point is distinctly chosen and while we see patterns on the levels of the original bid function, we do not observe any patterns on the derivative information (i.e. slope) extracted at the selected points. We acknowledge the existence of bidding frictions in the original data and highlight this observation for further work. Overall, we deem the selection of points to be of sufficient precision for a detailed study of the behaviour of the levels (and slope) of the bid functions (at the point $k = 3$, respectively[60]).

Finally, we emphasize that our methodology is scalable in the sense that we can select as many comparable points as desired[61]. We restrained ourselves to 5 points for computational reasons, more points would allow a more detailed functional analysis of the curves. In particular, more points between $k = 2$ and $k = 4$ would be of interest as these are relevant for equilibrium.

### 3.8.5 Technical details

#### 3.8.5.1 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 use a characteristic of the auction mechanism (the linear interpolation between consecutive bid points, for details see section 3.2.2) to our advantage and are able to transpose the observed bid function to one that suits our needs. This is done by adding linearly

[60] It does not make sense to look at the slope at points $k = 2, 4$ since these are defined as points of maximum change of the slope.

[61] See footnote 17 for details.

interpolated points at the unit cent level (corresponding to the minimum bidding unit). The kernel density estimation is then able to consistently 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.

#### 3.8.5.2 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 why we increase it is to ensure that the distribution of the first derivatives is uni-modal[62]. 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[63], thus we have a further incentive to 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 3.16 and 3.17. 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 3.16 shows the impact of smoothing on the estimates of the first derivative as compared to figure 3.17. 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

[62]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.

[63]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, 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.

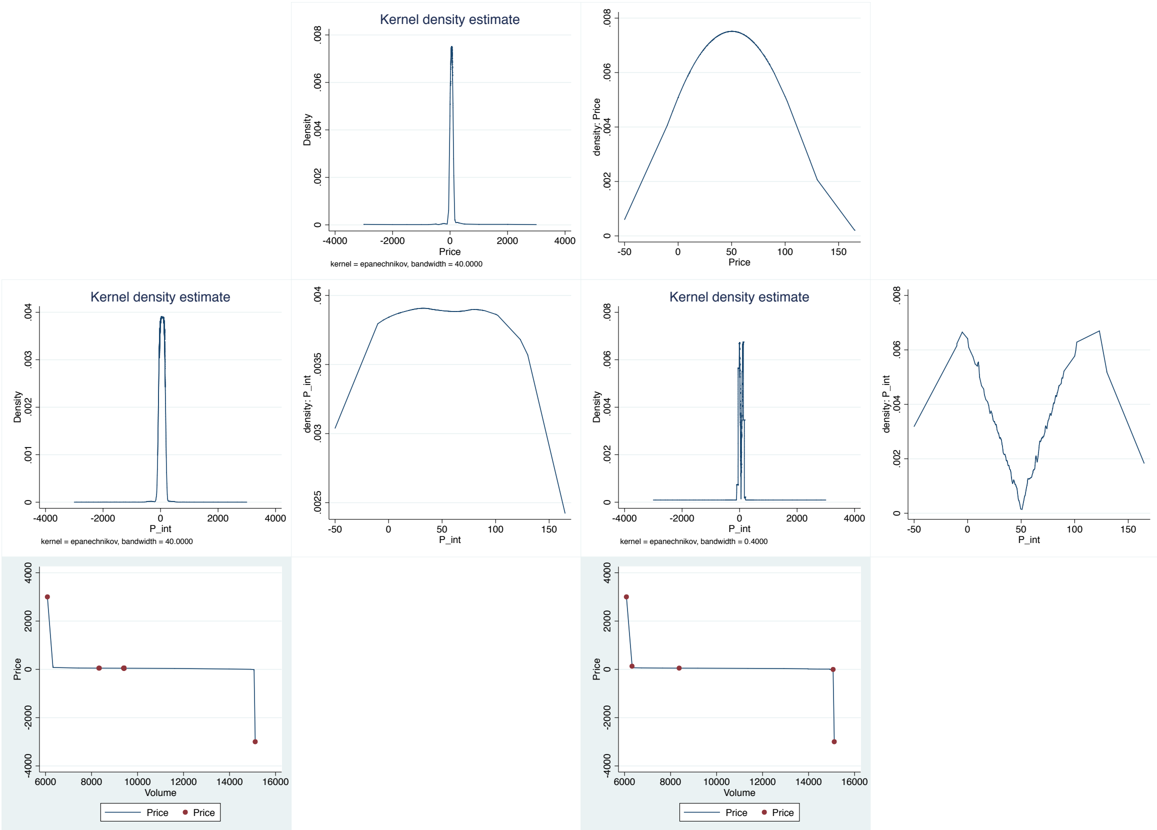



Figure 3.16: Comparison of bandwidths: Large bandwidth in first stage
*Note:* Large bandwidth in first stage (top row), large bandwidth in second stage (second row left), small bandwidth in second stage (second row right), Resulting selection of points for large bandwidth in stage one and two (bottow 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 3.16B).

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 satisfying robustness to noise in the bid function. We are aware that this subjective

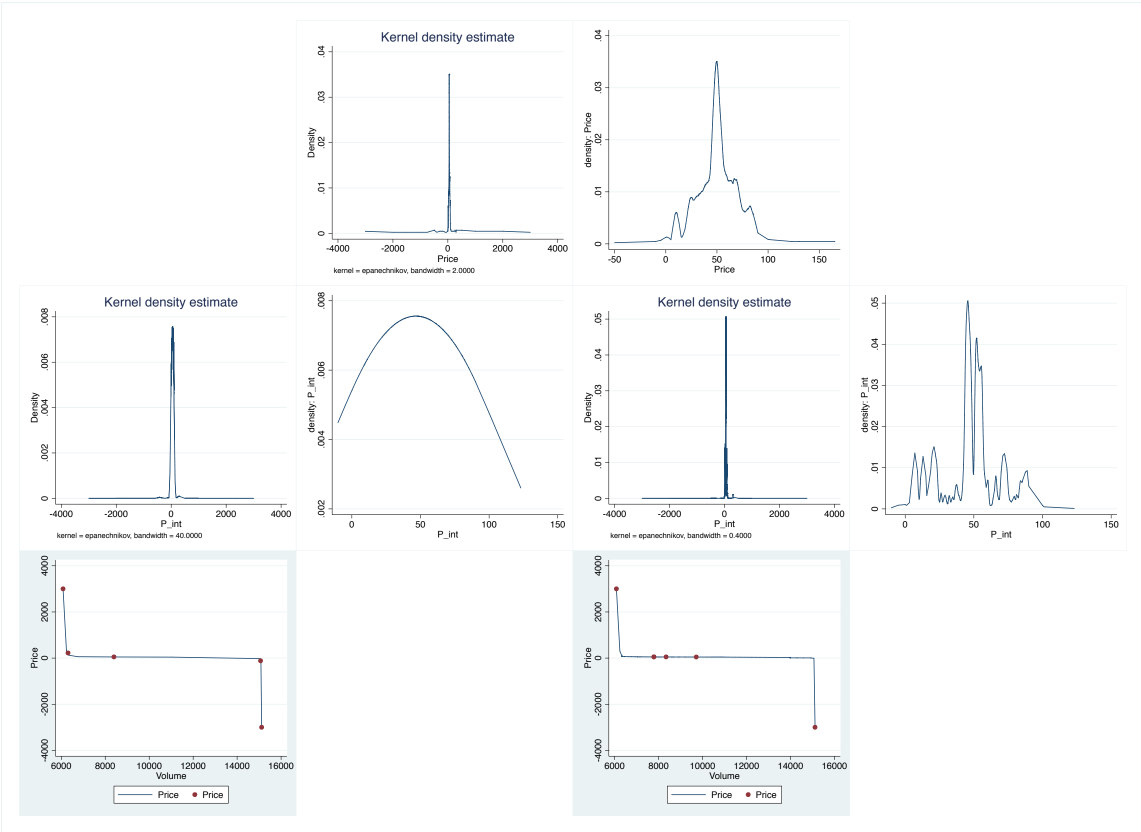



Figure 3.17: Comparison of bandwidths: Small bandwidth in first stage
*Note:* Small bandwidth in first stage (top row), large bandwidth in second stage (second row left), small bandwidth in second stage (second row right), Resulting selection of points for large bandwidth in stage one and small bandwidth in stage two (bottom row left, C) and selection of points for small bandwidth in stage one and two (bottow row right, D).

setting of the bandwidth is not without consequence for our work. However for computational reasons[64], we do not run a full robustness test on this choice ex-post.

### 3.8.6 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).

We screen for adjacent points that display quasi no variation in volumes. As an example, figure 3.18 shows a histogram of volumes differences over 2 selected points (from $k = 2$ to

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

$k = 4$) and reveals a positive mass point at zero, indicating outliers that do not display any volume variation between both points of maximum curvature on the same bid function. We use the histogram to identify and drop those outliers from our dataset.

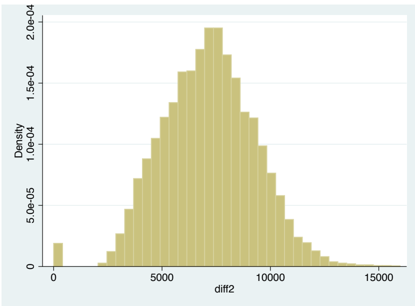



Figure 3.18: Histogram of volume variation between points
*Note:* The histogram shows the volume difference between points $k = 2$ and $k = 4$ of the same bid functions.

### 3.8.7 Summary statistics of selected points

| | Mean | Median | StdDev | Min | Max |
|---|---|---|---|---|---|
| Prices for $k = 1$ | -3,000.0 | -3,000.0 | 0 | -3,000 | -3,000 |
| Prices for $k = 2$ | -56.7 | -55.0 | 19 | -97 | 70 |
| Prices for $k = 3$ | 27.6 | 26.8 | 11 | -27 | 93 |
| Prices for $k = 4$ | 120.2 | 105.4 | 193 | -11 | 2,999 |
| Prices for $k = 5$ | 3,000.0 | 3,000.0 | 0 | 3,000 | 3,000 |

Table 3.8: Prices of selected demand points

| | Mean | Median | Std. Dev | Min | Max |
|---|---|---|---|---|---|
| Volumes for $k = 1$ | 13,328 | 13,222 | 2,213 | 4,990 | 23,254 |
| Volumes for $k = 2$ | 12,919 | 12,824 | 2,238 | 3,321 | 23,001 |
| Volumes for $k = 3$ | 8,779 | 8,664 | 2,028 | 1,958 | 18,335 |
| Volumes for $k = 4$ | 5,777 | 5,730 | 1,558 | 987 | 12,773 |
| Volumes for $k = 5$ | 5,031 | 4,968 | 1,467 | 914 | 11,301 |

Table 3.9: Volumes of selected demand points

| | Mean | Median | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Prices for $k = 1$ | -3,000.0 | -3,000.0 | 0 | -3,000 | -3,000 |
| Prices for $k = 2$ | -30.3 | -25.0 | 219 | -2,999 | 439 |
| Prices for $k = 3$ | 61.3 | 58.6 | 24 | 11 | 526 |
| Prices for $k = 4$ | 133.9 | 136.3 | 32 | 36 | 626 |
| Prices for $k = 5$ | 3,000.0 | 3,000.0 | 0 | 3,000 | 3,000 |

Table 3.10: Prices of selected supply points

| | Mean | Median | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Volumes for $k = 1$ | 3,721.7 | 3,526.0 | 1,344 | 618 | 10,594 |
| Volumes for $k = 2$ | 4,432.8 | 4,226.0 | 1,602 | 844 | 11,765 |
| Volumes for $k = 3$ | 8,467.2 | 8,365.5 | 1,814 | 3,431 | 20,932 |
| Volumes for $k = 4$ | 11,849.5 | 11,717.7 | 2,411 | 3,641 | 27,810 |
| Volumes for $k = 5$ | 14,390.6 | 14,142.0 | 3,052 | 6,580 | 35,356 |

Table 3.11: Volumes of selected supply points

## 3.9 Appendix: Methodology to extract and aggregate geographically dispersed information (Technical details on PLU$^R$)

We have two types of meteorological data: observations and forecasts. The methodology for each differs slightly.

### 3.9.1 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[65], 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 3.19.

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. 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

[65] There are 95 Départements in France

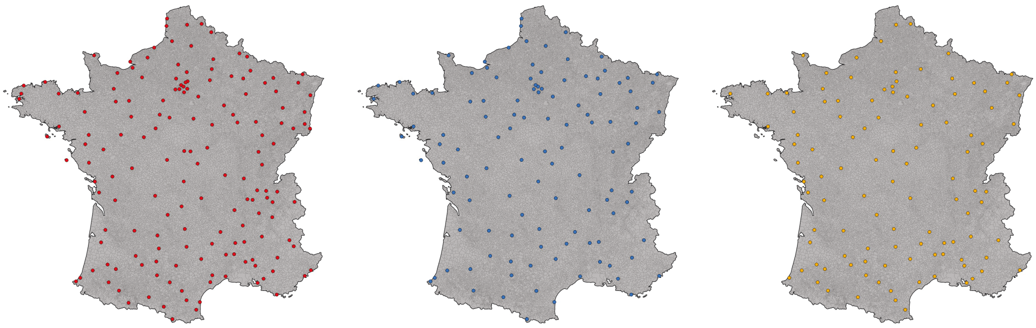


Figure 3.19: Weather stations for which we have hourly data.
*Note:* Left: temperature, center: wind speed, right: light intensity.

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 neighbour interpolation method, well known for its good balance between speed and accuracy. In short, through the use of a Voronoi algorithm (a method that divides the plane in regions "belonging" to certain points), one is able to define the natural neighbours of a point, that are then used to perform the interpolation using a ratio of surfaces as weights (see Fig 3.20 for more details).

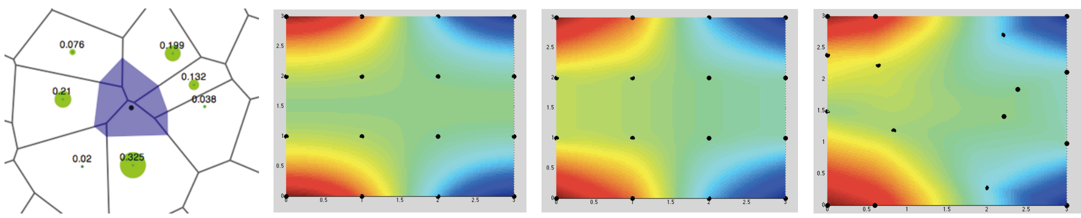



Figure 3.20: Steps in interpolation methodology on weather data
*Note:* Left: Voronoi's algorithm is applied once on the reference points highlighted in green to obtain the white surfaces, and a second time on the same points to which is added the query point in the center to obtain the new blue cell. 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 neighbour interpolation. Center right: interpolated surface with a reference set of 16 evenly organised points, represented in black. Right: interpolated surface with a reference set of 16 unevenly organised points, represented in black. From 16 points one is able to reconstruct the color mapped surfaces which aim at being able to reproduce the reference one, represented in the center left image.

### 3.9.2 Picture treatment to recover weather forecasts

Forecasts are obtained from the Global Forecast System (GFS), and come in the form of colormaps, as shown in Fig 3.21. We are going to illustrate our methodology on temperature data, but the exact same approach is performed on wind speed data. The general idea is that the pointwise precision is low ($2°C$ per color) but the overall map contains more precise topological data than a few tens of precise but sparse stations.

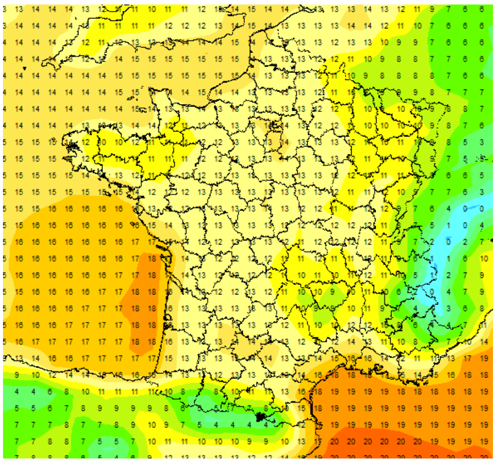


Figure 3.21: Example colourmap of temperature forecasts
*Note:* Temperature forecast from a simulation run by the GFS at 6 a.m. on the 3rd of November 2011, for a forecast at 22 p.m.

To extract the relevant data we first clean the color map from its irrelevant information, namely the temperature in numbers and the borders. Note that this step introduces a small amount of high spatial frequency noise, see Fig 3.22 left and center left.

Second, 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 3.22 center right.

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. We could perform an interpolation, but these methods are not well suited to such organised reference sets, here data points on curves representing iso-temperatures. In addition the first step introduced some spatial noise which we want to correct to some extent. We allow here our fitted surface to take different values than our data points. This allows us to define the rigidity of our fitted surface, i.e. a cost associated to spatial noise, and therefore reduce the importance of the high frequency noise introduced in the first step. The end result is presented in Fig 3.22 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 $2°C$.

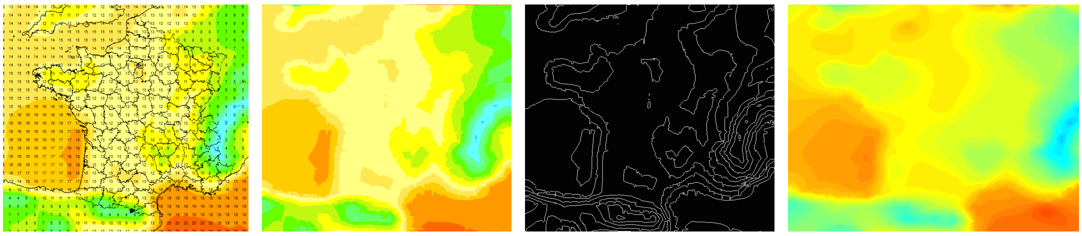


Figure 3.22: Steps of picture treatment methodology
*Note:* Left: reference image. Center left: borders and numbers are removed. Center right: edge recognition. Right: final fitted surface.

### 3.9.3 Autocorrelation lengthscale

We also use this dataset to build measures of the weather uncertainty. To do so we measure the autocorrelation lengthscale of our three weather variables of interest : temperature, wind speed and light intensity. This lengthscale measures the extent to which the weather variables are correlated spatially. We consider that the autocorrelation lengthscale is inversely proportional to uncertainty about the variable we are interested in. The reasoning is based on the idea that in order to forecast the effect of a weather variable on a national level, firms must aggregate the local weather information to a national level. Local correlation of the realisations of a variable is our measure of the uncertainty. When it is small, the variable is less spatially correlated, leaving more room for noise to blurr the anticipation of the impact of this variable on a national level. Conversely, when the autocorrelation 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 finitely 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 every possible pair of points in our dataset at a given point in time. For a given pair we compute its distance and its absolute difference in value (in black in Fig.3.23). 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.3.23).

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.3.23), and recover the exponential decay parameter as our lengthscale (in green in Fig.3.23). We perform this operation for every hour in our dataset and every weather variable. The results are a time series for the characteristic lengthscale of the weather parameters.

### 3.9.4 Aggregation of local information (Wind, Solar, Temperature)

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

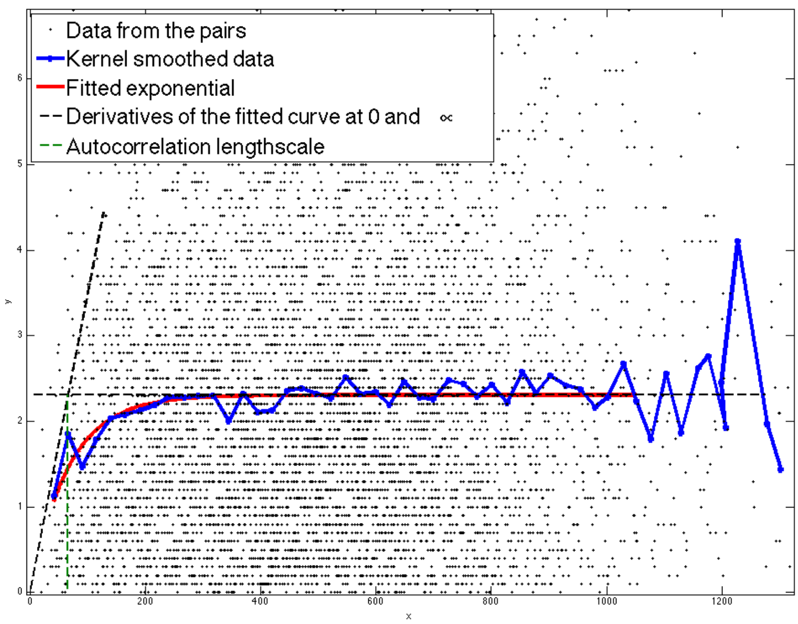



Figure 3.23: Autocorrelation lengthscale computation.

*Note:* 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 $\infty$. In green is the recovered autocorrelation lengthscale. The unit for the lengthscale is in km.

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, i.e. that they have the same response curve and height.

A typical response curve is represented in Fig. 3.24. 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[66] to obtain a rough estimate of the french average wind

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

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.

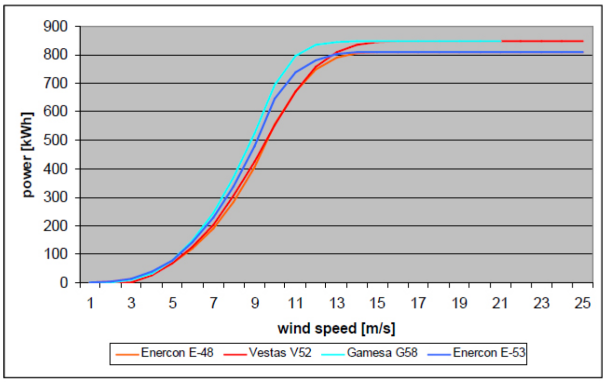



Figure 3.24: Typical response curves of different wind turbines

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 forecast computed by RTE, the french grid operator as shown in Fig.3.25. We stress here that our aim is two-fold: to link wind 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 reasonnable estimates (we obtain a correlation coefficient between our forecast and the observation of 0.85 where the grid operator obtains 0.97).

**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

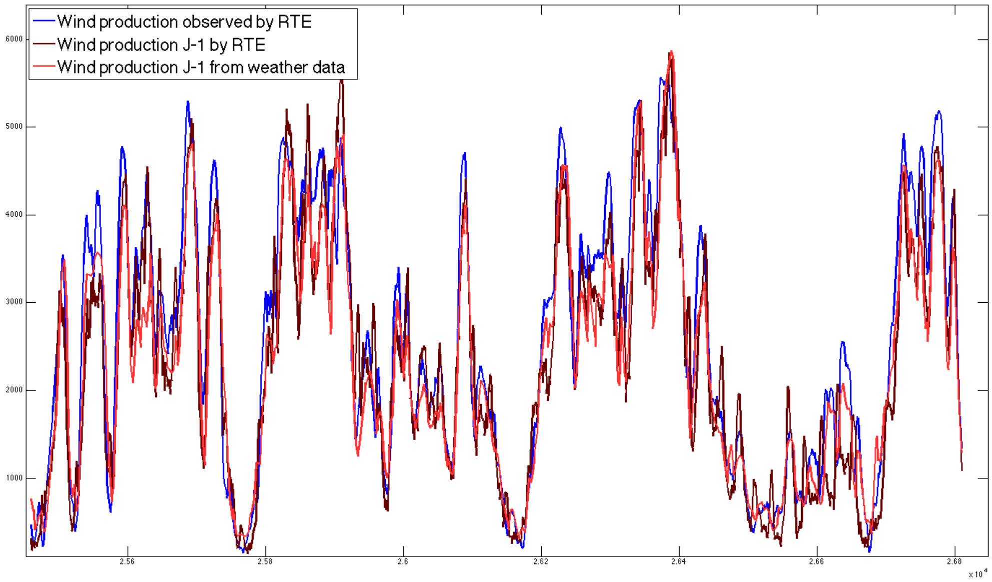



Figure 3.25: Wind production (observed and 2x predicted)
*Note:* All curves are hourly production data. The origin of the hours is the first of January 2011, and the production is in MWh. In blue: the observed wind production. In dark red: the day-ahead predictions from the grid operator. In light red: the day-ahead predictions from weather data.

electric consumption that can be attributed to lighting represents roughly 50 TWh where solar production is roughly 4 TWh[67].

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 $km$. France is formed of a little bit more than 36 000 of those communes.

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. 3.26.

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

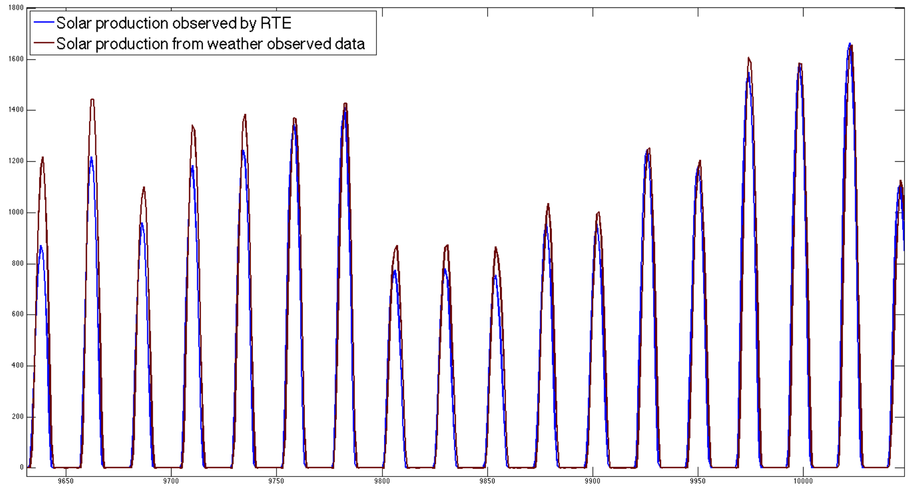



Figure 3.26: Solar production (observed and reconstructed)
*Note:* Hourly solar production in MWh (y-axis). The time origin is the first January 2011. In blue: observed production by RTE. In dark red: reconstructed production from observed weather data.

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 capture the net effect of both channels.

**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°$C$ below 15°$C$, as shown in Fig.3.27 sourced from RTE (2014), the French grid operator.

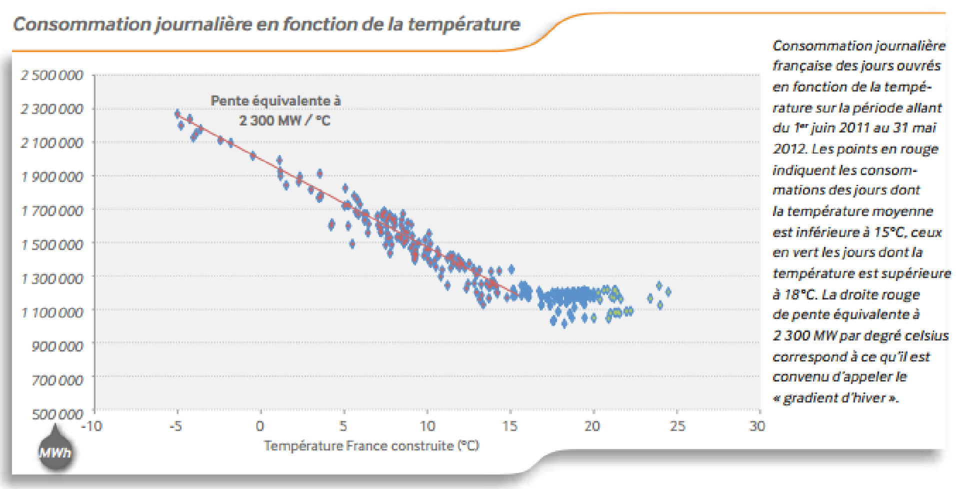



Figure 3.27: Daily electricity consumption in France as a function of the temperature

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*. 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 allows us to account at the country-level for the local impact of temperature on the electricity consumption.

### 3.9.5 Other controls

**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.

**Tempeff** We also build an effective temperature that does not account for the nonlinearity at 15°C following the same methodolgy as for Tempeff15 as a control.

**Roll_avgT**$H$ Variable capturing seasonal trends by using the rolling average temperature on temperature Tempeff (no kink) over the last $H$ hours, i.e. the last $H/24$ days.

**suncycle** Variable capturing intraday seasonality be measuring the intensity of sunlight as a percentage of the maximum daily observation. Midday is defined at the maximum sun intensity every day, i.e. Midday = max(Solar).
Thus, $\text{suncycle}_H = \text{Solar}_H$ / Midday.

**deltasun** Variable computed to proxy for dusk and dawn. It is computed as the absolute difference between $\text{suncycle}_H$ - $\text{suncycle}_{H-1}$.

**SolarRest** Solar represents estimates of solar production. Therefore, it is highly collinear with 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 3.12 gives the results of the regression.

| | 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 3.12: 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.

$$\text{PrevConsoH} = a + bX + u \tag{3.14}$$

In table 3.13 we give the output of regression 3.14 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%.

The signs and interpretation of the coefficients are exactly in line with the results of the demand estimation (in both the price and quantity dimension) for the point of inflection $k = 3$.

| | (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 3.13: "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.

Furthermore, 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).

**Gas and IT2** Gas turbines generate electricity using natural gas as a fuel. We thus proxy for its input price using a Gas variable for which we take the closing price for natural gas at 1 month on the London market (NBP). Electricity generation from gas is expensive and flexible. In general gas plants are only called upon to provide peak load electricity generation in moments of high demand. WE therefore compute an interaction term between Gas and an index for the hourly level of the demand. The index acts as a weight on the gas price. The weight is computed as the percentage demand level as compared to the maximum demand level observed in our dataset.

## 3.10 Appendix : Computational details and descriptives

### 3.10.1 Hard choices in the kernel-based PLU computation

In computing the multi-variate kernel based prediction of the uncertainty for a given auction, we select auctions of a sufficient degree of similarity. We base the forecast equation 3.5 on this subsample dataset. We thereby consider that firms use the forecasting equation only *locally* in the neighbourhood of the auction of interest.

In order to define the size of the neighbourhood of an auction, we have to explicitly specify the width of the kernel window used in selecting the respective subsamples. The trade-off involved is that we want to have small kernels for a precise computation of the PLU, while we want large kernels to make sure that we have a sufficient sample size in each kernel in order to derive meaningful statistics.

We choose to use a constant kernel window length with respect to each conditioning variable. We set the length of the window for each variable equal to $\frac{1}{3}$ of the variation of that variable. E.g. for Tempeff15, we observe a range of values from $-10^{o}$C to $14^{o}$C. The subsample used to compute the PLU$^{D}$ corresponding to a specific observation will consist of all observations that are within a range of $\pm 4^{o}$C of that observation for Tempeff15. The same logic is applied to selecting the neighbourhood with respect to all other conditioning variables. Table 3.14 gives descriptive statistics about the conditioning variables for the kernel and the explicit choice $m$, which determines the length of the kernel window for a variable $X_e$ using the formula $b_{X_e} = \frac{2}{m_{X_e}}$.

| $X_e$ | $m$ | Mean | Median | Std. dev. | Min | Max |
|---|---|---|---|---|---|---|
| Tempeff15 | 6 | 7.7 | 8 | 5 | -10 | 14 |
| Roll_Temp24 | 6 | 7.7 | 9 | 4 | -8 | 14 |
| Roll_Temp240 | 1 | 7.6 | 8 | 4 | -7 | 13 |
| suncycle | 6 | 0.3 | 0 | 0 | 0 | 1 |
| morning | 6 | 0.5 | 1 | 0 | 0 | 1 |
| deltasun | 6 | 0.1 | 0.1 | 0 | 0 | 0.4 |
| EWH | 6 | 0.3 | 0 | 0 | 0 | 1 |
| SolarRest | 6 | 5.4 | -1 | 364 | -1,337 | 2,241 |
| RteBlackBox | 6 | -0.0 | 37 | 4,755 | -16,966 | 18,209 |

Table 3.14: Variables used in the kernel based PLU$^{D}$ computation

*Note*: For the PLUv51, we have excluded the variable Roll_Temp240 from the conditioning in order to increase the size of each subsample used for the calculation of the observation specific PLU$^{D}$. Version 52 also conditions on the variable Roll_Temp240 using $m = 6$.

## 3.10.2 Descriptive Statistics

### 3.10.2.1 On realised market equilibria

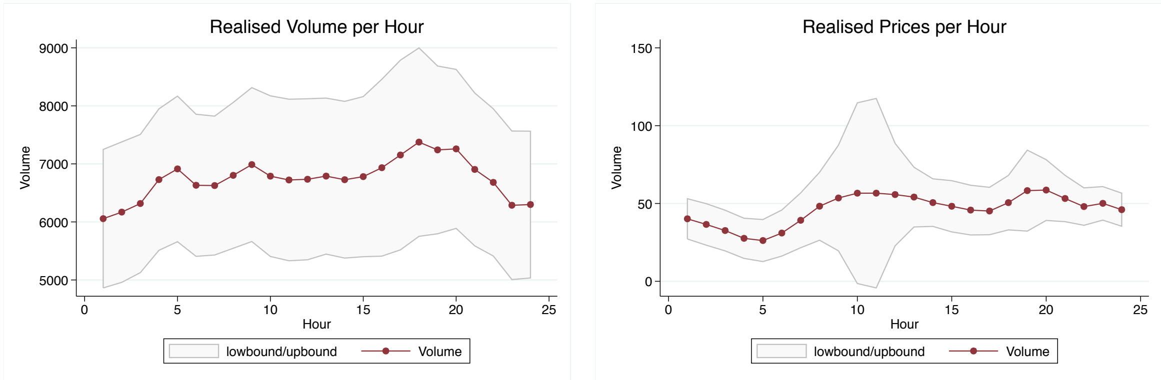



Figure 3.28: Realised volumes and prices per hour

*Note:* Plotted average realised Volume (left) and Price (right) per Hour with 95% confidence intervals.

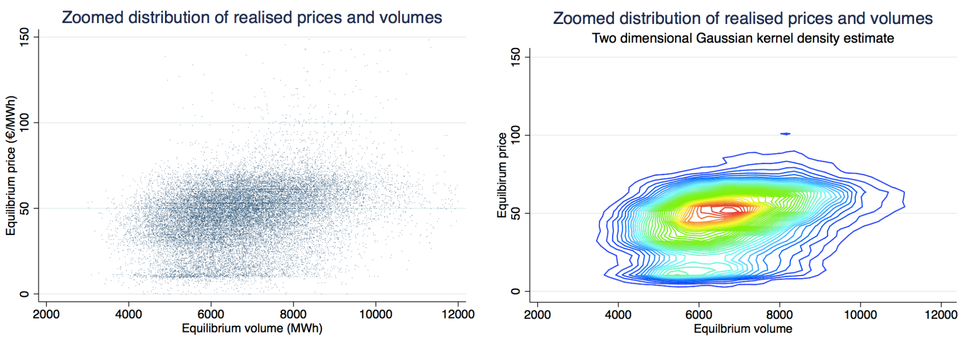



Figure 3.29: Distribution of observed market equilibria

*Note:* The warmer the colours of the heat map, the higher the frequency of realised price-quantity schedules. The colour legend is omitted for brevity, density changes between contours are of the order of $10^{-4}$.

### 3.10.2.2 On player bid functions

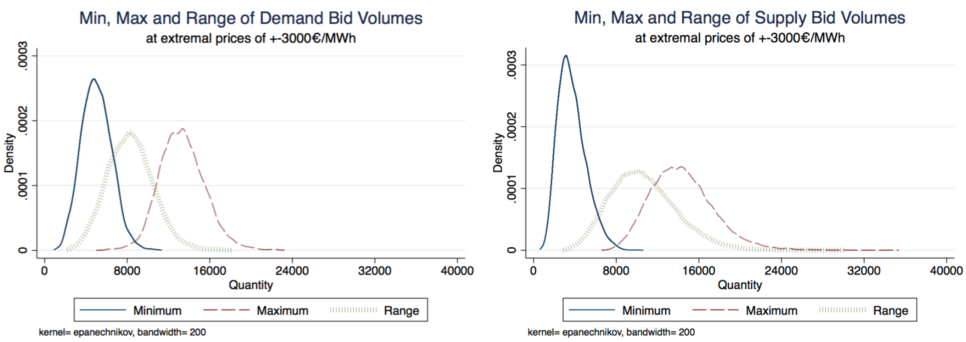



Figure 3.30: Min, max and range of production volume

*Note:* Distribution of minimum and maximum production volumes (and corresponding range) bid in an hourly auction.

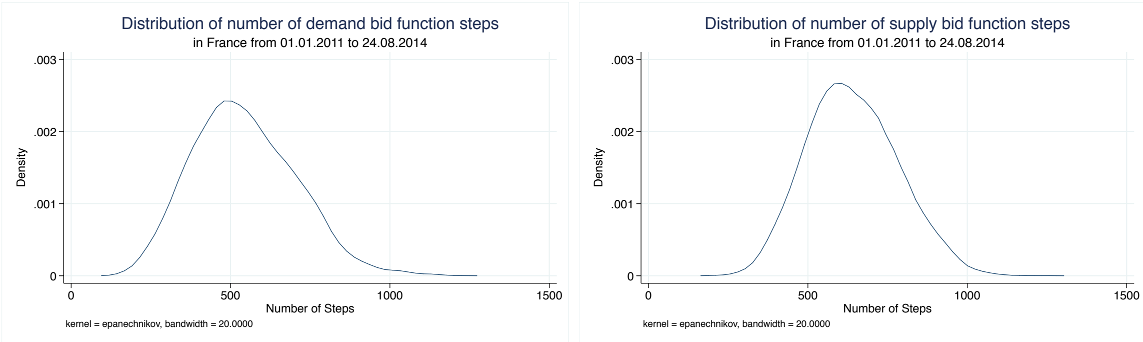



Figure 3.31: Distribution of number of bid function steps

#### 3.10.2.3 On exogenous factors

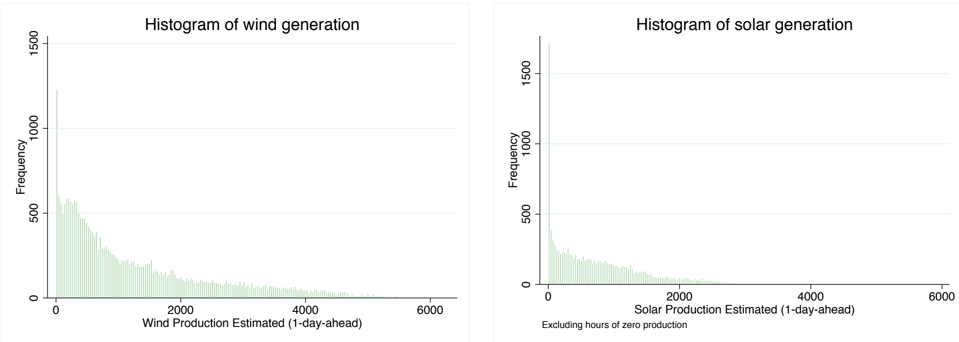



Figure 3.32: Histogram of predicted wind (left) and predicted solar (right) generation

# Bibliography


Allison, John R. and Mark A. Lemley (1998), "Empirical Evidence on the Validity of Litigated Patents." *American Intellectual Property Law Association Quarterly Journal.*

Anderson, Edward J and Huifu Xu (2005), "Supply Function Equilibrium in Electricity Spot Markets With Contracts and Price Caps." *Journal of Optimization Theory and Applications*, 124, 257–283.

Arrow, Kenneth (1962), "Economic Welfare and the Allocation of Resources for Invention." In *The rate and direction of inventive activity: Economic and social factors*, 609–626, Nber.

Atal, Vidya and Talia Bar (2010), "Prior Art: To Search or not to Search." *International Journal of Industrial Organization*, 28, 507–521.

Baye, Michael R and Heidrun C Hoppe (2003), "The Strategic Equivalence of Rent-Seeking, Innovation and Patent-Race Games." *Games and Economic Behavior*, 44, 217–226.

Belsunce, Henri de (2011), *Theory and Empirics of Multi-Unit, Uniform Price Auctions in the Domain of Energy Procurement in France*. Master's thesis, Paris School of Economics.

Berg, Sigbjorn Atle, Benzion Boukai, and Michael Landsberger (1999), "Bid Functions for Treasury Securities; Across Countries Comparison." Technical report, Citeseer.

Bergès, Alexis and David Martimort (2014), "Dynamics of the Electricity Day-Ahead Market : Supply Function Equilibria and Ramping Costs." *Working Paper.*

Bertoldi, Paolo and Bogdan Atanasiu (2007), "Electricity Consumption and Efficiency Trends in the Enlarged European Union, Status Report." Technical report, Institute for Environment and Sustainability, European Commission.

Bessen, James (2003), "Patent Thickets: Strategic Patenting of Complex Technologies." *Available at SSRN 327760.*

Bessen, James (2004), "Hold-up and Licensing of Cumulative Innovations with Private Information." *Economics Letters*, 82, 321–326.

Bolle, Friedel (1992), "Supply Function Equilibria and the Danger of Tacit Collusion: The Case of Spot Markets for Electricity." *Energy economics*, 14, 94–102.

Borenstein, Severin, James B. Bushnell, and Frank A. Wolak (2002), "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market ." *American Economic Review*, 92, 1376–1405.

Boukai, Benzion and Michael Landsberger (1998), "Market Bid Functions for Treasury Securities as Logistic Growth Curves." Technical report, Working Paper, Indiana University–Purdue University, Indianapolis, Indiana.

Caillaud, Bernard and Anne Duchêne (2011), "Patent Office in Innovation Policy: Nobody's Perfect." *International Journal of Industrial Organization*, 29, 242–252.

Calabresi, Guido and A Douglas Melamed (1972), "Property Rules, Liability Rules, and Inalienability: One View of the Cathedral." *Harvard Law Review*, 1089–1128.

Chiang, Tun-Jen (2012), "Forcing Patent Claims." *Michigan Law Review (Forthcoming)*, 113.

Choi, Jay Pil (1998), "Patent Litigation as an Information-Transmission Mechanism." *American Economic Review :*, 1249–1263.

Cohen, Wesley M, Richard R Nelson, and John P Wlash (2000), "Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not)." Technical report, National Bureau of Economic Research.

Commission of Experts for Research & Innovation, (EFI) (2015), "Research, innovation and technological performance in Germany - EFI Report 2015." Technical report, EFI.

Cotter, Thomas F (2009), "Patent Holdup, Patent Remedies and Antitrust Responses." *Journal of Corporation law*, 34, 08–39.

Cotter, Thomas F (2013), "The Comparative Law and Economics of Standard-Essential Patents and FRAND Royalties." *Texas Intellectual Property Law Journal, Forthcoming*, 13–40.

Crampes, Claude and Corinne Langinier (2002), "Litigation and Settlement in Patent Infringement Cases." *RAND Journal of Economics*, 258–274.

Department of Justice, DOJ (1997), "Business Review Letter of the MPEG-2 patent pool." *Business Review Letter, Antitrust Division.*

EPEX (2014), "Website of the European Power Exchange." URL `https://www.epexspot.com/en/`.

Farrell, Joseph, John Hayes, Carl Shapiro, and Theresa Sullivan (2007), "Standard Setting, Patents and Hold-up." *Antitrust Law Journal*, 603–670.

FTC, Federal Trade Commission (2003), "To Promote Innovation: The Proper Balance of Competition and Patent Law and Policy." Technical report.

Galasso, Alberto and Mark Schankerman (2010), "Patent Thickets, Courts, and the Market for Innovation." *The RAND Journal of Economics*, 41, 472–503.

Gallini, Nancy T. (2002), "The Economics of Patents: Lessons from Recent US Patent Reform." *The Journal of Economic Perspectives*, 16.2, 131–154.

Ganglmair, Bernhard, Luke M Froeb, and Gregory J Werden (2012), "Patent Hold-Up and Antitrust: How A Well-Intentioned Rule Could Retard Innovation." *The Journal of Industrial Economics*, 60, 249–273.

Geradin, Damien (2009), "Pricing Abuses by Essential Patent Holders in a Standard-Setting Context: A View from Europe." *Antitrust Law Journal*, 329–357.

Gilbert, Richard J and David MG Newbery (1982), "Preemptive Patenting and the Persistence of Monopoly." *The American Economic Review*, 514–526.

Glader, Marcus (2006), *Innovation Markets and Competition Analysis: EU Competition Law and US Antitrust Law*. Edward Elgar Publishing.

Goldstein, Larry M and Brian N Kearsey (2004), "Technology Patent Licensing: An International Reference on 21st Century Patent Licensing." *Patent Pools and Patent Platforms, Aspatore Inc*, 47.

Graevenitz, Georg von, Stefan Wagner, and Dietmar Harhoff (2011), "How to Measure Patent Thickets – A Novel Approach." *Economics Letters*, 111, 6–9.

Green, Richard J and David M Newbery (1992), "Competition in the British Electricity Spot Market." *Journal of political economy*, 929–953.

Hobbs, Benjamin F (2001), *The Next Generation of Electric Power Unit Commitment Models*, volume 36. Springer.

Hortacsu, Ali and Steven L Puller (2008), "Understanding Strategic Bidding in Multi-Unit Auctions: A Case Study of the Texas Electricity Spot Market." *The RAND Journal of Economics*, 39, 86–114.

Jaffe, Adam (2000), "The U.S. Patent System in Transition: Policy Innovation and the Innovation process." *Research Policy*.

Kamien, Morton I, Shmuel S Oren, and Yair Tauman (1992), "Optimal Licensing of Cost-Reducing Innovation." *Journal of Mathematical Economics*, 21, 483–508.

Kamien, Morton I and Yair Tauman (1986), "Fees versus Royalties and the Private Value of a Patent." *The Quarterly Journal of Economics*, 471–491.

Kastl, Jakub (2011), "Discrete Bids and Empirical Inference in Divisible Good Auctions." *The Review of Economic Studies*, 78, 974–1014.

Kim, Sung-Hwan (2004), "Vertical Structure and Patent Pools." *Review of Industrial Organization*, 25, 231–250.

Kitch, Edmund W (1977), "The Nature and Function of the Patent System." *Journal of law and economics*, 265–290.

Klemperer, Paul D and Margaret A Meyer (1989), "Supply Function Equilibria in Oligopoly under Uncertainty." *Econometrica: Journal of the Econometric Society*, 1243–1277.

Kortum, Samuel and Josh Lerner (1999), "What is Behind the Recent Surge in Patenting?" *Research policy*, 28, 1–22.

LaCommare, Kristina Hamachi and Joseph H Eto (2004), "Understanding the Cost of Power Interruptions to US Electricity Consumers." *Lawrence Berkeley National Laboratory*.

Langinier, Corinne and Philippe Marcoul (2007), "Patents, Search of Prior Art and Revelation of Information." *Working Paper*.

Layne-Farrar, Anne and Josh Lerner (2011), "To Join or not to Join: Examining Patent Pool Participation and Rent Sharing Rules." *International Journal of Industrial Organization*, 29, 294–303.

Layne-Farrar, Anne, A Jorge Padilla, and Richard Schmalensee (2007), "Pricing Patents for Licensing in Standard-Setting Organizations: Making Sense of FRAND Commitments." *Antitrust Law Journal*, 671–706.

Lemley, Mark A (2008), "Ignoring Patents." *Michigan State Law Review*, 19.

Lemley, Mark A and Ragesh K Tangri (2003), "Ending Patent Law's Willfulness Game." *Berkeley Technology Law Journal*, 18, 1085.

Lerner, Josh (2009), "The Empirical Impact of Intellectual Property Rights on Innovation: Puzzles and Clues." *The American Economic Review: Papers and Proceedings*, 99, 343–348.

Lerner, Josh and Jean Tirole (2004), "Efficient Patent Pools." *American Economic Review*, 94, 691–711.

Lerner, Josh and Jean Tirole (2013), "Standard-Essential Patents." Technical report, National Bureau of Economic Research.

Llobet, Gerard (2003), "Patent Litigation when Innovation is Cumulative." *International Journal of Industrial Organization*, 21, 1135–1157.

Means, Samuel Chase (2013), "The Trouble with Treble Damages: Ditching Patent Law's Willful Infringement Doctrine and Enhanced Damages." *University of Illinois Law Review*, 1999–2045.

Meibom, Peter, Christoph Weber, Rudiger Barth, and Heike Brand (2009), "Operational Costs induced by Fluctuating Wind Power Production in Germany and Scandinavia." *Renewable Power Generation, IET*, 3, 75–83.

Menell, Peter S and Suzanne Scotchmer (2007), "Chapter 19 Intellectual Property." *Handbook of Law and Economics, A. Mitchell Polinsky and Steven Shavell, Elsevier.*

Ménière, Yann and Sarah Parlane (2008), "Innovation in the Shadow of Patent Litigation." *Review of Industrial Organization*, 32.2, 95–111.

Merges, Robert P (1999a), "As many as Six Impossible Patent before Breakfast: Property Rights for Business Concepts and Patent System Reform." *Berkeley Technology Law Journal*, 14, 577.

Merges, Robert P (1999b), "Institutions for Intellectual Property Transactions: The Case of Patent Pools." *University of California at Berkeley Working Paper.*

Meurer, M. J. (1989), "The Settlement of Patent Litigation." *The RAND Journal of Economics*, 77–91.

Moore, Kimberly A (2004), "Empirical Statistics on Willful Patent Infringement." *Federal Circuit Bar Journal*, 14, 227.

Moore, Kimberly A. (2005), "Worthless Patents." Technical report, George Mason University School of Law Working Papers Series : 27.

MPEG-2 PPL (2009), "MPEG-2 Patent Portfolio License." On file with the author. For a summary see *http://www.mpegla.com/main/programs/M2/Pages/Agreement.aspx.*

Newbery, David M (1998), "Competition, Contracts and Entry in the Electricity Spot Market." *The RAND Journal of Economics*, 726–749.

Özcan, Rasim (2004), "The Logistic Function Approach to Discriminatory and Uniform Price Treasury Auctions." *Working Paper.*

Préget, Raphaële and Patrick Waelbroeck (2005), "Treasury Bill Auction Procedures: Empirical Perspectives from French Market Bid Functions." *Journal of International Money and Finance*, 24, 1054–1072.

Qian, Yi (2007), "Do National Patent Laws Stimulate Domestic Innovation in a Global Patenting Environment? A Cross-Country Analyis of Pharmaceutical Patent Protection, 1978-2002." *The Review of Economics and Statistics*, 89, 436–453.

Rapp, Richard T and Lauren J Stiroh (2002), "Standard setting and market power." *Washington DC: Joint Hearings of the United States Department of Justice and the Federal Trade Commission.*

Regibeau, Pierre and Katharine Rockett (2011), "Assessment of Potential Anticompetitive Conduct in the Field of Intellectual Property Rights and assessment of the Interplay between Competition Policy and IPR Protection." Technical report.

Reguant, Mar (2014), "Complementary Bidding Mechansims and Startup Costs in Electricity Markets." *The Review of Economic Studies*, 81, 1708–1742.

Reichl, Johannes, Michael Schmidthaler, and Friedrich Schneider (2013), "Power Outage Cost Evaluation: Reasoning, Methods and an Application." *Journal of Scientific Research & Reports*, 2, 249–276.

REN21, Steering Committee (2013), "Renewables 2013, Global Status Report." Technical report.

RTE (2014), "Website of RTE, the French grid operator." URL `http://www.rte-france.com/en/`.

Salgado-Ugarte, Isaias Hazarmabeth, Makoto Shimizu, and Toru Taniuchi (1994), "Exploring the Shape of Univariate Data Using Kernel Density Estimators." *Stata Technical Bulletin*, 3.

Salop, Steven C. and David T. Scheffman (1983), "Raising Rivals' Costs." *The American Economic Review*, 73, pp. 267–271.

Schmidt, Claudia (2010), *Refusal to License Intellectual Property Rights as Abuse of Dominance*. Peter Lang.

Schmidt, Klaus M (2014), "Complementary Patents and Market Structure." *Journal of Economics & Management Strategy*, 23, 68–88.

Seaman, Christopher B (2011), "Willful Patent Infringement and Enhanced Damages After in re Seagate: An Empirical Study." *Iowa Law Review*, 97, 417.

Sewalt, M and C De Jong (2003), "Negative Prices in Electricity Markets." *Commodities Now*, 2, 74–79.

Shapiro, Carl (2001), "Navigating the Patent Thicket: Cross Licenses, Patent Pools and Standard Setting." In *Innovation Policy and the Economy, Volume 1*, 119–150, MIT press.

Shapiro, Carl (2010), "Injunctions, Hold-Up and Patent Royalties." *American Law and Economics Review*, 280–318.

Silverman, Bernard W (1986), *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press.

Silverman, BW and JO Ramsay (2005), *Functional Data Analysis*. Springer.

Swanson, Daniel G and William J Baumol (2005), "Reasonable and Nondiscriminatory (RAND) Royalties, Standards Selection, and Control of Market Power." *Antitrust Law Journal*, 1–58.

Takenaka, Toshiko (2000), "Patent Infringement Damages in Japan and the United States: Will Increased Patent Infringement Damage Awards Revive the Japanese Economy." *Washington University Journal of Law and Policy*, 2, 309.

Torrance, Andrew W and Bill Tomlinson (2011), "Property Rules, Liability Rules, and Patents: One Experimental View of the Cathedral." *Yale Journal of Law and Technology*, 14, 138.

Vakili, Keyvan (2012), "Competitive Effects of Collaborative Arrangements: Evidence from the Effect of the MPEG-2 Pool on Outsiders' Innovative Performance." *DRUID conference submission*. Draft version 30.09.2012.

Wang, X Henry (1998), "Fee versus Royalty Licensing in a Cournot Duopoly Model." *Economics letters*, 60, 55–62.

Weniger, Johannes, Joseph Bergner, Tjarko Tjaden, and Volker Quaschning (2014), "Einfluss der Sonnenfinsternis im März 2015 auf die Solarstromerzeugung in Deutschland." Technical report, HTW Berlin.

White, Halbert (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica: Journal of the Econometric Society*, 817–838.

Wolak, Frank A (2007), "Quantifying the Supply-Side Benefits from Forward Contracting in Wholesale Electricity Markets." *Journal of Applied Econometrics*, 22, 1179–1209.

Wölfing, Nikolas (2013), *Interacting Markets in Electricity Wholesale: Forward and Spot, and the Impact of Emissions Trading.* Ph.D. thesis, Universit Paris 1 Panthon-Sorbonne & Paris School of Economics.

Wolfram, Catherine D (1998), "Strategic Bidding in a Multiunit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales." *The Rand Journal of Economics*, 703–725.

Wolfram, Catherine D (1999), "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review*, 805–826.

# Curriculum Vitae

| | |
|---|---|
| 2011 - 2015 | Promotionsprogramm Volkswirtschaftslehre<br>International Max Planck Research School for Competition and Innovation (Teil der Munich Graduate School of Economics)<br>Max Planck Institut für Innovation und Wettbewerb & Ludwig-Maximilians-Universität, München |
| 2015 | Forschungsaufenthalt<br>Institute of Intellectual Property of the Japan Patent Office, Tokyo, Japan |
| 2013 | Forschungsaufenthalt<br>GD Wettbewerb, Europäische Kommission, Brüssel, Belgien |
| 2013 | Forschungsaufenthalt<br>Centre for Research in Economics and Statistics (CREST), Paris, Frankreich |
| 2010 - 2011 | Master of Analysis and Policy in Economics (M2)<br>Paris School of Economics, Frankreich |
| 2009 - 2010 | Master of Quantitative Economics and Finance (M1)<br>Ecole Polytechnique, Paris, Frankreich |
| 2006 - 2009 | Bachelor of Science of Management Sciences<br>The London School of Economics and Political Science, England |
| 2005 | Irish Leaving Certificate<br>The King's Hospital School, Dublin, Irland |
| 1988 | Geboren in Frankfurt |