Université de Montréal

Essays in Econometrics and Energy Markets

par
David Benatia

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Mai, 2018

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :
Essays in Econometrics and Energy Markets

présentée par :
David Benatia

a été évaluée par un jury composé des personnes suivantes :

Raphaël Godefroy,	président-rapporteur
Marine Carrasco,	directrice de recherche
Étienne Billette de Villemeur,	co-directeur de recherche
Pierre-Olivier Pineau,	membre du jury
Steven L. Puller,	examinateur externe
François Perron,	vice-doyen, représentant du doyen de la FAS

Thèse acceptée le 28 mai 2018

à Charlotte, ma sœur Lucille, et mes parents Carole et Marc

Remerciements

Acknowledgments

Je remercie Marine Carrasco, ma directrice de recherche, pour sa confiance, son soutien, sa gentillesse, ses conseils, sa disponibilité, sa rigueur, et pour tout le temps passé à lire et commenter mon travail. Je suis très heureux d'avoir frappé à sa porte, et d'avoir bénéficié de sa supervision sur des thèmes de recherche qui m'ont passionné durant ces dernières années.

Je remercie aussi Étienne Billette de Villemeur, mon co-directeur de recherche, pour sa bienveillance, ses conseils et nos discussions lors de ses visites à Montréal. Je souhaiterais aussi exprimer ma profonde gratitude à Pierre-Olivier Pineau pour son enthousiasme et sa générosité, à Jean-Pierre Florens pour son soutien, son aide précieuse et sa confiance, et à Raphael Godefroy pour ses conseils et sa disponibilité.

La détermination nécessaire pour achever ce travail, je la dois en grande partie à celles et ceux qui m'ont soutenu et aux rencontres enrichissantes que j'ai faites pendant ces années d'études. Merci d'abord à Nick Johnstone, Stefan Ambec et Lukas de m'avoir conseillé et orienté vers Montréal, où j'y ai passé de très belles années. Je voudrais remercier le groupe d'économie environnementale de Montréal, notamment Hassan Benchekroun, Sophie Bernard, Robert Cairns, Gérard Gaudet, Pierre Lasserre, Justin Leroux, Martino Pelli, et Charles Séguin. Merci aussi à Derek Olmstead et Mathieu Marcoux pour leurs commentaires précieux. Je voudrais remercier Tuba Tunçel, pour son aide indispensable durant cette année de "job market", notamment lors d'un voyage épique jusqu'à Philadelphie à travers la tempête. Je dois aussi exprimer ma reconnaissance à mes collègues et amis: Somouaouga, Jonathan, Samuel, Gabriela, Sébastien, Kokouvi, Ada, Hervé, Louphou, Mohamed, Alex, Dominique et les autres étudiants du CIREQ, sans eux je n'aurais pu tenir la distance. Tout ceci aurait été tellement plus difficile sans mes amis de toujours: Adrian, Benjamin, Fabien, Martine et toute la bande; et ma famille: Maria, Maguy, Claudine, Nathalie, Véronique, et tous les autres que je ne cite pas. Je voudrais aussi remercier mes parents Carole et Marc, et ma soeur Lucille d'avoir toujours cru en moi et de m'avoir soutenu quand j'en avais besoin. Par dessus tout, je serais éternellement reconnaissant envers ma formidable partenaire de vie, Charlotte, venue vivre cette aventure à Montréal avec moi, pour son dynamisme contagieux, sa patience à rude épreuve, et tout le reste.

Pour finir, j'ai été très heureux d'étudier à l'Université de Montréal. Je voudrais exprimer ma gratitude à tous les professeurs et le personnel administratif du département de sciences économiques et du CIREQ. Je n'aurais pu faire de si longues études sans le soutien financier des bourses du département et du CIREQ, les contrats d'enseignement, les contrats d'assistant recherche, le fonds Bernard Landry, la générosité de la Chaire en gestion du secteur de l'énergie de HEC Montréal, l'Association Canadienne d'Économique de l'Énergie, mes parents, et Charlotte.

Je voudrais donc remercier toutes celles et tous ceux qui ont permis, de près ou de loin, de rendre ces années agréables et à ce travail d'aboutir.

Résumé

Cette thèse est organisée en trois chapitres où sont développées des méthodes d'analyse économique et économétrique des marchés de l'énergie.

Le Chapitre 1 propose une étude des incitations à la manipulation de marché générées par des opportunités de revenir sur ses engagements. Un modèle théorique est développé pour analyser le comportement d'un monopole face à une frange compétitive en présence d'une demande incertaine, de contraintes de capacité, et de possibilités de trahir ses engagements. Les entreprises se concurrencent avec des fonctions d'offre étant donnés leurs engagements. Le monopole revient sur ses engagements lorsqu'il retire sa production engagée en observant la réalisation de l'incertitude. Il peut ainsi exacerber son pouvoir de marché, réduire l'incertitude autour de la demande, et accroitre sa probabilité de devenir un offreur pivot. À l'équilibre, les stratégies d'offre dépendent du volume de production engagée et du coût d'opportunité de la retirer stratégiquement. En particulier, le monopole peut trouver profitable d'offrir sa production à des prix plus élevés lorsqu'il sait qu'il pourra revenir sur ses engagements si la demande est élevée. Finalement, cette stratégie est présentée comme un comportement de manipulation par perte, et des applications aux marchés de l'électricité sont discutées.

Dans le Chapitre 2, nous développons de nouveaux résultats pour les régressions fonctionnelles où le prédicteur $Z(t)$ et la réponse $Y(t)$ sont des fonctions d'espaces de Hilbert, indexés par le temps ou l'espace. Le modèle peut être compris comme une généralisation de la régression multivariée où le coefficient de régression est maintenant un opérateur inconnu Π. Nous proposons d'estimer l'opérateur Π par régularisation de Tikhonov, ce qui revient à appliquer une pénalité sur sa norme L^2. Nous dérivons le taux de convergence de l'erreur quadratique moyenne, la distribution asymptotique de l'estimateur, et développons des tests sur Π. Comme les trajectoires ne sont généralement pas complètement observables, nous considérons une sit-

uation où les données deviennent de plus en plus fréquentes (asymptotique de remplissage). Nous traitons aussi le cas où Z est endogène et des variables instrumentales sont utilisées afin d'estimer Π. Une application à la consommation d'électricité complète l'article.

Le Chapitre 3 propose une nouvelle approche pour l'analyse empirique des enchères à unités multiples, dans lesquelles les participants choisissent des fonctions d'offre ou de demande. Cette approche permet d'évaluer le pouvoir de marché des entreprises dans une cadre d'information privée, en évitant d'avoir à modéliser le mécanisme du marché. Elle repose sur des méthodes économétriques qui traitent les fonctions de mise comme des éléments aléatoires à valeurs fonctionnelles. Notamment, un estimateur fonctionnelle à variable instrumentale est développé. La méthode est appliquée au marché de l'électricité de l'état de New York sur des données micro-économiques de mises et de coûts à l'échelle des entreprises pour 2013-2015. J'estime le pouvoir de marché unilatéral des entreprises et compare les comportements observés aux comportements maximisant les profits sous information privée. Je trouve un faisceau d'indices sérieux de comportement optimal, qui suggère que les entreprises sont au courant de leur pouvoir de marché et se comportent en conséquence.

Mots-clés : Régression fonctionnelle, pouvoir de marché, marchés de l'électricité, problème contractuel, engagement, manipulation de marché, enchères empiriques, variables instrumentales, opérateur linéaire, régularisation de Tikhonov.

Abstract

This thesis is organized in three chapters which develop economic and econometric methods for the analysis of energy markets.

In Chapter 1, we study the incentives for market manipulations created by opportunities to renege on prior commitments. We develop a theoretical framework to analyze the behavior of a monopolist facing a competitive fringe in the presence of demand uncertainty, capacity constraints and reneging opportunities. The firms are assumed to compete in supply functions taking their commitments as sunk decisions. Reneging occurs when the monopolist withdraws its committed output upon observing the realization of demand. By doing so, it can exacerbate its market power, alleviate demand uncertainty, and be more likely to be pivotal. At equilibrium, supply strategies depend on the volume of committed output and the opportunity cost of reneging. In particular, the monopolist may find profitable to offer some of its market output at higher prices in the presence of reneging opportunities. Finally, we present strategic reneging as a loss-based manipulative conduct in a general framework and discuss applications to electricity markets.

In Chapter 2, we develop new estimation results for functional regressions where both the regressor $Z(t)$ and the response $Y(t)$ are functions of Hilbert spaces, indexed by the time or a spatial location. The model can be thought as a generalization of the multivariate regression where the regression coefficient is now an unknown operator Π. We propose to estimate the operator Π by Tikhonov regularization, which amounts to apply a penalty on the L^2 norm of Π. We derive the rate of convergence of the mean-square error, the asymptotic distribution of the estimator, and develop tests on Π. As trajectories are often not fully observed, we consider the scenario where the data become more and more frequent (infill asymptotics). We also address the case where Z is endogenous and instrumental variables are used to estimate Π. An application to the electricity consumption completes the paper.

Chapter 3 proposes a novel approach for the empirical analysis of multi-unit auctions, to which participants submit supply or demand functions observable by the researcher. The approach allows for the evaluation of firm-level market power in a private information setting, and avoids having to model the market mechanism. It relies on econometric methods that treat the observed bid functions as function-valued random elements. Notably, a functional instrumental variable estimator is developed. The method is applied to the New York electricity market using rich data on firm-level bids and marginal costs for 2013-2015. In this market, daily bids are disclosed three months later in order to limit strategic behaviors. I estimate firm-level market power and compare actual bidding behavior to profit-maximizing behavior under private information. I find consistent evidence of optimal bidding, suggesting that firms are well aware of their own market power and behave accordingly. Therefore, the late disclosure of bids is not sufficient to preclude firms from acting strategically, most likely due to the repeated nature of those auctions.
Keywords : Functional regression, market power, electricity markets, reneging, commitment, market manipulation, empirical auctions, instrumental variables, linear operator, Tikhonov regularization.

Table des matières

Table of contents

Liste des figures

List of figures

Liste des tableaux

List of tables

Introduction

"The free market may be humans' most powerful tool. But, like all very powerful tools, it is also a two-edged sword. [...] That means that we need protection against the problems [...] only a real fool would pretend that there are no disadvantages, or take no precautions against them."

Phishing for phools,
Akerlof and Shiller (2015)

Competition and free markets have long been praised as virtuous by economists. Nevertheless, the virtues of free markets rest on ideal conditions which are hardly ever met. In his Nobel Prize lecture, Tirole (2014) observes that "[a]las, competition is rarely perfect, markets fail, and market power must be kept in check". Market power, the ability of a firm to raise the market price substantially above cost, is a key market failure which has attracted a lot of attention from economists in the past decades, especially in industrial organization.

Market power is not the only cause of a firm's ability to alter market prices and profit from it. Market manipulations, that is anti-competitive behaviors e.g. through outright fraud or uneconomic trading, provide additional channels to raise market prices significantly above competitive levels. In a well acclaimed book, Nobel Prize laureates Akerlof and Shiller (2015) state that "competitive markets by their very nature spawn deception and trickery, as a result of the same profit motive that give us our prosperity."

Part of the role of economists hence consists of providing tools to diagnose and address the issues related to market power exertion and misconducts.

Modern economics widely acknowledges the importance of sound regulation, market design and adequate economic policies for well-functioning markets. They give rise to significant challenges, not only with regards to the developing industries based on new technologies, but also for the ever-important ones experiencing critical transitions, such as the energy sector.

Energy is key to life and a fundamental driver of socio-economic development. Our contemporaneous societies have become enormous consumers of energy in its diverse forms, and markets consist of the main mechanisms used to organize the production, transportation and consumption of energy commodities in the economy. Well-functioning energy markets are thus a cornerstone of sustained economic activity and growth.

The large and increasing reliance of energy consumption on electricity generation makes electricity markets critical platforms in our economies. Electricity is perceived as a means to increasing energy supply security and mitigating polluting emissions in the entire economy, notably through the large deployment of renewable power and electric transport.

The electricity industry has experienced a major paradigm shift in the 1990's, and considerably evolved since then. Electricity sectors evolved from vertically integrated geographic monopolies, either publicly owned or regulated as natural monopolies because of the large increasing returns to scale and sunk costs. Many developed countries have introduced comprehensive reform programmes aimed at restructuring their electricity industry through liberalization and privatization. The primary goal of the introduction of competition in the generation and retail segments is to achieve improved economic efficiency, while ensuring that an adequate share of the realised long-term social benefits is passed on to consumers through lower energy prices. In general, it has been perceived as welfare-enhancing (Joskow, 2008b).

Yet, significant structural factors render electricity markets prone to the exercise of market power and anti-competitive behavior. The most prevalent factors are the lack of price elasticity of demand and cost-effective storage options, binding network constraints, production capacity constraints, a limited number of suppliers and their repeated interactions not only within a given market but also across regional markets (Benatia and Koźluk, 2016).

In a large electricity market, e.g. that of New York or New England, market participants buy and sell billions of dollars' worth of electricity and related contracts annually. The potential welfare impacts of market power in electricity markets are consequently sizeable. It is hence of utmost importance to understand the behaviors of participants in electricity markets so as to be able to control the exercise of market power and prevent anti-competitive conducts. It is obviously not a trivial matter.

Liberalized electricity markets consist of large sets of rules, sophisticated allocation mechanisms and extremely system-dependent operational constraints such as the transmission network. Electricity market mechanisms can typically be understood as procurement auctions held ahead of actual delivery time, to which generators submit supply bids so as to signal their willingness to produce. Given a specific set of market rules, the system operator – i.e. the auctioneer – solves for the cost-minimizing production schedule subject to operational constraints in order to satisfy the demand at every moment of the day, while ensuring the security and reliability of supply. This wholesale market is often the centerpiece of a complex market environment which interlinks various markets for forward contracts, available capacity, transmission contracts, and ancillary services for frequency and voltage regulation.

An important aspect of those auctions is the fact that the strategies are entire functions. Wilson (1979) developed the seminal model of multi-unit auctions to which bidders submit supply or demand functions to signal their willingness to sell or buy different quantities of an homogeneous good. His crucial result is that those auction mechanisms can be manipulated by the bidders. Contrary to second-price sealed bid auctions for single items, multi-unit auctions are not strategy-proof as agents do not reveal their true valuation in equilibrium. This result underlines the importance of keeping market power in check.

The literature focusing on electricity markets is generally based on the supply function equilibrium (SFE) concept of Klemperer and Meyer (1989). A key result of the standard SFE is the ex-post optimality of equilibrium supply function strategies in the presence of additive random shocks as the only source of uncertainty. Green and Newberry (1992) is the first article to adapt this model to electricity markets. The authors use the SFE framework to study the deregulation of the England and Wales electricity industry in the early 1990's. A large literature has built on those initial contributions.

Nevertheless, the theoretical literature on electricity markets (Schöne, 2009) is often disconnected from empirical research. The existing economic theory of firms' behaviors in multi-unit auctions, such as the SFE model, is often based on smooth supply schedules, which are functional variables. Yet, despite the large amount of available data on electricity auctions, existing empirical methods lack a proper framework to account for the functional nature of the observed supply schedules. Dealing with functional variables in an empirical model complicates the analysis as the researcher is forced to address challenging problems related to the curse of dimensionality (Bellman, 1961). Functional data analysis (Ramsay and Silverman, 2005) offers a natural framework for the statistical analysis of random functions. It provides a building-block to develop econometric tools for the empirical analysis of strategies in multi-unit auctions, and thereby an opportunity to reconcile empirical research with the theory.

This thesis is organized in three chapters and aims at developing methods for the economic and econometric analyses of firms' behaviors in restructured electricity markets. The main research objective is two-fold: 1) to provide an understanding of the effects of the market structure and sources of uncertainties on the strategies of electricity producers and market outcomes, and; 2) to develop econometric techniques to evaluate market power and potential misconducts.

Chapter 1 develops a micro-economic framework to study the incentives for market manipulations created by opportunities to renege on prior commitments in a sequential market under imperfect competition. The analysis focuses on a monopolist competing in supply functions against a competitive fringe, taking its commitments as sunk decisions. The model accounts for demand uncertainty, capacity constraints and the ability to strategically renege on prior commitments. Reneging occurs when the monopolist withdraws its commitments upon observing the realization of demand. By doing so, it can exacerbate its market power, alleviate demand uncertainty, and be more likely to become pivotal. Strategic reneging is presented as a loss-based manipulative conduct and applications related to electricity markets are discussed. This type of misconducts is illustrated using examples of market manipulations that occurred in Alberta in 2010-2011. It shows that the regulatory tools for market power mitigation can lead to undesirable incentives for anti-competitive conducts.

Chapter 2 develops new estimation results for functional regressions where both the regressor and response are functions of Hilbert spaces. The model can be thought as a generalization of the multivariate regression where the regression coefficient is now an unknown operator. This chapter proposes an estimator of this operator based on Tikhonov regularization, which amounts to penalize its L^2 norm. Theoretical results are derived for both estimation and inference. The case of an endogenous predictor function is also addressed. The methods are illustrated through a numerical simulation study, and an application to forecasting electricity consumption in Ontario using weather data from 2010-2014.

Chapter 3 proposes a novel approach for the empirical analysis of multi-unit auctions, to which participants submit supply and demand functions observable by the researcher. The approach allows for the evaluation of firm-level market power in a private information setting, and avoids having to model the market mechanism – which can be extremely challenging, especially in electricity markets with location-based prices. It relies on econometric methods that treat the observed bid functions as function-valued random elements. An instrumental variable functional estimator is proposed, and some theoretical results are derived for both estimation and inference. The method is applied to the New York electricity market using rich data on firm-level bids and marginal costs for 2013-2015. In this market, daily bids are disclosed three months later in order to limit strategic behaviors. I estimate firm-level market power and compare actual bidding behavior to profit-maximizing behavior under private information. I find consistent evidence of optimal bidding, suggesting that firms are well aware of their own market power and behave accordingly. Therefore, the late disclosure of bids – although useful to limit the strategic information available to firms to some extent – is not sufficient to prevent strategic behaviors, most likely due to the repeated nature of those auctions.

Chapter 1

Strategic reneging in sequential markets under imperfect competition*

"- A strong night last night as we priced up units, Sun 5 came down early [...]. With the unit coming offline and Poplar Creek pricing up, prices jumped to $400 over the peak hours. Our portfolio benefited from a ton of length."
- Great job this first week. Some great value and it's clear we're learning a ton."

Discussion between one of TransAlta's Asset Optimizer and the Chief Operating Officer on November 20, 2010, the day after the plant outage (MSA, 2015).

*This chapter is co-authored with my co-advisor Étienne Billette de Villemeur (EQUIPPE, Université de Lille), and Pierre-Olivier Pineau (HEC Montréal). The authors would like to thank Derek Olmstead, and all participants of the Montréal Environment and Resource Economics Workshop, SCSE 2016 in Québec, CIREQ Ph.D. Students 2016 in Montréal, CEA 2016 in Ottawa, and IAEE 2016 in Bergen, for their valuable comments. Benatia gratefully acknowledges financial support from the Canadian Energy Economics Association.

1.1 Introduction

Opportunities to renege on prior commitments can create incentives to manipulate markets. Those incentives are especially relevant to imperfect markets with capacity constraints, such as electricity markets, because market-clearing prices may at times be very sensitive to unexpected supply or demand shocks.

In this paper, we study the incentives for market manipulations created by reneging opportunities in imperfect markets. We develop a theoretical framework to analyze the behavior of a monopolist facing a competitive fringe in the presence of demand uncertainty, capacity constraints and opportunities to renege on prior commitments. Firms are assumed to compete in supply functions, taking their commitments as sunk decisions, to satisfy a random and perfectly inelastic demand. Strategic reneging occurs when the monopolist withdraws its committed output upon observing the realization of demand in order to put upward pressure on the market price. The firm's market portfolio hence benefits from an artificially inflated price at the cost of the foregone profits from reneging.

In virtually all competitive markets, some of the exchanges are settled through forward commitments. Those commitments are generally contracts specifying a given price for a given output volume to be delivered. This contract price is however not directly tied to the market price, and can be used to hedge against its volatility.[1] At the time when the market price is settled, the firm faces its own committed output as competition.

The market duality created by the co-existence of the two prices may give room for market manipulations. A market participant who not only competes in the spot market but also holds commitments has both direct and indirect effects on the market outcomes. While its market strategy has a direct effect through the market mechanism, its commitments affect the net market demand which in turn affects the market outcome. By reneging on its commitments, the firm can increase the net market demand mechanically since –at least some of– the withdrawn output must be replaced in equilibrium. Reneging hence shifts the residual demand faced by the monopolist in the market. Such conduct can be described as a form of *residual demand*

[1] In a sequential market, the forward price is a signal for the expected spot price (Weber et al., 1981). Differences between the two are usually attributed to risk aversion (McAfee and Vincent, 1993), asymmetric shocks (Bernhardt and Scoones, 1994) and market power (Ito and Reguant, 2016).

manipulation.

Capacity constraints have an important role in this context. We assume that the committed output is tied to some production capacity through the contract. More specifically, if the monopolist reneges on its commitments the output is not produced and the associated capacity is left idle. That is, the committed output cannot be sold at the market price.

An illustrative case of strategic reneging took place in Alberta's electricity market in 2010. The Alberta Market Surveillance Administrator (MSA) accused TransAlta Corporation of market manipulations through strategically timed outages – and strategic under-production, or derating – of its coal-fired power plant's generation unit *"Sun 5"* in high demand periods. At the same time, the firm offered the output of its gas-fired capacity at *"Poplar Creek"* at increased prices in the energy market.[2] After due investigation, the regulator concluded that *"by timing the outages at its coal-fired units subject to [power purchase agreements] based on market conditions, rather than on operational conditions [...], TransAlta unfairly exercised its outage timing discretion [...] for its own advantage and made its own portfolio benefits paramount to the competitive operation of the market"* (MSA, 2015). The firm paid CAD 56 millions in settlement following these allegations.[3]

By withholding its coal-fired capacity from the market under claims of maintenance requirements, TransAlta reneged on its power purchase agreement (PPA) with an energy buyer. The PPA specified a price at which the output produced by this capacity must be sold to the other party.[4] The market could not adjust from the unanticipated supply shock caused by the plant coming off-line. An artificially inflated market price was hence the result of the scheme. MSA (2015) reveals that the firm even accounted for the penalties for non-delivery when designing its manipulative strategy.

[2]In the discussion cited in the epigraph, *Sun 5* refers to Sundance coal-fired plant's fifth generation unit located close to Edmonton whereas *Poplar Creek* refers to TransAlta's gas-fired power station near Fort McMurray. The matter of the investigation concerns four outage events involving two plants under PPAs with TransAlta which occurred between November 2010 and February 2011.

[3]This amount corresponds to USD 40.2 millions using exchange rates from December 2015.

[4]Coincidentally, the other party was a rival electricity supplier who was allowed to re-sale the energy at market prices. The strategic outage hence prevented the PPA buyers from competing in the market. In Alberta, the purpose of PPAs was initially to address the concentration of capacity held by TransAlta and two other players, which controlled about 90% of total generation capacity (MSA, 2015).

Following Carlton and Heyer (2008), this anti-competitive behavior is a form of *extensive conduct* as it consists to weaken competition to enhance the firm's profitability. In this case, the firm weakens the competition from its own output sold at forward prices. In contrast, the exercise of unilateral market power absent anti-competitive effects corresponds to an *extractive conduct* on behalf of the firm. Market power is a determinant of the profitability of strategic reneging.

The exercise of market power is often defined as any deliberately designed strategy used by a firm, or a group of firms, to alter market-clearing prices, and to profit from it.[5] The profitability of strategic reneging is plausibly larger for firms with market power. The most extreme form of market power arises when a firm's rivals are capacity constrained and its residual demand is price inelastic. One of the firm's possible strategy is thus to set the price at its maximum possible level. A firm in such a position is referred to as being pivotal (Genc and Reynolds, 2011).[6] A firm with a positive probability to be pivotal may find profitable to renege on its commitments to increase that probability. In this context, reneging weakens competition by forcing competitors to approach their capacity constraints. In our analysis, we study the incentives for market manipulations when the firm may be pivotal.

The scope of strategic reneging extends largely to all sequential markets, beyond this illustrative example in Alberta. A firm may commit to a production schedule in a forward market, then renege on this commitment under claims of technical failure when the uncertainty around the spot market is resolved. There are plenty of such opportunities in electricity markets. For example, a vertically-integrated firm may at times find profitable to under-forecast its demand so as to create an unexpected demand shock in the spot market.[7] Intermittent renewables may be subject to similar concerns. For instance, consider a market participant who also operates a wind farm under

[5]Remark that manipulative conducts involving uneconomic trading or outright fraud also aims at altering prices and do not require a large market share to succeed (Ledgerwood and Carpenter, 2012).

[6]In order to mitigate the impacts of even ephemeral pivotal suppliers, electricity market rules typically define a maximum price, or price cap.

[7]Demand-side participants are generally mandated to disclose their demand forecasts for day-ahead dispatch scheduling and reserve planning. A diversified firm participating on both sides of the market may have incentives to over- or under-forecast its actual demand depending on its net market position so as to affect the following day's reserve margins and production commitments. In 2006, the New York ISO implemented a procedure to identify chronic under-forecaster and curb such conduct (NYISO, 2015, 7-40 p.164).

a feed-in contract paying a fixed price per output unit. The firm may at times be better off to curtail its wind power production to benefit its market portfolio during periods of high prices. Interestingly, this strategy was considered by TransAlta in Alberta: "We believe there is the opportunity to economic dispatch our wind farms by *simply turning off during periods of high wind*" (MSA, 2015). This example is discussed further in Section 1.4.

This paper contributes to the literature on market manipulations. It is the first paper to study strategic reneging in sequential markets under imperfect competition. We disentangle the effects of demand uncertainty, capacity constraints and opportunities for manipulations on equilibrium strategies. Our theoretical framework allows to characterize how the supply strategies and market outcomes are affected by opportunities to renege on prior commitments.

Our main finding is that a strategic firm can benefit from reneging on its commitments from two channels. First, the firm can weaken competition from its own output sold at forward prices as well as from its competitors by forcing them to approach their capacity limit. By doing so, the firm extends its market power and increase its profitability through higher prices and possibly higher volume of market sales. Second, the firm can alleviate demand uncertainties and relax the monotonicity constraint on its supply strategy. We find that the profitability of reneging depends on the volume of committed output, its opportunity cost and the firm's market behavior. An important determinant of this profitability is the market position of the firm. We show that the ability to renege affects supply strategies in equilibrium. Specifically, the firm may increase its expected profits by selling some of its market output at higher prices and renege on its commitments for any realization of demand above a certain threshold. This conduct has potential to increase artificially the market price, the volume of market sales of the firm, and at the extreme, its probability to become pivotal. Under demand uncertainty, there always exists a demand threshold above which a monopolist finds profitable to renege on its commitments in order to become pivotal.

We also find an interesting side result. The presence of a pivotal supplier may render supply function strategies dependent on uncertainties in equilibrium, unlike predicted by the standard supply function equilibrium (SFE) model of Klemperer and Meyer (1989). This is due to a tradeoff between two competing strategies over a range of demand: acting as a monopolist on the elastic part of the residual demand or as a pivotal supplier on its inelastic part. It makes the monotonicity constraint on supply functions binding, hence the functions are not ex-post optimal anymore.

The paper is organized as follows. Section 1.2 presents the related literature. Section 1.3 unravels the model. Section 1.4 discusses reneging as a loss-based manipulative behavior in a general framework, as well as policy implications for intermittent renewables. Section 1.5 concludes.

1.2 Related literature

This section presents the related literature on market power, market manipulations and pivotal suppliers. This paper also relates to the literature on sequential markets in the presence of market power (Allaz and Vila, 1993; Bagnoli, Salant, and Swierzbinski, 1989; Coase, 1972; Ito and Reguant, 2016). However, the forward market is not endogenous in our setting and forward positions are hence considered as exogenous.

Market manipulations and market power

Exercises of market power and manipulative behaviors are strongly interrelated. We discuss those conducts with a focus on electricity markets. Exercises of market power include economic withholding, physical withholding, and transmission-related strategies.

First, economic withholding consists of supplying a firm's own potential output at very high prices so that it is not economically available except in extreme price periods. This extractive conduct is argued to allowing producers to recoup their investment costs and other fixed operation costs as part of the solution to the so-called "missing money problem", and as such does not necessarily constitute an illegal strategy in electricity markets (Harvey and Hogan, 2001; Olmstead and Ayres, 2014). For instance, Brown and Olmstead (2017) find that firms exercise substantial market power in high demand hours in Alberta's electricity market. However, the authors show that in absence of market power exertion, the market profits are generally insufficient to induce adequate investment in new capacity.

Second, physical withholding is also an extractive conduct. It consists in retaining production capacity out of the market so as to reduce the available excess capacity. It is typically associated with strategically timed plant maintenance schedules, as in the well-documented California electricity crisis of summer 2000 (Borenstein et al., 2002; Joskow and Kahn, 2002), and as such is considered a manipulative conduct. Besides strategic plant maintenance, physical withholding strategies may involve other channels. Schill and Kemfert (2011) use a numerical cournot model calibrated to the German electricity market to show that strategic firms tend to under-utilize pumped storage capabilities. Prüggler et al. (2011) find large potential revenues associated with under-utilization of storage and demand-response capacities for a dominant generation company. Their investigation is based on Ontario's electricity market data.

Third, transmission-related strategies can be considered as extensive conducts. Those aim at benefiting market portfolio positions from congestion rents in constrained networks. These rents are a consequence of the market segmentation created by congested networks. The simplest strategy involves predatory pricing in a particular network zone with the objective to foreclose competition in another zone so as to benefit from increased locational market power (Hogan, 1997). A textbook example of such conduct occurred in Alberta during 2011. TransAlta Corporation was fined for alleged impediments of electricity imports from British Columbia in order to inflate domestic market-clearing prices (MSA, 2012). More involved strategies are also used in nodal markets, where prices differ across transmission nodes depending on overall network congestions.

Manipulative conducts in electricity markets are increasingly related to financial markets. The introduction of virtual bidding in U.S. electricity markets has given birth to more sophisticated forms of market manipulations related to transmission networks. They generally aim at artificially amplifying spreads between nodal prices (Birge et al., 2014; FERC, 2012, 2013a,b,c). Virtual bidding is initially motivated by potential reductions of transaction costs through convergence bidding in multi-settlement nodal markets (Jha and Wolak, 2015). However, it also gives room for manipulations of derivative contracts based on transmission congestions (Ledgerwood and Pfeifenberger, 2013).

The form of market manipulations considered in this paper is part of a host of possible extensive conducts in electricity markets. As already explained, withholding of physical generation capacity is only one among many possible *manipulation tools*. Strategic virtual bidding or cross-border trading, withholding of transmission rights (Debia, Benatia, and Pineau, 2018), under-utilization of storage or demand-response capacities are means to alter the distribution of market-clearing prices in order to benefit a trading position. Ledgerwood and Carpenter (2012) present a general framework of market manipulations with examples taken from financial and commodity markets. They emphasize the use of uneconomic trading, i.e. dumping goods to lower a target price, as a manipulation tool which does not require a priori market power for the manipulator.

In Section 1.4, we present strategic reneging as a loss-based manipulative scheme within the framework of Ledgerwood and Carpenter (2012). Strategic reneging is a form of capacity withholding in a sequential framework. Recent empirical evidence of such conduct have been presented in the literature. For instance, Fogelberg and Lazarczyk (2014) find evidence of strategic physical withholding disguised as production failures in the Swedish electricity market, and Bergler et al. (2017) find similar evidence for the German-Austrian market. As regards with the TransAlta case mentioned earlier, the firm's strategy consisted in restraining production[8] from the coal-fired power plant in order to benefit its market portfolio.

As documented by MSA (2015), it need not be an easy task to fake (or overstate) production limitations for a thermal power plant. It should be easier to do so for intermittent renewable power plants, such as wind farms, which production constraints are inherently tied to the randomness of exogenous weather realizations. Unless the regulator is able to monitor these random production constraints accurately,[9] deliberate misforecasting of wind power consists in a relatively low-profile type of strategic reneging in electricity markets.

Pivotal suppliers

As first noted by Genc and Reynolds (2011), the presence of pivotal suppliers in capacity constrained market models is often neglected. Previous research did not consider how strategic firms may actually anticipate their own effects on their rivals' constraints in equilibrium. They provide the following definition:

Definition 1 (Pivotal firm)**.** *The firm is pivotal in the sense that it can move the market price to the price cap with positive probability by withholding output at prices below the cap.*

[8]The strategy also involved reducing output (derating the plant capacity) on some instances rather than completely shutting down the production units.

[9]In the New York electricity market, wind parks are required to submit production forecasts in the day-ahead auction (NYISO, 2016, 4-9 p.41).

In other words, a market participant's ability to exercise "pivotal market power" arises when all its rivals are capacity constrained and it can unilaterally set the market-clearing price to any desired level, below the price cap. Even though electricity markets are subject to capacity constraints, we have identified only a few articles on the role of pivotal suppliers, beside regional market reports. Genc and Reynolds (2011) propose an analysis of the set of symmetric SFE in the presence of pivotal suppliers. They show that as the market's excess capacity falls, the set of SFE shrinks and the most competitive equilibria vanish. Furthermore, as the demand distribution is increasingly more concentrated around the capacity limit, the probability that the price cap is reached goes to one. A key result of the standard SFE model of Klemperer and Meyer (1989) is the ex-post optimality of equilibrium supply function strategies in the presence of additive random demand shocks. Bosco et al. (2012, 2013) studies a vertically-integrated pivotal supplier in the Italian electricity market. They show evidence that it does not behave as predicted by a standard SFE model because the level of the price cap affects its optimal strategy. On a side note, Wolak (2009) argues that a firm in a pivotal position may not have incentives to actually exercise its market power depending on its forward contract position. The firm will not find profitable to exercise its pivotal position if it is a net buyer in the energy market. Forward contracts are often perceived as useful mechanisms to prevent the exercise of market power.[10]

1.3 The model

This section presents the general setup, timing, assumptions and the profit maximization problem. We consider four sets of assumptions. First, the benchmark case without uncertainty nor strategic reneging is presented. Second, the effect of reneging is analyzed with respect to the benchmark. Third, we single out the effect of uncertainty. Finally, we incorporate both ingredients and study how they interact with each other in equilibrium.

[10]It should be noted that, in practice, electricity system operators enforce rules to mitigate pivotal market power. The Pivotal Supplier Index (PSI) and the Residual Supply Index (RSI) are useful indices used to detect and identify pivotal suppliers for given demand levels. Once a pivotal firm is identified, the market regulator usually imposes mitigation procedures to prevent its from exercising unilateral market power.

1.3.1 The setup

Let us consider a market where firms compete in supply functions to satisfy the demand θ. The level of demand θ is a random variable distributed according to $F(\cdot)$ with support $[\underline{\theta}, \overline{\theta}]$. Moreover, the demand is perfectly price inelastic up to an exogenous price cap $\overline{P}$, or market reserve price – hence $F(\cdot)$ is independent of the market price p.[11]

There are two players: a monopolistic firm and a competitive fringe, respectively denoted m and f. Their short-run cost functions are specified as

$$C_m(q_m) = cq_m, \quad \forall q_m \in [0, K_m],$$

for the monopolist, and

$$C_f(q_f) = \begin{cases} cq_f & \text{if } 0 \leq q_f \leq \rho_f K_f \\ cq_f + \frac{b}{2}(q_f - \rho_f K_f)^2 & \text{if } \rho_f K_f \leq q_f \leq K_f, \end{cases}$$

for the fringe. K_i denotes firm i's capacity constraint and q_i denotes its output.[12] The cost parameter $\rho_f \in [0, 1]$ defines the share of the fringe's capacity subject to constant marginal cost $C_f'(q_f) = c > 0$. Production levels larger than $\rho_f K_f$ and up to the capacity limit are subject to an increasing marginal cost $C_f'(q_f) = c + b(q_f - \rho_f K_f)$.

Furthermore, the monopolist may have prior commitments. We assume that it has committed to deliver an output $K_e \geq 0$ at a unit profit profit $r \geq 0$ to some buyers, irrespectively of the realized market-clearing price. The monopolist's total production capacity is hence $K_e + K_m$.[13]

Timing and main assumptions

The timing of the game is as follows

1. Each firm $i \in \{m, f\}$ chooses its supply schedule $S_i(p)$.

2. The level of demand θ is realized according to $F(\cdot)$.

[11] This reserve price is often understood as a constraint imposed by the regulator to mitigate market power abuse – which coincides with the buyers' maximum willingness to pay, often referred to as the value of loss load in electricity markets.

[12] Alternatively, all the results derived in this paper can be obtained when both firms have 'hockey stick'-shaped marginal cost functions like the fringe does.

[13] K_e is assumed to be exogenous in this model.

3. The monopolist may choose to renege on its commitments by choosing $q_e \in [0, K_e]$.

4. The realized uniform market-clearing price p is such that the net demand $\theta - q_e$ is satisfied.

5. Payoffs π_m and π_f are realized.

We consider the following general assumptions.

Assumption 1 (Price cap)**.** *The price cap is such that the entire available capacity can profitably enter production, i.e.* $\bar{P} > c + b(1 - \rho_f)K_f$.

Assumption 2 (Competitive Fringe)**.** *The fringe reveals its marginal cost function, hence the monopolist is the only strategic player.*

Assumption 3 (Demand)**.** *We assume* $K_e < \underline{\theta} < \rho_f K_f + K_e$ *and* $\overline{\theta} = K_f + K_m + K_e$, *so that the firms are always net sellers in this market.*

The monopolist's profit maximization problem

The monopolist faces the residual demand function

$$D(p) = \theta - q_e - S_f(p), \tag{1.1}$$

where $\theta - q_e$ is net market demand and S_f is its rival's strategy. Under Assumption 2, we have $S_f(p) = C_f'^{-1}(p)$ and the corresponding inverse residual demand function faced by the monopolist writes

$$P(q_m, \theta - q_e) = \begin{cases} c & \text{if } q_m \geq \theta - q_e - \rho_f K_f \\ c + b(\theta - q_e - \rho_f K_f - q_m) & \text{if } q_m \in [\theta - q_e - K_f, \theta - q_e - \rho_f K_f] \\ \bar{P} & \text{if } q_m \leq \theta - q_e - K_f \end{cases} \tag{1.2}$$

The monopolist's expected profit maximization problem is given by

$$\begin{aligned} \underset{\{S_m(\cdot), q_e(\cdot)\}}{\text{Max}} \quad & E_\theta\Big[pS_m(p) - C_m(S_m(p)) + rq_e(\theta)\Big] \\ \text{s.t.} \quad & \begin{cases} q_m \geq 0 & \perp \underline{\lambda}_m, \quad q_m \leq K_m \quad \perp \overline{\lambda}_m \\ q_e \geq 0 & \perp \underline{\lambda}_e, \quad q_e \leq K_e \quad \perp \overline{\lambda}_e \end{cases}, \end{aligned} \tag{1.3}$$

where E_θ denotes the expectation operator with respect to the random demand θ. The capacity constraints' shadow costs are denoted $\underline{\lambda}_m$, $\overline{\lambda}_m$, $\underline{\lambda}_e$ and $\overline{\lambda}_e$.

1.3.2 Benchmark: no uncertainty and no reneging

Let us first consider the benchmark case without uncertainty nor strategic use of commitments. Market participants with perfect foresight over demand realizations can condition their supply strategies on the realized level of demand instead of the price level.

Assumption 4 (No uncertainty)**.** *The level of demand is deterministic and known by all players.*

Assumption 5 (No reneging)**.** *The monopolist cannot renege on its commitments, i.e.* $q_e = K_e$*.*

Under the above assumptions, the monopolist's profit maximization problem in (1.3) simplifies to

$$\begin{aligned} \underset{\{q_m(\cdot)\}}{\text{Max}} \quad & P(q_m(\theta), \theta - K_e) q_m(\theta) - C_m(q_m(\theta)) + rK_e \\ \text{s.t.} \quad & q_m(\theta) \geq 0 \perp \underline{\lambda}_m, \quad q_m(\theta) \leq K_m \perp \overline{\lambda}_m. \end{aligned} \tag{1.4}$$

For low demand levels, i.e. $\theta \leq \theta_c \coloneqq \rho_f K_f + K_e$, the fringe has sufficient constant marginal cost capacity $\rho_f K_f$ to satisfy the whole net demand $\theta - K_e$ at the competitive price c, implying $P'(q_m) = 0$. Therefore, the monopolist can at best make zero profit by engaging in Bertrand competition against the fringe.[14] For $\theta - K_e \in [\underline{\theta} - K_e, \theta_c - K_e]$, there is a continuum of possible equilibria where the fringe and the monopolist supply $\theta - K_e$ at price c. This corresponds to the competitive strategy q_m^c.

For higher values of demand, i.e. $\theta > \theta_c$, the fringe's constant marginal cost capacity is not sufficient. The strategic player's best response is to act as a monopolist over the elastic residual demand.[15] We denote this residual monopolist strategy as $q_m^\star$. As demand rises, the fringe's capacity constraint eventually becomes binding for some level of demand $\tilde{\theta}$ such that $\tilde{\theta} = K_e + K_f + q_m^\star(\tilde{\theta})$.

As demand increases beyond $\tilde{\theta}$ the fringe produces at full capacity $q_f = K_f$ and the monopolist has remaining idle capacity $K_m - q_m^\star(\tilde{\theta}) > 0$, assuming $K_m > (1 - \rho_f)K_f$. The monopolist is hence pivotal and can supply the inelastic residual demand $\theta - K_e - K_f - q_m^\star(\theta)$ at the price cap $\bar{P}$.

[14] This case relates to low demand periods in von der Fehr and Harbord (1993) (see Proposition 2. p.534.)

[15] This situation is similar to Klemperer and Meyer (1989), where market participants behave individually as residual monopolists.

Yet, there is no reason to expect this strategy to necessarily coincide with the best strategy.[16] The monopolist could profit from a pivotal position for lower demand levels, and earn possibly larger profits, by anticipating its own impact on the fringe's capacity constraint. In effect, choosing to offer a smaller quantity for demand levels below $\tilde{\theta}$ would force the fringe to meet its capacity constraint earlier. At the extreme, if the monopolist's strategy were to withhold its entire capacity, i.e. to produce $q_m = 0$ as long as $q_f < K_f$. The fringe's capacity would be exhausted as early as $\theta \geq \underaccent{\tilde}{\theta} := K_f + K_e$, i.e. the minimum level of demand at which the firm can become a pivotal supplier. The strategic player can hence choose to supply only the inelastic residual demand $\theta - \underaccent{\tilde}{\theta}$ at the highest possible price $\bar{P}$, as long as $K_m \geq \theta - \underaccent{\tilde}{\theta}$. This is however not necessarily the best strategy for the monopolist either. There are two competing natural strategies for $\theta \in [\underaccent{\tilde}{\theta}, \tilde{\theta}]$: either

(a) to act as a residual monopolist by pricing output using the standard inverse elasticity rule;[17] or

(b) to benefit from a pivotal position by supplying only the inelastic demand at the price cap.

In other words, the firm faces a trade-off between higher output with the residual monopolist strategy $q_m^\star$ and higher prices with the pivotal supplier strategy $q_m^\dagger$. Denote the indirect profit functions in both regimes by $\pi_m^\star(\theta) = \pi_m(q_m^\star(\theta))$ and $\pi_m^\dagger(\theta) = \pi_m(q_m^\dagger(\theta))$ respectively. Lemma 1 characterizes the monopolist's best strategy over the demand interval $[\underaccent{\tilde}{\theta}, \tilde{\theta}]$. Notably, it establishes existence, uniqueness and other properties of a threshold level of demand $\theta_1 \in [\underaccent{\tilde}{\theta}, \tilde{\theta}]$ which determines the optimal transition from $q_m^\star$ to $q_m^\dagger$.

Lemma 1 (Optimal Transition to Pivotal Position)**.** *Under Assumptions 1-5, the monopolist's best strategy over the demand interval $[\underaccent{\tilde}{\theta}, \tilde{\theta}]$ is such that:*

(a) There exists a unique $\theta_1 \in [\underaccent{\tilde}{\theta}, \tilde{\theta}]$ such that $\pi_m^\star(\theta_1) = \pi_m^\dagger(\theta_1)$. For $\theta \in [\underaccent{\tilde}{\theta}, \theta_1]$, the firm's best strategy is to act as a residual monopolist (i.e.

[16] The equilibria corresponding to this naive strategy are characterized in the appendix.

[17] In an oligopolistic setting, it could be any equilibrium strategy in a SFE as explained in Genc and Reynolds (2011).

with strategy $q_m^\star$), whereas it finds more profitable to act as a pivotal supplier (i.e. with strategy $q_m^\dagger$) for $\theta \in [\theta_1, \tilde{\theta}]$.

(b) *The larger the price cap, the smaller the demand threshold θ_1 above which the pivotal regime prevails in equilibrium.*

(c) *At the demand threshold θ, the monopolist reduces its output from $q_m^\star(\theta_1)$ to a strictly smaller quantity $q_m^\dagger(\theta_1)$.*

Proof. See Appendix. □

According to Lemma 1, this optimal transition occurs for the demand level such that the indirect profit functions in both regimes equalize. The price cap is an important determinant of this threshold: a larger cap will induce the firm to exercise its pivotal market power for lower demand levels, and reversely. Interestingly, when θ_1 is reached the firm reduces its output from $q_m^\star(\theta_1)$ to $q_m^\dagger(\theta_1)$ and rises the offer price from $P(q_m^\star(\theta_1))$ to $\bar{P}$. This strategy defines an inverse supply correspondence with respect to quantity as soon as $q_m^\star(\theta_1) > q_m^\dagger(\theta_1)$. The equilibria are characterized in Proposition 1.

Proposition 1 (Equilibria)**.** *Under Assumptions 1-5, all equilibria are such that:*

(a) *Competitive Regime ($\theta \in [\underline{\theta}, \theta_c]$): The fringe produces with the constant marginal cost capacity and both players make zero market profit.*

(b) *Monopoly Regime ($\theta \in (\theta_c, \theta_1]$): The fringe produces with its increasing marginal cost capacity and the strategic firm acts as a monopolist on its residual demand.*

(c) *Pivotal Regime ($\theta \in [\theta_1, \overline{\theta}]$): The fringe is capacity constrained and the strategic firm has a pivotal position as it supplies only the inelastic residual demand at the price cap.*

where $\tilde{\theta} = K_e + K_f + (1 - \rho_f)K_f$. The corresponding equilibrium strategies are such that:

$$S_1^\star(p) = \begin{cases} 0 & \text{if } p \in [0, c) \\ q_m \in [0, K_m] & \text{if } p = c \\ \frac{p-c}{b} & \text{if } p \in (c, c + bq_m^\star(\theta_1)] \\ K_m & \text{if } p = \bar{P} \end{cases}$$

Proof. See Appendix. □

Those results are illustrated in Figure 1.1a. The blue line represents the inverse residual demand function faced by the monopolist at net market demand $\theta_1 - K_e$. The red line represents the equilibrium inverse supply strategy of the monopolist. The red (resp. hatched) area corresponds to the market profits of the strategic player when acting as a pivotal player (resp. residual monopolist). It should be clear from this figure that the ex-post optimal supply function does not satisfy monotonicity in p. For comparison, Figure 1.1b illustrates the naive best strategy, denoted $S_0^{-1}(\cdot)$, which corresponds to the situation where the monopolist does not anticipate its effect on the fringe's capacity constraint.

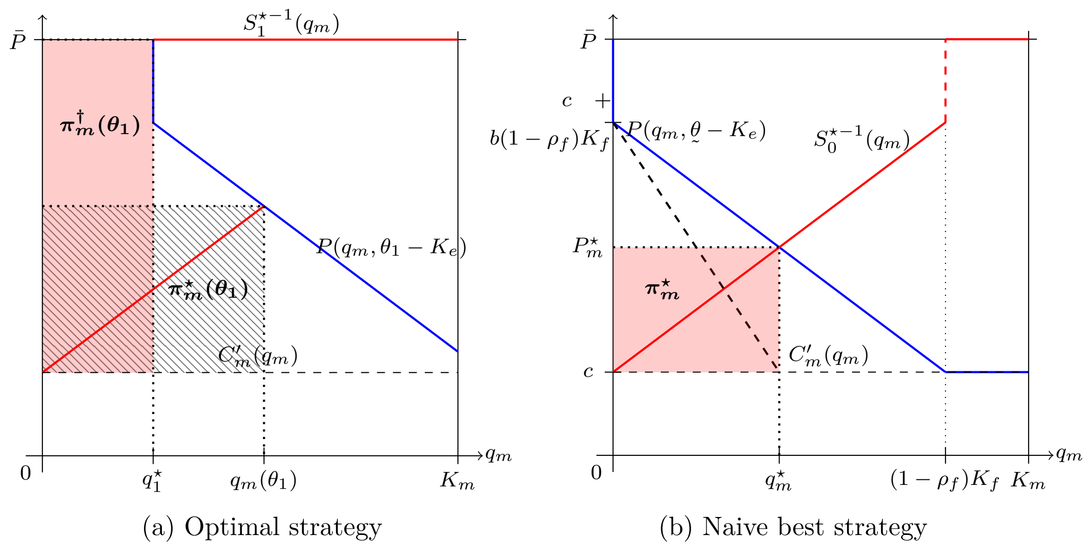


Figure 1.1: Optimal strategy without reneging and uncertainty

1.3.3 Reneging on commitments

We now turn to the case where the monopolist has the ability to renege on its commitments upon observing θ. First, we study the determinants of the profitability of this form of extensive conduct. Second, we characterize how optimal supply strategies are affected. Let us substitute Assumption 5 by the following.

Assumption 6 (Reneging on commitments)**.** *The monopolist is able to choose output level $q_e \in [0, K_e]$ upon observing the realized demand level θ.*

Under Assumption 6, the monopolist's profit function becomes

$$\pi_m(q_m, q_e) = P(q_m, q_e)q_m - C_m(q_m) + rq_e, \tag{1.5}$$

which total differentiation with respect to q_e gives

$$\frac{\partial P}{\partial q_e}q_m + \left(\frac{\partial P}{\partial q_m}q_m + P - C'_m(q_m)\right)\frac{dq_m}{dq_e} + r, \tag{1.6}$$

where $\frac{\partial P}{\partial q_e} = \frac{\partial P}{\partial q_m}$ since $P(q_m, q_e) = S_f^{-1}(\theta - q_e - q_m)$. A marginal reduction of the committed output to $q_e < K_e$ can be done at the cost of the foregone unit profit rate r. The strategic player's profit is affected through two channels.

- ★ First, it may affect the market-clearing price upwards, since $\frac{\partial P}{\partial q_e} \leq 0$. The *intensive margin* corresponds to the increased profit margin of the market output q_m sold by the monopolist. It coincides with the first term in (1.6).

- ★ Second, this output level may also increase since $\frac{dq_m}{dq_e} \geq 0$. In turn, it may enlarge market profits as the extra output is to be sold at an higher price $p - \frac{\partial P}{\partial q_m}q_m$. This second-order effect corresponds to the *extensive margin*. It is represented by the second term in (1.6).

Following Proposition 1, the monopolist's best strategy is a function of net market demand $\theta - q_e$ with $q_e = K_e$. For any value of $\theta - q_e$, we are assured that the best strategy is either to act competitively (q_m^c), as a residual monopolist ($q_m^\star$) or as a pivotal supplier ($q_m^\dagger$). Solving for the best strategy amounts as previously to finding the demand thresholds corresponding to the profit-maximizing transitions from one strategy to another. The only difference lies in the firm's ability to manipulate its residual demand through reneging. Reneging on commitments can serve two purposes:

(a) to increase net market demand *within* a regime, so as to push either the price or output upwards; or

(b) to *shift* the market equilibrium to a less competitive regime.

The feasibility and profitability depend on the level of demand, opportunity cost r, and ultimately the maximum profit margin $\bar{P} - c$.

If $q_e = 0$, the monopoly regime would arise as soon as $\theta \in [\theta_c - K_e, \theta_1 - K_e)$, and the pivotal regime for $\theta \geq \theta_1 - K_e$. We can decompose the demand's domain into five intervals for which reneging differs in scope as depicted in Figure 1.2. We briefly discuss the profitability of manipulations in each interval of demand.

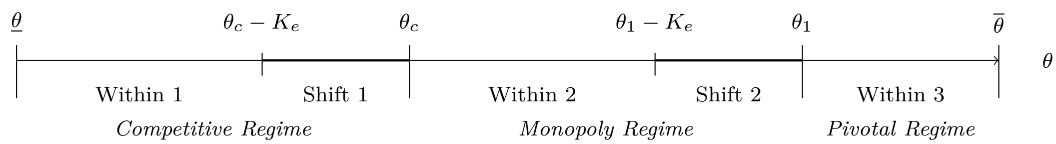


Figure 1.2: Market regimes as a function of demand

(Within 1) Within the competitive regime: For $\theta \in [\underline{\theta}, \theta_c - K_e)$, we have $\pi(q_m^c, q_e) = rq_e$. Since this profit function is increasing in q_e if $r > 0$, it is never profitable to withdraw its commitments within the competitive regime.

(Shift 1) Shift to the monopoly regime: In the absence of manipulations, the market is competitive for $\theta \in [\theta_c - K_e, \theta_c)$. However, if $q_e < K_e$, the monopoly regime prevails for $\theta \in [\theta_c - (K_e - q_e), \theta_c]$. To remain in the competitive regime provides zero market profit though a strictly positive profit from the exogenous capacity, $\pi_m^c(\theta, K_e) = rK_e$, whereas shifting to the monopoly regime by setting $q_e < K_e$ generates $\pi_m^\star(\theta, q_e) = P(q_m^\star, q_e)q_m^\star - C_m(q_m^\star) + rq_e$. From (Within 2) below, $q_e = 0$ when reneging is profitable in the monopoly regime. Therefore, it is profitable to shift from the competitive regime to the monopoly regime if for some $\theta \in [\theta_c - K_e, \theta_c]$, we have $\pi_m^\star(\theta, 0) - \pi_m^c(\theta, K_e) \geq 0$ where the equality holds at θ^M, and the inequality is strict for larger demand levels.

(Within 2) Within the monopoly regime: For $\theta \in [\theta_c, \theta_1 - K_e)$, the monopolist's best strategy is $q_m^\star$. Upon observing the realization of the inelastic demand θ, the firm seeks to maximize its profit as follows

$$\max_{q_e} \quad P(q_m^\star, q_e)q_m^\star - C_m(q_m^\star) + rq_e \quad \text{s.t.} \quad \{q_e \geq 0 \perp \underline{\lambda}_e, \quad q_e \leq K_e \perp \overline{\lambda}_e\}. \tag{1.7}$$

From the Envelope Theorem, differentiating with respect to q_e gives the first-order condition $\frac{\partial P(q_m^\star, q_e)}{\partial q_e} q_m^\star + r = \overline{\lambda_e} - \underline{\lambda_e}$. Taking the second-order derivative yields $\frac{\partial^2 P(q_m^\star, q_e)}{\partial q_e^2} q_m^\star \geq 0$ as $P(q_m, q_e) = S_f^{-1}(\theta - q_m - q_e)$ is a convex function of q_e. The convexity of the objective function leads us to investigate corner solutions. Therefore, reneging on commitments is profitable within the monopoly regime for any $\theta \in [\theta_c, \theta_1 - K_e)$ such that $\pi_m^\star(\theta, 0) - \pi_m^\star(\theta, K_e) \geq 0$. The equality holds at the threshold θ^M, and the inequality is strict for larger demand levels because this profit difference is increasing in θ.

(Shift 2) Shift to the pivotal regime: For $\theta \in [\theta_1 - K_e, \theta_1)$, the firm has a choice between setting $q_e = q_e^\star \in [\theta - K_f - q^\star, K_e]$ and earn the monopoly profit $\pi_m^\star(\theta, q_e^\star)$, or setting $q_e = q_e^\dagger \in [0, \theta - K_f - q^\star]$ so as to reap the pivotal supplier profit $\pi_m^\dagger(\theta, q_e^\dagger)$. Let $q_2^\star = \theta_2 - K_f - q_e$ denote the smallest output sold in the market when the pivotal regime is reached. There are two possible cases depending on the value of r. First, if r is such that full reneging is profitable for some θ in both the monopoly regime and the pivotal regime, then $q_e^\star = q_e^\dagger = 0$ for all $\theta \in [\theta_1 - K_e, \theta_1)$. The optimal $q_2^\star$ will coincide to $q_1^\star$ as it is derived from the equivalent profitability condition: $\pi_m^\dagger(\theta, 0) - \pi_m^\star(\theta, 0) \geq 0$. Second, if r is such that it is profitable to renege within the pivotal regime but not within the monopoly regime, the condition is changed into: $\pi_m^\dagger(\theta, 0) - \pi_m^\star(\theta, K_e) \geq 0$.

(Within 3) Within the pivotal regime: For $\theta \geq \theta_1$, the monopolist's best strategy is $q_m^\dagger$. Upon observing θ, the firm maximizes its profit as

$$\max_{q_e} \quad \bar{P}(\theta - q_e - K_f) - C_m(\theta - q_e - K_f) + rq_e \quad \text{s.t.} \quad \{q_e \geq 0 \perp \underline{\lambda}_e, \quad q_e \leq K_e \perp \overline{\lambda}_e\}. \tag{1.8}$$

The FOC is $-\bar{P} + C_m'(\theta - q_e - K_f) + r = \overline{\lambda_e} - \underline{\lambda_e}$ which is linear given the monopolist's constant marginal costs. The optimal solution corresponds to the corner solution where $q_e^\star(\theta) = 0$ if and only if $r \leq \bar{P} - c$. The monopolist will gradually supply the committed output for demand levels above $\overline{\theta} - K_e$. In the pivotal regime, any committed output withheld is exactly substituted by an extra market output produced by the monopolist, since all its rivals are capacity constrained. If the profit margin obtained in the market is lower than that of the committed output, i.e. $\bar{P} - c < r$, reneging on commitments is never profitable for the firm.

The discussion is summarized in Lemma 2 along with the characterization of optimal reneging in the absence of uncertainty.

Lemma 2 (Optimal Reneging)**.** *Under Assumptions 1-4 & 6, if manipulations are profitable for some θ^M, then they are also for all $\theta \geq \theta^M$. This demand threshold increases with the opportunity cost of manipulations, that is $\frac{\partial \theta^M}{\partial r} \geq 0$. In equilibrium, the monopolist's best strategy with respect to production q_e is such that:*

(a) If r is small, full reneging ($q_e = 0$) is profitable prior to the pivotal regime. There exists a unique $\theta^M \in [\underline{\theta}, \theta_1 - K_e]$. The lower (resp. upper) bound is reached for $r = 0$ (resp. $r = bq^\star_m(\theta_1) - b\frac{K_e}{4}$). The demand threshold for the pivotal strategy is $\theta_2 = \theta_1 - K_e$ with market position $q^\star_2 = q^\star_1$.

(b) If r is larger than $bq^\star_m(\theta_1) - b\frac{K_e}{4}$ though lower than $\bar{P} - c$, full reneging is profitable to reach the pivotal regime. There exists a unique $\theta^M \in (\theta_1 - K_e, \theta_1]$. The lower (resp. upper) bound is reached for $r = bq^\star_m(\theta_1) - b\frac{K_e}{4}$ (resp. $r = \bar{P} - c$). The demand threshold for the pivotal strategy is $\theta_2 = \theta^M$ with market position $q^\star_2 > q^\star_1$.

(c) If r is larger than $\bar{P} - c$, manipulations are never profitable. Thus $\theta^M > \overline{\theta}$, $\theta_2 = \theta_1$ and $q^\star_2 = q^\star_1$.

Proof. See Appendix. □

With regards to Lemma 2, market manipulations may affect the best supply strategy, depending on the value of r. Overall, the pivotal regime occurs for lower demand levels unless the exogenous price r is larger than the maximum profit margin the firm can obtain in the market $\bar{P} - c$. These results allow us to characterize the equilibria in presence of strategic reneging in Proposition 2.

Proposition 2 (Equilibria with Reneging)**.** *Under Assumptions 1-4 & 6, all equilibria are such that:*

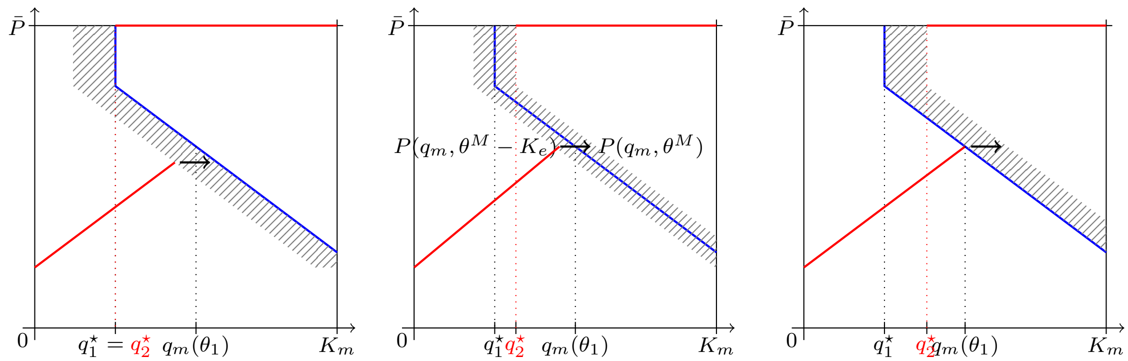
(a) For $r \leq \bar{P} - c$, there always exist profitable manipulations and all equilibria differ from those in Proposition 1. In particular, the demand threshold θ_2 above which the pivotal strategy prevails is lower and its corresponding market position $q_2^\star$ may be larger. Hence, the equilibrium strategies may be affected by the ability to renege on its commitments.

(b) For $r > \bar{P} - c$, manipulations through reneging are never profitable and all equilibria coincide to those in Proposition 1.

The corresponding equilibrium supply schedule strategies are such that:

$$S_2^\star(p) = \begin{cases} 0 & \textit{if } p \in [0, c) \\ q_m \in [0, K_m] & \textit{if } p = c \\ \frac{p-c}{b} & \textit{if } p \in (c, c + bq_m^\star(\theta_2)] \\ K_m & \textit{if } p = \bar{P} \end{cases}$$

Figure 1.3 illustrate the previous results. Figure 1.3a shows the case when r is small and reneging is profitable for low demand levels. The hatched area corresponds to the set of residual demand functions suppressed by reneging. When demand is such that the residual demand function belongs to this set, the monopolist will renege on its commitments by setting $q_e = 0$, which will shift the residual demand to the right by K_e. The arrow represents this shift. Figures 1.3b and 1.3c show the other cases when the opportunity cost is larger, and the equilibrium market output will be increased: $q_2^\star > q_1^\star$ compared to the benchmak. This is because the opportunity cost of manipulations is compensated by a larger market position when the price cap is attained, i.e. $q_2^\star > q_1^\star$. Note that this form of extensive conduct shrinks the range of demand over which the pivotal and monopoly strategies compete in equilibrium.



(a) r small (b) r intermediate (c) $r = \bar{P} - c$

Figure 1.3: Optimal strategy with reneging and without uncertainty

1.3.4 Demand uncertainty

Let us revert to Assumption 5 and assume that the demand is randomly distributed according to the known distribution $F(\cdot)$. The monopolist cannot condition its supply strategy on the level of demand anymore. Its supply function must now satisfy a monotonicity in p.

Suppose the firm chooses to supply a maximum amount q as a residual monopolist. The demand level at which this amount is exhausted is $\theta^-(q) = q_m^{\star -1}(q) = \rho_f K_f + K_e + 2q \geq \theta_c$, and the demand threshold above which the pivotal regime is reached becomes $\theta^+(q) = K_f + K_e + q \leq \tilde{\theta}$.[18] For demand levels $\theta \in [\theta^-(q), \theta^+(q)]$, the monopolist keeps its production level constant at q, and the market-clearing price is settled by the fringe. Let $\pi_m^w(\theta) = \pi_m(q) = P(q, \theta - K_e)q - C_m(q)$ be the profit function in this withholding regime. The monopolist's profit-maximization problem becomes

$$\max_q \int_{\underline{\theta}}^{\theta_c} \pi_m^c(\theta) dF(\theta) + \int_{\theta_c}^{\theta^-(q)} \pi_m^\star(\theta) dF(\theta) + \int_{\theta^-(q)}^{\theta^+(q)} \pi_m^w(\theta) dF(\theta) + \int_{\theta^+(q)}^{\bar{\theta}} \pi_m^\dagger(\theta) dF(\theta). \tag{1.9}$$

This *economic withholding* strategy is the only one that allows to enjoy a pivotal position prior to $\tilde{\theta}$. It is defined as follows.

[18] Following Lemma 4 in the appendix, $\theta^+(q) \leq \tilde{\theta}$ thus $q \leq (1 - \rho_f)K_f$.

Proposition 3. *(Economic withholding) Under demand uncertainty, the firm follows the strategy:*

1. *Act as a residual monopolist ($q_m^\star$) for all prices $p \in (c, q_m^{\star-1}(q^\star))$.*
2. *Keep production constant at $q^\star$ for all prices p such that $p \geq q_m^{\star-1}(q^\star)$ and the rival's capacity constraint slacks.*
3. *Act as a pivotal supplier ($q_m^\dagger$) for prices at which the rival's capacity constraint is binding.*

where $q^\star$ solves the profit maximization problem (1.9).

Let us assume that the distribution of θ is such that the second-order condition of (1.9) for a global maximum is satisfied. The optimal $q^\star$ is then characterized by the first-order condition given by

$$\Big(\pi_m^\dagger(\theta^+(q^\star)) - \pi_m^w(\theta^+(q^\star))\Big) f(\theta^+(q^\star)) = \int_{\theta^-(q^\star)}^{\theta^+(q^\star)} \frac{d\pi_m^w(\theta)}{dq^\star} dF(\theta). \tag{1.10}$$

This optimality condition states that $q^\star$ must result from the trade-off between the expected marginal benefit of economically withholding the quantity $(1-\rho_f)K_f - q^\star$, and its expected marginal cost. The former (left-hand-side) consists in the expected gains associated with an increases probability to become pivotal whereas the latter (right-hand-side) corresponds to the opportunity costs of keeping production fixed instead of pursuing the more profitable residual monopolist strategy.

The withholding range is the span of demand $[\theta^-(q^\star), \theta^+(q^\star)]$ over which the strategic player keeps its production fixed at $q^\star$. If the realized level of demand falls in this range, the firm would prefer to deviate ex-post to either the monopoly or the pivotal strategy.

In order to characterize the optimal $q^\star$, let us assume the following uniform distribution of demand. The previous discussion is then formalized in Proposition 4.

Assumption 7 (Sufficient condition)**.** *Assume $\theta \sim U[\underline{\theta}, \overline{\theta}]$.*

Proposition 4. *(Equilibria under Uncertainty) Under Assumptions 1, 2, 5 & 7, there exists a unique equilibrium where $q_3^\star \in [q_1^\star, q_m(\theta_1)]$ is characterized* *by the FOC (1.10), such that* $\theta_1 \in [\theta^-(q_3^\star), \theta^+(q_3^\star)]$. *Furthermore, this equilibrium strategy is not ex-post optimal for demand levels in the withholding regime, that is* $\theta \in [\theta^-(q_3^\star), \theta^+(q_3^\star)]$. *In particular, the strategic firm is ex-post strictly better off with* $q_m^\star$ *for realizations* $\theta \in [\theta^-(q_3^\star), \theta_1]$, *and strictly better off with* $q_m^\dagger$ *for realizations* $\theta \in [\theta_1, \theta^+(q_3^\star)]$. *Finally, as the demand distribution is more concentrated above (resp. below)* θ_1, *the probability to reach the price cap goes to one (resp. zero) and* $q_3^\star \to q_1^\star$ *(resp.* $q_3^\star \to q_m^\star(\theta_1)$*).*

The corresponding equilibrium supply schedule strategy is such that:

$$S_3^\star(p) = \begin{cases} 0 & \textit{if } p \in [0, c) \\ \frac{p-c}{b} & \textit{if } p \in (c, c + bq_3^\star] \\ K_m & \textit{if } p = \bar{P} \end{cases}$$

Proof. See Appendix. □

Proposition 4 suggests that the pivotal regime kicks in for larger demand levels that in the absence of uncertainty. This is because the profit loss in the withholding regime must be compensated by an increased profit in the pivotal regime, thus an increased output since the price cannot increase beyond its cap.

Figure 1.4 illustrates these results. The dotted lines around the grey zone denote the residual demand functions associated with demand levels $\theta^-(q_3^\star)$ and $\theta^+(q_3^\star)$, respectively. This zone corresponds to the withholding regime.

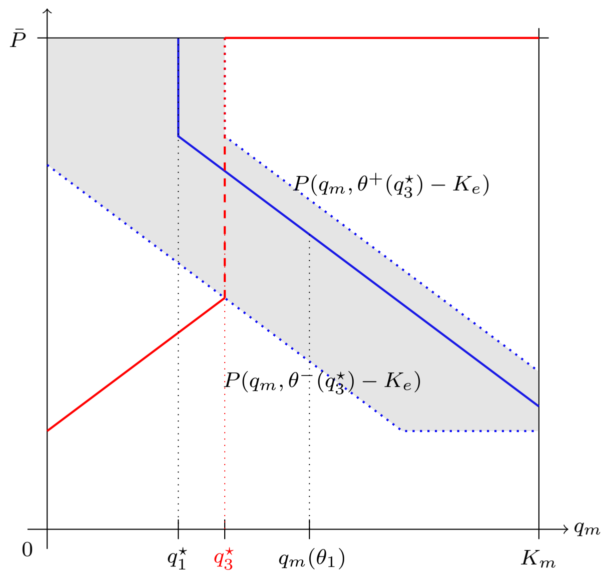


Figure 1.4: Optimal strategy under uncertainty without reneging

Cost of uncertainty and gains from economic withholding

Let us write the cost of uncertainty for the monopolist. It is defined as the difference between the ex-post and ex-ante optimal profits in (1.9) as

$$CU_m(q^\star) = \int_{\theta^-(q^\star)}^{\theta_1} \Big(\pi_m^\star(\theta) - \pi_m^w(\theta)\Big) dF(\theta) + \int_{\theta_1}^{\theta^+(q^\star)} \Big(\pi_m^\dagger(\theta) - \pi_m^w(\theta)\Big) dF(\theta). \tag{1.11}$$

From Proposition 4, both terms are strictly positive and the optimal $q^\star$ is such that this cost is minimized. In the absence of economic withholding, the cost of uncertainty simplifies to

$$CU_m((1-\rho_f)K_f) = \int_{\theta_1}^{\tilde{\theta}} \Big(\pi_m^\dagger(\theta) - \pi_m^\star(\theta)\Big) dF(\theta). \tag{1.12}$$

It corresponds to the expected opportunity cost of not being able to reach a pivotal position at the optimal demand level θ_1. On the other hand, the expected gains from economic withholding for the strategic firm corresponds to the difference in expected profits for $q = q^\star$ and q such that $\theta^-(q) = \tilde{\theta} = \theta^+(q)$ (i.e. $q = (1-\rho_f)K_f)$). That is, the withholding range is degenerate. This expected gain writes as

$$G_m(q^\star) = \int_{\theta^+(q^\star)}^{\tilde{\theta}} \Big(\pi_m^\dagger(\theta) - \pi_m^\star(\theta)\Big) dF(\theta) - \int_{\theta^-(q^\star)}^{\theta^+(q^\star)} \Big(\pi_m^\star(\theta) - \pi_m^w(\theta)\Big) dF(\theta). \quad (1.13)$$

It consists of two terms: *(i)* the expected profit from being able to reach a pivotal position at $\theta^+(q^\star) < \tilde{\theta}$ instead of acting as a residual monopolist; and *(ii)* the expected opportunity cost of foregoing the monopoly profit for the withholding profit. The optimal $q^\star$ also coincides with the maximizer of (1.13).

1.3.5 Reneging on commitments under demand uncertainty

In the presence of uncertainty, a firm may choose to renege on its commitments to complement economic withholding as a means to trigger the price cap before $\tilde{\theta}$. Manipulations can essentially alleviate the cost of uncertainty associated with economic withholding, and reduce the threshold demand level above which the pivotal position prevails in equilibrium.

Let us assume that both Assumption 6 and Assumption 7 hold, so that reneging is possible in the presence of demand uncertainty. The optimal strategy is fully characterized by the amount of economic withholding $(1-\rho_f)K_f - q$ and the demand threshold θ^M above which manipulations are profitable. The monopolist can shift its residual demand to attenuate the effects of a random demand and thereby ensure larger ex-post profits when the realized demand falls in the withholding range. We draw on the results from Lemma 2 to focus our analysis on the three main cases:

(a) The opportunity cost r is small enough to make manipulations profitable before the transition to the pivotal regime;

(b) r is large enough to affect the choice of $q^\star$ at the margin; and

(c) r is so large that manipulations are never profitable.

We avoid a number of possible cases by replacing Assumption 1 with the more restrictive Assumption 8. This allows to neglect the case where $q^\star$ is so small and K_e so large that $\theta^-(q^\star, K_e) > \theta^+(q^\star, 0)$, i.e. manipulations offset the entire withholding regime in equilibrium. The resulting market regimes are depicted in Figure 1.5.

Assumption 8 (Simplification)**.** *The price cap is such that* $\bar{P} > c + b\frac{((1-\rho_f)K_f - K_e)^2}{(1-\rho_f)K_f - 2K_e} > b(1-\rho_f)K_f$*, and the committed output is not too large with respect to the increasing marginal cost capacity of the fringe, i.e.* $(1-\rho_f)K_f > 2K_e$*.*

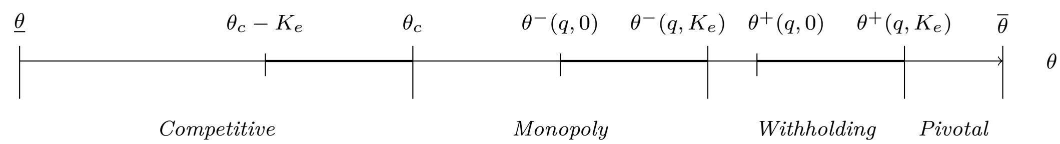


Figure 1.5: Market regimes as a function of demand under uncertainty

Suppose reneging on its commitments is not profitable unless used to shift to the pivotal regime and beyond. That is, $\pi^w_m(\theta, K_e) \geq \pi^w_m(\theta, 0)$ and $\pi^\dagger_m(\theta, 0) \geq \pi^\dagger_m(\theta, K_e)$ but $\pi^\dagger_m(\theta, 0) \geq \pi^w_m(\theta, K_e)$. The first two conditions imply $r \in [dq, \bar{P} - c]$, and the last condition holds for all $\theta \geq \theta^M(q) = K_f + \frac{r - bq}{\bar{P} - c - bq} K_e + \frac{b((1-\rho_f)K_f - q)}{\bar{P} - c - bq} q$, where feasibility requires $\theta^M(q) \in [\theta^+(q, 0), \theta^+(q, K_e)]$. Hence, the monopolist's expected profit maximization problem is given by

$$
\begin{aligned}
\max_q \int_{\underline{\theta}}^{\theta_c} \pi^c_m(\theta, K_e) dF(\theta) &+ \int_{\theta_c}^{\theta^-(q,K_e)} \pi^\star_m(\theta, K_e) dF(\theta) + \int_{\theta^-(q,K_e)}^{\theta^M(q)} \pi^w_m(\theta, K_e) dF(\theta) \\
&+ \int_{\theta^M(q)}^{K_m + K_f} \pi^\dagger_m(\theta, 0) dF(\theta) + \int_{K_m+K_f}^{\bar{\theta}} \pi^\dagger_m(\theta, \theta - K_m - K_f) dF(\theta) \\
&\text{s.t.} \quad \theta^M(q) \geq \theta^+(q, 0) \perp \mu,
\end{aligned}
$$

and the optimal $q^\star$ is characterized by the following FOC

$$
\mu(1 - \theta^{M'}(q^\star)) = \int_{\theta^-(q^\star, K_e)}^{\theta^M(q^\star)} \frac{d\pi^w_m(\theta, K_e)}{dq} dF(\theta). \tag{1.14}
$$

This optimality condition states that the withholding range $[\theta^{-}(q^\star, K_e), \theta^{M}(q^\star)]$ must be as narrow as possible with $q^\star$ such that $\theta^{M}(q^\star) = \theta^{+}(q^\star, 0)$. Proposition 5 characterizes the equilibria.

Proposition 5 (Equilibria with Reneging and Uncertainty). *Under Assumptions 2, 6, 7 and 8,*

(a) If r is small, full reneging ($q_e^\star = 0$) is profitable for low demand levels and economic withholding is left unchanged ($q_4^\star = q_3^\star$).

(b) For larger values of r although lower than $\bar{P} - c$, full reneging is still profitable but the optimal amount of economic withholding is increased ($q_4^\star \leq q_3^\star$), if K_e is not too large.

(c) It is not profitable to renege on its commitments only if the opportunity cost is infinite, i.e. as $r \to \infty$, $q_e^\star \to K_e$ and $q_4^\star \to q_3^\star$.

The corresponding equilibrium supply schedule strategy is such that:

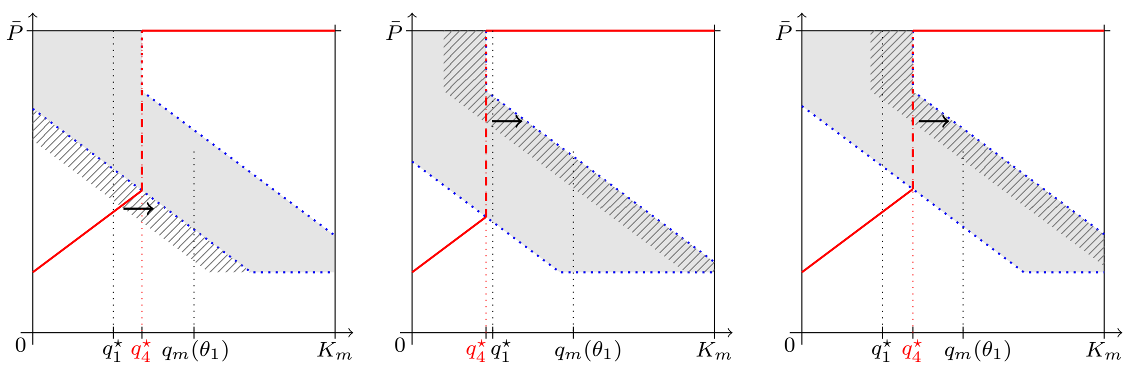
$$S_4^\star(p) = \begin{cases} 0 & \text{if } p \in [0, c) \\ \frac{p-c}{b} & \text{if } p \in (c, c + bq_4^\star] \\ K_m & \text{if } p = \bar{P} \end{cases}$$

Proof. See Appendix. □

As long as r is finite, reneging will always be profitable because of the discontinuous transition in the optimal profit function from the withholding regime to the pivotal regime. The monopolist can smooth this transition by reneging on its commitments. This explains the absence of the left term of (1.10) in (1.14). The results in Proposition 5 provide theoretical aspects about how the monotonicity constraint is relaxed by reneging on committed output.

These results are illustrated in Figure 3.3. The hatched areas correspond to the set of residual demand functions suppressed by reneging in equilibrium. As soon as demand is such that the residual demand belongs to this set, the firm will renege on its commitments so as to shift the function to the right, as represented by the arrow. When r is small, full reneging is optimal before the withholding regime is reached and the market position coincides to that without reneging (see Proposition 4). As r increases, the market position changes and reneging can be used to suppress the high end of the withholding range. Notably, economic withholding is increased as reneging allow to shrink the withholding range, making them complementary strategies to maximize expected profits.

As shown in Figure 1.6b (resp. Figure 1.6c), the conjunction of reneging and economic withholding can result in a lower (resp. higher) market position with respect to the benchmark level $q_1^\star$, and thus to $q_2^\star$ (with no uncertainty but opportunities to renege). This suggests that in the presence of manipulation opportunities, uncertainty has ambiguous effects on market output in some instances. However, reneging on commitments shifts the distribution of net market demand to the right and thereby makes less competitive regimes more likely to arise. The bottom line is that the firm should be expected to price up its market output through increased economic withholding.



(a) r small (b) r intermediate (c) $r = \bar{P} - c$

Figure 1.6: Optimal strategy with reneging and uncertainty

1.4 Discussion

In this section, we discuss the insights delivered by the model. First, we present strategic reneging as a manipulative conduct and explain how such conduct can be identified. Second, we discuss the implications for intermittent renewable power plants.

1.4.1 A loss-based manipulative behavior

As exposed throughout this paper, the strategic reneging on prior commitments can be used as a manipulative scheme. This scheme consists in incurring a loss, i.e. the foregone revenues from withdrawn commitments, in order to affect the market to earn larger overall profits. Such conduct can be understood as a loss-based manipulative behavior.

Anatomy of the manipulative scheme

Following the framework of Ledgerwood and Carpenter (2012), the anatomy of a market manipulation can be decomposed into a trigger, a target and a nexus.[19] In the present paper, the trigger consists in the commitments on which the firm will renege ex-post upon observing the realization of demand, i.e. $K_e - q_e$. This trigger is associated with a per-unit loss r, hence the loss of the manipulative scheme is $r(K_e - q_e)$. The objective of the scheme is to use the trigger to benefit a portfolio position based on the target. The nexus is the mechanism by which the trigger affects the target.

In this manipulation, the target has two components: the market price p and the market output q_m sold by the firm. The nexus is the causal channel through which reneging shifts the residual demand, and in turn can increase the market price p by Δp and the firm's market output q_m by Δq. Following (1.6), this manipulation is profitable if

$$\Delta p q_m + [(\Delta p + p)\Delta q - \Delta C] \geq r(K_e - q_e), \tag{1.15}$$

where ΔC represents the total cost increase due to the output increase Δq following reneging. The term $(\Delta p + p)\Delta q - \Delta C$ in (1.15) is a second-order effect, unless the firm is already a pivotal supplier since the price cannot go beyond its cap.

The profitability of this loss-based manipulation scheme crucially depends on 4 elements:

- ★ the elasticities of the demand and supply functions which determine Δp and Δq.
- ★ the market position q_m, which acts as leverage;

[19] This terminology relates to that used by the FERC: tool (trigger), target and benefiting position (nexus).

- ★ the per-unit profit loss r; and,
- ★ the commitments to be reneged upon later $K_e - q_e$, which acts as the trigger size.

The identification of this manipulation necessitates information on each of those elements. Interestingly, a profitable manipulation of this type does not require a large market share. It requires the capability to capitalize on the market's inability to adjust to unexpected changes, which essentially boils down to the effect captured by Δp.

Identifying strategic reneging

Proving a manipulative behavior is not a trivial task. It entails to provide evidence of the ability and intent to manipulate the market, as well as the effective creation of an "artificial price" through the actions of the alleged manipulator. The identification of a manipulation hence often revolves around the identification of Δp.

The evaluation of its profitability is key to the identification of a manipulation. Consider that a firm has reneged on its commitments under claims of a production failure. Let us assume that Δp can be estimated and the leverage, unit profit loss and trigger are observed. A simple profitability evaluation is done by comparing the first-order gains to the loss, i.e. $\Delta p q_m \geq r(K_e - q_e)$. If the gains outweigh the loss, there is evidence that this event has been profitable to the firm.

Unfortunately, an estimate of Δp may be the subject of contention. Furthermore, benefiting from the event may not be a satisfactory proof of intent to manipulate. Additional evidence may be collected through audits performed ex-post, or sometimes from email exchanges or tapes, e.g. the famous recording of two Enron traders following the California power system crisis.[20]

The theoretical results derived from our model deliver additional potential red flags that may help to identify a manipulation. According to the model, such an event is more likely to be of strategic nature when:

(a) the leverage and the trigger size are sufficiently large, because it would never be profitable to manipulate prices otherwise;

(b) the per-unit loss is not too large;

[20] https://www.cbsnews.com/news/enron-traders-caught-on-tape/.

(c) the firm has a positive probability to be pivotal, because the scheme could potentially aim at increasing this probability;

(d) the market outcome is less competitive than anticipated ex-ante, i.e. a regime shift may have been caused by the event; and,

(e) the firm prices its market output above usual levels.

The elements (c) and (d) could explain an important price jump Δp if the firm reneged on its commitments at a moment where the market was particularly illiquid. Additional proof of intent can be delivered by element (e): if the event coincides to a moment where the firm's market output was offered at higher prices, it may be evidence of a strategic move. An empirical illustration of this result can be found in the discussion cited in the epigraph of this paper. The trader indicates that *Poplar Creek* – a gas-fired power plant – was pricing up during the hours where *Sun 5* – a coal-fired power plant's unit under PPA – was called off-line.

Welfare implications and inefficiencies

A loss-based manipulative conduct has welfare implications for all market participants. The artificially inflated price created by the manipulation benefits all suppliers with a net selling position in the market, that is not only the manipulator but also all of its competitors. On the other hand, all participants with a net buying position see their welfare reduced due to the inflated price.

This conduct also creates inefficiencies because the withdrawn output would have been produced at a lower cost that the output produced to replace it. Moreover, the price increase may reduce consumption if the demand function is not perfectly inelastic. The manipulation generates a welfare loss embedding both of those effects.

Penalty

The per-unit penalty t that would allow to prevent any profitable strategic reneging is such that

$$\Delta p q_m + [(\Delta p + p)\Delta q + \Delta C] \leq (r + t)(K_e - q_e). \tag{1.16}$$

Neglecting the second-order term and rearranging yield

$$t \geq \Delta p \frac{q_m}{K_e - q_e} - r, \tag{1.17}$$

which corresponds to a lower bound to the profit rate earned from each reneged unit created by the manipulative scheme. Even though it may be difficult to prove an intent to manipulate the market, this lower bound provides a way to estimate disgorgement penalties. This is in line with the FERC's statement that the disgorgement amount "need only be a reasonable approximation of profits causally connected to the violation".[21]

1.4.2 The example of intermittent renewable power plants

In restructured electricity markets, intermittent renewable power plants are often granted dispatch priority and financial support from renewable policies. This support can take the form of contracts paying a fixed per-unit price independent of the wholesale market price signals, such as feed-in tariffs. The main argument in favor of protecting renewable plants from market price signals is to reduce market-related uncertainty for investors. There are two paradigms surrounding the integration of intermittent renewable plants in electricity systems: *(i) market-integration*: to expose them to price signals as for conventional power plants; or *(ii) market-segmentation*: to rely on exogenous payments linked to support schemes, e.g. feed-in tariffs.

Both forms of market design can generate situations for strategic reneging. Whatever the design, a wind power plant will be committed to a production schedule following a given wind forecast. Obviously, there may be forecast errors and actual production will eventually deviate from the committed schedule. Strategic reneging on this commitment is possible by two means. First, it is possible to over- or under-forecast output, so that the actual output will deviate from commitments in the desired direction. Second, the producer can choose to produce less than its actual capacity by curtailing production when profitable for its market portfolio. Both strategic conducts can be claimed to be forecast errors. The latter can also be disguised as a production failure or an unexpected outage.

[21] City Power Marketing LLC and K. Stephen Tsingas, Docket No. IN15-5-000, Paragraph 271, 152 FERC Paragraph 272.

Nevertheless, the incentives to manipulate the market differ across designs. The main difference lies in the opportunity cost of reneging: the supplier will either forego the exogenous fixed price, or the market price. Remark that both forms of market design can accommodate penalties to curb incentives for such extensive conduct.

Large diversified incumbents should typically be more prone to such conduct under both designs than small specialized producers. Moreover, the market-integration of small specialized firms, such as fringe wind farms, may have unsuspected positive aspects. Those are demonstrated by the empirical analysis of the Iberian electricity market of Ito and Reguant (2016). The authors show that while dominant firms tend to withhold output – of both conventional and wind power plants – from the day-ahead market to generate a premium, the competitive wind producers counteract this premium formation by arbitraging between the sequential markets. This arbitrage is possible because wind power producers can sell more than their actual forecasts in the day-ahead market and buy back the difference in the real-time market. This analysis suggests that a dominant firm may find profitable to under-forecast its wind output in the day-ahead market so as to have the option to either increase or decrease its output sold in the real-time market to benefit its own portfolio position.

As shown in this paper, the penalty scheme for preventing last minute deviations must depend on the net market position. This is however not sufficient to alter the dominant firm's incentives to withhold its wind output from the day-ahead market during high price periods. The regulators should be able to evaluate the firm's forecasts quality and to impose penalties based on their net market position.[22] On the other hand, the arbitrage service provided by fringe wind farms illustrates what constitutes an acceptable – and even desirable – form of residual demand manipulations in sequential markets. Sound penalties must hence also depend on whether the conduct increases the total market costs.

[22] A procedure in line of that for preventing demand under-forecasters may be necessary. The centralization of forecasts by the market operator can provide a solution to reduce the productive and environmental inefficiencies generated by renewable power under-production.

1.5 Conclusion

We develop an analytical framework to study strategic reneging as a manipulative conduct in sequential markets under imperfect competition. By being able to renege on its prior commitments upon observing the demand realization, the firm can weaken competition, exacerbate its market power, and increase its probability to be pivotal. Our results suggest that the equilibrium supply strategy may be affected by the ability to renege on its forward commitments. In addition, manipulations are found to become more likely as the level of demand approaches the overall capacity limit.

Strategic reneging leads to higher profits for the manipulator through possibly both an inflated market-clearing price and a larger volume of market sales. On the other hand, it creates inefficiencies by allocating more costly resources to replace the output withdrawn through reneging.

The form of market manipulations considered in this paper is only one among a variety of possible extensive conducts. Yet, it allows to underlines the common economic mechanism behind manipulative schemes in the general framework of Ledgerwood and Carpenter (2012). Our analysis yields important insights to identify potential misconducts in this context.

Nevertheless, our theoretical framework neglects three important aspects. First, we focus on a monopolistic market whereas having multiple strategic players could bring additional insights. Second, the forward positions are taken as exogenous. Finally, we neglect a dynamic component of strategic reneging which is the additional value for the firm to learn about the elasticities of demand and supply.[23]

[23] "... it's clear we're learning a ton." (MSA, 2015).

Chapter 2

Functional linear regression with functional response*

"Functional data analysis has a long historical shadow, extending at least back to the attempts of Gauss and Legendre to estimate a comet's trajectory."

Ramsay and Silverman (2005).

*This chapter is co-authored with my advisor Marine Carrasco (Université de Montréal), and Jean-Pierre Florens (Toulouse School of Economics). It has been published in the Journal of Econometrics: https://doi.org/10.1016/j.jeconom.2017.08.008. The authors thank the editors, two referees, Stéphane Bonhomme, Samuele Centorrino, Jan Johannes, and the participants of the 6th French Econometrics Conférence, CESG 2016, and the Workshop on Inverse Problems (Heidelberg, 2016) for helpful comments. Carrasco gratefully acknowledges financial support from NSERC. The data and MATLAB codes can be found at https://davidbenatia.com.

2.1 Introduction

With the increase of storage capability, continuous time data are available in many fields including finance, medicine, meteorology, and microeconometrics. Researchers, companies, and governments look for ways to exploit this rich information. In this paper, we develop new estimation results for functional regressions where both the regressor $Z(t)$ and the response $Y(t)$ are functions of an index such as the time or a spatial location. Both $Z(t)$ and $Y(t)$ are assumed to belong to Hilbert spaces. The model can be thought as a generalization of the multivariate regression where the regression coefficient is now an unknown operator Π. An interesting feature of our model is that $Y(t)$ depends not only on contemporaneous $Z(t)$ but also on past and future values of Z.

We propose to estimate the operator Π by Tikhonov regularization, which amounts to apply a penalty on the L^2 norm of Π. The choice of a L^2 penalty, instead of L^1 used in Lasso, is motivated by the fact that - in the applications we have in mind - there is no reason to believe that the relationship between Y and Z is sparse. We derive the rate of convergence of the mean-square error (MSE) and the asymptotic distribution of the estimator for a fixed regularization parameter α and develop tests on Π. In some applications, it would be interesting to test whether $Y(t)$ depends only on the past values of Z or only on contemporaneous values of Z. If the application is on network and t refers to the spatial location, our model could describe how the behavior of a firm $Y(t)$ depends on the decision of neighboring firms $Z(s)$. Testing properties of Π will help to characterize the strategic response of firms.

Often, the full trajectories are not observed but only a discretized version is available. This case raises specific challenges which will be addressed in the scenario where the data become more and more frequent (infill asymptotics).

We also consider the case where Z is endogenous and instrumental variables are used to estimate Π. To the best of our knowledge, the model with functional response and endogenous functional regressor has never been studied before. We derive an estimator based on Tikhonov regularization and show its rate of convergence.

There is a large body of work done on linear functional regression where the response is a scalar variable Y and the regressor is a function. Some recent references include Cardot, Ferraty, and Sarda (2003), Hall, Horowitz et al. (2007), Horowitz and Lee (2007), Darolles, Fan, Florens, and Renault (2011), Crambes, Kneip, and Sarda (2009), and Florens and Van Bellegem

(2015). In contrast, only a few researchers have tackled the functional linear regression in which both the predictor Z and the response Y are random functions. The object of interest is the estimation of the conditional expectation of Y given Z. In this setting, the unknown parameter is an integral operator. This model is discussed in the monographs by Ramsay and Silverman (2005), Ferraty and Vieu (2006), and Horváth and Kokoszka (2012). Cuevas, Febrero, and Fraiman (2002) consider a fixed design setting and propose an estimator of Π based on interpolation. Yao, Müller, Wang et al. (2005) consider the case where both predictor and response trajectories are observed at discrete and irregularly spaced times. Their estimator is based on nonparametric estimators of the principal components. Park and Qian (2012) use functional principal components to estimate a regression where both the response and independent variables are densities. Crambes and Mas (2013) consider also a spectral cut-off regularized inverse and derive the asymptotic mean square prediction error which is then used to derive the optimal choice of the regularization parameter. Our model is also related to the functional autoregressive (FAR) model studied by Bosq (2000) , Kargin and Onatski (2008), and Aue, Norinho, and Hormann (2015), among others. The estimation methods used in these papers are based on functional principal components and differ from ours. Antoch et al. (2010) use a FAR to forecast the electricity consumption. In their model, the weekday consumption curve is explained by the curve from the previous week. The authors use B-spline to estimate the operator.

Our contribution is to develop an estimator which can be expressed as products of matrices and vectors. It does not require choosing a basis or estimating eigenfunctions. It involves only one smoothing parameter. We derive the rate of convergence of our estimator under some source condition which is common in the inverse problem literature. Interestingly, we do not need to impose any restriction on the multiplicity of eigenvalues.

The paper is organized as follows. Section 2.2 introduces the model and the estimators. Section 2.3 derives the rate of convergence of the MSE. Section 2.4 presents the asymptotic normality of the estimator for a fixed regularization parameter. Issues relative to the choice of the regularization parameter are discussed in Section 2.5. Discrete observations are addressed in Section 2.6. Section 2.7 considers an endogenous regressor. Section 2.8 presents simulation results. Section 2.9 presents an application to the electricity market where the dependent variable is the electricity consumption and the independent variable is the temperature. The proofs are collected in

Appendix. Data and Matlab programs used in this paper are available at: https://davidbenatia.com/.

2.2 The model and estimator

2.2.1 The model

We consider a regression model where both the predictor and response are random functions. We observe pairs of random trajectories (y_i, z_i) $i = 1, 2, ..., n$ with square integrable predictor trajectories z_i and response trajectories y_i. They are realizations of random processes (Y, Z) with zero mean functions and unknown covariance operators. The extension to the case, where the mean is unknown but estimated, is straightforward. The arguments of Y and Z are denoted t which may refer to the time, a location or a characteristic such as the age or income of an agent.

We assume that Y belongs to a separable real Hilbert space $\mathcal{E}$ equipped with an inner product $\langle , \rangle$ and Z belongs to a separable real Hilbert space $\mathcal{F}$ equipped with an inner product $\langle , \rangle$ (to simplify notations, we use the same notation for both inner products even though they usually differ).

The model is

$$Y = \Pi Z + U \tag{2.1}$$

where U is a zero mean random element of $\mathcal{E}$ and Π is a nonrandom Hilbert-Schmidt[1] operator from $\mathcal{F}$ to $\mathcal{E}$. The regressor Z is said to be exogenous if $cov\,(Z, U) = 0$. Z will be assumed exogenous in Sections 2 to 6. This assumption will be relaxed in Section 2.7.

For illustration, consider the following example

$$\begin{aligned} \mathcal{E} &= \left\{ g : \int_{\mathcal{S}} g\,(t)^2\, dt < \infty \right\}, \\ \mathcal{F} &= \left\{ f : \int_{\mathcal{T}} f\,(t)^2\, dt < \infty \right\} \end{aligned}$$

where $\mathcal{S}$ and $\mathcal{T}$ are some intervals of $\mathbb{R}$. An operator Π from $\mathcal{E}$ to $\mathcal{F}$ is a Hilbert-Schmidt if and only if it admits a representation as an integral operator such that

$$(\Pi \varphi)\,(s) = \int_{\mathcal{T}} \pi\,(s, t)\, \varphi\,(t)\, dt$$

[1] K is Hilbert-Schmidt if $\sum_j \langle K\phi_j, K\phi_j \rangle < \infty$ for any basis (ϕ_j).

for any $\varphi \in \mathcal{F}$ and

$$\int_S \int_{\mathcal{T}} |\pi(s,t)|^2 \, dtds < \infty.$$

π is referred to as the kernel of the operator Π. Note that our theory covers the case where $\mathcal{S}$ and $\mathcal{T}$ are subsets of $\mathbb{R}^p$ for $p > 1$ so that the index s and t may be vectors. Moreover, we are not limited to L^2 spaces.

Consider the case where $\mathcal{E}$ and $\mathcal{F}$ are Sobolev spaces:

$$\begin{aligned}
\mathcal{E} &= \left\{ g \in \mathcal{C}^1(\mathcal{S}) : \int_{\mathcal{S}} g(t)^2 \, dt + \int_{\mathcal{S}} g(t)'^2 \, dt < \infty \right\}, \\
\mathcal{F} &= \left\{ f \in \mathcal{C}^1(\mathcal{T}) : \int_{\mathcal{T}} f(t)^2 \, dt + \int_{\mathcal{S}} f(t)'^2 \, dt < \infty \right\}
\end{aligned}$$

where $\mathcal{C}^1(\mathcal{S})$ denotes the space of all complex valued functions defined on $\mathcal{S}$ with continuous first derivatives. $\mathcal{E}$ is a Hilbert space with inner product $\langle g_1, g_2 \rangle = \int_{\mathcal{S}} (g_1(t) g_2(t) + g_1'(t) g_2'(t)) \, dt$. The same is true for $\mathcal{F}$. Let Π be an integral operator from $\mathcal{F}$ to $\mathcal{E}$, $(\Pi\varphi)(s) = \int_{\mathcal{T}} \pi(s,t) \varphi(t) \, dt$ for all $\varphi \in \mathcal{F}$. Then Π is a Hilbert-Schmidt operator if and only if

$$\int_S \int_{\mathcal{T}} \left(|\pi|^2 + \left| \frac{\partial \pi}{\partial s} \right|^2 + \left| \frac{\partial \pi}{\partial t} \right|^2 + \left| \frac{\partial^2 \pi}{\partial s \partial t} \right|^2 \right) dtds < \infty.$$

Our theory covers the case of Sobolev spaces, the main difference with L^2 spaces is the way the inner products are defined.

The main model we have in mind is the model where Π is an integral operator and (2.1) takes the form:

$$Y(s) = \int_{\mathcal{T}} \pi(s,t) Z(t) \, dt + U(s).$$

In this model, $Y(s)$ depends not only on $Z(s)$ but also on all the $Z(t)$, for $t \neq s$. The object of interest is the estimation of the operator Π using a panel of observations (Y_i, Z_i), $i = 1, 2, ..., n$. Below, we describe three examples of applications of Model (2.1).

Example of application 1. Electricity market.

Let $Y_i(t)$ be the electricity consumption for the province of Ontario in Canada for day i at hour t, and $Z_i(t)$ be the average temperature for the same province for day i at hour t. Recent research (McLaughlin et al., 2011;

Yu et al., 2013) reveals that the optimal consumption path of electricity over one day may depend not only on the current temperature but also on past and future temperatures. We could model this relationship as a multivariate regression where the dependent variable is the 24×1 vector of electricity consumptions at different hours of the day and the independent variable is a 72×1 vector of temperatures. Estimating this model by ordinary least squares (OLS) would require inverting a 72×72 matrix which would likely be near singular. The variance of the resulting OLS estimator would be very large. Moreover, if the frequency of the observations increases, it is expected that the successive observations will become more and more correlated. It is therefore natural to assume that the observations result from the discretization of a curve. Model (2.1) seems to be an attractive alternative to the multivariate regression in this setting. This application will be continued in Section 9.

Example of application 2. Technological spillovers in productivity.

There has been an increase interest in the economic literature on quantifying interactions and spillovers among firms (see Manresa, 2013 and references therein). Assume that $Y(s)$ is the log of the average output of firms with industry code s and $Z(t)$ is the log of the average R&D expenditures for firms with industry code t. Model (2.1) would permit to characterize how firms from one sector benefit from R&D advancements done in adjacent sectors. From a practical point of view, note that industry code goes from 1 to 6 digits. A higher number of digits correspond to a thinner grid on the continuum of industry codes. To obtain an index between 0 and 1, we normalize the industry codes t by dividing them by the largest existing industry code with the same number of digits. As in Manresa (2013), we suggest using repeated observations over time to estimate the model, so that the index i corresponds to a time period. This example will not be pursued here.

Example of application 3. Risk neutral density[2].

The risk neutral density (RND) plays a crucial role in the pricing of financial derivatives. Let S_t denote the value of an asset at time t. The price $g_t(K,T)$ of a European option with the strike price K and the maturity T is equal to the expected present value of the future payoff $g(S_T, K)$ under the risk-neutral density:

$$g_t(K,T) = \frac{1}{R_{t,T}} \int_0^\infty g(S_T, K) f(S_T, T; S_t, t) \, dS_T \tag{2.2}$$

[2]We thank a referee for suggesting this example.

where $f(S_T, T; S_t, t)$ is the conditional RND and $R_{t,T} = e^{\int_t^T r_s ds}$ where r_t is the risk-free rate. For a European call option, we have

$$\begin{aligned} g_t(K,T) &= C_t(K,T), \\ g(S_T,K) &= \max(S_T - K, 0). \end{aligned}$$

It is well known that (2.2) does not hold exactly because of measurement errors and the incompleteness of the market (see Gourieroux and Jasiak, 2001, p.322), therefore we add an error term to (2.2) and obtain an equation equivalent to (1). Using a set of call prices with different strike prices K and maturity times T, we can write (2.2) as

$$Y_t(T, S_t) = \int_0^\infty \pi(T, S_t, s) Z(s) ds + U(T, S_t) \tag{2.3}$$

where $Y_t(T, S_t) = R_{t,T}C_t(K,T), \pi(T, S_t, s) = f(s, T; S_t, t), Z(s) = \max(s - K, 0).$ To estimate this model, one could use the same data as in Ait-Sahalia and Lo (1998). They focus on S&P 500 index options traded at the Chicago Board Options Exchange. They use price data on every option traded during 1993, which results in a sample of 14,431 options after cleaning the data. By observing so many options and maturity times, we should be able to recover most of the supports of S_t and T. Note that the first index in π is bidimensional (T, S_t). Our theory covers this case. Here $Z(s)$ is deterministic, so we are in a fixed design setting. The error term U may be correlated across observations but this issue is not addressed in the present paper[3].

Many papers have proposed nonparametric estimators of the RND before, among others Jackwerth and Rubinstein (1996), Ait-Sahalia and Lo (1998), Garcia and Gençay (2000), Y. et al. (2001), and Bondarenko (2003). Most of these papers start with the relation

$$f(s, T; S_t, t) = \frac{1}{R_{t,T}} \frac{\partial^2 C_t}{\partial x^2}\bigg|_{x=S_T}$$

and then rely on some smoothing. Here, we propose to use regularization to deal with the ill-posedness of the inverse problem (2.2) directly.

[3] Another issue is that the data may not be stationary because the stock price S_t is not stationary. One way to make it stationary is to divide by the strike price K, see Garcia et al. (2010).

2.2.2 The estimator

We denote V_Z the operator from $\mathcal{F}$ to $\mathcal{F}$ which associates to functions $\varphi \in \mathcal{F}$:

$$V_Z\varphi = E\left[Z\left\langle Z,\varphi\right\rangle\right].$$

Note that, as Z is centered, V_Z is the covariance operator of Z. We denote C_{YZ} the covariance operator of (Y,Z). It is the operator from $\mathcal{F}$ to $\mathcal{E}$ such that

$$C_{YZ}\varphi = E\left[Y\left\langle Z,\varphi\right\rangle\right]$$

Using (2.1), we have

$$\begin{aligned} cov\left(Y,Z\right) &= cov\left(\Pi Z+U,Z\right) \\ &= \Pi cov\left(Z,Z\right)+cov\left(U,Z\right). \end{aligned}$$

Hence, we have the following relationships:

$$C_{YZ} = \Pi V_Z, \tag{2.4}$$

$$C_{ZY} = V_Z\Pi^* \tag{2.5}$$

where Π^* is the adjoint of Π and C_{ZY} is defined as the operator from $\mathcal{E}$ to $\mathcal{F}$ such that

$$C_{ZY}\psi = E\left[Z\left\langle Y,\psi\right\rangle\right]$$

for any ψ in $\mathcal{E}$. Note that C_{ZY} is the adjoint of C_{YZ}, C_{YZ}^*.

First we describe how to estimate Π^* using (2.5). The unknown operators V_Z and C_{ZY} are replaced by their sample counterparts. The sample estimate of V_Z is

$$\hat{V}_Z\varphi = \frac{1}{n}\sum_{i=1}^{n} z_i\left\langle z_i,\varphi\right\rangle$$

for $\varphi \in \mathcal{F}$. The sample estimate of C_{ZY} is

$$\hat{C}_{ZY}\psi = \frac{1}{n}\sum_{i=1}^{n} z_i\left\langle y_i,\psi\right\rangle$$

for $\psi \in \mathcal{E}$. An estimator of Π^* cannot be obtained directly by solving $\widehat{C}_{ZY} = \widehat{V}_Z\Pi^*$ because the initial equation $C_{ZY} = V_Z\Pi^*$ is an ill-posed problem in the sense that V_Z is invertible only on a subset of $\mathcal{E}$ and its inverse is not continuous. Note that $\hat{V}_Z$ has finite rank equal to n and hence is not invertible. A Moore-Penrose generalized inverse would not help because it would not be continuous. To stabilize the inverse, we need to use some regularization scheme. We adopt Tikhonov regularization (see Kress, 1999 and Carrasco et al., 2007).

The estimator of Π^* is defined as

$$\hat{\Pi}^*_\alpha = \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZY} \tag{2.6}$$

and that of Π is defined by

$$\hat{\Pi}_\alpha = \hat{C}_{YZ} \left(\alpha I + \hat{V}_Z\right)^{-1} \tag{2.7}$$

where α is some positive regularization parameter which will be allowed to converge to zero as n goes to infinity. The estimators (2.6) and (2.7) can be viewed as generalization of ordinary least-squares estimators. They also have an interpretation as the solution to an inverse problem.

2.2.3 Link with inverse problems

At this stage, it is useful to make the link with the inverse problem literature. Let $\mathcal{H}$ be the Hilbert space of linear Hilbert-Schmidt operators from $\mathcal{F}$ to $\mathcal{E}$. The inner product on $\mathcal{H}$ (see Pedersen, 1989) is

$$\langle \Pi_1, \Pi_2 \rangle_{\mathcal{H}} = tr\left(\Pi_1^* \Pi_2\right) = tr\left(\Pi_2^* \Pi_1\right).$$

Dropping the error term in (2.1), we obtain, for the sample, the equation

$$\hat{r} = K\Pi \tag{2.8}$$

where $\hat{r} = (y_1, ..., y_n)'$ and K is the operator from $\mathcal{H}$ to $\mathcal{E}^n$ such that $K\Pi = (\Pi z_1, ..., \Pi z_n)'$. The inner product on $\mathcal{E}^n$ is

$$\langle f, g \rangle_{\mathcal{E}^n} = \frac{1}{n} \sum_{i=1}^n \langle f_i, g_i \rangle_{\mathcal{E}}$$

with $f = (f_1, .., f_n)'$ and $g = (g_1, ..., g_n)'$. Let us check that $\hat{\Pi}_\alpha$ is a classical Tikhonov regularized inverse of the operator K:

$$\hat{\Pi}_\alpha = \left(\alpha I + K^* K\right)^{-1} K^* \hat{r}.$$

We need to find K^*. We look for the operator B from $\mathcal{F}$ to $\mathcal{E}$ solution of

$$\langle K\Pi, f\rangle_{\mathcal{E}^n} = \langle \Pi, B\rangle_{\mathcal{H}}. \tag{2.9}$$

Note that

$$\begin{aligned}\langle \Pi, B\rangle_{\mathcal{H}} &= tr\left(B^*\Pi\right) \\ &= \sum_j \langle B^*\Pi\varphi_j, \varphi_j\rangle_{\mathcal{F}} \\ &= \sum_j \langle \Pi\varphi_j, B\varphi_j\rangle_{\mathcal{E}}\end{aligned}$$

where (φ_j) is an orthonormal basis of $\mathcal{F}$. On the other hand,

$$\langle K\Pi, f\rangle_{\mathcal{E}^n} = \frac{1}{n}\sum_i \langle \Pi z_i, f_i\rangle_{\mathcal{E}}.$$

Using $z_i = \sum_j \langle z_i, \varphi_j\rangle \varphi_j$, we obtain

$$\begin{aligned}\langle K\Pi, f\rangle_{\mathcal{E}^n} &= \frac{1}{n}\sum_i \left\langle \sum_j \langle z_i, \varphi_j\rangle \Pi\varphi_j, f_i\right\rangle_{\mathcal{E}} \\ &= \sum_j \left\langle \Pi\varphi_j, \frac{1}{n}\sum_i \langle z_i, \varphi_j\rangle f_i\right\rangle_{\mathcal{E}}.\end{aligned}$$

It follows from (2.9) that $B\varphi_j = \frac{1}{n}\sum_i \langle z_i, \varphi_j\rangle f_i$ for all j and hence

$$B\varphi = \left(K^* f\right)\varphi = \frac{1}{n}\sum_i \langle z_i, \varphi\rangle f_i.$$

for all φ in $\mathcal{E}$.

Now, we have

$$K^*K\Pi = \frac{1}{n}\sum_i \langle z_i, \varphi\rangle \Pi z_i = \Pi\widehat{V}_Z$$

and

$$K^*\hat{r} = \frac{1}{n}\sum_i \langle z_i, .\rangle y_i = \hat{C}_{YZ}.$$

It follows that

$$\begin{aligned}\hat{\Pi}_\alpha &= \left(\alpha I + K^*K\right)^{-1} K^*\hat{r} \\ &= \hat{C}_{YZ}\left(\alpha I + \hat{V}_Z\right)^{-1}.\end{aligned}$$

In summary, we established that $\hat{\Pi}_\alpha$ is the Tikhonov regularized inverse of K in Equation (2.8).

The estimator $\hat{\Pi}_\alpha$ is also a penalized least-squares estimator:

$$\begin{aligned}\hat{\Pi}_\alpha &= \arg\min_{\Pi} \sum_{i=1}^{n} \|y_i - \Pi z_i\|^2 + \alpha \|\Pi\|_{HS}^2 \\ &= \arg\min_{\Pi} \sum_{i=1}^{n} \|y_i - \Pi z_i\|^2 + \alpha \sum \tilde{\mu}_j^2\end{aligned}$$

where $\tilde{\mu}_j$ are the singular values of the operator Π.

2.2.4 Identification

It is easier to study identification from the viewpoint of Equation (2.5). Let $\mathcal{H}$ be the space of Hilbert-Schmidt operators from $\mathcal{E}$ to $\mathcal{F}$. Let T be the operator from $\mathcal{H}$ to $\mathcal{H}$ defined as

$$TH = V_Z H \text{ for } H \text{ in } \mathcal{H}.$$

According to (2.5), Π^* is identified if and only if T is injective.

V_Z injective implies T injective. Indeed, we have

$$\begin{aligned} TH &= 0 \\ &\Leftrightarrow V_Z H = 0 \\ &\Leftrightarrow V_Z H\psi = 0, \forall \psi \\ &\Leftrightarrow H\psi = 0, \forall \psi \end{aligned}$$

by the injectivity of V_Z. Hence $H = 0$. It turns out that T is injective if and only if V_Z is injective. This can be shown by deriving the spectrum of T.

First, we show that T is self-adjoint. The adjoint T^* of T satisfies

$$\langle TH, K\rangle = \langle H, T^* K\rangle$$

for arbitrary operators H and K of $\mathcal{H}$. We have

$$\begin{aligned}\langle TH, K\rangle &= tr\left(\left(TH\right)^* K\right) \\ &= tr\left(\left(V_Z H\right)^* K\right) \\ &= tr\left(H^* V_Z K\right).\end{aligned}$$

Hence, $T^*K = V_Z K = TK$. Therefore, T is self-adjoint.

The spectrum of T is also closely related to that of V_Z. Let $(\mu_j,\, H_j)_{j=1,2....}$ denote the eigenvalues and eigenfunctions of T and $(\lambda_j, \varphi_j)_{j=1,2,...}$ be the eigenvalues and eigenfunctions of V_Z so that $V_Z\varphi_j = \lambda_j\varphi_j$. H_j is necessarily of the form, $H_j = \varphi_j \langle \iota, . \rangle$ where ι is the 1 function in $\mathcal{E}$. Then,

$$\begin{aligned} TH_j &= V_Z\varphi_j \langle \iota, . \rangle \\ &= \lambda_j\varphi_j \langle \iota, . \rangle \\ &= \lambda_j H_j. \end{aligned}$$

So that the eigenvalues of T are the same as those of V_Z.

In summary, a necessary and sufficient condition for the identification of Π is that V_Z is injective. When V_Z is not injective, our estimator will typically converge to $\Pi_{N^\perp}$, the projection of Π on the orthogonal to the null space of T. This case is discussed in Section 2.3.

An example of non injective V_Z is given by an integral operator with degenerate kernel:

$$(V_Z\varphi)(s) = \int_0^1 v(s,t)\,\varphi(t)\,dt$$

where

$$v(s,t) = \sum_{j=1}^{J} a_j(s)\, b_j(t)$$

for some finite J and a_j linearly independent. Then, φ belongs to the null space of V_Z if and only if $\int_0^1 b_j(t)\,\varphi(t)\,dt = 0$. 0 is an eigenvalue of V_Z associated with the elements of the null space of V_Z, hence V_Z is not injective.

2.2.5 Computation of the estimator

To show how to compute $\hat{\Pi}_\alpha^*$ explicitly, we multiply the left and right hand sides of (2.6) by $\left(\alpha I + \hat{V}_Z\right)$ to obtain

$$\begin{aligned} \hat{C}_{ZY}\psi &= \left(\alpha I + \hat{V}_Z\right)\hat{\Pi}_\alpha^*\psi \Leftrightarrow \\ \frac{1}{n}\sum_{i=1}^{n} z_i \langle y_i, \psi \rangle &= \alpha\hat{\Pi}_\alpha^*\psi + \frac{1}{n}\sum_{i=1}^{n} z_i \left\langle z_i, \hat{\Pi}_\alpha^*\psi \right\rangle. \end{aligned} \tag{2.10}$$

Then, we take the inner product with z_l, $l = 1, 2, ..., n$ on the left and right hand side of (2.10), to obtain n equations:

$$\frac{1}{n}\sum_{i=1}^{n}\langle z_l, z_i\rangle \langle y_i, \psi\rangle = \alpha \left\langle z_l, \hat{\Pi}_\alpha^* \psi\right\rangle + \frac{1}{n}\sum_{i=1}^{n}\langle z_l, z_i\rangle \left\langle z_i, \hat{\Pi}_\alpha^* \psi\right\rangle, \; l = 1, 2, ..., n, \tag{2.11}$$

with n unknowns $\left\langle z_i, \hat{\Pi}_\alpha^* \psi\right\rangle$, $i = 1, 2, ..., n$. Let M be the $n \times n$ matrix with (l, i) element $\langle z_l, z_i\rangle / n$, v the $n-$vector of $\left\langle z_i, \hat{\Pi}_\alpha^* \psi\right\rangle$ and w the $n-$vector of $\langle y_i, \psi\rangle$. (2.11) is equivalent to

$$Mw = (\alpha I + M)\, v.$$

And $v = (\alpha I + M)^{-1} Mw = M (\alpha I + M)^{-1} w$. For a given ψ, we can compute

$$\begin{aligned} \hat{\Pi}_\alpha^* \psi &= \frac{1}{\alpha n}\sum_{i=1}^{n} z_i \left(\langle y_i, \psi\rangle - \left\langle z_i, \hat{\Pi}_\alpha^* \psi\right\rangle\right) \\ &= \frac{1}{\alpha n}\underline{z}' \left(I - M (\alpha I + M)^{-1}\right) w \\ &= \frac{1}{n}\underline{z}' (\alpha I + M)^{-1} w \end{aligned} \tag{2.12}$$

where $\underline{z}$ is the $n-$vector of z_i.

Now, we explain how to estimate $\Pi\varphi$ for any $\varphi \in \mathcal{F}$. Taking the inner product with φ in the left and right hand sides of (2.12), we obtain

$$\begin{aligned} \left\langle \varphi, \hat{\Pi}_\alpha^* \psi\right\rangle &= \frac{1}{\alpha n}\sum_{i=1}^{n}\langle \varphi, z_i\rangle \left(\langle y_i, \psi\rangle - \left\langle z_i, \hat{\Pi}_\alpha^* \psi\right\rangle\right) \Leftrightarrow \\ \left\langle \hat{\Pi}_\alpha \varphi, \psi\right\rangle &= \frac{1}{\alpha n}\sum_{i=1}^{n}\langle \varphi, z_i\rangle \left\langle y_i - \hat{\Pi}_\alpha z_i, \psi\right\rangle \end{aligned}$$

for all $\psi \in \mathcal{E}$. This implies

$$\hat{\Pi}_\alpha \varphi = \frac{1}{\alpha n}\sum_{i=1}^{n}\langle \varphi, z_i\rangle \left(y_i - \hat{\Pi}_\alpha z_i\right). \tag{2.13}$$

Hence, to compute $\hat{\Pi}_\alpha \varphi$, we need to know $\hat{\Pi}_\alpha z_i$. From (2.7), we have

$$\alpha \hat{\Pi}_\alpha + \hat{\Pi}_\alpha \hat{V}_Z = \hat{C}_{YZ}.$$

Applying the l.h.s and r.h.s to z_i, $i = 1, 2, ..., n$, we obtain

$$\begin{aligned} \alpha\hat{\Pi}_\alpha z_i + \hat{\Pi}_\alpha \hat{V}_Z z_i &= \hat{C}_{YZ} z_i \Leftrightarrow \\ \alpha\left(\hat{\Pi}_\alpha z_i\right)(t) + \frac{1}{n}\sum_{j=1}^{n}\left(\hat{\Pi}_\alpha z_j\right)(t)\left\langle z_j, z_i\right\rangle &= \frac{1}{n}\sum_{j=1}^{n} y_j(t)\left\langle z_j, z_i\right\rangle, \; i = 1, 2, ..., n \end{aligned} \tag{2.14}$$

For each t, we can solve the n equations with n unknowns $\left(\hat{\Pi}_\alpha z_j\right)(t)$ given by (2.14) and deduct $\hat{\Pi}_\alpha \varphi$ from (2.13). Let us denote $\underline{y}(t)$, $\underline{z}(t)$, $\underline{\varsigma}(t)$ the $n \times 1$ vectors with ith element $y_i(t)$, $z_i(t)$, and $\hat{\Pi}_\alpha z_i(t)$ respectively. The solution $\underline{\varsigma}(t)$ of Equation (2.14) is equal to

$$\underline{\varsigma}(t) = (\alpha I + M)^{-1} M \underline{y}(t).$$

It follows from (2.13) that $\hat{\Pi}_\alpha$ is an integral operator with degenerate kernel equal to

$$\hat{\pi}_\alpha(s, t) = \frac{1}{n}\underline{y}(s)'(\alpha I + M)^{-1}\underline{z}(t). \tag{2.15}$$

Note that this calculation is exact, we did not rely on any discretization. We see that to compute the kernel of $\hat{\Pi}_\alpha$, we only need to compute the elements $\langle z_l, z_i\rangle / n$ of the matrix M.

The prediction of Y_i is given by

$$\hat{y}_i = \hat{\Pi}_\alpha z_i.$$

2.2.6 Alternative estimators

In this section, we briefly expose the existing alternative estimators of the operator Π in a functional linear regression with functional response. In these papers, the operator kernel is defined slightly differently from ours. To fix the notation, we consider the estimation of β in the following model

$$Y(t) = \int \beta(s, t) Z(s)\, ds + U(s). \tag{2.16}$$

In our notation, $\Pi(t, s) = \beta(s, t)$.

Ramsay and Silverman (2005) propose two ways for estimating β. Both methods rely on an approximation of β using basis functions $\{\eta_j : j \geq 1\}$ and

$\{\theta_l : l \geq 1\}$ which could be for instance spline or Fourier bases. In the first version, β is estimated by

$$\beta_{JL}(s,t) = \sum_{j=1}^{J} \sum_{l=1}^{L} b_{jl} \eta_j(s) \theta_l(t) = \eta(s)' B\theta(t)$$

for some small numbers J and L. Truncating the η basis permits to avoid over-fitting whereas truncating the θ basis insures that the prediction is smooth. The estimator of b_{jl} is obtained by minimizing

$$||y - \int z(s) \beta_{JL}(s,.) \, ds||^2 = \int \sum_{i=1}^{n} \left[y_i(t) - \int z_i(s) \beta_{JL}(s,.) \, ds \right]^2 dt.$$

In the second approach proposed by Ramsay and Silverman (2005), J and L are chosen to be large and smoothness is ensured by adding some roughness penalties to the objective function:

$$\min_{\beta} ||y - \int z(s) \beta_{JL}(s,.) \, ds||^2 + \lambda_s \int \int [L_s \beta_{JL}(s,t)]^2 \, ds dt + \lambda_t \int \int [L_t \beta_{JL}(s,t)]^2 \, ds dt$$

where λ_s and λ_t are the penalization parameters and L_k denotes a differential operator to be applied to $\beta(s,t)$ with respect to $k \in \{s,t\}$ only. Ramsay and Silverman (2005) use cross-validation methods for choosing the penalization parameters λ_s and λ_t.

Instead of using arbitrary bases $\{\eta_j\}$, $\{\theta_l\}$, a popular approach consists in using the functional principal components which correspond to the eigenfunctions of the covariance operators V_Z and V_Y of Z and Y respectively. Let ψ_j and ϕ_l be the eigenfunctions of V_Z and V_Y respectively. Let λ_j be the eigenvalues of V_Z. Estimators $\hat{\lambda}_j$, $\hat{\psi}_j$ and $\hat{\phi}_l$ of λ_j, ψ_j and ϕ_l respectively are given by the eigenvalues and eigenfunctions of the sample counterparts of V_Z and V_Y. Then, an estimator of $\beta(s,t)$ is given by

$$\hat{\beta}(s,t) = \sum_{j=1}^{J} \sum_{l=1}^{L} \frac{\hat{\rho}_{jl}}{\hat{\lambda}_j} \hat{\psi}_j(s) \hat{\phi}_l(t) \tag{2.17}$$

where

$$\hat{\rho}_{jl} = \int \int \hat{C}_{ZY}(s,t) \hat{\psi}_j(s) \hat{\phi}_l(t) \, ds dt$$

and $\hat{C}_{ZY}(s,t)$ is an estimator of $E(Z(s)Y(t))$. The simplest estimator is given by the empirical covariance

$$\hat{C}_{ZY}(s,t) = \frac{1}{n}\sum_{i=1}^{n} Z_i(s)\, Y_i(t).$$

The resulting kernel is

$$\hat{\beta}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J} \frac{\left\langle Z_i, \hat{\psi}_j \right\rangle}{\hat{\lambda}_j} \hat{\psi}_j(s) \sum_{l=1}^{L} \left\langle Y_i, \hat{\phi}_l \right\rangle \hat{\phi}_l(t). \tag{2.18}$$

Yao et al. (2005) rely on scatterplot smoothing to compute $\hat{\rho}_{jl}$ in (2.17) and use leave-one-out cross-validation to select the number of included eigenfunctions J and L. Park and Qian (2012) and Crambes and Mas (2013) use a slightly different estimator:

$$\hat{\beta}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J} \frac{\left\langle Z_i, \hat{\psi}_j \right\rangle}{\hat{\lambda}_j} \hat{\psi}_j(s)\, Y_i(t).$$

For comparison, using Equation (2.6), our estimator of β can be written as

$$\hat{\beta}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\left\langle Z_i, \hat{\psi}_j \right\rangle}{\hat{\lambda}_j + \alpha} \hat{\psi}_j(s)\, Y_i(t).$$

So the only differences are that the sum over j is not truncated and $\hat{\lambda}_j$ in the denominator is replaced by $\hat{\lambda}_j + \alpha$. Crambes and Mas (2013) derive the rate of convergence of their estimator under the assumption that λ_j is a convex function of j.

Bosq (2000) considers a model similar to (2.16) where the regressor is the lagged value of the dependent variable. Let Z_i be a stationary process in a Hilbert space. The autoregressive model of order one writes

$$Z_{i+1}(t) = \int \beta(s,t)\, Z_i(s)\, ds + U_i(t). \tag{2.19}$$

Bosq (2000) proposes an estimator of the form (2.18) and establishes its consistency under the assumption that all eigenvalues are distinct. Its rate of convergence depends on the difference between successive eigenvalues.

In summary, our estimator presents some advantages compared to existing estimators: (a) It involves only one smoothing parameter α. (b) Computing our estimator using formula (2.15) is simple and does not require the estimation of eigenfunctions. (c) In Section 2.3, we derive the properties of our estimator under a source condition which implicitly imposes some restrictions on the decay rate of eigenvalues but does not rule out multiple eigenvalues. (d) The extension in Section 2.7 allows for endogenous regressors.

2.3 Rate of convergence of the MSE

In this section, we study the rate of convergence of the mean square error (MSE) of $\hat{\Pi}_\alpha^*$. Several assumptions are needed.

Assumption 1. *The observations* (U_i, Y_i, Z_i), $i = 1, 2, ..., n$ *are independent, identically distributed as* (U, Y, Z) *where* Y *and* U *are random processes of a separable Hilbert space* $\mathcal{E}$ *and* Z *is a random process of a separable Hilbert space* $\mathcal{F}$. *Moreover,* $E(U_i) = 0$, $cov(U_i, Z_i) = 0$, $cov(U_i, U_j | Z_1, Z_2, ..., Z_n) = 0$ *for all* $i \neq j$ *and* $= V_U$ *for* $i = j$ *where* V_U *is a nonrandom trace-class operator.*

Assumption 2. Π *belongs to* $\mathcal{H}(\mathcal{F}, \mathcal{E})$ *the space of Hilbert-Schmidt operators.*

Assumption 3. V_Z *is a trace-class operator and* $\left\| \hat{V}_Z - V_Z \right\|_{HS}^2 = O_p(1/n)$.

Assumption 4. *There is a Hilbert-Schmidt operator* R *from* $\mathcal{E}$ *to* $\mathcal{F}$ *and a constant* $\beta > 0$ *such that* $\Pi^* = V_Z^{\beta/2} R$.

Assumption 1 imposes that U_i is homoskedastic and Z_i is exogenous. The stronger condition of strict exogeneity, $E(U_i | Z_i) = 0$, is not needed here. It will be used only briefly in Section 4.

An operator K is trace-class if $\sum_j \langle K\phi_j, \phi_j \rangle < \infty$ for any basis (ϕ_j). If K is self-adjoint positive definite, it is equivalent to say that the sum of the eigenvalues of K is finite. Given V_U is a covariance operator, V_U is trace-class if and only if $E\left(\|U_i\|^2\right) < \infty$.

The notation $\|\|_{HS}$ refers to the Hilbert-Schmidt norm of operators, $\|K\|_{HS}^2 \equiv \sum_j \langle K\phi_j, K\phi_j \rangle < \infty$ for any basis (ϕ_j). An operator K is Hilbert-Schmidt (noted HS) if $\|K\|_{HS}^2 < \infty$. If K is self-adjoint positive definite, the condition $\|K\|_{HS}^2 < \infty$ is equivalent to the condition that the eigenvalues of K are square summable. A sufficient condition for $\left\|\hat{V}_Z - V_Z\right\|_{HS}^2 = O_p\left(1/n\right)$ is that Z_i is a i.i.d. random process and $E\left(\|Z_i\|^4\right) < \infty$, see Proposition 5 of Dauxois, Pousse, and Romain (1982).

Assumption 4 is a source condition needed to characterize the rate of convergence of the MSE. Moreover, it guarantees that Π^* belongs to the orthogonal of the null space of V_Z denoted $\mathcal{N}\left(V_Z\right)$. Given this condition, there is no need to impose $\mathcal{N}\left(V_Z\right) = \{0\}$ to get the identification.

The source condition can be rewritten in terms of the spectral decomposition of V_Z. Let (λ_j, φ_j) be the eigenvalues (ordered in decreasing order) and eigenfunctions of V_z. The source condition of Assumption 4 is equivalent to assume that

$$\sum_{j=1}^{\infty} \frac{\langle \Pi^* \varphi, \varphi_j \rangle^2}{\lambda_j^{\beta}} < \infty \text{ for all } \varphi \in \mathcal{E}.$$

As V_Z is a compact operator, λ_j go to zero as j goes to infinity. So this condition imposes that the eigenvalues of λ_j decline to zero not too fast relatively to the Fourier coefficients $\langle \Pi^* \varphi, \varphi_j \rangle$.

Because V_Z is a compact integral operator, its inverse is a differential operator. Hence, the larger β in Assumption 4, the smoother the operator Π^* is. To illustrate the connection between Assumption 4 and smoothness conditions, let us consider two simple examples.

1. Example 1. Let $\mathcal{E} = \mathcal{F} = L^2\left[0,1\right]$ and $\left(V_Z \varphi\right)(s) = \int_0^1 \min(s,t)\varphi\left(t\right) dt$. If $\beta = 2$, then the source condition of Assumption 4 is satisfied if and only if for all $\varphi \in L^2\left[0,1\right]$, $\left(\Pi^* \varphi\right)(0) = 0$, $\left.\frac{\partial (\Pi^* \varphi)(t)}{\partial t}\right|_{t=1} = 0$ and $\left(\Pi^* \varphi\right)(t)$ is twice differentiable with respect to t.

2. Example 2. We consider now the case where V_Z has a gaussian kernel, that is $\left(V_Z \varphi\right)(s) = \int_{-\infty}^{\infty} \exp\left\{-\frac{(s-t)^2}{\sigma^2}\right\} \varphi\left(t\right) dt$ and assume $\beta = 2$. In this case, $\left(\Pi^* \varphi\right)(t)$ is infinitely differentiable with respect to t for all $\varphi \in \mathcal{E}$.

The conditional MSE is defined by

$$MSE = E\left(\left\|\hat{\Pi}_\alpha - \Pi\right\|_{HS}^2 | Z_1, .., Z_n\right).$$

Our definition of the MSE makes sense only if $\hat{\Pi}_\alpha$ is a Hilbert-Schmidt operator. We establish this property in the following proposition.

Proposition 1. *Under Assumptions 1 and 3,* $\hat{\Pi}_\alpha$ *belongs to* $\mathcal{H}(\mathcal{F}, \mathcal{E})$ *for all* $\alpha > 0$.

Now we turn our attention to the rate of convergence of the MSE.

Proposition 2. *Assume Assumptions 1 to 4 hold.*
If $\beta > 1$*, then* $MSE = O_p\left(\frac{1}{n\alpha} + \alpha^{\beta\wedge 2}\right)$.
If $\beta < 1$*, then* $MSE = O_p\left(\frac{\alpha^\beta}{n\alpha^2} + \alpha^\beta\right)$.

Remark 3.1. Proposition 2 shows that the MSE exhibits the usual trade-off between the variance decreasing in α and the bias increasing in α. Taking the optimal α which equates the rates of the squared bias and variance, we obtain the following rates:

- for $\beta > 1$, $\alpha \sim n^{-1/(1+\beta\wedge 2)}$, $MSE \sim n^{-\beta\wedge 2/(1+\beta\wedge 2)}$,
- for $\beta < 1$, $n\alpha^2 \to \infty$, $MSE \sim \alpha^\beta$.

Remark 3.2. We see that the rate of convergence of the bias does not improve when $\beta > 2$. This is due to the well-known saturation effect of Tikhonov regularization (see for instance, Engl, Hanke, and Neubauer, 1996). This rate could be improved by using iterated Tikhonov regularization.

Remark 3.3. Proposition 2 does not impose any restriction on the multiplicity of the eigenvalues. This is a major difference with the estimators based on the truncation. The consistency proof for these estimators requires that the estimated eigenvalues are consistent estimators of the true eigenvalue. When the order of multiplicity of an eigenvalue is greater than one, there is no such consistent estimator. On the other hand, our proof does not rely on the spectral decomposition of the operator. In that sense, Tikhonov regularization is more robust to situations where eigenvalues are multiple than methods based on the truncation.

Now we consider the case where the operator T defined in Section 2.2.4 is not injective and Π is not identified. We can decompose the operator Π as $\Pi_N + \Pi_{N^\perp}$ where $\Pi_N \in \mathcal{N}(T)$ and $\Pi_{N^\perp} \in \mathcal{N}(T)^\perp$. $\Pi_{N^\perp}$ corresponds to the minimal least squares solution of the inverse problem. We will show below that under appropriate assumptions, $\hat{\Pi}_\alpha$ converges to $\Pi_{N^\perp}$. To derive this result, we replace Assumption 4 by the following assumption where Π^* is replaced by $\Pi^*_{N^\perp}$.

Assumption 4'. There is a Hilbert-Schmidt operator R from $\mathcal{E}$ to $\mathcal{F}$ and a constant $\beta > 0$ such that $\Pi^*_{N^\perp} = V_Z^{\beta/2} R$.

Proposition 3. *Assume Assumptions 1-3 and 4' hold.*
$E\left(\left\|\hat{\Pi}_\alpha - \Pi_{N^\perp}\right\|^2_{HS} |Z_1, .., Z_n\right) = O_p\left(\frac{1}{n\alpha^2} + \alpha^{\beta\wedge 2}\right).$

Remark 3.4. Under our assumptions, Π can not be estimated consistently because there is no information about Π_N in the data. However, $\Pi_{N^\perp}$ is consistently estimated.

Remark 3.5. Observe that the lack of identification results in a larger bias and hence a slower rate of convergence of the MSE. Here, the optimal α satisfies $\alpha \sim n^{-1/(2+\beta\wedge 2)}$ and the rate for the MSE is $n^{-\beta\wedge 2/(2+\beta\wedge 2)}$. Florens, Johannes, and Van Bellegem (2011) studied the effect of nonidentification in the context of nonparametric instrumental regression and found a reduced rate of convergence as we found here.

Remark 3.6. The saturation effect of Tikhonov persists in the non identified case.

2.4 Asymptotic normality for fixed α and tests

In this section, we will derive the asymptotic distribution of $\hat{\Pi}^*_\alpha$ when α is fixed and then develop a test for the null hypothesis $H_0 : \Pi = \Pi_0$ where Π_0 is some known operator.

When α is fixed, $\hat{\Pi}^*_\alpha$ is not consistent and keeps an asymptotic bias. It is useful to define Π^*_α the regularized version of Π^* :

$$\Pi^*_\alpha = (\alpha I + V_Z)^{-1} V_Z \Pi^*.$$

We have

$$\begin{aligned}
\hat{\Pi}_\alpha^* - \Pi_\alpha^* &= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZU} \\
&\quad + \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z \Pi^* - (\alpha I + V_Z)^{-1} V_Z \Pi^* \\
&= (\alpha I + V_Z)^{-1} \hat{C}_{ZU} \\
&\quad + \left[\left(\alpha I + \hat{V}_Z\right)^{-1} - (\alpha I + V_Z)^{-1}\right] \hat{C}_{ZU} \\
&\quad + \alpha \left(\alpha I + \hat{V}_Z\right)^{-1} \left(V_Z - \hat{V}_Z\right) (\alpha I + V_Z)^{-1} \Pi^* \\
&= (\alpha I + V_Z)^{-1} \hat{C}_{ZU} \qquad (2.20) \\
&\quad + \alpha (\alpha I + V_Z)^{-1} \left(\hat{V}_Z - V_Z\right) (\alpha I + V_Z)^{-1} \Pi^* \qquad (2.21) \\
&\quad + O_p\left(\frac{1}{n}\right).
\end{aligned}$$

As n goes to infinity, $\hat{\Pi}_\alpha^* - \Pi_\alpha^*$ converges to zero and is $\sqrt{n}-$asymptotically normal. The first two terms of the r.h.s are $O_p\left(1/\sqrt{n}\right)$ and will affect the asymptotic distribution. This distribution is not simple. We are going to characterize it below. The notation $\otimes$ denotes the functional tensor product defined as $(x \otimes y)(f) = \langle x, f \rangle y$.

From Equations (2.20) and (2.21), neglecting the $O_p(1/n)$ term, we have

$$\begin{aligned}
\hat{\Pi}_\alpha^\star - \Pi_\alpha^\star &= (\alpha I + V_Z)^{-1} \hat{C}_{ZU} + \alpha(\alpha I + V_Z)^{-1}(\hat{V}_Z - V_Z)(\alpha I + V_Z)^{-1}\Pi^\star \\
&= (\alpha I + V_Z)^{-1}\frac{1}{n}\sum_i (u_i \otimes z_i) + \alpha(\alpha I + V_Z)^{-1}\frac{1}{n}\sum_i (z_i \otimes z_i - V_Z)(\alpha I + V_Z)^{-1}\Pi^\star \\
&= \frac{1}{n}\sum_i \left(u_i \otimes (\alpha I + V_Z)^{-1} z_i + \alpha \Pi (\alpha I + V_Z)^{-1} z_i \otimes (\alpha I + V_Z)^{-1} z_i\right) \\
&\quad - \alpha(\alpha I + V_Z)^{-1} V_Z (\alpha I + V_Z)^{-1}\Pi^\star \\
&= \frac{1}{n}\sum_i \left(u_i \otimes (\alpha I + V_Z)^{-1} z_i + \alpha \Pi (\alpha I + V_Z)^{-1} z_i \otimes (\alpha I + V_Z)^{-1} z_i\right) \\
&\quad - \alpha \mathrm{E}[\Pi(\alpha I + V_Z)^{-1} Z \otimes (\alpha I + V_Z)^{-1} Z] \\
&= \frac{1}{n}\sum_i \left(u_i \otimes \tilde{z}_i + \alpha \Pi \tilde{z}_i \otimes \tilde{z}_i - \alpha \mathrm{E}[\Pi \tilde{Z} \otimes \tilde{Z}]\right) \\
&= \frac{1}{n}\sum_i \left((u_i + \alpha \Pi \tilde{z}_i) \otimes \tilde{z}_i - \alpha \mathrm{C}_{\tilde{Z}\Pi\tilde{Z}}\right),
\end{aligned}$$

where $\tilde{Z} \equiv (\alpha I + V_Z)^{-1}Z$ and $\tilde{z}_i \equiv (\alpha I + V_Z)^{-1}z_i$. The first equality makes use of the definition of the empirical covariance operators using tensor products. The second line uses the elementary properties $K(Y \otimes X) = Y \otimes KX$ and $(Y \otimes X)K = K^\star Y \otimes X$ for $X \in \mathcal{F}, X \in \mathcal{E}$ and $K \in \mathcal{H}$. The third line uses the definition of V_Z. The interchange of the expectation operator and $(\alpha I + V_Z)^{-1}$ is allowed since the latter is a bounded linear operator. By Banach inverse theorem, the inverse of a bounded linear operator is itself linear and bounded (see, for instance, Rudin, 1991). The last equality holds since the functional tensor product distributes over addition.

The covariance operator of $\hat{\Pi}^\star_\alpha - \Pi^\star_\alpha$ is an operator which maps the space of Hilbert-Schmidt operators from $\mathcal{E}$ to $\mathcal{F}$, denoted $\mathcal{G}$, into itself. Such an operator may be difficult to write explicitly. Fortunately, the properties of tensor products of infinite-dimensional Hilbert-Schmidt operators defined on separable Hilbert spaces are well-known,[4] and may be used like in Dauxois et al. (1982) to write explicitly the covariance operator of an infinite-dimensional Hilbert-Schmidt random operator. The tensor product $\Pi_1 \tilde{\otimes} \Pi_2$ for $(\Pi_1, \Pi_2) \in \mathcal{G}^2$ is a mapping from $\mathcal{G}$ into itself, hence $\Pi_1 \tilde{\otimes} \Pi_2$ is an element of the Hilbert space of Hilbert-Schmidt operators from $\mathcal{G}$ to $\mathcal{G}$ equipped with the Hilbert-Schmidt inner product. For $T = \varphi \otimes \psi \in \mathcal{G}$, $\Pi_1 = X \otimes Z \in \mathcal{G}$ and $\Pi_2 = Y \otimes W \in \mathcal{G}$, this tensor product is equivalently defined as :

(i) $(\Pi_1 \tilde{\otimes} \Pi_2)T = \langle T, \Pi_1 \rangle_{\mathcal{G}} \Pi_2 \in \mathcal{G}$

(ii) $\Big((X \otimes Z) \tilde{\otimes} (Y \otimes W)\Big)(\varphi \otimes \psi) = \Big((X \otimes Y)\varphi\Big) \otimes \Big((Z \otimes W)\psi\Big),$
$\forall \varphi, X, Y \in \mathcal{E}, \psi, Z, W \in \mathcal{F},$

Based upon definition (i), the covariance operator of Π_1 and Π_2 naturally writes as

$$E\Big[\langle \cdot, \Pi_1 - \mathrm{E}[\Pi_1] \rangle_{\mathcal{G}} \Big(\Pi_2 - \mathrm{E}[\Pi_2]\Big)\Big] = \mathrm{E}\Big[\Big(\Pi_1 - \mathrm{E}[\Pi_1]\Big) \tilde{\otimes} \Big(\Pi_2 - \mathrm{E}[\Pi_2]\Big)\Big]. \quad (2.22)$$

Furthermore, to show asymptotic normality, we shall use the classical central limit theorem for i.i.d. processes in separable Hilbert spaces. The following is stated as Theorem 2.7 in Bosq (2000) and is reproduced here for clarity.

[4] See, for instance, (Vilenkin, 1968) p.59-65.

Theorem 1. *(Bosq, 2000) Let* $(Z_i, i \geq 1)$ *be a sequence of i.i.d.* $\mathcal{F}$*-valued random variables, where* $\mathcal{F}$ *is a separable Hilbert space, such that* $E\|Z_i\|^2 < \infty$*,* $E(Z_i) = \bar{Z}$ *and* $V_Z = V$*, then one has*

$$\frac{1}{\sqrt{n}}\sum_i (Z_i - \overline{Z}) \xrightarrow{d} \mathcal{N}(0, V),$$

We are now geared to derive the asymptotic covariance operator of interest in its general form, under some standard assumptions.

Proposition 4. *Assume* (U_i, Z_i) *i.i.d.,* $\Omega_\alpha < \infty$*,* $E\|Z_i\|^4 < \infty$*,* $E\|U_i\|^2\|Z_i\|^2 < \infty$*, then*

$$\sqrt{n}(\hat{\Pi}^\star_\alpha - \Pi^\star_\alpha) \xrightarrow{d} \mathcal{N}(0, \Omega_\alpha), \tag{2.23}$$

where the asymptotic covariance operator[5] Ω_α *for fixed* α *is given by*

$$\Omega_\alpha = E\Bigg[\Bigg((U + \alpha\Pi\tilde{Z}) \otimes \tilde{Z}\Bigg)\tilde{\otimes}\Bigg((U + \alpha\Pi\tilde{Z}) \otimes \tilde{Z}\Bigg)\Bigg] - \alpha^2 C_{\tilde{Z}\Pi\tilde{Z}}\tilde{\otimes}C_{\tilde{Z}\Pi\tilde{Z}}, \tag{2.24}$$

which simplifies to

$$\Omega_0 = E\Bigg[\Big(U \otimes V_Z^{-1}Z\Big)\tilde{\otimes}\Big(U \otimes V_Z^{-1}Z\Big)\Bigg], \tag{2.25}$$

when $\alpha \to 0$*.*

Remarks. The results of Proposition 4 are derived without imposing the strict exogeneity assumption, $E[U_i|Z_i] = 0$. If this assumption is satisfied, the asymptotic covariance operator in (2.24) simplifies into

$$\Omega_\alpha = \mathrm{E}\Bigg[\Bigg(U\otimes\tilde{Z}\Bigg)\tilde{\otimes}\Bigg(U\otimes\tilde{Z}\Bigg)\Bigg] + \alpha^2\mathrm{E}\Bigg[\Bigg(\Pi\tilde{Z}\otimes\tilde{Z}\Bigg)\tilde{\otimes}\Bigg(\Pi\tilde{Z}\otimes\tilde{Z}\Bigg)\Bigg] - \alpha^2 C_{\tilde{Z}\Pi\tilde{Z}}\tilde{\otimes}C_{\tilde{Z}\Pi\tilde{Z}}.$$

In econometrics, we are often interested in testing the significance of estimates and produce confidence bands. However, there is no obvious meaningful way to perform standard significance tests using the derived asymptotic

[5]The kernel of Ω_α can be written as

$$\omega_\alpha(s,t,r,\tau) = \mathrm{E}\Big[\big(U(s)+\alpha\Pi\tilde{Z}(s)\big)Z(t)\big(U(r)+\alpha\Pi\tilde{Z}(r)\big)Z(\tau)\big)\Big] - \alpha^2\mathrm{E}\Big[\Pi\tilde{Z}(s)\tilde{Z}(t)\Big]\mathrm{E}\Big[\Pi\tilde{Z}(r)\tilde{Z}(\tau)\Big],$$

covariance. Indeed, for fixed α the estimated residuals will be biased and one must specify $\Pi^\star$. On the other hand, if we assume $\alpha \to 0$, an estimator of (2.25) may be uninformative since V_Z^{-1} does not necessarily exist. A more practical approach would be to keep α fixed to obtain an estimate of $(\alpha I + V_Z)^{-1}$ and use it to derive an estimator of (2.25). Other statistical tests may involve applying a test operator to Ω_α.

We want to test the null hypothesis: $H_0 : \Pi = \Pi_0$ where Π_0 is known. A simple way to test this hypothesis is to look at $\hat{C}_{ZY} - \hat{V}_Z \Pi_0^*$. Under H_0, this operator equals $\hat{C}_{ZU}$ and should be close to zero. Moreover, under H_0,

$$\sqrt{n}\left(\hat{C}_{ZY} - \hat{V}_Z \Pi_0^*\right) \xrightarrow{d} \mathcal{N}\left(0, K_{ZU}\right)$$

where

$$K_{ZU} = E\left[(u \otimes Z)\, \widetilde{\otimes}\, (u \otimes Z)\right]$$

(see Dauxois et al., 1982).

Let $\{f_j, g_j : j = 1, 2, ..., q\}$ be a set of test functions, then

$$\begin{bmatrix} \sqrt{n}\left\langle \left(\hat{C}_{ZY} - \hat{V}_Z \Pi_0^*\right) f_1, g_1 \right\rangle \\ \vdots \\ \sqrt{n}\left\langle \left(\hat{C}_{ZY} - \hat{V}_Z \Pi_0^*\right) f_q, g_q \right\rangle \end{bmatrix}$$

converges to a multivariate normal distribution with mean 0_q and covariance matrix the $q \times q$ matrix Σ with (j, l) element:

$$\begin{aligned} \Sigma_{jl} &= E\left[\left\langle \sqrt{n}\hat{C}_{ZU} f_j, g_j \right\rangle \left\langle \sqrt{n}\hat{C}_{ZU} f_l, g_l \right\rangle\right] \\ &= \langle g_j, V_Z g_l \rangle \langle f_j, V_U f_l \rangle \end{aligned}$$

where the second equality follows from (U_i, Z_i) iid and the homoskedasticity assumption given in Assumption 1. This covariance matrix can be easily estimated by replacing V_Z and V_U by their sample counterpart. The appropriately rescaled quadratic form converges to a chi-square distribution with q degrees of freedom which can be used to test H_0. The test functions could be cumulative normals as in Conley et al. (1997) or could be normal densities with same small variance but centered at different means. Note that here we assumed Π_0 completely specified under H_0. If instead, Π_0 depends on some unknown parameters θ which need to be estimated, the asymptotic variance Σ needs to be adjusted to take into account the estimation error.

2.5 Data-driven selection of α

The estimator involves a tuning parameter, α, which needs to be selected. It can be chosen as the solution to

$$\min_{\alpha} \frac{1}{\alpha} \left\| \hat{V}_Z \hat{\Pi}_{\alpha}^{*} - \hat{C}_{ZY} \right\|_{HS}^{2},$$

see Engl et al. (1996) p.102 or

$$\min_{\alpha} tr\left(\hat{\Pi}_{\alpha} \hat{\Pi}_{\alpha}^{*}\right) \left\| \hat{V}_Z \hat{\Pi}_{\alpha}^{*} - \hat{C}_{ZY} \right\|_{HS}^{2},$$

see Engl et al. (1996) Proposition 4.37 where $\hat{\Pi}_{\alpha}^{*}$ in $\left\| \hat{V}_Z \hat{\Pi}_{\alpha}^{*} - \hat{C}_{ZY} \right\|_{HS}^{2}$ is obtained by iterated Tikhonov.

Another possibility is to use leave-one-out cross-validation

$$\min_{\alpha} \frac{1}{n} \sum_{j} \left\| y_i - \hat{\Pi}_{\alpha}^{(-i)} z_i \right\|^{2}$$

where $\hat{\Pi}_{\alpha}^{(-i)}$ has been computed using all observations except for the ith one. Centorrino (2016) studies the properties of the leave-one-out cross-validation for nonparametric IV regression and shows that this criterion is rate optimal in mean squared error. This method is also used in a binary response model by Centorrino and Florens (2015). Various data-driven selection techniques are compared via simulations in Centorrino et al. (2017).

An alternative approach would be to use a penalized minimum contrast criterion as in Goldenshluger et al. (2011). This could lead to a minimax-optimal estimator (Comte and Johannes, 2012).

2.6 Discrete observations

2.6.1 Effect of discretization

In this section, to simplify the exposition, we will refer to the arguments of (y_i, z_i), t, as time even though it could refer to a location or other characteristic. Suppose that the data (y_i, z_i) are not observed in continuous time but at discrete (not necessarily equally spaced) times within a fixed time span $\mathcal{T} \subset \mathbb{R}$ (for example $\mathcal{T} = [0, 1]$). It is necessary to construct pairs of curves (y_i^m, z_i^m), $i = 1, 2, ..., n$ such that $y_i^m \in \mathcal{E}$ and $z_i^m \in \mathcal{F}$, that can be evaluated at any desired t in the domain $\mathcal{T}$. These approximate curves can be obtained with some interpolation or smoothing procedure using step functions, kernel smoothing or splines for instance (see Ramsay and Silverman, 2005).[6]

Let the subscript m correspond to the smallest number of discrete measurements across $i = 1, 2, ..., n$ which defines the sampled domain $\mathcal{T}_m$ of the smoothed data $(y_i^m, z_i^m)_{i=1}^n$. We consider infill asymptotics where m grows with the sample size n (m stands for $m(n)$). That is, we assume that for all $t \in \mathcal{T}$, $Pr(t \in \mathcal{T}_m) \to 1$ as $n \to \infty$ hence $\mathcal{T}_m \to \mathcal{T}$ as $n \to \infty$. Intuitively, it means that new discrete measurements become available at values of t between the existing ones, as the sample size increases, in such a way that the sampled domain becomes dense and coincides with the true domain at the limit.

Using the smoothed observations, we compute the corresponding estimators of V_Z and C_{ZY} denoted $\hat{V}_Z^m$, $\hat{C}_{ZY}^m$ and the estimator of Π^* denoted $\hat{\Pi}_\alpha^{m*}$:

$$\hat{\Pi}_\alpha^{m*} = \left(\alpha I + \hat{V}_Z^m\right)^{-1} \hat{C}_{ZY}^m.$$

To assess the rate of convergence of $\hat{\Pi}_\alpha^{m*}$, we add the following conditions which guarantee that the discretization error is negligible with respect to the estimation error.

Assumption 5. $\|z_i^m - z_i\| = O_p(f(m))$ *and* $\|y_i^m - y_i\| = O_p(f(m))$.

Assumption 6. *We have*

$$\frac{f(m)}{\alpha n} = o\left(\alpha^{\beta \wedge 2}\right)$$

as n *and* $m = m(n)$ *go to infinity.*

Proposition 5. *Under Assumptions 1 to 6, the MSE of* $\hat{\Pi}_\alpha^{m*} - \Pi^*$ *has the same rate of convergence as that of the MSE of* $\hat{\Pi}_\alpha^* - \Pi^*$ *in Proposition 2 as* n *and* $m = m(n)$ *go to infinity.*

This proposition shows that using discrete observations does not affect the rate of convergence of the estimators as long as the discretization error is negligible with respect to the estimation error.

[6]Interpolation is used when the discrete measurements along the curves are errorless, whereas smoothing may be used to remove some observational errors.

2.6.2 Relation with the Ridge estimator

In order to clarify some of the merits of using our functional estimator, we show how it relates to the standard ridge estimator. Suppose we have at hands n pairs of discretized $L^2[0,1]$ functions (y_i^m, z_i^m), for $i \in \{1, 2, ..., n\}$, each evaluated at m equispaced points $t = \frac{1}{m}, \frac{2}{m}, ..., 1$. Let $\underline{z}$ and $\underline{y}$ be the $n \times m$ matrices with (i,j) element $\left(z_i\left(\frac{j}{m}\right)\right)$ and $\left(y_i\left(\frac{j}{m}\right)\right)$, $i = 1, ..., n$, $j = 1, ...m$, respectively. From (2.15), the kernel of $\hat{\Pi}_\alpha$ can be written as

$$\hat{\pi}_\alpha = \underline{z}'(\alpha n I_n + nM)^{-1}\underline{y}, \tag{2.26}$$

If the $n \times n$ matrix M with (l,i) element $\frac{\langle z_l, z_i\rangle}{n} = \frac{1}{n}\int_0^1 z_l(t)z_i(t)dt$ is approached by the estimator $\hat{M} = \frac{\underline{z}\underline{z}'}{n}$ with (l,i) element $\frac{1}{nm}\sum_{j=1}^{m} z_l(\frac{j}{m})z_i(\frac{j}{m})$ $\forall l, i \in \{1, 2, ..., n\}$, we obtain the approximation defined by

$$\tilde{\hat{\pi}}_\alpha = \underline{z}'\left(\alpha n I_n + \underline{z}\underline{z}'\right)^{-1}\underline{y}. \tag{2.27}$$

Let ϕ_j denote the eigenvectors of $\underline{z}'\underline{z}$ associated with eigenvalues λ_j^2 and ψ_j denote the eigenvectors of $\underline{z}\underline{z}'$ associated with eigenvalues λ_j^2. For any vector v of dimension n, we have $v = \sum_j \langle v, \psi_j\rangle\psi_j$ and hence $\underline{z}'v = \sum_j \lambda_j\langle v, \psi_j\rangle\phi_j$ since $\underline{z}'\psi_j = \lambda_j\phi_j$. Therefore, we can write

$$(\alpha n I_m + \underline{z}'\underline{z})^{-1}\underline{z}'v = \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha n}\langle v, \psi_j\rangle\phi_j, \tag{2.28}$$

and

$$\begin{aligned}\underline{z}'(\alpha n I_n + \underline{z}\underline{z}')^{-1}v &= \sum_j \frac{1}{\lambda_j^2 + \alpha n}\langle v, \psi_j\rangle\underline{z}'\psi_j \\ &= \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha n}\langle v, \psi_j\rangle\phi_j.\end{aligned} \tag{2.29}$$

Combining (2.28) and (2.29) yields the identity

$$\underline{z}'\left(\alpha n I_n + \underline{z}\underline{z}'\right)^{-1} = (\alpha n I_m + \underline{z}'\underline{z})^{-1}\underline{z}', \tag{2.30}$$

which allows to rewrite (2.26) as the standard ridge estimator defined as

$$\hat{\pi}_\alpha^r = (\alpha n I_m + \underline{z}'\underline{z})^{-1}\underline{z}'\underline{y}. \tag{2.31}$$

Therefore, for L^2 functions, the ridge estimator (2.31) corresponds to an approximation of the functional estimator (2.26) based on $\hat{M} = \frac{\underline{z}\underline{z}'}{n}$. The key difference is the inversion of an $m \times m$ matrix, rather than an $n \times n$ matrix for the functional estimator. As the number of discrete measurements m increases, the required inversion of the $m \times m$ matrix $(\alpha n I_m + \underline{z}'\underline{z})$ in (2.31) becomes increasingly more computationally demanding, whereas the computation of the functional estimator remains unaltered. In other words, our estimator enables a more precise and flexible approximation of matrix M,[7] and is faster to compute as soon as $m > n$.

2.7 Case where Z is endogenous

It is frequent in economics that the regressor Z is endogenous. In that case, the least squares estimator studied in Section 2 is not consistent. In this section, we propose an estimator based on instrumental variables (IV) W that satisfy $\text{cov}(U, W) = 0$. For illustration, assume we want to estimate the price elasticity of electricity. To do so, we use as dependent variable Y the log consumption of electricity and as independent variable Z the log price of electricity. For big consumers (i.e. big firms), the electricity price might be endogenous. A reliable instrument would be given by W, the wind speed. Indeed wind influences the price because a stronger wind increases the electricity production of wind turbines but can be considered as exogenous because one does not control wind.

Under the assumption, $\text{cov}(U, W) = 0$, it follows that

$$C_{YW} = \Pi C_{ZW} \tag{2.32}$$

where $C_{YW} = E\left(Y \langle W, . \rangle\right)$ and $C_{ZW} = E\left(Z \langle W, . \rangle\right)$. Similarly, we have

$$C_{WY} = C_{WZ}\Pi^* \tag{2.33}$$

where $C_{WZ} = E\left(W \langle Z, . \rangle\right)$.

We need the following identification condition:

Assumption 1′. *The observations (U_i, W_i, Y_i, Z_i), $i = 1, 2, ..., n$ are independent, identically distributed as (U, W, Y, Z) where U, W, and Y are random processes of a separable Hilbert space $\mathcal{E}$ and Z is a random process*

[7] In the MATLAB programs, we implement a fast numerical method to construct M based on the trapezoidal rule.

of a separable Hilbert space $\mathcal{F}$. Moreover, $E(U_i) = 0$, $cov(U_i, W_i) = 0$, $cov(U_i, U_j | Z_1, ..., Z_n, W_1, ..., W_n) = 0$ for all $i \neq j$ and $= V_U$ for $i = j$ where V_U is a nonrandom trace-class operator.

Assumption 7. *C_{WZ} is injective.*

Assumption 1' imposes that the instruments W_i are exogenous.

Under Assumption 7, Π is uniquely defined from (2.32). To see this, assume that there are two solutions Π_1 and Π_2 to (2.32). It follows that $(\Pi_1 - \Pi_2) C_{ZW} = 0$ or equivalently $C_{WZ}(\Pi_1^* - \Pi_2^*) = 0$. Hence the range of $(\Pi_1^* - \Pi_2^*)$ belongs to the null space of C_{WZ}. However, under Assumption 7, the null space of C_{WZ} is reduced to zero and thus the range of $(\Pi_1^* - \Pi_2^*)$ is equal to zero. It follows that $\Pi_1^*\varphi - \Pi_2^*\varphi = 0$ for all φ, hence $\Pi_1^* = \Pi_2^*$.

To construct an estimator of Π^*, we first apply the operator C_{ZW} on the l.h.s and r.h.s of Equation (2.33) to obtain

$$C_{ZW}C_{WY} = C_{ZW}C_{WZ}\Pi^*.$$

Note that $C_{ZW} = C_{WZ}^*$ and therefore the operator $C_{ZW}C_{WZ}$ is self-adjoint. The operators C_{ZW}, C_{WZ}, and C_{WY} can be estimated by their sample counterparts. The estimator of Π^* is defined by

$$\hat{\Pi}_\alpha^* = \left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1} \hat{C}_{ZW}\hat{C}_{WY}. \tag{2.34}$$

Similarly, the estimator of Π is given by

$$\hat{\Pi}_\alpha = \hat{C}_{YW}\hat{C}_{WZ}\left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}.$$

Now, we explain how to compute $\hat{\Pi}_\alpha^*$ in practice. From (2.34), we have

$$\left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)\hat{\Pi}_\alpha^*\psi = \hat{C}_{ZW}\hat{C}_{WY}\psi.$$

Note that

$$\begin{aligned}
\hat{C}_{ZW}\hat{C}_{WY}\psi &= \frac{1}{n^2}\sum_{i,j}\langle y_j, \psi\rangle \langle w_i, w_j\rangle z_i, \\
\hat{C}_{ZW}\hat{C}_{WZ}\hat{\Pi}_\alpha^*\psi &= \frac{1}{n^2}\sum_{i,j}\left\langle z_j, \hat{\Pi}_\alpha^*\psi\right\rangle \langle w_i, w_j\rangle z_i.
\end{aligned}$$

Taking the inner product with z_l yields n equations

$$\begin{aligned} & \alpha \left\langle z_l, \hat{\Pi}_\alpha^* \psi \right\rangle + \frac{1}{n^2} \sum_{i,j} \left\langle z_j, \hat{\Pi}_\alpha^* \psi \right\rangle \langle w_i, w_j \rangle \langle z_l, z_i \rangle \\ = & \frac{1}{n^2} \sum_{i,j} \langle y_j, \psi \rangle \langle w_i, w_j \rangle \langle z_l, z_i \rangle, \; l = 1, 2, ..., n \end{aligned}$$

with n unknowns $\left\langle z_j, \hat{\Pi}_\alpha^* \psi \right\rangle$, $j = 1, 2, ..., n$. Let v be the $n-$ vector of $\left\langle z_j, \hat{\Pi}_\alpha^* \psi \right\rangle$, y_ψ be the $n-$ vector of $\langle y_j, \psi \rangle$ and Q be the $n \times n$ matrix with (l, j) element equal to $\frac{1}{n^2} \sum_i \langle w_i, w_j \rangle \langle z_l, z_i \rangle$. We obtain $v = (\alpha I + Q)^{-1} Q y_\psi = Q (\alpha I + Q)^{-1} y_\psi$.

Hence, for each ψ, $\hat{\Pi}_\alpha^* \psi$ can be computed from

$$\begin{aligned} \hat{\Pi}_\alpha^* \psi &= \frac{1}{\alpha} \left[\hat{C}_{ZW} \hat{C}_{WY} \psi - \hat{C}_{ZW} \hat{C}_{WZ} \hat{\Pi}_\alpha^* \psi \right] \\ &= y_\psi' (\alpha I + Q)^{-1} \xi \end{aligned}$$

where ξ is the $n-$ vector with ith element $\frac{1}{n^2} \sum_j \langle w_i, w_j \rangle z_i$. The computation of $\hat{\Pi}_\alpha \varphi$ can be done using the same approach as in Section 2.2.

Assumption 8. *$C_{ZW} C_{WZ}$ is a trace-class operator and* $\left\| \hat{C}_{ZW} \hat{C}_{WZ} - C_{ZW} C_{WZ} \right\|_{HS}^2 = O_p (1/n)$.

Assumption 9. *There is a Hilbert-Schmidt operator R from $\mathcal{E}$ to $\mathcal{F}$ and a constant $\beta > 0$ such that $\Pi^* = (C_{ZW} C_{WZ})^{\beta/2} R$.*

Similar to Assumption 4, Assumption 9 is a source condition which permits to characterize the rate of convergence of $\hat{\Pi}_\alpha^*$. It also guarantees that Π^* belongs to the orthogonal of the null space of C_{WZ} and hence, under Assumption 9, Assumption 7 is not needed to identify Π^*.

We decompose $\hat{\Pi}_\alpha^* - \Pi^*$ in the following manner:

$$\begin{aligned} & \hat{\Pi}_\alpha^* - \Pi^* \\ = & \left(\alpha I + \hat{C}_{ZW} \hat{C}_{WZ} \right)^{-1} \hat{C}_{ZW} \hat{C}_{WY} - \Pi^* \\ = & \left(\alpha I + \hat{C}_{ZW} \hat{C}_{WZ} \right)^{-1} \hat{C}_{ZW} \hat{C}_{WU} \qquad (2.35) \\ & + \left(\alpha I + \hat{C}_{ZW} \hat{C}_{WZ} \right)^{-1} \hat{C}_{ZW} \hat{C}_{WZ} \Pi^* - (\alpha I + C_{ZW} C_{WZ})^{-1} C_{ZW} C_{WZ} \Pi^* \qquad (2.36) \\ & + (\alpha I + C_{ZW} C_{WZ})^{-1} C_{ZW} C_{WZ} \Pi^* - \Pi^*. \qquad (2.37) \end{aligned}$$

Proposition 6. *Under Assumptions 1', 2, 8, and 9, the MSE of* $\hat{\Pi}_{\alpha}^{*} - \Pi^{*}$ *has the same rate of convergence as in Proposition 2.*

Remark 7.1. Proposition 6 shows that our estimator based on instrumental variables is consistent and its rate is the same as in the exogenous case.

Remark 7.2. To the best of our knowledge, Model (1) with endogeneity has never been studied before, therefore we are the first to provide a consistent estimator in that setting.

Remark 7.3. In the case where C_{WZ} is not injective, one could replace Assumption 9 by Assumption 9' below.

Assumption 9′. *Let* $\Pi = \Pi_N + \Pi_{N^\perp}$ *where* $\Pi_N \in \mathcal{N}(C_{WZ})$ *and* $\Pi_{N^\perp} \in \mathcal{N}^\perp(C_{WZ})$. *There is a Hilbert-Schmidt operator R from* $\mathcal{E}$ *to* $\mathcal{F}$ *and a constant* $\beta > 0$ *such that* $\Pi_{N^\perp}^{*} = (C_{ZW}C_{WZ})^{\beta/2} R$.

Under Assumption 9', we could establish a result similar to that of Proposition 3, namely that $\hat{\Pi}_{\alpha}^{*}$ converges to $\Pi_{N^\perp}^{*}$ with the same rate of convergence as before.

2.8 Simulations

This section consists of a simulation study of the estimators presented earlier. Let $\mathcal{E} = \mathcal{F} = L^2[0,1]$ and $\mathcal{S} = \mathcal{T} = [0,1]$. Π is an integral operator from to $L^2[0,1]$ to $L^2[0,1]$ with kernel $\pi(s,t) = 1 - |s-t|^2$.[8] We consider an Ornstein-Uhlenbeck process with zero mean and mean reversion rate equal to one to represent the error function. It is described by the differential equation $dU(s) = -U(s)ds + \sigma_u dG_u(s)$, for $s \in [0,1]$ and where G_u is a Wiener process and σ_u denotes the standard deviation of its increments dG_u. Note that this error function is stationary.

We study the model

$$Y_i = \Pi Z_i + U_i, \qquad i = 1, ..., n$$

[8]Simulations have also been performed using different kernels. In particular, we have considered multiple kernels, allowing to include multiple functional predictors in a single functional model. Results suggest that the performance of the estimator is analogous in "multivariate" functional linear regression.

in two different settings. First, we consider design functions uncorrelated to the error functions ($cov(U, Z) = 0$), then investigate the case where Z is endogenous ($cov(U, Z) \neq 0$).

2.8.1 Exogenous predictor functions

We consider the design function

$$Z_i(t) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i) + \Gamma(\beta_i)} t^{\alpha_i - 1}(1-t)^{\beta_i - 1} + \eta_i$$

for $t \in [0,1]$, with $\alpha_i, \beta_i \sim U[2,5]$ and $\eta_i \sim N(0,1)$, for all $i = 1, ..., n$. These predictor functions are probability density functions of some random beta distributions over the interval $[0,1]$, with an additive gaussian term.

The numerical simulation is performed as follows:

1. Construct both a pseudo-continuous interval of $[0,1]$, denoted $\mathcal{T}$, consisting of 1000 equally-spaced discrete steps, and a discretized interval of $[0,1]$, denoted $\tilde{\mathcal{T}}$, consisting of only 100 equally-spaced discrete steps.

2. Generate n predictor functions $z_i(t)$ and error functions $u_i(s)$, where $t, s \in \mathcal{T}$ so as to obtain pseudo-continuous functions.

3. Generate the n response functions $y_i(s)$ using the specified model where $s \in \mathcal{T}$.

4. Generate the sample of n discretized pairs of functions $(\tilde{z}_i, \tilde{y}_i)$ by extracting the corresponding values of the pairs (z_i, y_i) for all $t, s \in \tilde{\mathcal{T}}$.

5. Estimate Π using the regularization method on the sample of n pairs of functions $(\tilde{z}_i, \tilde{y}_i)$ (discrete points are connected by linear interpolation to obtain curves) and a fixed smoothing parameter $\alpha = .01$.

6. Repeat steps 2-5 100 times and calculate the MSE by averaging the quantities $\|\hat{\Pi}_\alpha - \Pi\|^2_{HS} = \int_{\tilde{\mathcal{T}}}\int_{\tilde{\mathcal{T}}} (\hat{\pi}_\alpha(s,t) - \pi(s,t))^2 dtds$ over all repetitions.

All numerical integrations are performed using the trapezoidal rule (i.e. piecewise linear interpolation) although it is possible to use other quadrature rules (such as another Newton–Cotes rule or adaptive quadrature). In addition, the simulations of the stochastic processes for the error terms are constructed using the Euler-Maruyama method for approximating numerical solutions to stochastic differential equations.

Figure 2.1 shows 10 discretized predictor functions (z_i), Ornstein-Uhlenbeck error functions for $\sigma_u = 1$ (u_i), response functions (y_i) and an example of a response function for various values of σ_u.

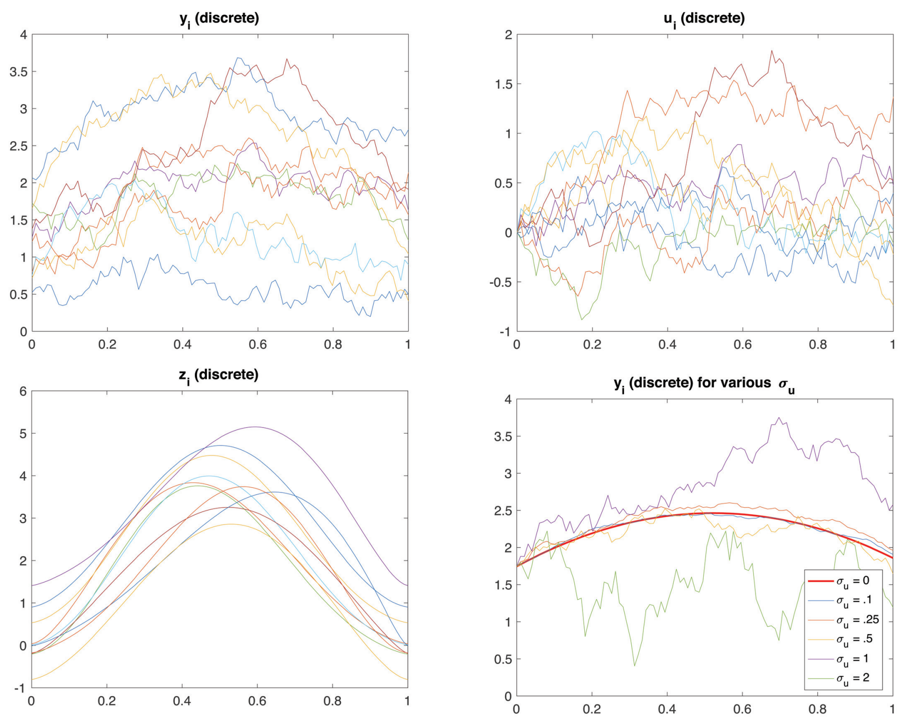


Figure 2.1: Examples of simulated functions (top left: discretized y_i; top right: discretized u_i for $\sigma_U = 1$, bottom left: discretized z_i, bottom right: a single y_i for various σ_u).

Table 1 reports the MSE for 4 different sample sizes ($n = 50, 100, 500, 1000$) and 5 values of the standard deviation parameter ($\sigma_u = 0.1, 0.25, 0.5, 1, 2$). Naturally, the use of a fixed smoothing parameter $\alpha = .01$ that is independent of the sample size prevents the MSE from converging towards zero. In fact, the MSE converges to $\|\Pi - \Pi_\alpha\|^2_{HS}$, which is a measure of the squared bias introduced by the regularization method.[9] The last two columns of Table 1

[9]The magnitude of this bias depends on both the design functions and the value of α

report the true global (R^2) and extended local ($\tilde{R}^2$) functional coefficients of determination, defined as

$$R^2 = \frac{\int_{\mathcal{S}} \text{var}(E[Y(s)|Z])ds}{\int_{\mathcal{S}} \text{var}(Y(s))ds} = \frac{\int_{\mathcal{S}} \text{var}(\Pi Z(s))ds}{\int_{\mathcal{S}} \text{var}(Y(s))ds}$$

$$\tilde{R}^2 = \int_{\mathcal{S}} \frac{\text{var}(E[Y(s)|Z])ds}{\text{var}(Y(s))ds} = \int_{\mathcal{S}} \frac{\text{var}(\Pi Z(s))ds}{\text{var}(Y(s))ds},$$

which are directly related to those proposed in Yao, Müller and Wang (2005).[10]

Std errors	Empirical MSE				Squared bias	Coef. of d.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$\|\Pi - \Pi_\alpha\|^2_{HS}$	R^2	$\tilde{R}^2$
$\sigma_u = 0.1$	.0154	.0135	.0126	.0124	.0095	.995	.995
	(.0027)	(.0017)	(.0008)	(.0005)			
$\sigma_u = 0.25$	.0291	.0205	.0138	.0130	.0095	.976	.976
	(.0098)	(.0063)	(.0022)	(.0013)			
$\sigma_u = 0.5$	.0773	.0438	.0194	.0156	.0095	.910	.911
	(.0363)	(.0193)	(.0057)	(.0028)			
$\sigma_u = 1$	.2909	.1354	.0371	.0257	.0095	.712	.724
	(.1789)	(.0659)	(.0161)	(.0089)			
$\sigma_u = 2$	.9128	.4755	.1245	.0668	.0095	.383	.423
	(.5495)	(.2607)	(.0660)	(.0378)			

Note: Standard deviations are reported in parentheses.

Table 2.1: Simulation results: Mean-Square Errors over 100 replications

Simulations results are in line with the theoretical results. We observe that, for a fixed α, the MSE decreases as the sample size grows. Further, the coefficients of determination decrease as the standard deviation of the error increases, since the estimation is made more difficult. As a result, the MSE grows with σ_u.

since $\Pi_\alpha = (\alpha I + V_Z)^{-1} V_z \Pi$. We perform Monte-Carlo simulations to approximate the regularized operator Π_α using 100 random samples of 1000 z_i's.

[10]These true coefficients are approximated by their mean values using 1000 random functions over 100 simulations. In practice (when the true Π is unknown) it is possible to use a consistent estimators of those coefficients by using $\hat{\Pi}_\alpha$ and the sample counterpart of variance operators.

For illustration purposes, we provide two sets of surface plots. Figure 2.2 shows 3D-plots of the actual kernel (top-left), the regularized kernel (top-right), their superposition (bottom-left) and the bias computed as their difference (bottom-right). The Tikhonov regularization appears to introduce most of the bias on the edges of the kernel.

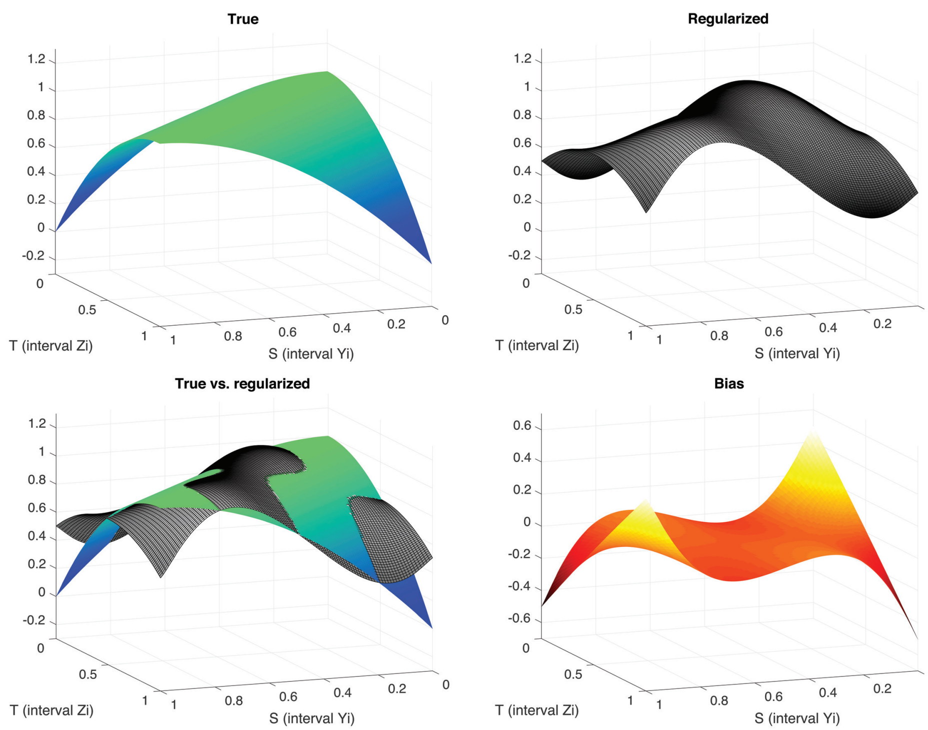


Figure 2.2: True kernel vs. regularized kernel (top left: True; top right: Regularized, bottom left: True vs. regularized, bottom right: Bias)

Figure 2.3 shows the mean estimated kernel for $n = 500$ and Ornstein-Uhlenbeck errors with $\sigma_u = 1$ (top-left), against the true kernel (bottom-left), against the regularized kernel (top-right), and its mean errors with respect to the true kernel (bottom-right). One may observe that the mean estimate is relatively close to the regularized kernel. However it does not perform well on the edges when compared to the true kernel[11].

[11] A possible solution to the boundary effect would be to estimate the kernel on a larger support than necessary and then truncate.

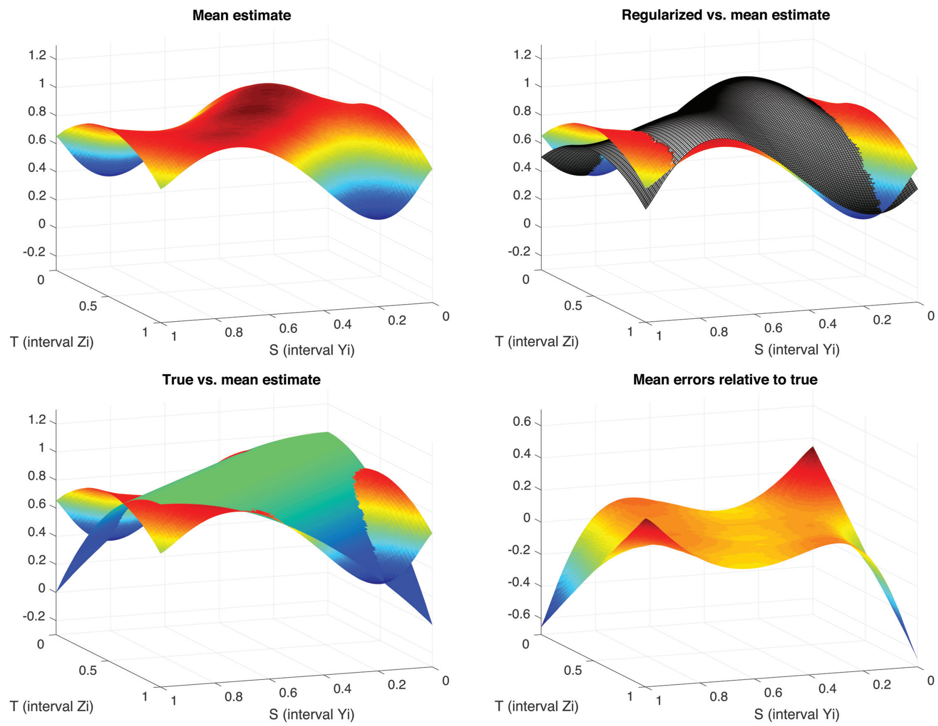


Figure 2.3: True kernel vs. mean estimate (100 runs with $n = 500$, $\sigma_u = 1$) (top left: Mean estimate, top right: Regularized vs. mean estimate, bottom left: True vs. mean estimate, bottom right: Mean errors)

Let us now turn to the case where Z is endogenous.

2.8.2 Endogenous predictor functions

We consider the design function

$$Z_i(t) = bW_i(t) + \xi_i(t),$$

where $\xi_i(t) = aU_i(t) + c\epsilon_i(t)$ and the instrument w_i is defined as

$$W_i(t) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i) + \Gamma(\beta_i)} t^{\alpha_i - 1}(1-t)^{\beta_i - 1} + \eta_i$$

for $t \in [0, 1]$, $\alpha_i, \beta_i \sim U[2, 5]$ and $\eta_i \sim \mathcal{N}(0, 1)$, for all $i = 1, ..., n$. Moreover, U_i and ε_i are Ornstein-Uhlenbeck processes with standard deviation parameters $\sigma_u = \sigma_\varepsilon = 1$. It is easily shown that ξ_i is also an Ornstein-Uhlenbeck process with unit mean-reversion rate described by the differential equation $d\xi(t) = -\xi(t) + \sqrt{a^2\sigma_u^2 + c^2\sigma_\varepsilon^2} dG_\xi(t)$. We further assume $a = 1$, $b \in [0,1]$ and c such that $\int_{\mathcal{S}} \text{var}(Y(s))ds$ is unchanged as b varies.[12] Hence, the choice of b amounts to that of the instrument's strength.

The numerical simulation design is slightly modified so as to incorporate the generation of the instruments W and the dependence between Z and U:

1. Construct both a pseudo-continuous interval of $[0,1]$, denoted $\mathcal{T}$, consisting of 1000 equally-spaced discrete steps, and a discretized interval of $[0,1]$, denoted $\tilde{\mathcal{T}}$, consisting of only 100 equally-spaced discrete steps.

2. Generate n instrument functions $w_i(t)$ and error functions $u_i(s)$ and $\varepsilon_i(s)$, where $t, s \in \mathcal{T}$ so as to obtain pseudo-continuous functions.

3. Generate n predictor functions $z_i(t)$ using the design specified above, where $t, s \in \mathcal{T}$ so as to obtain pseudo-continuous functions.

4. Generate the n response functions $y_i(s)$ using the specified model where $s \in \mathcal{T}$.

5. Generate the sample of n discretized pairs of functions $(\tilde{w}_i, \tilde{z}_i, \tilde{y}_i)$ by extracting the corresponding values of the pairs (w_i, z_i, y_i) for all $t, s \in \tilde{\mathcal{T}}$.

6. Estimate Π using the regularization method on the sample of n triplets of functions $(\tilde{w}_i, \tilde{z}_i, \tilde{y}_i)$ and a fixed smoothing parameter $\alpha = .01$.

7. Repeat steps 2-5 100 times and calculate the MSE by averaging the quantities $\|\hat{\Pi}_\alpha - \Pi\|_{HS}^2 = \int_{\tilde{\mathcal{T}}}\int_{\tilde{\mathcal{T}}} (\hat{\pi}_\alpha(s,t) - \pi(s,t))^2 dtds$ over all repetitions.

Table 2 reports the MSE for 4 different sample sizes ($n = 50, 100, 500, 1000$) and 4 values of b when estimating the model without accounting for the endogeneity of Z using the estimator described in Section 2.2.2. Unsurprisingly, the estimation errors are important. The squared bias is smaller to that of the previous design and decreases with b. The last two columns report R^2

[12] This assumption allows to keep the variance of Y stable when varying instrument strength. It implies $c = \sqrt{1 + (1-b^2)\frac{\int_{\mathcal{S}}(\text{var}(\Pi W(s))ds}{\int_{\mathcal{S}}\text{var}(\Pi\varepsilon(s))ds}}$.

and $\tilde{R}^2$ for the full model. They are relatively stable since $\int_S \text{var}(Y(s))ds$ is fixed.

Instr. strength	Empirical MSE				Squared bias	Coef. of deter.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$\|\Pi - \Pi_\alpha\|^2_{HS}$	R^2	$\tilde{R}^2$
$b = 0.25$	2.4834	1.4642	.4690	.3214	.0060	.5144	.5461
$(c = 2.3)$	(.4678)	(.2011)	(.0435)	(.0317)			
$b = 0.5$	2.3346	1.4504	.5826	.4541	.0027	.5140	.5450
$(c = 1.96)$	(.4014)	(.2416)	(.0679)	(.0460)			
$b = 0.75$	2.1858	1.5363	.8535	.7529	.0011	.5294	.5591
$(c = 1.55)$	(.4825)	(.2974)	(.1027)	(.0640)			
$b = 1$	2.4219	2.0547	1.6583	1.6310	.0006	.5633	.5919
$(c = 1)$	(.5305)	(.3525)	(.1581)	(.1121)			

Note: Standard deviations are reported in parentheses.

Table 2.2: Non-IV estimator: Mean-Square Errors over 100 replications

We now turn to the simulations results for the instrumental variable estimator described in Section 2.7. Table 3 reports the MSE's along with R^2 and the squared regularization biases. Squared biases are fairly small in this setup. This is related to the covariance operator of the predictor functions. R^2_{FS} denotes the first-stage regression's coefficient of determination. It shows how b relates to the instrument's strength. Naturally, weaker instruments are associated with larger MSE's, although the spread seems to vanish rather quickly in this setup.

For comparisons with the exogenous case, we provide a final set of surface plots. Figure 2.4 shows 3D-plots of the mean IV estimated kernel (top-left), the mean non-IV (top-right), the superposition of the mean IV and the true kernels (bottom-left) and the mean estimation errors computed as the difference between the true kernel and the mean IV estimate (bottom-right). Note that the mean IV estimate is relatively close to the actual kernel, whereas the estimate when neglecting endogeneity exhibits a large bias.

Instr. str.	Empirical MSE				Squared bias	Coef. of d.	
	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$\|\Pi - \Pi_\alpha\|_{HS}^2$	R^2	R_{FS}^2
$b = 0.25$	.2383	.1710	.0752	.0542	.0060	.0175	.0246
$(c = 2.3)$	(.2019)	(.1422)	(.0779)	(.0209)			
$b = 0.5$	.1040	.0619	.0315	.0276	.0027	.0737	.1092
$(c = 1.96)$	(.0859)	(.0349)	(.0099)	(.0053)			
$b = 0.75$	.0682	.0444	.0242	.0216	.0011	.1767	.2683
$(c = 1.55)$	(.0364)	(.0203)	(.0044)	(.0028)			
$b = 1$	.0466	.0330	.0211	.0199	.0006	.3287	.5048
$(c = 1)$	(.0244)	(.0138)	(.0029)	(.0021)			

Note: Standard deviations are reported in parentheses.

Table 2.3: IV estimator: Mean-Square Errors over 100 replications

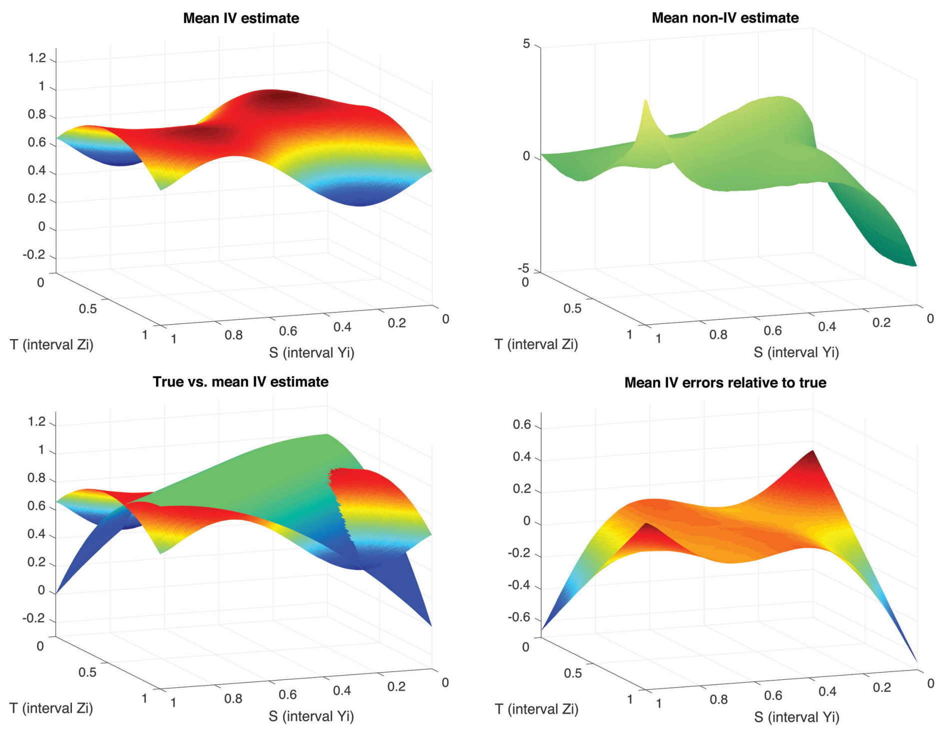


Figure 2.4: True kernel vs. mean IV estimate (100 runs with $n = 500$, $\sigma_u = 1$ and $b = 0.75$) (top left: Mean estimated IV; top right: Mean estimated non-IV, bottom left: True vs. mean IV estimate, bottom right: Mean IV errors)

2.9 Application

2.9.1 Introduction

This section presents an empirical application of the functional linear regression model with functional response to the study of the dynamics between daily electricity consumption and temperature patterns. There is a tradition of applications to electricity in functional data analysis (Ferraty and Vieu, 2006; Antoch et al., 2010; Andersson and Lillestol, 2010; Liebl et al., 2013), although mostly focused on forecasting. Rather, we propose an application illustrating the usefulness of our estimator for inference.

If electricity users were to optimize their consumption with respect to indoor air temperature on a real-time basis, the dynamics of aggregate electricity demand would depend on both past and future weather realizations. Let us consider the reduced-form model for the daily aggregate electricity demand trajectory Y as a function of the outside air temperature pattern Z, given by

$$Y(s) = \pi_0(s) + \int_{\mathcal{T}} \pi_1(s,t)Z(t)dt + U(s), \tag{2.38}$$

where Y, π_0, Z and U are L^2 functions of time indices s and t, although defined over possibly different time intervals $\mathcal{S}$ and $\mathcal{T}$. We will consider the case where $\mathcal{S} \subset \mathcal{T}$. That is, aggregate electricity consumption at time s, $Y(s)$, may possibly depend on current ($t = s$), past ($t < s$) and future ($t > s$) temperature levels $Z(t)$.

Abstracting from the uncertainty surrounding future realizations,[13] this reduced-form model can be related to the Euler equation of a continuous time model with partial adjustment dynamics based on the literature on dynamic linear rational expectations models (Muth, 1961; Kennan, 1979; Hansen and Sargent, 1980).[14]

[13] A model like $Y(s) = \pi_0(s) + \int_0^s \pi_1(s,t)Z(t)dt + \int_s^1 \pi_2(s,t)E[Z(t)|I_s]dt$, which explicitly accounts for expectations (where I_s denotes the information set at time s), could be estimated with our method, although with significant departures from the original model.

[14] The electrical engineering literature on home energy management develops rational expectations models for smart thermostats that optimize real-time energy consumption with respect to past weather measurements and available forecasts, given individual preferences settings (McLaughlin et al., 2011; Yu et al., 2013).

This model can be estimated using the procedure developed in this paper if one has an *i.i.d.* sample of functional observations $\{y_i, z_i\}_{i=1,\dots,n}$, and if u_i can be presumed a mean-zero *i.i.d.* random functional error process. We propose to estimate the model using daily patterns extracted from aggregate electricity consumption and temperature trajectories at the level of the Canadian province of Ontario. Albeit using successive days may induce autocorrelation in the error which is not taken into account in our theory, we believe that our estimator is still consistent in this setting.

The remainder of the section describes institutional features of Ontario's electricity market, then provides details on data construction before performing a preliminary data analysis. Finally, estimation results are interpreted using contour plots and compared to OLS estimates.

2.9.2 Electricity demand in Ontario

Prior to presenting the data set construction, let us present some facts about the electricity market in Ontario. The market defines two categories of consumers. First, small consumers (residential end-users and small businesses) are billed for electricity usage by their local distribution company. The vast majority pays fixed time-of-use rates which are updated from season to season. The Ontario's energy markets regulator (Ontario Energy Board), defines the winter period from November 1 to April 30 and has two daily peak periods: one in the morning (7am-11am) and the other after worktime (5 pm-7 pm), a mid-peak period (11 am-5 pm) and an off-peak period (7 pm-7 am). The remaining part of the year is defined as the summer period and mid-peak and peak periods are reversed with respect to winter.

Second, large consumers[15] (large businesses and the public sector) are subject to the wholesale market price, determined on an hourly basis in a uniform-price multi-unit auction subject to operational constraints. The wholesale market price being quite volatile, some large consumers choose to go with retail contractors to avoid market risk exposure, although the bulk of electricity trade goes through the wholesale market. Furthermore, all large consumers must also pay the monthly *Global Adjustment* which represents other charges related to market, transport and regulatory operations.

Consequently, it is difficult to evaluate the extent to which aggregate

[15] Roughly speaking, businesses are considered large when their electricity bills exceed $2,000 per month.

electricity consumption depends on the wholesale price and the time-of-use price. Furthermore, short-term demand for electricity is typically perceived as being very price inelastic (see for instance Faruqui et al., 2013). For these reasons, we decide to abstract from the price effects.

2.9.3 Preliminary Data Analysis

The original data set consists of hourly observations of real-time aggregate electricity consumption and weighted average temperature in Ontario from January 1, 2010 to September 30, 2014. Hourly power data for Ontario are publicly available on the system operator's website, http://www.ieso.ca. Hourly province-wide temperature values have been constructed from hourly measurements at 77 weather stations in Ontario,[16] publicly available on Environment Canada's website, http://climat.meteo.gc.ca/.

Let $Z(t)$ denote our measure of temperature at time t for the entire province. It is constructed in four steps. First, we match a set of 41 Ontarian cities (of above 10,000 inhabitants)[17] to their three nearest weather stations. Second, we compute a weighted average using a distance metric. Third, we obtain $Z(t)$ as a weighted average of cities' temperatures, where weights are defined by each city's relative population. The constructed province-wide hourly temperature variable is formally defined by

$$Z(t) = \sum_{c} \gamma_c \{ \sum_{w(c)} \rho_{w(c)} Z_{w(c)}(t) \}, \qquad \forall i, h$$

where $\gamma_c = \frac{Pop_c}{(\sum\limits_j Pop_j)}$ is city c's weight, $\rho_{w(c)} = \frac{((lat_c - lat_{w(c)})^6 + (lon_c - lon_{w(c)})^6)^{-1}}{\sum\limits_{l(c)} ((lat_c - lat_{l(c)})^6 + (lon_c - lon_{l(c)})^6)^{-1}}$ is station w's weight for city c's temperature average, lat denotes latitude, lon longitude, and $Z_w(t)$ is the temperature measurement at station w in hour t.[18] Finally, we use robust locally weighted polynomial regression on the constructed temperature series in order to smooth implausible jumps, which are most likely due to measurement errors. Table 1 reports descriptive statistics for hourly electricity consumption and our constructed measure of

[16]The complete data set contained 139 weather stations although once matched to neighboring cities, only 77 were found relevant.

[17]Those cities represent 85.3% of the province's population as of 2011.

[18]The distance metric is the sum of differences in geographic coordinates to the sixth power. The exponent is chosen so as to put arbitrarily more weight on nearby weather stations with respect to those located further away from the cities.

	Hourly temperature (C)				Hourly consumption (GW)				
Year	Mean	SD	Min	Max	Mean	SD	Min	Max	Obs
2010	8.27	9.17	-17.12	29.13	16.23	2.59	10.62	25.07	8760
2011	8.02	9.37	-17.96	31.15	16.15	2.45	10.76	25.43	8760
2012	9.32	8.67	-14.32	30.57	16.09	2.40	10.99	24.60	8784
2013	7.55	9.36	-17.40	28.88	16.07	2.39	10.77	24.49	8760
2014	7.58	11.02	-19.84	26.52	16.08	2.36	10.71	22.77	6552
All	8.18	9.49	-19.84	31.15	16.12	2.45	10.62	25.43	41616

Table 2.4: Descriptive statistics

temperature. Unsurprisingly, we observe some correlation between annual consumption peaks and maximum temperatures.

Figure 2.5 displays the relationships between electricity consumption and our constructed temperature measure for the periods spanning from November 1 to April 30 (blue), May 1 to May 31 (black), June 1 to September 15 (red) and September 16 to October 31 (black). The plots show evidence of a relatively linear relationship between electricity consumption and temperature for winter and summer months. On the other hand, the relation is much flatter in October and May.

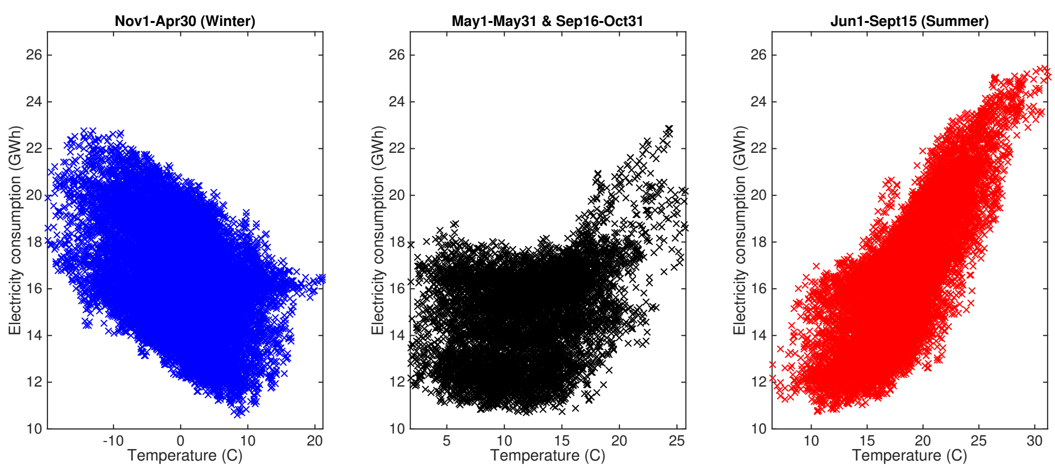


Figure 2.5: Electricity consumption (GWh) vs. temperature (C)

The original data series are presented in Figure 2.6. The plots in Figure 2.5 and Figure 2.6 suggest that power usage is more sensitive to warm weather than cold weather. Probably because cooling can be more energy-intensive than heating, but more importantly because electricity represents a small share of heating fuel for residential users in Ontario. For this latter reason, we will focus the analysis on summer months, i.e. the period from June 1 to September 15.

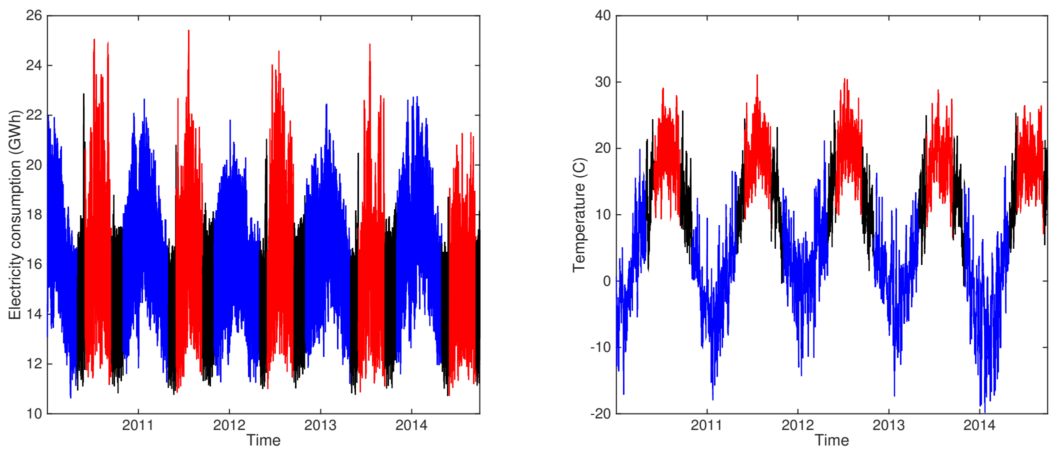


Figure 2.6: Entire data series

Hence, we extract the summer periods for each year from 2010 to 2014 and proceed to construct the functional data sample. It consists of 368 daily trajectories of 25 discrete observations (from midnight to midnight) for the dependent variable and a three-day window of 73 observations for the predictor variable. This window is chosen so that the dependent variable will always be regressed on at least 24 lagged hours and 24 future hours. One can expect significant correlation between successive hourly temperature measurements.

We discard weekends as well as statutory holidays in Ontario in summer months. For ease of interpretation, the temperature variable is transformed into cooling degrees defined as $Z_c(t) = Z(t) - \min_{Z \in \text{Summer}} (Z(t))$, with $\min_{Z \in \text{Summer}} (Z(t)) = 6°C$. The estimation sample (years 2010-2013) consists of 295 functional observations, and is presented in Figure 2.7. The data used for out-of-sample prediction (year 2014) contains 73 functional observations, and is shown in Figure 2.8. The bold lines show the sample mean trajectories. Peak mean summer temperatures occurs around 3 pm, whereas peak mean consumption occurs between 4 pm to 5 pm.

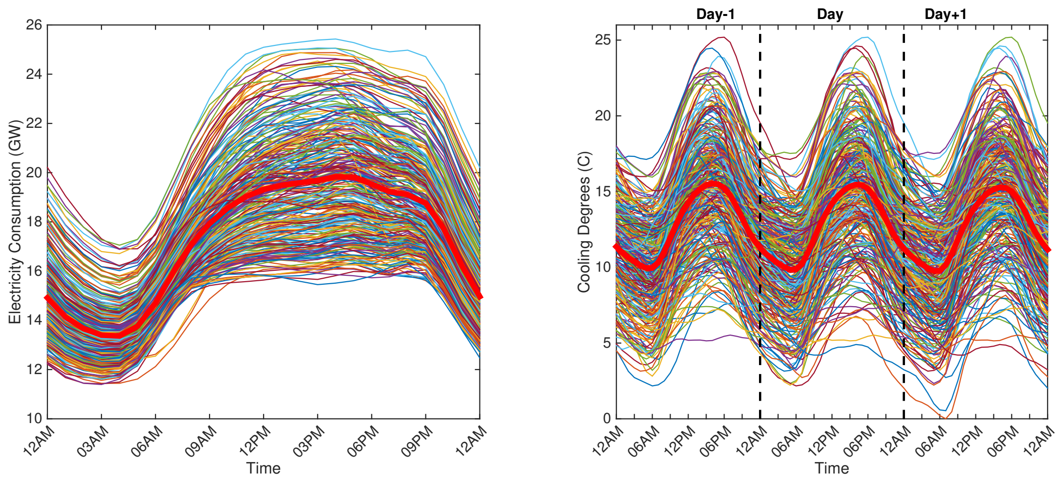


Figure 2.7: Estimation sample (295 functional observations): Y_i's (left) and Z_i's (right)

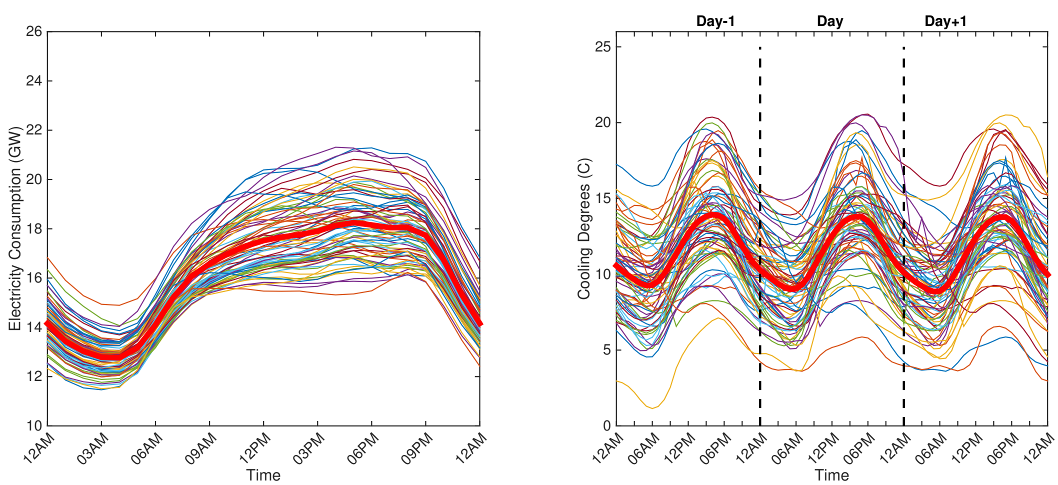


Figure 2.8: Prediction sample (73 functional observations): Y_i's (left) and Z_i's (right)

2.9.4 The functional model and estimation results

Following the preceding discussion, we specify $\mathcal{E} = L^2[0, 24]$ and $\mathcal{F} = L^2[-24, 48]$ with $\mathcal{S} = [0, 24]$ and $\mathcal{T} = [-24, 48]$. The functional linear regression model of interest is given by

$$Y_i(s) = \pi_0(s) + \int_{-24}^{48} \pi_1(s,t) Z_i(t) dt + \sum_{j \in J} \beta_j d_{ij} + U_i(s), \tag{2.39}$$

where $Y_i(s)$ is the aggregate electricity consumption at time $s \in \mathcal{S}$ for day i, $\pi_0(s)$ is a constant function, $Z_i(t)$ is the temperature at time $t \in \mathcal{T}$, and $U_i(s)$ is a zero-mean error term. We also include a set of J binary variables d_{ij} aimed at capturing unobserved seasonalities, in particular for years, months of the year, weekdays and hours of the day. The object of interest is the kernel π_1, which characterizes the dynamic relation between electricity consumption for air conditioning needs and temperature patterns. Since we are interested in daily electricity consumption patterns, the cross-sectional unit i denotes a daily functional observation.

We compare the performance of our estimator with that of the OLS estimator of the discrete analogous model given by

$$Y_i(s) = \pi_0(s) + \sum_{t=-24}^{48} \pi_1(s,t) Z_i(t) + \sum_{j \in \mathcal{J}} \beta_j d_{ij} + U_i(s). \tag{2.40}$$

To assess the performance of our estimator, we do an out-of-sample prediction exercise. First, data from 2010 to 2013 are used to estimate Π, let $\hat{\Pi}_\alpha$ be the resulting estimator. Then, the prediction of Y_i over 2014 is given by $\hat{Y}_i = \hat{\Pi}_\alpha Z_i$ where Z_i is the actual temperature observed in 2014. The Mean Squared Prediction Error (MSPE) for 2014 is computed as $\sum_i \int \left(\hat{Y}_i(s) - Y_i(s)\right)^2 ds$ where Y_i is the actual electricity consumption. Table 2.5 reports MSPE associated with four alternative estimators: (a) our functional estimator computed with α_{CV}, obtained with the leave-one-out CV method based on 2010-2013 data and (b) with α_{BP} which minimizes the MSPE for 2014, (c) the OLS estimator based on hourly predictors and (d) the OLS estimator based on only 24 predictors (3-hourly data). We find that the functional estimator performs better in terms of MSPE. The condition number[19] reported in Table 5 shows that the matrix $Z'Z/n$ is

[19] The condition number is the ratio of the largest eigenvalue on the smallest eigenvalue.

	Func Reg		OLS	
	α_{CV}	α_{BP}	hourly	3 hourly
α	0.15	3.40		
MSPE	16.14	15.85	28.39	17.20
cond(Z'Z/n)			1.40e+06	3.71e+03

Table 2.5: Mean Squared Prediction Errors for Summer 2014

severely ill-conditioned. This explains the poor performance of OLS. The cross-validation criterion and the MSPE are plotted as functions of the tuning parameter in Figure 2.9.

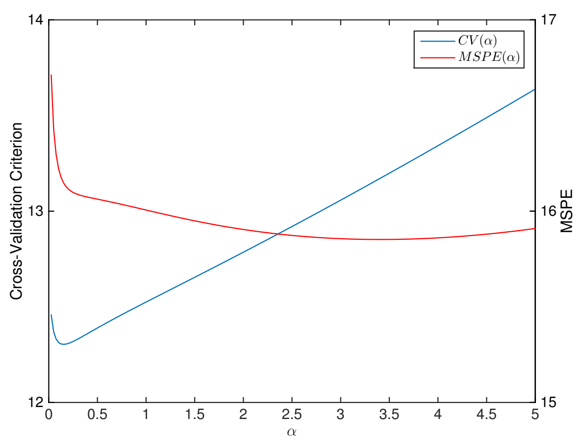


Figure 2.9: CV criterion (blue) and MSPE (red) as functions of α

Figure 2.10 displays contour plots of the 3-hourly OLS kernel estimate (left) and the regularized kernel estimate for α_{CV} (right). In order to ease interpretation of the results, we plot indicative dashed lines to separate out the daily windows and add a diagonal so as to emphasize the contemporaneous relation between the functions of interest. Estimates may be read both horizontally and vertically. The effects of the entire temperature pattern on electricity consumption at a given hour of the day is observed horizontally, whereas the effect of temperature at a specific time upon the daily electricity consumption pattern is read vertically. It is important to emphasize the ceteris paribus nature of those point estimates. Each corresponds to the additional effect of a slight increase in temperature at a specific time, e.g. 12 pm, on the electricity demand at a given time, e.g. 13 pm, *holding everything else (i.e. the temperature at any other point in time) constant*. The magnitudes of the correlation are indicated using colors from dark red to white with corresponding values given in the legend (lighter color corresponds to stronger correlation).

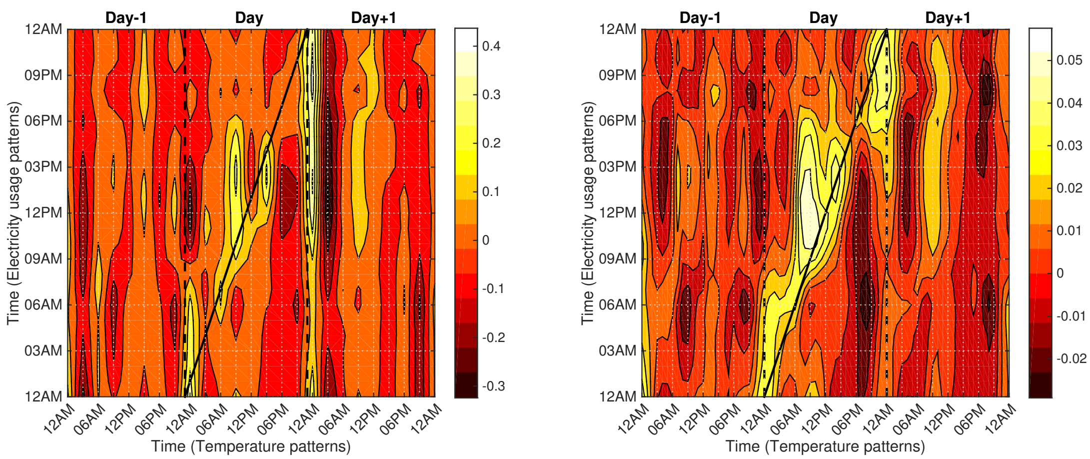


Figure 2.10: 3 hourly OLS kernel estimate (left) and α_{CV}-regularized kernel estimate (right)

In this simple application to real-world data, OLS estimates are clearly too unstable to allow reliable interpretation of the results, even when reducing the set of predictors by a factor of 3. On the other hand, the α_{CV}-regularized kernel estimate appears somewhat undersmoothed. That is, in this application, the cross-validation method delivers too little smoothing with respect to the optimal smoothing for out-of-sample prediction. In comparison, the α_{BP}-regularized kernel estimate shown in Figure 2.11 is smoother, and allows to uncover interesting insights about the relationship under study.

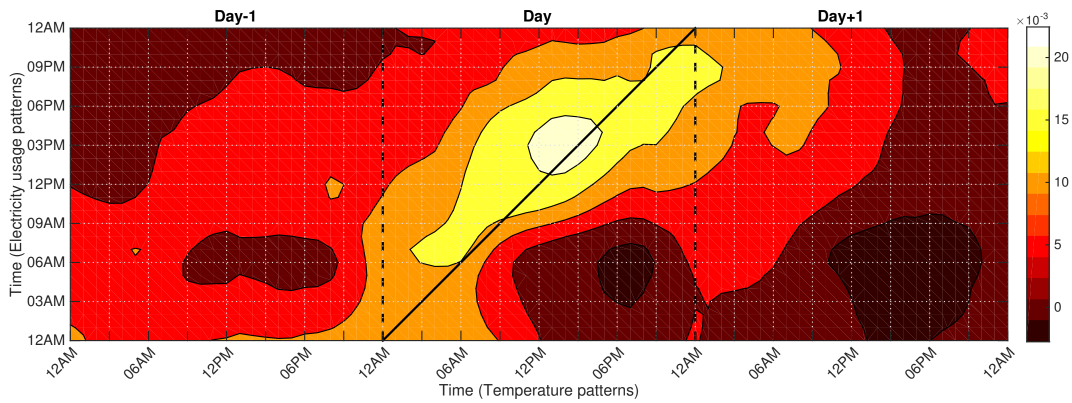


Figure 2.11: α_{BP}-Regularized Kernel Estimate

The kernel estimate suggests that the contemporaneous correlation between temperature and electricity consumption (i.e. coefficients along the indicative diagonal) is the largest from 12 pm to 4 pm (light yellow area), which almost coincides with the Ontario Energy Board's definition of summer peak period (11 am-5 pm). The relation is found of smaller magnitudes for periods 9 am-11 am and 4 pm-10 pm (yellow area), and even more so for 10 pm-9 am (orange area).

The estimated correlations with past and future temperatures (i.e. off-diagonal coefficients) beyond a 24-hour window are found relatively small (in darker areas). Therefore, the dynamic relationship between the variables is mainly characterized by the estimates in the lighter areas spanning around the indicative diagonal in Figure 2.11. For instance, the outside air temperature at noon positively correlates with electricity consumption from 9 am to 9 pm that day. This possibly indicates the presence of both lag and anticipatory effects of outdoor temperature.

We observe that outdoor temperature has lasting effects on electricity consumption. A plausible explanation relates to the law of motion of indoor air temperature, which mainly depends on buildings insulation characteristics, the outdoor air temperature trajectory and indoor heating and cooling use. In the absence of sufficient air conditioning along the day, the indoor temperature level will eventually converge to the outdoor level. Since most people do not consume much air conditioning at home while being away, they eventually have to increase their consumption later on during the day in order to get back to their preferred level. Hence an increase in temperature at 9 am may positively affect consumption at 12 pm through its dynamic effect on indoor temperature.

We also find correlation with future temperature values, which are more difficult to interpret. For instance, consumption at 9 pm positively correlates with nighttime and next day temperature levels until 12 pm (yellow, orange and red areas). These estimates suggest that end-users may increase evening air conditioning in anticipation of higher nighttime temperature levels. A plausible underlying motivation may be to accumulate coolness so as to ensure a desired indoor temperature level at night and in the morning.

If agents were real-time optimizers, like smart thermostats, with rational expectations based on weather forecasts, the presence of partial dynamic adjustments of indoor temperature would create both lag effects and forward-looking dependence. The reality may however be different and one should keep in mind that these estimates are subject to a regularization bias. Also, we acknowledge that these results may be subject to attenuation bias due to the ad-hoc construction of our temperature variable. Finally, the model considers aggregate measures of temperature and electricity consumption over the province of Ontario, the estimated parameters may therefore capture spatial correlation along with the temporal effects.

In order to assess the predictive power of the model, we plot $|\hat{u}_i(s)|/Y_i(s)$, the ratio of the residuals (in absolute terms) to the dependent variable, and the out-of-sample predictions using α_{BP} against actual trajectories for a workweek (Monday 16/6 to Friday 20/6) in summer 2014 in Figure 2.12. We find that the model provides a reasonable prediction (4 % mean absolute prediction error).

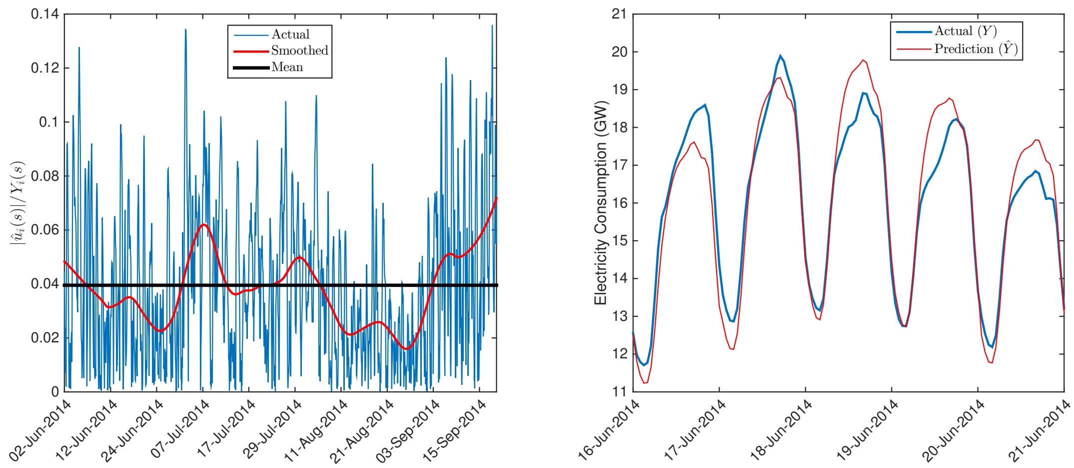


Figure 2.12: $|\hat{u}_i(s)|/Y_i(s)$ (left) and out-of-sample prediction (right)

2.9.5 Conclusion of the application

In conclusion, the estimation results indicate positive correlations between electricity consumption and past, current and future temperature levels. The correlation with future values may be due to anticipatory behavior of agents, who look at the temperature forecast before deciding on whether to leave the air conditioning on that day, for instance. It may also be due to a measurement error in the temperature variable, or even spatial correlations captured through the temporal dimension from the use of aggregate variables. In general, the estimated coefficients vary across hours of the day, and the effect of temperature appears to persist for about 12 hours. The model allows to obtain a good out-of-sample prediction of electricity consumption in the summer of 2014 using data for summers of 2010 to 2013.

The growing deployment of smart-metering technologies in electricity systems creates a need for micro-econometric models able to capture the dynamic behaviors of end-users with respect to market fundamentals. Such models allow to take advantage of the large amount of data so as to provide new insights to practitioners and policymakers about the behavior of consumers. In particular, information with regards to the end-users' behavior with respect to changes in weather or prices is valuable to local distribution companies since it can contribute to improve their demand-side bidding strategies in day-ahead markets, as suggested by Patrick and Wolak (2001). Access to smart meters data being limited, we have proposed instead a model of aggregate electricity consumption as a function of past and future outside air temperature. Nonetheless, we expect the functional regression framework to provide a valuable avenue for taking advantage of the large data sets from individual households' smart meters.

2.10 Conclusion

In this paper, we considered two functional regression models with functional response. In the first model, the regressor was an exogenous function. In the second model, the regressor was an endogenous function and identification relied on instrumental variables. For both models, we have proposed an estimator which is simple to implement, depends on only one smoothing parameter, and is robust to situations where eigenvalues are multiple. Moreover, the indexes t and s could be vectors and our framework allows for multiple regressors.

Various extensions would be interesting to investigate.

- We assumed that we observed a random sample, hence the data were independent. Many relevant applications involve observations which are cross-sectionally correlated.
- When the regressor Z is not observed, Z needs to be estimated first, which induces an error-in-variable. This case was considered in Park and Qian (2012).
- We discussed only the case where we penalized the HS norm of Π, the extension to other norms including norms of the derivatives of Π would be useful in the applications where the interest lies in the derivatives of Π.
- In our framework, we could relatively easily impose restrictions on Π (estimation under constraint).
- Instead of Tikhonov, one could consider the iterative regularization scheme called Landweber-Fridman (see Kress, 1999).
- The derivation of an oracle inequality for the choice of α is left for future research.

Chapter 3

Functional econometrics of multi-unit auctions: an application to the New York electricity market*

"Alas, competition is rarely perfect, markets fail, and market power must be kept in check."

Jean Tirole, Nobel Prize Lecture,
December 8, 2014.

*I would like to thank Marine Carrasco, Jean-Pierre Florens, Étienne Billette de Villemeur, Mathieu Marcoux, Natalia Fabra, Matthew White, Kokouvi Tewou, Samuel Gingras, Jonathan Créchet and the participants of the seminars at the Bank of Canada, Ryerson University, Sciences Po, Universidad Carlos III, Bristol, Tilburg, CREST, HEC Montréal, Polytechnique Montréal, CIREQ Econometrics Conference 2017, CIREQ PhD Students 2017 in Montréal, the Montréal Environment and Resource Economics Workshop, the Marcel Dagenais Econometrics Seminar, and SCSE 2017 in Ottawa for their valuable comments. Data and MATLAB codes will be made available on the author's website.

3.1 Introduction

Multi-unit auctions are auction mechanisms where multiple units of some homogeneous good are sold. They are widely used in wholesale electricity markets to procure energy from power plants, by governments to sell treasury bonds, and for other purposes as well. In electricity markets, each supplier submits an entire *supply function*, generally in the form of several price-quantity pairs, to signal its willingness to produce different quantities of energy.

Unfortunately, multi-unit auctions are not strategy-proof mechanisms: agents do not reveal their true value in equilibrium (Ausubel et al., 2014).[1] It is hence of utmost importance to understand strategic bidding in those auctions in order to address market power concerns.[2]

A participant's optimal bidding strategy in multi-unit procurement auctions depends crucially on its residual demand function, i.e. the demand function net of its rivals' supply functions. Therefore, one could suspect that a strategic bidder may easily infer its residual demand by observing the past bids of its competitors. This is not so easy in practice, for two main reasons. First, the auctioneer often discloses bids after a time lag in order to limit strategic behaviors. Second, residual demand functions are subject to uncertainty: bidders usually have private information, and the market mechanism itself may also introduce uncertainty.

This paper brings two principal contributions. First, I propose a novel approach for the empirical analysis of multi-unit auctions based on functional econometric methods. The approach allows for the evaluation of firm-level market power in a private information setting, and avoids having to model the market mechanism. It relies on a functional instrumental variable (IV) linear estimator which I develop in this paper. Second, the method is applied to the New York electricity market using rich data on firm-level bids and marginal costs for 2013-2015. I estimate firm-level market power and compare actual bidding behavior to profit-maximizing behavior under private information. I find consistent evidence that firms act strategically despite the late disclosure of bids, private information and a complex market environment.

[1] Uniform-price multi-unit auctions can be understood as second-price auctions for single objects, except that the attractive results from auction theory does not extent to the multi-unit setting.

[2] This paper studies unilateral market power, that is the ability of a firm to unilaterally raise market prices above competitive levels, and to profit from this price increase.

This paper seeks to understand whether firms bid optimally in this limited information setting. The disclosure of historical bids provides firms with valuable strategic information. In a repeated auction setting, such as in wholesale electricity markets where firms bid everyday, early disclosure may also permit competitors to coordinate tacitly onto the most profitable equilibrium. The regulation authorities attempt to limit the strategic information available to firms in order to prevent those issues. For instance, in the New York's electricity market, bids are disclosed under anonymous identification numbers three months after each daily auction. I will show that it is relatively easy to lift the anonymity of bids for most firms.[3] However, this time lag should still, in principle, prevents firms from acting as if they could observe their residual demand function.

The real-world functioning of multi-unit auctions may depart substantially from its canonical setting. This is especially true about electricity markets, which have become increasingly complex market platforms interlinked to financial markets. Market participants typically hold financial positions, such as forward contracts, to hedge against the volatility of wholesale electricity prices. Such contracts consist in a private information and introduce uncertainty around the residual demand function. The price received by firms in the market may also deviate significantly from the theoretical uniform-price[4] that clears the market because of congestions in the transmission network. In many electricity markets, a generator is paid its locational price, which depends on its location on the network, and the state of that network. All producers receive the same price only in the absence of congestions and transmission losses. Consequently, the transmission network affects which rival bids matter in the determination of a firm's residual demand function. This is another source of uncertainty to firms, which compounds with the uncertainty created by private information. For empirical researchers, it may be difficult to model the effect of this uncertainty onto bidding behaviors.

[3]Discussions with market designers at ISO-New England – a closely related electricity market – confirmed that the industry participants know the identity of firms behind anonymous bidders.

[4]There are two main formats of multi-unit auctions: pay-as-bid (i.e. discriminatory) and uniform-price. The ranking of the two formats is ambiguous with respect to both efficiency and revenue maximization, essentially because firms do not reveal their true value in equilibrium (Ausubel et al., 2014). In this paper, I focus on the uniform-price format, which is used in New York.

There exist two main empirical approaches to analyze multi-unit auctions. The researcher must either: be able to replicate the allocation mechanism – which requires to model the transmission network when necessary – so as to be able to simulate the effect of uncertainty on strategies; or be ready to make an assumption to alleviate this effect. As regards to the effect of forward contracts, the usual assumption is to restrict bidding strategies to be additively separable (AS) in the firm's private information (Hortaçsu and Puller, 2008). Nevertheless, for some markets such as the New York State's electricity market, the former approach may be exceedingly difficult to implement while the AS restriction may simply not be satisfied.

I propose a novel approach that relies on Wilson's (1979) key result that a market participant in such auctions only cares about the effects of its bids on the distribution function of market-clearing prices. This approach avoids the difficulties associated with modelling transmission constraints, as well as making assumptions like the AS restriction that may not apply in some settings.

The recent developments of functional estimators in the econometric literature (Benatia, Carrasco, and Florens, 2017) provide a novel framework for the analysis of multi-unit auctions. Strategies are function-valued variables, such as supply or demand functions observable by the econometrician, and can be treated as function-valued random elements. The functional approach hence takes full advantage of the richness of the data from those auctions. Thus far, the existing estimation methods either rely on moment conditions evaluated at the observed prices and quantities (Wolak, 2000), or reduce the dimensionality of observed bid functions by linearization (Hortaçsu and Puller, 2008).

This paper develops functional econometric methods for the analysis of empirical bidding behavior in multi-unit auctions. I present the functional linear regression model and propose an IV estimator. Inference and hypothesis testing procedures in functional spaces are then developed. Following Hortaçsu and Puller (2008), the analysis is based on the optimality condition from a static model of strategic bidding in multi-unit auctions. I develop a method to estimate a firm's optimal bid function according to this model in three steps. First, I estimate non-parametrically the price distribution functions separately for several periods. Second, I use the functional estimator to regress the sample of distribution functions onto the observed bid functions. Bid functions are simultaneously determined with the price distribution in equilibrium. Hence, they are considered as endogenous functional regressors. I use forward contracts are instrumental variables because they are predetermined and belong to the firm's private information set. The resulting functional parameter has a clear economic interpretation: at each price, it corresponds to the effect of a firm's quantity bid onto the probability of observing that price in equilibrium. This method avoids any parametric assumption on the price distribution and corresponds to a first-order approximation of the effect of interest. Finally, I solve for the optimal bid functions using the optimality condition evaluated at the estimated parameters. The resulting optimal bid functions can then be compared to the observed functions, and eventually be used to estimate the entire marginal cost functions.

The method is applied to a novel dataset of the New York State's electricity market for the period 2013-2015. The dataset is constructed from public data. I develop an algorithm to de-anonymize the bids and recover firm-level supply functions. Then, I construct firm-level marginal cost functions using methods from the economic literature (Mansur, 2007). The functional tests consistently reject the AS restriction for all firms. The above estimation method yields optimal bid functions close to the observed ones. I consider this result as evidence that the firms behave closely to the optimum of the private information setting, in spite of the late disclosure of anonymized bids, private information and a complex market environment. It suggests that firms are well aware of the effect of their own supply function onto the price distribution, and behave accordingly to maximize profits. Therefore, late bid disclosure should not be considered sufficient to limit strategic behaviors. It is most likely due to the repeated nature of those auctions. In effect, firms can infer their market power using only their own private information and the past realizations of equilibrium prices.

The remaining of the paper is organized as follows. The related literature is presented in Section 3.2. In Section 3.3, I describe the New York state electricity market. Section 3.4 derives the economic model of bidding behavior. In Section 3.5, the functional econometric methods are developed. The application to New York is in Section 3.6. Section 3.7 concludes.

3.2 Related literature

This paper is related to three strands of research: the literature on auctions, the analyses of market power in restructured electricity markets, and the literature on functional data analysis.

Wilson's (1979) share auction model is the first theoretical analysis of multi-unit auctions. Firms are assumed to compete in continuously differentiable demand schedules to buy shares of a perfectly divisible good.[5] Those auctions are found to be manipulable by the bidders. This analysis gave rise to three main streams of research. Back and Zender (1993) is the first account of a theoretical analysis of Treasury auction formats, where bidders have private information and submit differentiable demand schedules. Klemperer and Meyer (1989) developed the concept of supply function equilibrium (SFE) which have then been extensively used to study electricity procurement auctions. Oligopolists are assumed to compete in supply functions to satisfy a demand function that is subject to additive random shocks. This game admits a multiplicity of equilibria, but has the attractive feature that each equilibrium is ex-post profit-maximizing.[6] von der Fehr and Harbord (1993) initiated a modelling framework with step bid functions, hence non-differentiable, which subsequently served as a competing approach to study electricity markets. A tentative reconciliation was recently offered by Holmberg, Newbery, and Ralph (2013) who show that as the step size of bid functions decrease to zero the equilibria coincide to the corresponding SFE. Kastl (2012) provide a similar reconciliation in a private information setting.

The literature on empirical auctions goes back to the pioneering work of Paarsch (1992) on parametric estimation of single-item auction models. The first nonparametric estimation was proposed by Elyakime et al. (1994) using a simulation approach. Guerre, Perrigne, and Vuong (2000) is the most influential article on the nonparametric estimation of single-object auctions. The main insight is to use the distribution of bids and the optimality condition of the economic model to identify the distribution of independent private values. By using the first-order condition, the authors avoid computing the Bayesian-Nash equilibrium of the game, which is analytically intractable in many auction models.[7]

---

[5] The theoretical analysis is illustrated by considering the effect of changing the selling mechanism of tract leases for oil and gas exploration from a sealed-bid first-price auction to a share auction.

[6] This result disappears in many cases. For instance, Benatia, Billette de Villemeur, and Pineau (2018) show that optimal supply functions are not ex-post optimal in the presence of a pivotal supplier and a price cap.

[7] Armantier, Florens, and Richard (2008) propose a technique to approximate the Bayesian Nash equilibria that cannot be solved analytically. Armantier and Sbaï (2006) applies the technique in the multi-unit auction context to compare treasury auction formats when bidders are asymmetric using French data.

Hortaçsu (2002) is among the first empirical analysis of multi-unit auctions. The author uses the model of Wilson (1979) and draws on the simulation idea of Elyakime et al. (1994) and the first-order approach of Guerre et al. (2000) to estimate the conditional distribution of equilibrium prices in the Turkish treasury auctions. The key idea of the approach is to resample the rival bid functions and intersect the simulated residual supply function to the observed demand bids to obtain the conditional distribution. Hortaçsu and McAdams (2010) investigate the mechanism choice in the Turkish treasury auctions in the published version of this initial work.[8] There is also a substantial literature on empirical multi-unit auctions in electricity markets. Those analyses are most often oriented towards the evaluation of market power. An early example is the adoption of the setting of Klemperer and Meyer (1989) in Green and Newberry (1992) to analyze the restructuring of the British electricity industry.

As mentioned in the introduction, there are two main empirical approaches to multi-unit auctions. The former aims at estimating structural parameters, such as the conditional distribution function in Hortaçsu (2002), or marginal cost parameters in Wolak (2000, 2003, 2007). This approach requires the researcher to be able to replicate the market mechanism, and to be ready to assume that agents behave optimally. Wolak (2000) pioneered this method by assuming that firms behave as expected profit maximizers and determine the equilibrium prices and quantities at the intersection of observed residual demand functions and bid functions. Wolak (2003) derives identification results for this approach and estimates generation unit-level marginal costs in the Australian market. Related papers have extended this method to cases where the actual residual demand function differs from the observed one because of unobserved forward contracts (Gans and Wolak, 2008), complementary bidding parameters (Reguant, 2014), or binding transmission constraints (Ryan, 2013). The methods aim at modelling the effect of uncertainty on bidding strategies by resampling observed bids of "similar days", and compute the counterfactual prices and quantities using an algorithm that approximates the market mechanism. This resampling approach yields moment conditions around equilibrium values.

The second approach attempts to alleviate the effect of uncertainty. The objective is to evaluate the empirical behavior of firms. Hortaçsu and Puller

[8]Hortaçsu and McAdams (2017) propose a recent survey of the literature on multi-object auctions, including multi-unit mechanisms.

(2008) restrict bidding strategies to be additively separable in their private information: forward contracts. This AS restriction allows to consider the uncertainty surrounding forward positions as additive shocks on the residual demand functions, and thereby fall back onto the framework of Klemperer and Meyer (1989). The authors develop a testing procedure based on linearized version of the observed bid functions to investigate this restriction. They compute the optimal supply functions under the AS restriction using calculated marginal cost functions, and evaluate the extent to which the bidding behavior of firms in the real-time electricity market of Texas deviates from this benchmark. The results suggest that large players with substantial stake on the market behave close to optimal whereas smaller firms tend to bid sub-optimally. In the same line of research, Sioshansi and Oren (2007) seek to evaluate the predictive power of the SFE model, hence ruling out private information, for the same market and find similar results. Mercadal (2016) studies the role of financial arbitrageurs in the Midwestern ISO electricity market. This market is subject to significant transmission limitations and locational pricing. The market is segmented into submarkets based on the correlation between locational prices using unsupervised hierarchical clustering. The residual demand functions are then constructed separately for each submarket. The AS restriction is finally imposed in order to compute the optimal bids and interpret the deviations from the observed bids.

This paper is mostly related to this second strand of the literature, in particular to Hortaçsu and Puller (2008). I propose a method to investigate empirical bidding behavior without relying on the AS restriction, and without modelling how network congestions segment the market. Nevertheless, the method may also yield non-parametric estimates of marginal cost parameters by imposing the optimality condition.

The methods presented in this paper relates to the recent literature on functional econometrics. Wolak (2003) acknowledged the innovative work of Florens (2003) on inverse problems in structural econometrics as a way to develop non-parametric extensions of his estimation method. Any large-dimensional problem in econometrics can be stated as an inverse problem where the object of interest is a function. Kernel density estimation, density deconvolution (Carrasco and Florens, 2011), GMM with a continuum of moment conditions (Carrasco and Florens, 2000), and nonparametric instrumental variable regression (Darolles et al., 2011) are important examples of inverse problems. Carrasco, Florens, and Renault (2007) present a detailed introduction to a variety of estimation methods for functional equations which are used to describe linear inverse problems. The methods essentially rely on regularization techniques to solve the functional equations. Inference procedures for these problems are collected in Carrasco, Florens, and Renault (2014). This literature also develops methods for functional data analysis, i.e. the statistical analysis of function-valued random elements. Functional data consist of curves, like supply functions in multi-unit auctions. The statistical literature on functional data is well documented in the monographs of Ramsay and Silverman (2005) for linear regression models and Ferraty and Vieu (2006) for non-parametric estimation. Florens and Van Bellegem (2015) propose a linear IV estimator for a regression model with a functional independent variable and a scalar-valued dependent variable. Benatia et al. (2017) study a more general model where the dependent is also a functional variable. This paper extends the pointwise functional linear regression model of Ramsay and Silverman (2005) to the IV setting.

There is a substantial interest for applications related to electricity in the literature on functional data. Related to multi-unit auctions, Aneiros et al. (2013) use non-parametric functional methods to forecast a firm's residual demand functions in the Spanish electricity market. In a similar vein, Pelagatti (2013) proposes to forecast a firm's supply function in the Italian electricity auctions using linear functional regression. However, forecasting methods may fall short under late bid disclosure.

The functioning of the New York electricity market is presented in the next section.

3.3 The New York State electricity market

The New York State electricity system is a restructured electricity market, and is designed to ensure competitive electricity prices and the reliability of the grid.[9] It consists of markets for installed capacity, energy, transmission congestion contracts and ancillary services. The New York Independent System Operator (NYISO) operates the markets, provides transmission service and administers the scheduling of power plants since 1999. This section focuses on how bidding occurs in the energy market organized by the NYISO. I attempt to provide an accurate description, though not exhaustive, of the actual market mechanism. This description is sought to highlight why the

[9]Restructured electricity markets typically ally market-based prices for energy and cost-based regulation for grid reliability.

method proposed in this paper helps better understand relevant features of this industry.

The energy market. The energy market provides a mechanism to allocate electricity production and consumption on a daily basis. On the supply-side, generating facilities may sell energy directly into the market or engage in a bilateral transaction with a purchaser. Similarly, load serving entities (LSE) and other buyers may purchase energy from the market and/or be party to a bilateral contract with a supplier. The energy market is run as a two-settlement process: the day-ahead market (DAM) which takes place one day prior to physical production, and the real-time market (RTM) which is essentially used to adjust for forecast errors and unexpected plant outages. The empirical analysis focuses on the DAM, which accounts for about 94% of energy sales, while the remaining 6% is traded through the RTM. Bilateral contracts account for roughly 40% of DAM energy exchanges.[10] Both markets are organized as two-sided multi-unit uniform price auctions. Firms may participate in the DAM and/or the RTM by submitting generation/load bids to sell/buy energy for each hour of the day. On the demand-side, buyers are required to submit hourly demand forecasts for the next day in the DAM.[11]

Energy bids can be of various types: load bids, transaction bids and generation bids. Load bids allow energy buyers to signal their willingness to pay using a demand curve. The resulting empirical aggregate demand curve is however very inelastic since load bids represent only a small share of total demand. Transaction bids permit firms to import and/or export electricity with surrounding electricity systems.[12] Finally, each generating asset can submit generation bids to signal their willingness to produce using up to 12 price-quantity steps for each hour of the day.[13] In practice, bids seldom vary

[10]See *NYISO Markets: New York's Marketplace for Wholesale Electricity* (http://www.nyiso.com/public/webdocs/mediaroom/publicationspresentations/OtherReports/OtherReports/NYISO%20Markets%20-%20New%20Yorks%20Marketplace%20for%20Wholesale%20Electricitv.pdf).

[11]NYISO enforces punishment procedures for chronic under-forecasters. For more information, refer to the NYISO's Market Participants User's Guide.

[12]The bids allow up to 12 price-quantity steps. The NYISO is interconnected to Quebec (Hydro-Quebec), Midwestern ISO, New England (ISO-NE), Ontario (IESO) and the PJM regional system (Pennsylvania, New Jersey, Maryland).

[13]Generation bids may also include unit commitment parameters used to indicate the upper and lower operating limits, as well as a minimum revenue requirement to signal non-linear cost components such as start-up costs. Reguant (2014) finds that this lat-

across hours of a given day. Quantities must satisfy the generation's capacity limit, whereas prices must respect the price floor of −1000 USD and price cap of 1000 USD.[14] Anonymous bids, associated with a time-consistent *Masked Gen ID* and *Masked Bidder ID*, are disclosed three months after the auction took place.

Bilateral transactions, day-ahead load forecasts and energy bids must be submitted on the day prior to physical production.[15] These are used as inputs for NYISO's day-ahead dispatch algorithm: the Security-Constrained Unit Commitment (SCUC) software package, which solves for the least-cost production allocation among generators for each hour of the next day under operational and transmission constraints. The SCUC outputs the respective hourly day-ahead schedules for each generator and LSE as well as day-ahead locational-based marginal prices (LBMP). Each asset is thus committed to follow its respective scheduled hourly production during the next day. Generators are paid the LBMP associated with their respective location on the network for every MWh supplied, whereas LSEs pay the zonal price for every MWh consumed. The zonal price is calculated as the load-weighted average LBMP for each of the market's 11 demand zones.

Market zones and locational prices. The zones are denoted by letters from A to K and depicted in Figure 3.1. Table 3.1 shows descriptive statistics on the installed capacity, fuel shares, demand and prices by zone. The zones have different demand profiles. Generating assets also spread unevenly across the network, resulting in dissimilar zonal supply curves. The northern zones (A-F) are typically more hydro-intensive whereas the southern zones (G-K), essentially rely on fossil-fuel and nuclear power generation. The existence of significant transmission limitations in conjunction with these types of geographical heterogeneities is an important motivation for locational pricing market designs. Examples of zonal hourly prices on December 27th, 2017 are shown in Figure 3.2.

---

ter parameter have had positive welfare implications in the Spanish market, in terms of reduced market power. They are neglected here.

[14] Price cap and floors are used to mitigate market power abuse.

[15] The day-ahead market also accept virtual bids, which consist of a financial instrument for participants to hedge against market risks, and allow firms – in particular trading institutions such as banks – outside the market to participate (Birge et al., 2014; Jha and Wolak, 2015). Virtual bids are essentially used to arbitrage between the DAM and RTM.

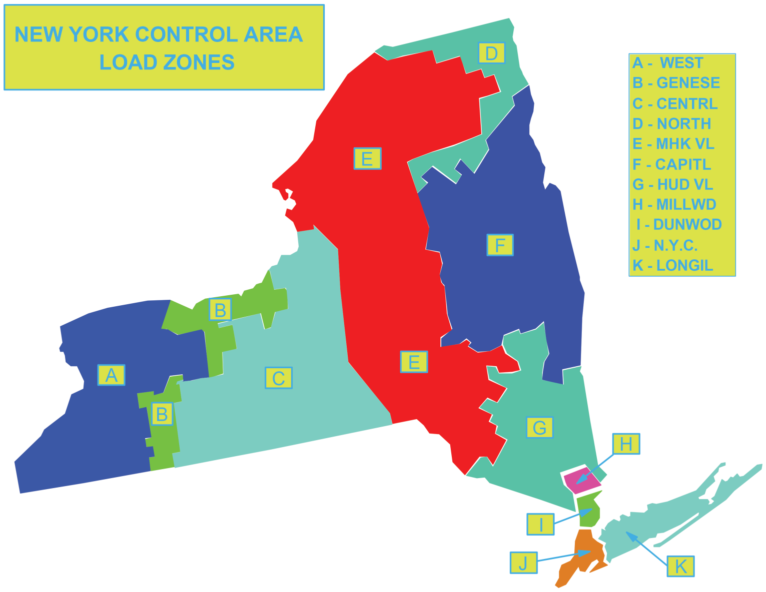


Figure 3.1: New York State electricity market zone map (source: NYISO)

Zone	Name	Capacity	Thermal	Nuclear	Mean D	Min D	Max D	Mean P
A	West	5210.8	38%	0%	1761	1143	2758	38.28
B	Genesee	854.9	20%	72%	1093	656	1995	37.05
C	Central	7391.0	53%	38%	1763	1084	2871	38.75
D	North	2180.7	19%	0%	597	343	958	34.80
E	Mohawk Valley	1263.5	26%	0%	805	443	1266	39.68
F	Capital	5400.1	69%	0%	1329	802	2375	49.88
G	Hudson Valley	2741.2	96%	0%	1083	592	2276	49.26
H	Millwood	2370.7	3%	97%	317	61	695	49.62
I	Dunwoodie	0.2	0%	0%	701	400	1499	49.54
J	New York City	11197.9	100%	0%	6053	3754	11441	50.83
K	Long Island	5712.1	99%	0%	2481	1447	5603	59.99
Total		44323.1	68%	13%	17984	11260	33449	47.48

NOTES. This table reports installed capacity (MW) and descriptive statistics for demand (D in MWh) and price (P in USD) by zone, along with the share of thermal (this includes biomass-fired plants) and nuclear capacity. The remaining share contains hydro, solar and wind power facilities. The mean price for the row 'total' is the demand-weighted average price across zones (sources: NYISO "Gold Book" 2014 and NYISO website).

Table 3.1: NYISO capacity and demand by zone

The LBMP is equal to the marginal cost of supplying an additional MW

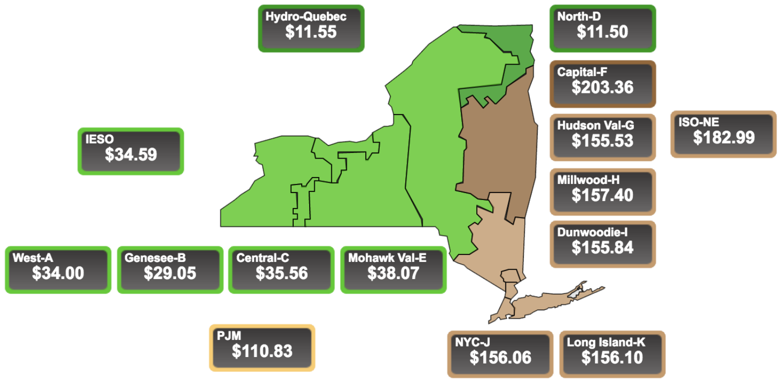


Figure 3.2: Zonal prices map (source: NYISO, December 27th, 2017)

of electricity at that location, or node. The locational prices are calculated as the sums of three components: the energy price, the marginal losses and the marginal congestion. The energy price is equal to the marginal cost of supplying an extra MW at the NYISO-selected reference network location, while respecting transmission constraints.[16] The marginal losses and marginal congestion components at a node signal the extent to which additional generation at the node will increase or decrease system-wide transmission losses and congestions, respectively. Whenever transmission constraints are binding, production schedules will deviate from the least-cost schedules had there been unlimited transmission capacity and zero loss.[17] The locational prices will hence differ across nodes depending on the state of the network. In the presence of congestions between two areas, prices are settled as if there were two separate markets with limited exchange capacity. Therefore, a generator in a congested area will generally receive higher prices due to the scarcity of available generation.

The modelling of this market mechanism requires an engineering model of least-cost dispatch that accounts for transmission limitations,[18] which can be particularly challenging to develop.[19]

[16]This reference node is located in zone E.

[17]Transmission losses occur when power is transported in wires because of Joule heating.

[18]Note that the stylized formulations of an alternative-current network through direct-current approximations neglect important aspects of market power in an electrical network (Bautista, Anjos, and Vannelli, 2007).

[19]As an illustration of computational complexity, the SCUC software takes several hours

Forward contracts. Forward contracts are an important source of private information in electricity markets. Buyers and sellers in the wholesale market mainly use futures to hedge against the volatility of market-clearing prices. Those contracts are generally financial "swaps" specifying a quantity to be delivered/consumed at a fixed price for a given period of time, usually a calendar month. They can typically be discriminated at the zonal level and between off-peak and peak time periods over the contract length. Contracts can be negotiated bilaterally or through a broker on a financial platform, such as NYMEX or ICE,[20] up to five years before the term.

The other markets. The purpose of capacity markets is to ensure that the scarcity of generation capacities with respect to future demand is adequately signalled to investors.[21] The installed capacity market is operated on a monthly basis. Each month, firms participate to a multi-unit uniform price procurement auction. Buyers must fulfil capacity requirements while the sellers bid their generation capacity. Those requirements are based on the projected demand for the following month. All procured generators are obliged to offer energy in the energy market during the following month. This market hence provides a commitment mechanism which ensures that sufficient generation capacity is available.[22] In addition, the market of transmission congestion contracts provides a financial instrument to hedge against the effect of transmission limitations. Finally, ancillary services are provided by generators and other system assets through the bidding process, or at cost-based rates.

In the empirical analysis, I focus exclusively on generation bids in the day-ahead market, and avoid the exercise of modelling transmission constraints. This approach relies on the assumptions that the other markets function adequately and that other bid types can be considered of second-order when

to run.

[20]For instance, see http://www.cmegroup.com/trading/energy/electricity/nyiso-zone-a-5-mw-peak-calendar-month-day-ahead-lbmp-swap-futures.html or https://www.theice.com/products/6590386/NYISO-Zone-C-Day-Ahead-Peak-Fixed-Price-Future

[21]Capacity markets are advocated as a solution to the "missing money" problem by supplementing revenues for generators that seldom enter production in order to induce investment (Joskow, 2008a).

[22]More precisely, there are three sequential capacity markets (one bi-annual and two monthly) used to ensure the availability of resource requirements and reduce the volatility of "capacity prices".

studying the behavior of suppliers in the day-ahead market.

Market structure and other considerations. The structure of this market can be summarized as follows. There are 70 producers in the electricity market, where the 10 largest firms accounted for 70% of total installed capacity as of 2014. Table 3.2 reports the installed capacity by zone of the 10 largest firms participating to the supply-side of the DAM. The New York Power Authority (NYPA), Long Island Power Authority (LIPA), and Consolidated Edison (ConEd) are the largest utilities in NYISO.[23] NYPA operates more than 5.5GW of hydro capacity in the state and about 1GW of natural gas-fired capacity in New York City. Entergy and Exelon operate nuclear power plants, and the rest of the firms essentially have natural gas and oil power plants. The fact that the largest firms endowed with marginal technologies, i.e. generators that set the market-clearing price, are located downstate New York – where supply is tight and demand is large and inelastic – naturally raise concerns about the exercise of market power in this area.

Firm	Capacity	Share	Capacity by zone
1. New York Power Authority	6731	15%	A(46%) D(16%) F(18%) G(1%) J(16%) K(3%)
2. Long Island Power Authority	5216	12%	K(100%)
3. NRG Power Marketing	4118	9%	A(15%) C(44%) J(41%)
Ã§ 4. Entergy Nuclear Power Marketing	3193	7%	C(28%) H(72%)
5. Consolidated Edison	2594	6%	D(4%) F(4%) G(48%) J(44%)
6. TC Ravenswood	2557	6%	J(100%)
7. Exelon Generation Company	2547	6%	B(25%) C(75%)
8. Astoria Generating Company	1771	4%	J(100%)
9. Athens Generating Company	1323	3%	F(100%)
10. Astoria Energy	1300	3%	J(100%)

NOTES. This table reports installed capacity by zone and market share of the 10 largest supplying firms in the New York electricity market (source: NYISO "Gold Book" 2014).

Table 3.2: Firms with the largest installed capacity

The price of natural gas is the main source of variations of the firm-level

[23]Those utilities are subject to cost-of-service regulation. The regulated rates include a *power supply charge* based on the cost of electricity bought from generators through contracts and the wholesale market (https://www.psegliny.com/page.cfm/Account/ServiceAndRates/PowerSupplyCharge).

short-run marginal costs of production in this market. The price variations can be attributed to a variety of changes in the fundamentals of the global energy market, but also to local congestions in the delivery of natural gas in the state of New York. Further details about cost components, including environmental regulations, are given in Section 3.6.1.

Market power mitigation. It is finally important to understand the limitations to strategic bidding in this market. The market is subject to market power mitigation procedures which are automated in the dispatch algorithm. The key measure consists in an evaluation of generation bids with respect to their reference level, which is basically an estimation of the generator's marginal cost established by the NYISO. The aim is to assess potential economic withholding. Bids that are substantially beyond their reference levels are mitigated automatically down to their reference levels.[24] Although firms should not be expected to bid consistently down to their marginal costs, they are refrained by design from exerting too much market power.[25]

3.4 The economic model

In this section, I present an economic model of strategic behavior in the day-ahead market. The objective of this model is to derive a benchmark that accounts for the uncertainty faced by the firms. I follow the uniform price multi-unit auction model of static profit-maximization of Wilson's (1979) share auction setting, like exposed in Hortaçsu and Puller (2008).

Everyday, N firms compete in a uniform-price multi-unit auction to supply electricity for each hour $h \in \{1, ..., 24\}$ of the next day. A capacity market, held ahead of the energy auction, ensures that firms are committed to offer a given capacity $K_{i,h}$ in each hour. Each firm $i \in \{1, ..., N\}$ holds a forward position $(QC_{i,h}, PC_{i,h})$ where $QC_{i,h}$ is the contracted sales offered at price $PC_{i,h}$. At the time of bidding, these contracts are considered as sunk

[24] More precisely, mitigation procedures include both conduct and impact tests. The former attempts to evaluate whether a firm's bid is competitive "enough" while the latter evaluates the bid impact on market prices. Mitigation occurs when a bid judged potentially anticompetitive is assessed to have a “substantial impact” on the market price.

[25] Most mitigation happens in the DAM, and in particular in New York City. In 2014, mitigation occured only in 5% of hours for about 25MW on average. For additional details, refer to the 2014 *State of Market Report for the NYISO Markets by Potomac Economics.*

decisions and consist in private information. Each firm i has a cost function $C_i(q)$ of class $\mathcal{C}^2$, about which the rival firms have full information. Cost functions may be assumed as common knowledge in the context of electricity markets because fuel prices, plants technology and plants ownership are publicly available information. This will be discussed further in the empirical study.

The electricity demand in each hour is subject to a stochastic additive term ε_h such that the demand function is $\tilde{D}_h(p) = D_h(p) + \varepsilon_h$. Each firm i bids simultaneously a supply function $S_{i,h}(p, QC_{i,h})$ that is twice continuously differentiable. For every hour, the equilibrium price $p_h^\star$ is defined by the market-clearing condition

$$\sum_{i=1}^{N} S_{i,h}(p_h^\star, QC_{i,h}) = \tilde{D}_h(p_h^\star), \tag{3.1}$$

and each firm receives the ex-post hourly profit

$$\pi_{i,h} = p_h^\star S_{i,h}(p_h^\star, QC_{i,h}) - C_i(S_{i,h}(p_h^\star, QC_{i,h})) - (p_h^\star - PC_{i,h})QC_{i,h}. \tag{3.2}$$

This profit has two components. The hourly profit from market sales is $p_h^\star S_{i,h}(p_h^\star, QC_{i,h}) - C_i(S_{i,h}(p_h^\star))$, and the adjustments due to the forward position represented by $-(p_h^\star - PC_{i,h})QC_{i,h}$. Whenever $p_h^\star > PC_{i,h}$, the firm must pay this price difference to its contracted buyers for the quantity covered by the contract, $QC_{i,h}$. Conversely, when $p_h^\star < PC_{i,h}$, this difference is paid to the firm by its customers.

When a firm chooses its supply function, the principal source of uncertainty is the market clearing price $p_h^\star$. Following Klemperer and Meyer (1989), in the absence of private information and with additive demand shocks, there exists a set of supply function equilibria. Each equilibrium in this set is such that supply functions are ex-post profit-maximizing. This is because, given a firm's rival supply functions there is a one-to-one mapping between equilibrium prices and realizations of the random additive demand shock. This results in a unique profit-maximizing price-quantity pair for each possible shock, keeping rivals' bids fixed. Thus, the firm can calculate its rivals' equilibrium supply function from their cost functions, provided that it can anticipate which equilibrium is played. Nevertheless, forward positions are a source of private information in this setting. Therefore, the uncertainty surrounding prices also comes from the unknown forward positions of all rival firms. I will now characterize the Bayesian-Nash equilibrium of the game in this private information setting.

Let us consider that the firms' strategies are of the form $S_{i,h}(p, QC_{i,h})$. Assume that rival firms play their equilibrium bidding strategies $\{S_{j,h}(p, QC_{j,h}), j \in -i\}$, then firm i's expected profit maximization problem over the range of possible prices $\mathcal{P}$ is given by

$$\max_{S_{i,h}} \int_{\mathcal{P}} (pS_{i,h}(p) - C_i(S_{i,h}(p)) - (p - PC_{i,h})QC_{i,h})\, dF_{i,h}(p|S_{i,h}(p), QC_{i,h}), \tag{3.3}$$

where $F_{i,h}$ is the cumulative distribution function of the market-clearing price in hour h, conditional on firm i's supply function $S_{i,h}$ and forward position $QC_{i,h}$. This conditional distribution is endogenously determined in equilibrium and embeds all relevant information and beliefs that affect the realization of the market-clearing price from the point of view of firm i. $F_{i,h}$ can be understood as an equilibrium selection mechanism that attributes probabilities to each possible equilibrium.

I integrate (3.3) by parts in order to rewrite it into an equivalent variational problem and then derive the Euler-Lagrange condition[26] given by

$$p - C_i'(S^\star_{i,h}(p)) = (S^\star_{i,h}(p) - QC_{i,h})\frac{F_s(p|S^\star_{i,h}(p), QC_{i,h})}{F_p(p|S^\star_{i,h}(p), QC_{i,h})}, \tag{3.4}$$

where

$$F_s(p|S^\star_{i,h}(p), QC_{i,h}) = \frac{\partial F_{i,h}(p|S^\star_{i,h}(p), QC_{i,h})}{\partial S(p)} \tag{3.5}$$

$$F_p(p|S^\star_{i,h}(p), QC_{i,h}) = \frac{\partial F_{i,h}(p|S^\star_{i,h}(p), QC_{i,h})}{\partial p}. \tag{3.6}$$

This first-order condition characterizes the (pointwise) optimal supply function, which depends crucially on the conditional distribution of prices. F_p represents the conditional probability density function of equilibrium prices, assuming it is well defined, and F_s measures the firm's ability to affect the distribution by changing its supply function. A firm with such an ability can increase the probability to face prices strictly above p by reducing its supplied quantity at price p, and reversely. The optimality condition (3.4)

[26] Wilson (1979) and Hortaçsu and Puller (2008) derive this first-order condition in a more general setting where firms are allowed to be risk-averse.

states that the optimal mark-up $p - C_i'(S^\star_{i,h}(p))$ depends on the firm's market position and ability to affect the price distribution. The mark-ups are positive for prices at which the firm is a net seller, that is for p such that $S^\star_{i,h}(p) - QC_{i,h} > 0$, and negative for prices at which it is a net buyer, that is for p such that $S^\star_{i,h}(p) - QC_{i,h} < 0$.

Assuming that the density is strictly positive for every prices $p \in \mathcal{P}$, it is optimal for a firm to reveal its entire marginal cost function, i.e. $p = C_i'(S^\star_{i,h}(p))$ for all p, if and only if it has no ability to affect the equilibrium price distribution, i.e. if $F_s(p) = 0$ for all p. In any other cases, the firm will set some positive mark-ups in absolute terms. The term $\frac{F_s(p|S^\star_{i,h}(p),QC_{i,h})}{F_p(p|S^\star_{i,h}(p),QC_{i,h})}$ can hence be interpreted as a measure of the firm's ability to exercise market power. It also embeds the firm's beliefs about which equilibria is played by other firms. The set of all firm's Euler-Lagrange condition characterizes the Bayesian-Nash equilibrium of this game.

A key identification result pointed out by Hortaçsu and Puller (2008) with regards to forward positions is given in the proposition below.

Proposition 1 (Identification of forward positions ((Hortaçsu and Puller, 2008))**.** *From* (3.4)*,* $S^\star_{i,h}(p) = QC_{i,h}$ *implies* $p = C_i'(S^\star_{i,h}(p))$*. Therefore, a firm's forward position is identified by* $S^{\star-1}_{i,h}(QC_{i,h}) = C_i'(QC_{i,h})$*.*

This result means that the contracted quantity is such that the inverse supply function intersects the marginal cost function. It allows to infer the firm's forward positions from its supply and marginal cost functions. This will be used in the application.

The principal focus of the empirical analysis consists in evaluating how firms behave with respect to this benchmark. This requires to estimate (3.5) and (3.6) in order to calculate the optimal bid function. The simultaneous determination of $F_{i,h}$ and $S^\star_{i,h}$ in equilibrium however gives rise to an endogeneity issue that will need to be addressed. In order to circumvent this estimation, that will proved to be difficult, Hortaçsu and Puller (2008) propose to restrict the supply function strategies to be additively separable in the private information possessed by the firms. That is, strategies must be such that $S_i(p, QC_i) = a_i(p) + b_i(QC_i)$. This testable restriction to the space of strategies permits to simplify the characterization of equilibrium strategies. The authors prove that under this additive separability (AS) hypothesis, the optimality condition in (3.4) becomes

$$p - C_i'(S_i(p)) = (S_i(p) - QC_i)\frac{1}{-RD_i'(p)} \tag{3.7}$$

where $RD_i'(p)$ is the derivative of the ex-post residual demand function faced by the firm. The residual demand function is defined as $RD_i(p) = \tilde{D}(p) - \sum_{j\neq i} S_j(p)$, and thus $RD_i'(p) = \tilde{D}'(p) - \sum_{j\neq i} S_j'(p)$. $RD_i(p)$ is assumed to be observable by the firm. Denoting the residual demand's price elasticity by $\varepsilon_{RD}(p) = p\frac{RD'(p)}{RD(p)}$, and given that $S_i(p^\star) = RD_i(p^\star)$ I can rewrite (3.7) at the equilibrium price $p^\star$ as

$$\frac{p^\star - C_i'(S_i(p^\star))}{p^\star} = \frac{(S_i(p^\star) - QC_i)}{S_i(p^\star)}\frac{1}{-\varepsilon_{RD}(p^\star)}. \tag{3.8}$$

This condition resembles the inverse elasticity pricing rule of the standard monopoly. It shows that the firm acts as a monopolist on its residual demand function. The resulting profit is earned for each output unit away from its contracted position. Similarly, if (3.5) and (3.6) are known, the elasticity ε_F corresponding to the implicit residual demand function induced by F_s and F_p is given by

$$\frac{1}{-\varepsilon_F(p^\star)} = \frac{S_i(p^\star)}{p^\star}\frac{F_s(p^\star|S_i, QC_i)}{F_p(p^\star|S_i, QC_i)}. \tag{3.9}$$

The AS restriction results in equilibrium strategies that coincide with those in Klemperer and Meyer (1989). This is because the uncertainty associated with unknown forward positions does not affect the slope of the residual demand functions anymore. This uncertainty only shifts the residual demand functions and thus has an effect equivalent to additive demand shocks.

This assumption provides a useful benchmark to evaluate how firms behave in multi-unit auctions. If one is convinced by its empirical validity, then it is possible to study the deviations of observed bids from the predicted optimal behavior. Hortaçsu and Puller (2008) show evidence of sub-optimal bidding (in light of this model) in the balancing market of Texas and identifies a list of possible explanations, notably the presence of significant participation costs. Yet, it is important to acknowledge that firms may not be able to enforce the reaction functions characterized by (3.4) and (3.7) in practice. There are reasons to suspect that all firms will exhibit at least some idiosyncratic errors. Some firms may also behave sub-optimally.

First, the optimality conditions may hold at best on average. The nature of the bid functions restricts the strategy space and imposes the presence of errors in that model. These errors allows to rationalize step bid functions in the continuously differentiable model. Furthermore, the market operator discloses anonymized bid data after a time lag in order to prevent collusion. The firm must therefore invest resources to forecast its residual demand or the conditional distribution function of prices. Forecast errors are bound to occur. Second, the regulatory framework may lead to sub-optimal bidding behaviors. Market power mitigation mechanisms are used to prevent strategic conduct by controlling mark-ups. Also, some producers are former utilities subject to regulation aimed at limiting rent-seeking behaviors. Consequently, firms may under- or over-react with respect to rival bids because of the constrained strategic space, forecast errors and regulation.

In the next section, I present the econometric methods that will be used in the empirical analysis based on this model.

3.5 Econometric methods

This section is organized as follows. First, I present the functional linear regression model and propose an extension to the IV setting. This will consist of the main tool of analysis. Second, I propose three statistical tests to investigate the first-order condition under the AS hypothesis in (3.7). Third, these tests are adapted to test the AS restriction using functional regression estimates. Fourth, I develop a functional approach to estimate the partial derivatives F_s and F_p in (3.5) and (3.6) in order to calculate a benchmark of bidding behavior using (3.4).

3.5.1 Functional linear regression

Let us consider a regression model where both the dependent Y and independent variable Z are random functions of an index variable p taking values in a space $\mathcal{P}$. The index p could denote time, a spatial location, or price over a specified continuum. In the application presented in Section 3.6, Y corresponds to the cumulative distribution function of equilibrium prices, and Z is a firm's supply function.

The econometrician observes pairs of random functions (y_t, z_t) $t = 1, ..., T$ which are realizations of random processes (Y, Z) with zero mean functions and unknown covariance operators.[27] Y and Z are assumed to belong to the Hilbert space of square-integrable functions, denoted $\mathbb{L}^2$. This space is equipped with the inner product $\langle , \rangle$ defined $\forall \phi_1, \phi_2 \in \mathbb{L}^2$ as $\langle \phi_1, \phi_2 \rangle = \int_{\mathcal{P}} \phi_1(p)\phi_2(p)dp \in \mathbb{R}$. This inner product induces the $\mathbb{L}^2$-norm defined for any $\phi \in \mathbb{L}^2(\mathcal{P})$ as $||\phi||^2 = \int_{\mathcal{P}} \phi(p)^2 dp$.

The functional regression model is

$$Y(p) = Z(p)\beta(p) + U(p), \tag{3.10}$$

where U is a zero-mean random element of $\mathbb{L}^2$ and β is a nonrandom functional parameter of interest. The model specifies a linear relation between the observed functions. Specifically, $Y(p)$ can only depend on $Z(p)$ and not $Z(p')$ for $p' \neq p$.[28] Each $\beta(p)$ could be estimated separately for each p in a given grid using ordinary least squares or seemingly unrelated regression. However, this method will result in a functional parameter that is not smooth and thereby hardly interpretable. This may not be satisfactory in many applications. Functional estimation permits instead to control for the degree of smoothness, hence for the order of differentiability, to be imposed on the estimated function.

Exogenous regressor

This section shows how to obtain the estimator of β developed in Ramsay and Silverman (2005) from an econometric viewpoint. Let us suppose that the orthogonality assumption $E[Z(p)U(p)] = 0$ holds for all $p \in \mathcal{P}$. Consider the expansion of the functional parameter β onto a set of K basis functions[29] such that $\beta_K(p) = \boldsymbol{\Phi_K}(p)^\top \boldsymbol{b_K}$, where $\boldsymbol{\Phi_K}$ is a K-dimensional vector of basis functions and $\boldsymbol{b_K}$ is a vector of K expansion coefficients, and $\top$ denotes the transpose. Combining the orthogonality condition and the basis expansion

[27] The extensions to non-zero mean functions and multiple regressors are straightforward.

[28] Benatia et al. (2017) develop a more general estimator in the context where $Y(p)$ is allowed to depend on all values along Z. In their framework, functions are allowed to belong to more general functional spaces such as Sobolev spaces.

[29] Basis functions can be splines, as in the application, or power series, Fourier series, Hermite polynomials, wavelets, among others, depending on the space of functions. In the econometrics literature, the choice of the basis functions is referred to as the choice of a sieve space, that is a space of functions that will be used to approximate the function space of interest.

gives

$$E[Z(p)Y(p)] = E[Z(p)^2]\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b_K}, \tag{3.11}$$

for all $p \in \mathcal{P}$. Thus, the expansion coefficients must satisfy the equality

$$\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)E[Z(p)Y(p)]dp = \left[\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)E[Z(p)^2]\boldsymbol{\Phi_K}(p)^\top dp\right] \times \boldsymbol{b_K}. \tag{3.12}$$

For K small, the inversion of the $K \times K$ matrix in square brackets on the RHS would yield a consistent least-square estimator of $\boldsymbol{b_K}$. However, as K increases, the dimension of this matrix increases and its inverse may become unstable. This is because K controls the number of parameters in the regression. It is possible to stabilize the inverse using regularization techniques (Carrasco et al., 2007). Ramsay and Silverman (2005) opt for the Tikhonov-regularized estimator given by

$$\hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}} = \left[\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)\sum_{t=1}^{T} z_t(p)^2\boldsymbol{\Phi_K}(p)^\top dp + \lambda \boldsymbol{I}_K\right]^{-1} \int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)\sum_{t=1}^{T} z_t(p)y_t(p)dp, \tag{3.13}$$

where the expectation terms in (3.12) are replaced by their empirical counterparts, $\lambda \geq 0$ is the regularization parameter and $\boldsymbol{I}_K$ is the K-dimensional identity matrix. As λ increases, the variance of the estimated function is reduced at the cost of an increased bias. The choice of this parameter value is important as it corresponds to the arbitrage between bias and variance. Furthermore, the penalty resulting from a positive value of λ imposes smoothness restrictions on β. It is easy to show that (3.13) is a penalized least square estimator corresponding to the solution of

$$\min_{\boldsymbol{b_{K,\lambda}}} ||T^{-1}\sum_{t=1}^{T}\left(y_t(p) - z_t(p)\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b_{K,\lambda}}\right)||^2 + \lambda||\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b_{K,\lambda}}||^2. \tag{3.14}$$

The estimator of β is thus a penalized sieves estimator (Chen, 2011) given by

$$\hat{\beta}_{K,\lambda}(p) = \boldsymbol{\Phi_K}(p)^\top \hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}}, \tag{3.15}$$

and depends on the choice of the sieves space (or basis system), the number of functions K and the regularization parameter λ. I will always use the B-spline basis system throughout the paper.[30] The choice of the order of splines, K and λ will be discussed later.

[30]Spline functions are piecewise polynomial functions. B-splines are basis splines that can be used to express any spline function as a series.

Ramsay and Silverman (2005) propose to extend this penalization to higher order derivatives in order to impose smoothness on the curvature of the functional parameter β. It can be useful if the economic theory suggests that the object of interest must satisfy specific differentiability restrictions. This is obtained by solving the minimization problem

$$\min_{\boldsymbol{b}} ||T^{-1}\sum_{t=1}^{T}\Big(y_t(p) - z_t(p)\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b_{K,\lambda}}\Big)||^2 + \lambda||L^m\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b_{K,\lambda}}||^2, \quad (3.16)$$

where L^m is the differential operator of order m, i.e. $L^m\phi(p) = \frac{d^m\phi(p)}{dp^m}$. It yields the solution in (3.13) though with the different penalty matrix $\boldsymbol{R_K} = \langle L^m\boldsymbol{\Phi_K}, L^m\boldsymbol{\Phi_K^\top}\rangle$ instead of $\boldsymbol{I_K}$.[31] Since $\boldsymbol{\Phi_K}$ is a vector of orthonormal basis functions, $\boldsymbol{R_K}$ corresponds to the identity matrix of dimension K only if $m = 0$. In the application, it will be necessary to impose the first-order differentiability of the functional parameter of interest, that is $m = 1$ will be chosen in all estimations.

In order to avoid overfitting the model, it is important to keep in mind the degree of freedom. Assuming each independent variable y_t is expanded onto K_y basis functions, and K_z dependent variables are used, then the degree of freedom is $T \times K_y - K \times K_z$.

Endogenous regressor

In many economic applications, like in the application to the New York state's electricity auctions presented in this paper, the regressor function may be suspected to be endogenous. That is the orthogonality condition may not be satisfied: $E[Z(p)U(p)] \neq 0$, for some p. This section extends the previous estimator to the instrumental variable (IV) setting.

Let us assume that the econometrician has a instrument W satisfying the exclusion restriction $E[W(p)U(p)] = 0\ \forall p$, and being correlated with the endogenous regressor such that $E[W(p)Z(p)] \neq 0\ \forall p$. Consider a similar expansion $\beta_K(p) = \boldsymbol{\Phi_K}(p)^\top\boldsymbol{b_K}$. The orthogonality condition yields

$$E[W(p)Y(p)] = E[W(p)Z(p)]\boldsymbol{\Phi_K}(p)^\top\boldsymbol{b_K}, \quad (3.17)$$

[31] Throughout the paper, I use the following notations: for $\boldsymbol{\Phi_K} = (\phi_1, ...\phi_K)^\top$ and $\boldsymbol{\Psi_L} = (\psi_1, ...\psi_L)^\top$ two vectors of functions, $\langle\boldsymbol{\Phi_K}, \boldsymbol{\Psi_L^\top}\rangle$ denotes the $K \times L$ matrix with (k, l) element $\langle\phi_k, \phi_l\rangle$.

for all $p \in \mathcal{P}$, which in turn implies that

$$E[Z(p)W(p)]^\top E[W(p)Y(p)] = E[Z(p)W(p)]^\top E[W(p)Z(p)]\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b_K}. \tag{3.18}$$

Thus, the expansion coefficients must satisfy the equality

$$\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)E[Z(p)W(p)]^\top E[W(p)Y(p)]dp = \left[\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)E[Z(p)W(p)]^\top E[W(p)Z(p)]\boldsymbol{\Phi_K}(p)^\top dp\right] \times \boldsymbol{b_K}. \tag{3.19}$$

The IV analog of the estimator in (3.13) is hence given by

$$\hat{\boldsymbol{b}}^{\boldsymbol{IV}}_{\boldsymbol{K},\boldsymbol{\lambda}} = \left[\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)\left(\sum_{t=1}^{T} z_t(p)w_t(p)\right)\left(\sum_{t=1}^{T} w_t(p)z_t(p)\right)\boldsymbol{\Phi_K}(p)^\top dp + \lambda \boldsymbol{I_K}\right]^{-1} \times \int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)\left(\sum_{t=1}^{T} z_t(p)w_t(p)\right)\left(\sum_{t=1}^{T} w_t(p)y_t(p)\right)dp, \tag{3.20}$$

as the solution of the penalized minimization problem

$$\min_{\boldsymbol{b}^{\boldsymbol{IV}}_{\boldsymbol{K},\boldsymbol{\lambda}}} ||T^{-1}\sum_{t=1}^{T} w_t(p)\Big(y_t(p) - z_t(p)\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b}^{\boldsymbol{IV}}_{\boldsymbol{K},\boldsymbol{\lambda}}\Big)||^2 + \lambda||\boldsymbol{\Phi_K}(p)^\top \boldsymbol{b}^{\boldsymbol{IV}}_{\boldsymbol{K},\boldsymbol{\lambda}}||^2. \tag{3.21}$$

The penalty matrix can be adapted as in the exogenous setting to penalize the m^{th} order derivative of the functional parameter. Remark that β may be identified even if W is a scalar-valued random variable, provided that $E[WZ(p)] \neq 0$ and $E[WU(p)] = 0$, for all p.

This model is closely related to Benatia et al. (2017), with an additional regularization scheme from the basis expansion. I do not prove the consistency of the estimator but I show how to compute the asymptotic covariance operator of β in the next section. Additional asymptotic results will be present in a future version of this work.

Asymptotic normality for fixed λ and K

In Section 3.5.3, I will present testing procedures on the shape of β. These procedures rely on estimates of its covariance operator $\mathcal{B}$. Let us denote $\boldsymbol{b_{K,\lambda}}$ the regularized expansion coefficients defined as

$$\boldsymbol{b_{K,\lambda}} = \left[\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)E[Z(p)^2]\boldsymbol{\Phi_K}(p)^\top dp + \lambda \boldsymbol{R_K}\right]^{-1}\left[\int_{\mathcal{P}} \boldsymbol{\Phi_K}(p)E[Z(p)^2]\boldsymbol{\Phi_K}(p)^\top dp\right]\boldsymbol{b_K}.$$

Let us write

$$\begin{aligned}\hat{\beta}_{K,\lambda} - \beta &= \hat{\beta}_{K,\lambda} - \beta_{K,\lambda} + \beta_{K,\lambda} - \beta \\ &= \boldsymbol{\Phi}_{\boldsymbol{K}}^{\top}(\hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}} - \boldsymbol{b}_{\boldsymbol{K},\boldsymbol{\lambda}}) + (\boldsymbol{\Phi}_{\boldsymbol{K}}^{\top}\boldsymbol{b}_{\boldsymbol{K},\boldsymbol{\lambda}} - \beta),\end{aligned} \tag{3.22}$$

where the second term is a bias that vanishes only when $\lambda \to 0$ and $K \to \infty$. Let us define the pointwise product of two functions as $(\phi_1 \cdot \phi_2)(p) = \phi_1(p)\phi_2(p)$, such that $V_Z = \int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p)E[Z(p)^2]\boldsymbol{\Phi}_{\boldsymbol{K}}(p)^{\top}dp = \langle \boldsymbol{\Phi}_{\boldsymbol{K}}, E[Z \cdot Z] \cdot \boldsymbol{\Phi}_{\boldsymbol{K}}^{\top}\rangle$ and $\hat{V}_Z$ denotes its empirical counterpart. Lemma 3 characterizes the $\sqrt{T}$-asymptotic behavior of $\hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}} - \boldsymbol{b}_{\boldsymbol{K},\boldsymbol{\lambda}}$.

Lemma 3 (Asymptotic normality of expansion coefficients)**.** *Assume K and λ fixed, (U, Z) i.i.d., $\mathcal{B}_{K,\lambda} < \infty$, $E||Z||^4 < \infty$, and $E||U||^2||Z||^2 < \infty$, then*

$$\sqrt{T}\Big(\hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}} - \boldsymbol{b}_{\boldsymbol{K},\boldsymbol{\lambda}}\Big) \xrightarrow{d} N(0, B_{K,\lambda}), \tag{3.23}$$

as $T \to \infty$.

The asymptotic distribution of $\hat{\beta}_{K,\lambda} - \beta_{K,\lambda}$ is characterized in Proposition 2.

Proposition 2 (Asymptotic normality)**.** *Assume K and λ fixed, (U, Z) i.i.d., $\mathcal{B}_{K,\lambda} < \infty$, $E||Z||^4 < \infty$, and $E||U||^2||Z||^2 < \infty$, then*

$$\sqrt{T}(\hat{\beta}_{K,\lambda} - \beta_{K,\lambda}) \xrightarrow{d} N(0, \mathcal{B}_{K,\lambda}), \tag{3.24}$$

as $T \to \infty$. The asymptotic covariance operator $\mathcal{B}_{K,\lambda}$ has a kernel $\sum_{k,l=1}^{K} B_{k,l}\phi_k(p)\phi_l(p')$ where $B_{k,l}$ is the (k,l)-element of the asymptotic covariance matrix $B_{K,\lambda}$ in Lemma 3.[32] *Its expression is given in the proof in the appendix.*

Statistical tests on $\hat{\beta}_{K,\lambda}$ require an estimator of the covariance operator. A practical estimator relies on the following approximation for λ small

$$\begin{aligned}B_{K,\lambda} =& E\Big[\langle (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1}\boldsymbol{\Phi}_{\boldsymbol{K}}, Z \cdot U\rangle\langle U \cdot Z, \boldsymbol{\Phi}_{\boldsymbol{K}}^{\top}(V_Z + \lambda \boldsymbol{R})^{-1}\rangle\Big] + o(\lambda) \\ \approx& (V_Z + \lambda \boldsymbol{R})^{-1}\int_{\mathcal{P}}\int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p)E\Big[Z(p)U(p)U(p')Z(p')\Big]\boldsymbol{\Phi}_{\boldsymbol{K}}^{\top}(p')dpdp'(V_Z + \lambda \boldsymbol{R})^{-1},\end{aligned} \tag{3.25}$$

[32] That is, for $\delta \in \mathbb{L}^2$ then $\mathcal{B}_{K,\lambda}\delta(p) = \int_{\mathcal{P}} \sum_{k,l=1}^{K} B_{k,l}\phi_k(p)\phi_l(p')\delta(p')dp'$.

which simplifies into

$$B_{K,\lambda} \approx (V_Z + \lambda \boldsymbol{R_K})^{-1} \int_{\mathcal{P}} \int_{\mathcal{P}} \boldsymbol{\Phi_K}(p) E\Big[Z(p)Z(p')\Big] \sigma^2(p,p') \boldsymbol{\Phi}_{\boldsymbol{K}}^{\top}(p') dp dp' (V_Z + \lambda \boldsymbol{R_K})^{-1}, \tag{3.26}$$

under the conditional homoskedasticity assumption: $E[U(p)U(p')|Z] = \sigma^2(p,p')$, which will be assumed throughout the analysis. The estimator is obtained by replacing the expectation terms by their empirical counterparts. I am now fully equipped to propose statistical tests based on functional regression estimates.

3.5.2 Two-sample functional tests

In this section, I introduce three functional tests of mean equality. I derive theoretical results and propose a numerical simulation to obtain guidance on which test should be preferred for empirical applications. Specifically, the tests will be used in the application to investigate the first-order condition under additive separability holds in expectations. In the following section, I will show how those testing procedures can be based on functional estimates from the above model.

Assume $Y, X \in \mathbb{L}^2(\mathcal{P})$. From (3.7) and the discussion in Section 3.4, let us define $Y(p) = p - C_i'(S_i(p))$ and $X(p) = (S_i(p) - QC_i)\frac{1}{-RD_i'(p)}$ for all $p \in \mathcal{P}$ for a given firm i. Assuming the firm behaves optimally on average and the AS restriction is satisfied then the null hypothesis

$$H_0 : E[Y(p) - X(p)] = 0, \quad \forall p \in \mathcal{P} \tag{3.27}$$

should hold. Suppose an i.i.d. sample of functions $\{y_t, x_t\}_{t=1}^T$ is available. I will consider three alternative tests following Carrasco et al. (2014).

Statistical tests

Let us first consider the Cramer-von Mises test statistic

$$CvM_T = ||T^{-1/2} \sum_t (y_t - x_t)||^2, \tag{3.28}$$

where $||\cdot||^2$ denotes the $\mathbb{L}^2$ norm.[33]

[33]The functional tests proposed here are based on the $\mathbb{L}^2$ inner product. There exist other functional tests based on different metrics, e.g. the Kolmogorov-Smirnov test uses the $\mathbb{L}^\infty$ norm.

An alternative testing strategy uses a test function $\psi \in \mathbb{L}^2(\mathcal{P})$ to obtain a standard distribution. Consider the test statistic

$$J_{1,T} = \frac{T^{-1/2}\langle \sum_t (y_t - x_t), \psi \rangle}{\langle \hat{C}\psi, \psi \rangle^{-1/2}} \tag{3.29}$$

where $\hat{C}$ is a consistent estimator of the asymptotic covariance operator of $T^{-1/2}\sum_t (y_t - x_t)$. This operator writes $C = E[(Y-X)\langle (Y-X), \cdot \rangle]$ and a consistent estimator of its kernel is given by

$$\hat{c}(p,p') = T^{-1}\sum_{t=1}^{T}\Big(y_t(p) - x_t(p)\Big)\Big(y_t(p') - x_t(p')\Big). \tag{3.30}$$

The power of $J_{1,T}$ can be improved by using a set of linearly independent test functions $\psi_1, \psi_2, ..., \psi_Q$ as given by

$$J_{Q,T} = V_Q^\top S_Q^{-1} V_Q, \tag{3.31}$$

where V_Q is a vector of length Q with q element $T^{-1/2}\langle \sum_t (y_t - x_t), \psi_q \rangle$ and S_Q is the $Q \times Q$ diagonal matrix with (q,q) element $\langle \hat{C}\psi_q, \psi_q \rangle$, with $q = 1, ..., Q$. Throughout the simulations and application, I will use the test functions defined as follows. Given Q, for any $q \in \{1, ..., Q\}$, ψ_q is defined over $[0,1]$ as

$$\psi_q(p) = \begin{cases} 1 & \text{if } p \in [\frac{q-1}{Q}, \frac{q}{Q}] \\ 0 & \text{otherwise.} \end{cases} \tag{3.32}$$

The test functions consist of Q partitions of equal length of the unit line.[34] Proposition 3 presents the asymptotic distributions of the test statistics.

Proposition 3 (Tests distribution)**.** *Under the assumptions* (Y, X) *i.i.d.,* $C < \infty$*,* $E||Y - X||^2 < \infty$*,* Q *is fixed and* $H_0 : E(Y - X) = 0$*, the test statistics are such that*

$$i) \quad CvM_T \overset{d}{\to} \sum_{l=1}^{+\infty} \lambda_l \chi_l^2(1), \tag{3.33}$$

where λ_l*'s are the eigenvalues of* C*, the asymptotic covariance operator of* $T^{-1/2}\sum_t (y_t - x_t)$*, and* χ_l^2 *denotes independent chi-square random variables,*

$$\begin{aligned} ii) \quad & J_{1,T} \overset{d}{\to} N(0,1), \text{ and} \\ iii) \quad & J_{Q,T} \overset{d}{\to} \chi^2(Q) \end{aligned} \tag{3.34}$$

as $T \to \infty$*.*

[34] Lehmann and Romano (2007) develop a chi-squared test of uniformity based on similar partitions.

Remark that the distribution of CvM_T is not standard. It is defined as the infinite sum of independent chi-square random variables weighted by the eigenvalues of the covariance operator C. The numerical implementation of this test is as follows. First, the eigenvalues of C are estimated by computing the empirical eigenvalues of $\hat{C}$ expressed as a square matrix for a discrete grid of p. Second, the eigenvalues are sorted in descending order. Finally, I use the numerical approach of Imhof (1961) to compute the p-values of this distribution.[35] The main idea is to truncate the eigenvalues below a given threshold to approximate the infinite sum. Numerical integration is performed using Simpson's rule.

It is important to evaluate the asymptotic power of those tests. Proposition 4 provides results on the consistency of the tests. It shows that CvM_T is the only test that is consistent against any alternative, whereas the two other statistics are not consistent against some specific alternatives that depend on the chosen test functions. Consider the following example to illustrate this result. Assume (Y,X) is a i.i.d sequence of $\mathbb{L}^2$ functions defined over the unit interval, such that $E[Y(p)] = 0$ for all p, $E[X(p)] = -1$ for $p < 1/2$ and $E[X(p)] = 1$ for $p \geq 1/2$. The test function $\psi(p) = 1$ for all p yields a test statistic $J_{1,T}$ that wrongly accepts $H_0 : E[Y - X] = 0$ as $T \to \infty$.

Proposition 4 (Tests consistency)**.** *Under the assumptions* (Y, X) *i.i.d.,* $C < \infty$*,* $E||Y - X||^2 < \infty$*,* Q *is fixed, and* $H_1 : E(Y - X) \neq 0$*, as* $T \to \infty$*,*

i) CvM_T *is consistent against all alternatives,*

ii) For any ψ*,* $J_{1,T}$ *is only consistent against alternatives such that* $E[\langle Y, \psi\rangle] \neq E[\langle X, \psi\rangle]$*.*

iii) For any $\psi_1, ..., \psi_Q$*,* $J_{Q,T}$ *is only consistent against alternatives such that* $E[\langle Y, \psi_q\rangle] \neq E[\langle X, \psi_q\rangle]$*, for at least one* $q \in \{1, ..., Q\}$*.*

In the limiting case where $Q \to \infty$, the test $J_{Q,T}$ is consistent against any alternative. However, increasing Q too much decreases the limiting power. If Q is large then the limiting power is equal to the size of the test.[36] Therefore, the choice of Q results in a trade-off between asymptotic consistency and power.

---

[35] I develop a MATLAB counterpart of Farebrother's (1990) PASCAL implementation of this method.

[36] See Lemma 14.3.1 and the following discussion in Lehmann and Romano (2007) on uniformity tests.

Simulations for two-sample tests

This section presents numerical simulations aimed at investigating the empirical size and power of the tests. Under H_0, the functions are specified as

$$\begin{aligned} Y_t(p) &= Z_{0,t}(p) + U_{Y,t}(p) \\ X_{0,t}(p) &= Z_{0,t}(p) + U_{X,t}(p) \end{aligned} \tag{3.35}$$

for $p \in \mathcal{P} = [0,1]$. Under the alternative Y_t is unchanged but X_t becomes

$$X_{1,t}(p) = Z_{1,t}(p) + U_{X,t}(p), \tag{3.36}$$

where

$$Z_{0,t}(p) = \frac{\Gamma(\alpha_{0,t} + \beta_{0,t})}{\Gamma(\alpha_{0,t}) + \Gamma(\beta_{0,t})} p^{\alpha_{0,t}-1}(1-p)^{\beta_{0,t}-1} + \eta_t \tag{3.37}$$

with $\alpha_{0,t} \sim iid\ U[3,6]$, $\beta_{0,t} \sim iid\ U[1,4]$, and $\eta_t \sim iid\ N(0,1)$ for all $t = 1,...,T$. $Z_{1,t}$ is specified as (3.37) although with $\alpha_{1,t} \sim iid\ U[2.7, 5.7]$ and $\beta_{1,t} \sim iid\ U[1.5, 4.5]$. These functions are probability density functions of random beta distributions with an additive Gaussian error term. The functions $U_{Y,t}(p)$ and $U_{X,t}(p)$ correspond to independent Ornstein-Uhlenbeck processes with zero mean and mean reversion rate $\theta = 10$.[37] It is described by the differential equation $dU(p) = -\theta_u U(p)dp + \sigma_u dG_u(p)$ for $p \in [0,1]$ and G_u is a Wiener process where $\sigma = 1$ is the standard deviation of its increments $dG_u(p)$.

This simulation design evaluates the power of the tests CvM_T, $J_{1,T}$, $J_{2,T}$, $J_{5,T}$ and $J_{10,T}$ in terms of their ability to detect average differences in the shape (α_0 and α_1) and scale (β_0 and β_1) of the random gamma densities. The numerical simulation is performed with $10,000$ random draws of functions evaluated on the pseudo-continuous interval $[0,1]$ of 100 equally-spaced discrete steps, for 4 sample sizes: $T = 50$, $T = 100$, $T = 200$ and $T = 500$. Table 3.3 presents the empirical size and power.

[37] Simulation results remain similar with more persistent error processes.

Test	$T=50$ Size	$T=50$ Power	$T=100$ Size	$T=100$ Power	$T=200$ Size	$T=200$ Power	$T=500$ Size	$T=500$ Power
CvM	4.4	20.9	4.8	53.4	5.2	94.5	5.0	100.0
J_1	4.5	4.7	4.9	5.0	5.1	5.0	5.1	4.7
J_2	6.8	12.1	7.0	24.3	7.0	62.8	7.0	99.8
J_5	11.1	32.8	10.9	67.8	11.2	97.5	11.1	100.0
J_{10}	14.6	46.7	14.5	81.9	14.9	99.3	14.9	100.0

NOTES. This table reports empirical size and power (in percentage points) of the proposed statistical tests for a nominal test size of 5%. Power is evaluated at the empirical size, hence is not size-corrected.

Table 3.3: Simulations results

Results suggest that CvM_T is correctly sized and has the most power with respect to the other tests under this alternative. $J_{1,T}$ has correct size but is not powerful even for large sample sizes. The tests based on multiple test functions are oversized and have relatively good power. Theoretical and numerical results suggest that CvM_T should be preferred in applications.

3.5.3 Additive separability tests

In this section, I show how those functional testing procedures, as well as multivariate alternative methods, may be used to investigate the additive separability hypothesis. A numerical simulation is then presented to compare the empirical size and power of the test of Hortaçsu and Puller (2008) with respect to the functional methods.

The additive separability of firm i's supply function S_i in its private information QC_i holds if and only if $\frac{\partial S_i(p, QC_i)}{\partial p}$ does not depend on QC_i. Let us drop subscript i. Assume $S \in \mathbb{L}^2(\mathcal{P})$, $QC \in \mathbb{R}$ and suppose an i.i.d. sample $\{S_t, QC_t\}_{t=1}^T$ is available. The tests will be based on the null hypothesis

$$H_0 : E\Big[\frac{\partial^2 S(p, QC)}{\partial p \partial QC}\Big] = 0 \quad \forall p \in \mathcal{P}. \tag{3.38}$$

Hortaçsu-Puller test

To test this, Hortaçsu and Puller (2008) propose the following two-step procedure.

In a first step, each observed supply function S_t is linearized by regressing the quantities $\{S_t(p_0), ..., S_t(p_L)\}$ over a grid of prices $\{p_0, ..., p_L\}$ as in

$$S_t(p_l) = a_t + b_t p_l + u_l. \tag{3.39}$$

The slope coefficients are collected to form a sample $\{b_t\}_{t=1}^{T}$. In a second step, these slope coefficients are regressed onto contract quantities QC_t using

$$b_t = c + dQC_t + v_t. \tag{3.40}$$

The Hortaçsu-Puller (HP) test of the AS restriction consists in testing the null hypothesis: $H_0 : d = 0$ using a t-statistic, henceforth denoted t_{HP}.

This testing procedure relies on the assumption that supply functions are linear.

Penalized spline smoothing

Alternatively, it is possible to smooth the functions so as to obtain their derivatives, and then use the functional regression model to regress the derivative functions onto the variable of interest.

There is a handful of methods such as kernel smoothing or local polynomial regressions which may be used to compute the derivative functions. However, penalized spline smoothing is a more convenient smoothing technique when used prior to functional data analysis.[38] It consists in using splines to solve for the smooth function $\tilde{S}(\cdot)$ which minimizes the sum of the integrated mean squared error (IMSE) and a roughness penalty term, as given by

$$\min_{\tilde{S}} ||S(p) - \tilde{S}(p)||^2 + \lambda ||L^m \tilde{S}||^2, \tag{3.41}$$

where λ is a smoothing parameter. Choosing $m = 1$ means that the penalty is used to enforce the smoothness of the first derivative of $\tilde{S}$. This minimization problem is then solved in two steps. First, the function is expanded onto a set of B-spline basis functions[39] $\{\phi_1, ..., \phi_K\}$ as

$$\tilde{S}(p) = \sum_{k=1}^{K} c_k \phi_k(p) = \boldsymbol{\phi}_{\boldsymbol{K}}^{\top} \boldsymbol{c}_{\boldsymbol{K}}, \tag{3.42}$$

[38] A thorough treatment of penalized spline smoothing is presented in Ramsay and Silverman (2005). They briefly discuss the advantage of the method at the beginning of Chapter 5.

[39] The interested reader is referred to Wahba (1990) for an introduction to spline models.

where $\boldsymbol{c_K}$ and $\boldsymbol{\phi_K}$ are the K-vectors of expansion coefficients and basis functions, respectively. If S is observed over a grid $\{p_0, ..., p_L\}$, then (3.41) can be approximated in a second-step as

$$\min_{\boldsymbol{c_K}} (\boldsymbol{S} - \boldsymbol{\Phi_K c_K})^\top (\boldsymbol{S} - \boldsymbol{\Phi_K c_K}) + \lambda \boldsymbol{c_K^\top R_K c_K}, \tag{3.43}$$

where $\boldsymbol{S}$ is a L-vector with elements $S(p_l)$, $\boldsymbol{\Phi_K}$ is a $L \times K$ matrix with (l,k)-element $\phi_k(p_l)$ and $\boldsymbol{R_K}$ is a $K \times K$ penalty matrix with (k_1,k_2)-element $\langle L^m \phi_{k_1}, L^m \phi_{k_2} \rangle$. This minimization problem has the least-square solution

$$\boldsymbol{\hat{c}_K} = (\boldsymbol{\Phi_K^\top \Phi_K} + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K^\top S}, \tag{3.44}$$

which in turn yields the smoothed function $\hat{S}(p) = \boldsymbol{\phi_K^\top \hat{c}_K}$ and its first-order derivative given by

$$\hat{S}'(p) = \sum_{k=1}^{K} \hat{c}_k \phi'_k(p). \tag{3.45}$$

This method necessitates four choices: the basis system, the number of basis functions, the regularization parameter λ and the order of the differential operator. Ramsay and Silverman (2005) recommend to penalize the $m+2$ derivative when the m derivative is the object of interest. Hence, I will impose a penalty on the third-order derivative and use B-spline basis functions of order four. The number of basis and the regularization parameter will be chosen graphically.[40] Remark that penalizing the second-order derivative with large value of λ yields linearized functions similar to those used in the HP test. In practice, a small λ should be preferred.

Functional tests

The first step of all of the proposed tests consists in estimating $\hat{S}'_t$ for all t to construct a sample $\{\hat{S}'_t\}_{t=1}^T$. In a second step, consider the functional linear regression model

$$\hat{S}'_t(p) = \alpha(p) + \beta(p) QC_t + \epsilon_t(p) \tag{3.46}$$

where β is a functional parameter of interest and ϵ_t is a functional error term. QC_t can be interpreted as a constant function of p. The functional estimator presented earlier will be applied to (3.46).

[40] Alternatively, one could use data-driven selection methods such as leave-one out or generalized cross-validation.

In this case, testing the AS restriction amounts to specify

$$H_0 : \beta(p) = \beta_0(p), \quad \forall p \in \mathcal{P}, \tag{3.47}$$

where $\beta_0(p) = 0$ for all p.[41]

The second step consists in testing this hypothesis using the test statistics presented in Section 3.5.2. Formally, H_0 can be tested using

$$CvM_T = ||T^{1/2}(\hat{\beta}_{K,\lambda} - \beta_0)||^2 \xrightarrow{d} \sum_{l=1}^{+\infty} \lambda_l^{\mathcal{B}_{K,\lambda}} \chi_l^2(1), \tag{3.48}$$

$$J_{1,T} = \frac{T^{-1/2}\langle \hat{\beta}_{K,\lambda} - \beta_0, \psi \rangle}{\langle \hat{\mathcal{B}}_{K,\lambda}\psi, \psi \rangle^{-1/2}} \xrightarrow{d} N(0,1), \tag{3.49}$$

and

$$J_{Q,T} = V_Q^\top S_Q^{-1} V_Q \xrightarrow{d} \chi^2(Q), \tag{3.50}$$

where V_Q is a vector of length Q with elements $T^{-1/2}\langle \hat{\beta}_{K,\lambda} - \beta_0, \psi_q \rangle$ and S_Q is the $Q \times Q$ diagonal matrix with elements $\langle \hat{B}\psi_q, \psi_q \rangle$, with $q = 1, ..., Q$. The results of Proposition 3 and Proposition 4, and the discussion with respect to the trade-off implied by the choice of Q are still valid in this case.

Wald-type tests

It is also possible to construct Wald-type tests based on $\hat{\beta}_{K,\lambda}$. Consider the M-vector of point estimates evaluated at equally-spaced points, then the Wald statistic in

$$\begin{aligned} W_{M,T} =& (\hat{\beta}(p_1) - \beta_0(p_1), ..., \hat{\beta}(p_M) - \beta_0(p_M))\hat{\Sigma}_M^\dagger(\hat{\beta}(p_1) - \beta_0(p_1), ..., \hat{\beta}(p_M) - \beta_0(p_M))^\top \\ &\xrightarrow{d} \chi^2(M) \end{aligned} \tag{3.51}$$

consists of an alternative testing strategy for H_0, where $\hat{\Sigma}_M^\dagger$ is the Moore-Penrose generalized inverse of the estimated covariance matrix of the vector of M parameters. As M grows, the estimated covariance matrix becomes singular – the limit case being when M is infinite – and its inverse is not stable. The singularity of this matrix comes from the correlation across parameters. Hence a feasible test would consist in choosing a small value for

[41] Remark that the regularized null hypothesis is the same since $\beta_{K,\lambda,0} = \beta_0$ in this specific case.

M, although the spacing between values of p remains ad-hoc. The generalized inverse yields a more stable estimate.[42]

The choice of M results in trade-off between consistency and power similar to that of Q. The result of Proposition 4 for $J_{Q,T}$ can be easily adapted to $W_{M,T}$. That is, $W_{M,T}$ is a consistent test against all alternatives such that $\beta_1(p_m) \neq \beta_0(p_m)$ for at least one m. Yet, as M increases and becomes large, the asymptotic power of the test is approximately equal to its size (Lehmann and Romano, 2007).

Simulations of additive separability tests

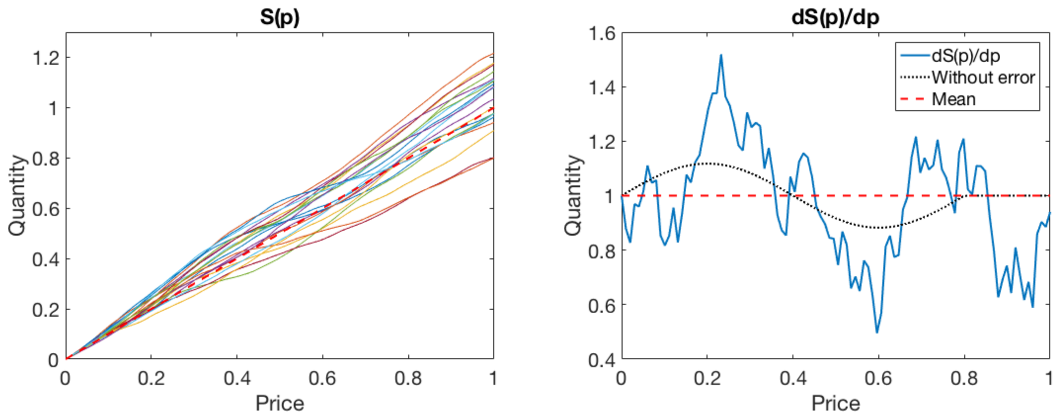
The test based on linearized functions may have low power against a variety of important alternatives. Let us consider the following specification of the derivative function:

$$S'_t(p) = 1 + \beta_0(p)QC_t + \epsilon_t(p), \tag{3.52}$$

where $\beta_0(p) = 0$ under H_0 and $\beta_0(p) = 0.5\sin(2.5p \times \pi)$ if $p \in [0, 0.8]$, and $\beta_0(p) = 0$ if $p \in (0.8, 1]$ under the alternative H_1. Assume that $QC_t = 0.3 + 0.1 \times \xi_t$, with ξ_t being a standard normal error, and ϵ_t is an Ornstein-Uhlenbeck process, as described in Section 3.5.2, with zero mean, mean reversion rate equal to 10 and standard deviation of the increments of the Wiener process equal to 1. Figure 3.3a shows a sample of 20 simulated supply functions.

It may seem natural to use linearization since the functions look relatively linear. The simulations will prove that to be a mistake. Figure 3.3b displays an example of the simulated derivative function $S'(p)$ under the alternative. The dotted line corresponds $1+\beta_0(p)QC_t$ under H_1. It illustrates the assumed effect of QC onto $S'(p)$ under the alternative. This specification is motivated by the empirical results obtained in Section 3.6.2.

[42] A presumably better approach is proposed by Dufour and Valéry (2016). They develop regularized Wald-type tests for inference with a singular covariance matrix in a general context. Their tests are based on a spectral cut-off-type estimator of the inverse covariance matrix. I do not pursue this approach here.



(a) Sample of 20 simulated functions $S(p)$ (b) An illustrative simulation of $S'(p)$

Figure 3.3: Simulated sample under H_1

Once the derivative functions are generated, they are numerically integrated to recover the functions S that serve as observations. The functional testing procedures begin by estimating the derivative using penalized spline smoothing with 30 spline functions of order 4 and a penalty $\lambda = 10^{-7}$ on the third order derivative. Then, the functional linear regression model is estimated to obtain $\hat{\beta}$ using 25 splines of order 3 and $\lambda = 0$.

The simulation results for $10,000$ replications and 4 sample sizes are presented in Table 3.4. The table shows empirical size and power. The integrated mean square error (IMSE) of the estimate functional parameter is shown in the last row. It is defined as

$$IMSE(\hat{\beta}) = T^{-1} \sum_{t=1}^{T} ||\hat{\beta}_t - \beta||^2. \tag{3.53}$$

The values illustrate the consistency of the estimator. The simulated size and power suggest that t_{HP} has very low power against this alternative, whereas all other tests based on the functional estimates exhibit some reasonable power. This is because t_{HP} has no power against alternatives in which the effect of QC on the linearized function integrates to zero over the range of prices. Similarly to the previous section, J_1 has low power, but the other functional tests perform well in terms of power. The size of the tests based multiple test functions J_Q, and the Wald-type tests W_M, appear to converge to the power as Q and M increase. All the results confirm that the test based on CvM_T should be preferred in applications.

Test	$T = 50$		$T = 100$		$T = 200$		$T = 500$	
	Size	Power	Size	Power	Size	Power	Size	Power
CvM	6.5	55.5	6.5	87.4	6.9	99.6	7.0	100.0
J_1	5.6	5.6	5.1	5.1	5.4	5.4	5.2	5.2
J_2	6.2	34.7	5.8	61.0	5.7	90.8	5.6	100.0
J_5	9.5	60.3	8.7	88.6	8.3	99.6	8.3	100.0
J_{10}	16.0	73.4	14.6	94.4	14.1	99.9	14.1	100.0
W_5	3.7	38.9	2.6	69.8	2.1	97.4	1.6	100.0
W_{10}	60.9	87.3	51.4	95.3	46.2	99.8	43.9	100.0
W_{30}	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
W_{all}	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
t_{HP}	2.5	0.9	2.3	0.6	2.8	0.3	2.5	0.1
	H_0	H_1	H_0	H_1	H_0	H_1	H_0	H_1
IMSE	0.07	0.07	0.03	0.03	0.02	0.02	0.01	0.01

NOTES. This table reports empirical size and power (in percentage points) of the proposed statistical tests for a nominal test size of 5%. Power is evaluated at the empirical size, hence is not size-corrected.

Table 3.4: Simulations results

3.5.4 Estimation of firm-level market power and optimal bids

If private information is found to have a significant effect on bidding strategies, then one needs an estimation procedure for the ratio of partial derivatives in (3.4) in order to understand empirical bidding behavior.

This section proposes an alternative benchmark for the optimal bidding behavior that is not based on the AS restriction, but rather on a first-order approximation of the effect of each firm's supply function on the daily market-clearing price distribution. First, I present the estimation procedure. Then I discuss identification considerations.

Estimation procedure

The estimation procedure consists of three steps. First, I estimate the daily cumulative distribution functions (CDF) of hourly prices. Second, the CDFs are regressed onto supply functions and controls. Third I use the resulting estimates to compute the optimal supply functions identified by the optimality condition (3.4) for a given marginal cost function. The construction of the marginal cost function is discussed in Section 3.6.1.

Estimation of the CDFs of prices. For every hour of each day in the dataset, one would like to observe the CDF of equilibrium prices. Two difficulties arise. First, the hourly CDFs are not observable. However, firms hourly bid functions are essentially constant in any given day in the New York's DAM.[43] I assume that each firm submits a single bid function per day, to obtain a single estimate of the distribution function per day by pooling hourly prices. Second, firms that operate facilities located apart from one another may receive different prices due to congestion rents, as explained in Section 3.3. In order to deal with this latter difficulty, I assume that a firm with generation facilities at various locations is paid the average price between the different locational prices, weighted by its installed capacity at each location.[44]

Suppose one has at hands a sample of T daily samples of hourly prices $p_{h,t}$, with $h = 1, ..., 24$, T daily supply functions S_t and vectors of (possibly functional) control variables $\boldsymbol{Z}_t$. The estimation is as follows.

Define a discretized grid $\mathcal{P}$ of 1000 equally-spaced prices in $[-1000, 1000]$. This grid will be used whenever the functions need to be evaluated numerically. For each day $t = 1, ..., T$, I estimate the price distribution function $F_t(p)$ with the sample of 24 hourly prices $\{p_{1,t}, ..., p_{24,t}\}$ using the kernel estimator of the distribution function given by

$$\hat{F}_t(p) = \frac{1}{24} \sum_{h=1}^{24} \int_{-1000}^{p} G(\frac{q - p_{h,t}}{b}) dq, \qquad (3.54)$$

where G is a Gaussian kernel and b is a bandwidth parameter.[45] This allows to collect the associated density functions $\hat{f}_t$ corresponding to (3.6) which will be necessary to estimate optimal bid functions. In the application, I will use

[43]For example, for Astoria Energy, the IMSE between the supply function at 6 pm and at all other hours for each day is $\frac{1}{1095}\sum_{t=1}^{1095}\frac{1}{23}\sum_{h\neq 18}\frac{1}{\overline{P}-\underline{P}}\int_{\underline{P}}^{\overline{P}}(\frac{S_{t,18}(p)}{K_{t,18}} - \frac{S_{t,h}(p)}{K_{t,h}})^2 dp < 10^{-4}$.

[44]For computational convenience, I weight zonal prices according to the firm's respective installed capacity in each zone instead of using locational prices directly.

[45]I use kernel estimation methods (Reiss, 1981; Hill, 1985) rather than the maximum penalized likelihood method based on the empirical distribution function (Silverman, 1982) as the latter although useful when dealing with functional data requires an iterative procedure that is more computationally demanding.

$b = 10$.[46] Under the assumption that firms choose a unique bid function per day, these daily distribution functions can be interpreted as being conditional on the supply function and private information. This results in T daily conditional distribution functions $\hat{F}_t$ and density functions $\hat{f}_t$.

Estimation of the market power parameter. Assume that $F_t, S_t \in \mathbb{L}^2$ are function-valued random elements. A first-order approximation of $F_t(p|S_t(p))$ around the equilibrium bid $S_t^\star(p)$ gives

$$F_t(p|S_t(p)) = \left(F_t(p, S_t^\star(p)) - \frac{\partial F_t(p|S_t^\star(p))}{\partial S(p)} S_t^\star(p)\right) + \frac{\partial F_t(p|S_t^\star(p))}{\partial S(p)} S_t(p) + \eta_t(p), \tag{3.55}$$

where $\eta_t(p)$ denotes an approximation error. Assuming further that $E\left[\eta_t(p)|S_t\right] = 0$ yields

$$E\left[F_t(p|S_t(p))|S_t\right] = \overline{F}_t(p) + \pi_t(p)S_t(p), \tag{3.56}$$

where the *intercept* function $\overline{F}_t(p)$ denotes the term between parentheses in (3.55) and $\pi_t(p) = E[\frac{\partial F_t(p|S_t^\star(p))}{\partial S(p)}]$ is the functional parameter of interest. The marginal effect of S_t hence corresponds to the average effect of a change of supplied quantity around the equilibrium onto the probability distribution function. As $S_t(p)$ increases, the firm offers an increased quantity at price p which should increase the probability to observe price realizations below or equal to p, i.e. it should be $\pi(p) \geq 0$. Remark that if $S_t(p) \in [0, 1]$, then $\pi(p)$ gives the probability change for an increased offered quantity of 1% of the firm's available capacity.

Firm-level market power may not be constant over a long time period, hence pooling all days for three years may not yield meaningful estimates. Considering that it is relatively stable over a month, I propose to estimate π_t for every day using a rolling window of 31 days. For each day $t = 16, ..., T - 16$, I estimate the functional linear regression[47] model

$$\hat{F}_d(p) = \pi_t(p)S_d(p) + \boldsymbol{\gamma}_t(p)\boldsymbol{Z}_d(p) + U_d(p), \tag{3.57}$$

using the sample $\{\hat{F}_d, S_d, \boldsymbol{Z}_d\}_{d=t-15}^{d=t+15}$ constructed by pooling functions from day $t - 15$ to $t + 15$. $\boldsymbol{Z}_d$ represents (possibly functional) control variables

[46] This choice has no important effect on the resulting estimates.

[47] An alternative approach consists in estimating the fully functional model $\hat{f}_d(p) = \int \pi_t(p, q)S_d(q)dq + \boldsymbol{\gamma}_t(p)\boldsymbol{Z}_d(p) + U_d(p)$ using the procedure of Benatia et al. (2017) although it would require to estimate a much larger number of parameters. I do not pursue this approach here.

observed by the firm and that affect the price distribution. This results in $T - 30$ functional estimates $\hat{\pi}_t$, which consist of estimates of the partial derivative (3.5) as

$$\partial F_t \widehat{(p|S_t)} / \partial S_t = \hat{\pi}_t(p). \tag{3.58}$$

Scaling, smoothing and penalties. I now discuss practical considerations for estimation. First, it is crucial to rescale the variables in high-dimensional statistical models. Functional regression models are no exception. Since distribution functions take values in $[0, 1]$, I divide all variables by their maximum value in the sample so that they are defined on a similar scale. For example, a firm's supply functions are rescaled as

$$S_t^{scaled}(p) = \frac{S_t(p)}{K_{max}}, \tag{3.59}$$

where K_{max} is the maximum capacity offered by the firm in the sample.

Second, empirical supply functions are step functions hence non-differentiable. This is typically a problem if one is interested in the counterfactual price distribution. Supply functions are smoothed using penalized spline smoothing prior to estimation in order to ensure that any counterfactual $F(\cdot)$ is differentiable. The same reason motivates the choice of penalties on the first-order derivatives of the parameters. Note that this penalty is not sufficient for a counterfactual $F(\cdot)$ to be a CDF.

Third, the regularization parameter used in the penalty term is chosen using hold-out validation. There exist many methods to choose tuning parameters, such as generalized cross-validation or K-fold cross-validation, among others. Some of them are discussed in Benatia et al. (2017). Hold-out validation consists in choosing the value of λ_t that satisfies

$$\min_{\lambda_t \geq 0} ||\hat{F}_t - \tilde{\hat{F}}_t^{\lambda_t}||^2, \tag{3.60}$$

where $\tilde{\hat{F}}_t^{\lambda_t}$ is the predicted function based on the estimation sample left without observation at t: $\{\hat{F}_d, S_d, \boldsymbol{Z}_d\}_{d=t-15,\ldots,t-1,t+1,\ldots,t+15}$. The main advantage of hold-out CV is to be computationally fast while being able to target the function of interest.

Estimation of the optimal supply function. For each t, I solve for the optimal supply function $S_t^\star$ satisfying the first-order condition

$$p - c_t'(S_t^\star(p)) = (S_t^\star(p) - QC_t)\frac{\hat{\pi}_t(p)}{\hat{f}_t(p)}. \tag{3.61}$$

The empirical cost function may however not satisfy the conditions for the solution of (3.61) to hold for all p under the monotonicity constraint. I consider a recursive solution method that allows to enforce the monotonicity constraint while being able to solve the problem quickly over a given grid $\mathcal{P}$. The problem is solved separately for quantities above and below QC, as described in Algorithm 1 in the appendix. The same algorithm is used to compute the optimal supply function proposed by Hortaçsu and Puller (2008), denoted $S^{\star HP}$, by substituting $\frac{\hat{\pi}_t(p)}{\hat{f}_t(p)}$ with $\frac{1}{-RD_t'(p)}$ in the objective function. The computation of standard errors is described in appendix C.2. The procedure neglects the randomness of $\hat{f}(\cdot)$[48] and the asymptotic covariance operator of $\hat{\pi}$ is subject to a regularization bias. The resulting confidence interval should thus be interpreted with caution.

Having set the method to estimate the optimal bid function, I discuss identification issues in the next section.

Identification

There are two principal identification issues that need to be addressed: the endogeneity of supply functions, and the lack of identification for prices at both tails of the distribution.

Endogeneity. An important difficulty arises from the endogeneity of the supply function with respect to the distribution of equilibrium prices. Strategic firms naturally have anticipations about this distribution when choosing their supply function. F and S_i are hence simultaneously determined in equilibrium. This can be observed as follows. Suppose that firm i knows its own π_t and has anticipations ξ_t that are unobservable to the econometrician. Neglecting the control variables, the distribution function is

$$F_t(p) = \pi_t(p)S_t(p) + \xi_t(p) + U_t(p), \tag{3.62}$$

[48] $\hat{f}(\cdot)$ takes small values on the tails of the price distribution, hence a Fieller confidence interval may offer yield better coverage probabilities.

where U_t is a functional error term that represents approximation and prediction errors. Those are assumed to be mean-zero and unobserved by the firm so that $E[S_t(p)U_t(p)] = 0$. The probability density of p anticipated by the firm is thus $f_t(p) = \pi_t'(p)S_t(p) + \pi_t(p)S_t'(p) + \xi_t'(p)$. Substituting this expression in the first-order condition gives

$$p - c_t'(S_t(p)) = (S_t(p) - QC_t)\frac{\pi_t(p)}{\pi_t'(p)S_t(p) + \pi_t(p)S_t'(p) + \xi_t'(p)}, \tag{3.63}$$

which makes clear that S_t and ξ_t are correlated in equilibrium.

The sign of this correlation is ambiguous. An anticipated positive shock $\xi_t(p)$ on the distribution function at price p means that prices equal or below p are more likely. The firm faces a trade-off: increase its supply at p to enlarge the expected volume of sales, or decrease it so as to reduce the expected price drop through by shifting the price distribution. If the latter effect prevails, then one should expect that the equilibrium supply function is negatively correlated to this error term.

Therefore, the sign and magnitude of the bias of the estimates obtained when neglecting this endogeneity will depend crucially on this correlation. If negative, then the estimated parameter $\hat{\pi}_t$ with expansion coefficients given by

$$\begin{aligned} \hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}} =& \left[\int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p) \sum_{t=1}^{T} S_t(p)^2 \boldsymbol{\Phi}_{\boldsymbol{K}}(p)^\top dp + \lambda \boldsymbol{I}_K \right]^{-1} \\ & \times \int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p) \left(\sum_{t=1}^{T} S_t(p) F_t(p)(p) + \sum_{t=1}^{T} S_t(p)\xi_t(p) \right) dp. \end{aligned} \tag{3.64}$$

will be too small with respect to its true value.

A *valid* instrument would affect the price distribution only through the firm's supply function, and not through the unobserved shocks in ξ_t. I consider the firm's contracted sales QC_t and lagged supply function S_{t-1} as possible instruments. When choosing its supply function, both are taken as sunk decisions by the firm. Additionally, they have no direct effect onto the current day's realizations of equilibrium prices. Most importantly, they belong to the firm's private information set, and therefore cannot be used by rival firms to condition their bidding strategies. Therefore, I consider that the two orthogonality conditions $E[QC_t(p)\xi_t(p)] = 0$ and $E[S_{t-1}(p)\xi_t(p)] = 0$ are satisfied.

Beside these exclusion restrictions, the two instruments may be considered as satisfying an *inclusion restriction* (Reiss and Wolak, 2007), that is $QC_{i,t}$ and $S_{i,t-1}(p)$ belong to the equation that characterize the endogenous variable $S_{i,t}(p)$. As for forward positions, the optimality condition establishes that supply functions depend on the forward quantity. If the AS restriction is violated then contracted sales correlate with the slope of supply functions. This may provide sufficient variations to identify the parameter of interest. Furthermore, the lagged supply function embeds unobserved information about firm-specific costs which can be considered as exogenous shifters of the supply bids. It provides a potentially strong functional instrumental variable as it should correlate sufficiently with the current period's supply function even under the AS restriction.[49]

No identification around prices at the tails of the distribution. It should be clear that π is not identified for prices at the tails of the distribution function. Small prices that never occur will always be assigned a value $F_t(p) = 0$, whereas large prices that never occur will be such that $F_t(p) = 1$, for all t. The absence of sufficient variations in the distribution function around those values prevents identification.[50] Thus, I estimate the model for prices between the lowest and highest percentiles of the price distribution obtained from pooling hourly prices for the entire dataset. This yields a discrete grid of prices $\mathcal{P}_2$ that is consistent across all estimations for a given firm.

Nevertheless, this grid may be too large to derive the daily optimal supply function. Calculating the optimal supply function of day t requires the density function $\hat{f}_t$ and $\hat{\pi}_t$. Remark that $\hat{f}_t(p)$ being in the denominator in the optimality condition, it cannot be equal to zero. I define a daily price grid $\mathcal{P}_t = [\overline{p}_t, \underline{p}_t]$ such that $\underline{p}_t = \min\{p \in \mathcal{P} | \hat{f}_t(p) > 10^{-4}\}$ and $\overline{p}_t = \max\{p \in \mathcal{P} | \hat{f}_t(p) > 10^{-4}\}$. The analysis of estimates and optimal bids is done with respect to this range $\mathcal{P}_t$.

[49] Alternative instrumental variables may be the capacity commitments from the capacity markets. However, complementary bidding mechanisms, such as minimum revenue requirements, directly affect the determination of equilibrium prices and hence are not valid instruments.

[50] This is also true if the firm's supply functions are kept constant over the estimation sample.

3.6 Application to the New York electricity market

One of the most important concern in liberalized electricity markets is the exercise of producer market power. I use the above methods to evaluate the bidding strategies of suppliers in the New York day-ahead electricity market using rich data on bids and costs.

This section is organized as follows. Section 3.6.1 presents the data. It provides details on the bid de-anonymization algorithm, the construction of firm-level marginal cost functions, descriptive statistics on bids and the estimation of market-clearing price CDF. Section 3.6.2 presents the empirical results, including the tests of optimal bidding under additive separability, the tests of the additive separability restriction, the estimated measures of market power, and finally the resulting optimal bids and counterfactual expected profits. I also present a non-parametric approach to the estimation of marginal costs under optimal bidding. Section 3.6.3 concludes the application.

3.6.1 Data

This study uses data on firm-level energy bids, marginal production costs, and market-clearing prices and quantities. I construct a novel dataset with daily information on bids and plants in the New York State for the period 2013-2015. The collection of this data and the construction of firm-level datasets require significant efforts. I identify two main difficulties.

First and foremost, energy bids are anonymized by the NIYSO to prevent anti-competitive bidding behavior. Each generating unit or facility participating to the auction process has a time-consistent *Masked Gen ID*, while each bidder submitting bids for generating units has a time-consistent *Masked Bidder ID*. Anonymized bid data are publicly disclosed on NYISO website three months later.[51] Firms often have multiple generating units and bidders. Uncovering the identity of the company behind each anonymous identification number may thus appear as a daunting task. Yet, I will show that it can be done with reasonable accuracy using a matching algorithm based on publicly available information provided by the NYISO on existing generating facilities. Bid anonymity should hence not hinder the collection of strategic

[51] Accessible at `http://mis.nyiso.com/public/P-27list.htm`.

information about costs and past bids on competitors. However, the late disclosure of bids ensures that recent bids and forward contracts remain private information.

Second, plant-level marginal costs may, in principle, be considered as valuable private information for producers. Nevertheless, the Energy Information Administration (EIA) and the Environmental Protection Agency (EPA) provide extensive data on regulated power plants in the United States. This information can be used to construct estimates of marginal cost functions once matched to the data on existing generating facilities in New York.

Interestingly, those two difficulties have been the subject of a recent ruling of the New York Public Service Commission (PSC). The PSC ruled in Case 11-M-0294 that "lightly regulated" electric companies be required to file annual operation reports with regards to their costs, revenues and profits. When the electric companies filed their reports, they censored virtually all relevant information under claims of trade secrets. The companies argued that if this private information were made publicly available, then competitors may infer marginal cost functions and use it to implement anti-competitive bidding strategies.

McCullough Research, an energy consulting firm, investigated this argument on behalf of the New York State Legislative Assembly and concluded that, with regards to the anonymization of bid data: *"Any competent analyst can quickly map the actual plant name to the Masked Bidder IDs [...]"*.[52] The general conclusion of the consultants is that *"Information in the Lightly Regulated Annual Reports has not been shown to cause economic harm; the information is widely available; the competitive worth of the Annual Reports is negligible; the cost of deriving it is low; it can be developed easily by third parties;[...]"*.[53] In light of these conclusions, the de-anonymization of bids does not seem as difficult as first thought, and can be partly validated by this report. The PSC finally ruled in favor of the public disclosure of the annual reports. This conclusion establishes that marginal costs should not be considered as private information.

The rest of the data used in this application, such as equilibrium prices and quantities, is available on NYISO website.[54] As explained before, equi-

[52] Affidavit of Robert McCullough, August 24, 2015, page 8.

[53] Affidavit of Robert McCullough, August 24, 2015, page 2.

[54] Zonal day-ahead market prices are available at `http://mis.nyiso.com/public/P-2Alist.htm`, zonal day-ahead demand forecasts are available at `http://mis.nyiso.com/public/P-7list.htm` and zonal load commitments are available at `http://mis.`

librium prices are established at each node of the network by the dispatch algorithm using energy bids and zonal demand forecasts. Zonal prices are then calculated as load-weighted averages of nodal prices within each zone. I use zonal demand forecasts as equilibrium quantities for the DAM, and neglect imports and exports to other regions: ISO-NE, IESO, PJM, Quebec. The latter is not an issue since the equilibrium is not computed explicitly.

The next sections present the method to de-anonymize bid data, to match cost data from the EIA and EPA to New York's power plants so as to construct firm-level datasets for the empirical analysis, and estimate the CDFs of equilibrium prices.

Supply bids de-anonymization

The objective of de-anomyzing the supply bids is to reconstruct firm-level supply functions. I propose a matching algorithm to perform this task. In a first stage, the algorithm matches each firm to a Masked Bidder ID. In a second stage, it pairs each of a firm's facility to a Masked Gen ID (henceforth GENID) associated with the Masked Bidder ID matched in the first stage. The identification relies on the observation that firms have heterogeneous portfolios of generating assets. This feature is especially true for the largest firms.

The algorithm compares public data to anonymized data. I collect data on each firm's portfolio of generating facilities from the NYISO's "Gold Book", a publicly available annual report that provides detailed information on market participants and existing power plants. For each firm $i \in \mathcal{I}$ participating in the wholesale electricity market,[55] Table III of the "Gold Book" 2014 contains data on each of its facility $j \in J_i$: name, location, in-service date, nameplate capacity ($CAP^{NP}_{i,j}$), summer and winter capacity ratings (CAP^{SUM}_{ij} and $CAP^{WIN}_{i,j}$), whether it also generates heat (co-gen), the generation type (e.g. combustion, steam, wind turbines), the fuel types (e.g. natural gas, kerosene, water) and its net energy output in 2013. Each station, or facility, is associated to a unique plant ID, henceforth PTID. For each firm i with J_i generation facilities, I construct 3 ordered sets that describes its capacity portfolio:

$$PORTFOLIO^H_i = \{CAP^H_{i,1}, ..., CAP^H_{i,J^H_i}\}, \tag{3.65}$$

nyiso.com/public/P-59list.htm. The individual plant production schedules are not available.

[55] Firms are referred to owner, operator and/or billing organization in the "Gold Book".

for $H \in \{NP, SUM, WIN\}$ and such that $CAP^H_{i,1} \geq ... \geq CAP^H_{i,J^H_i}$.

On the other hand, I have anonymous hourly bid data organized by Masked Bidder ID for 2013. I approximate the nameplate capacity of each GENID l by taking its maximum quantity bid in the sample $(\widehat{CAP}^{NP}_l)$. The same is done to approximate summer and winter ratings by taking the maximum quantity bid over months from June to September $(\widehat{CAP}^{SUM}_l)$ and November to February $(\widehat{CAP}^{WIN}_l)$, respectively. For each Masked Bidder ID $k \in \mathcal{K}$ with L_k GENIDs, I use those measures to construct 3 ordered sets describing its portfolio capacity:

$$\widehat{PORTFOLIO}^H_k = \{\widehat{CAP}^H_{k,1}, ..., \widehat{CAP}^H_{k,L^H_k}\}, \tag{3.66}$$

for $H \in \{NP, SUM, WIN\}$, and such that $\widehat{CAP}^H_{i,1} \geq ... \geq \widehat{CAP}^H_{k,L^H_k}$. This yields the anonymized counterparts of (3.65).

The matching stragegy relies on combining various pseudometrics[56] computed from the above measures of firm and bidder portfolios. For all $H \in \{NP, SUM, WIN\}$ and each pair $(i,k) \in \mathcal{I} \times \mathcal{K}$, I first compute the two pseudometrics

$$d^H_1(i,k) = 4 \times \frac{|\sum_{j=1}^{J^H_i} CAP^H_{i,j} - \sum_{k=1}^{L^H_k} CAP^H_{k,l}|}{\sum_{j=1}^{J^H_i} CAP^H_{i,j}} \tag{3.67}$$

$$d^H_2(i,k) = 4 \times \frac{|J^H_i - L^H_k|}{J^H_i}, \tag{3.68}$$

which respectively compares the relative total capacity and number of plants.[57] I also compute a pseudometric $d^H_3(i,k)$ measuring the differences between $PORTFOLIO^H_i$ and $\widehat{PORTFOLIO}^H_k$ even for cases where $J^H_i \neq L^H_k$ using Algorithm 2 in appendix C.3.

The core of the matching procedure is formally described in Algorithm 3 in appendix C.3. The algorithm feeds on the set of firms $\mathcal{I}$ and anonymous bidders $\mathcal{K}$ to output a set of pairs of firm-bidder. Specifically, it uses matching probabilities based on logistic transformations of various combinations of the

[56] They measure the distance between two sets but do not satisfy symmetry and the triangle inequality for being considered as formal metrics.

[57] The scalar value is used as a scaling factor.

pseudometrics in order to find the closest bidder among the set of possible candidates for each given firm. A set of constraints is imposed to ensure that all matches make sense: individual and average matching probabilities across criteria must not be too small unless one of the criteria is exactly matched.

The final step consists in pairing PTIDs to GENIDs between matched portfolios. Given that each firm $i \in \mathcal{I}_I$ is paired to bidder $k_i \in \mathcal{K}_I$. I pair the first $\min(J_i, K_{k_i})$ elements of the ordered sets $PORTFOLIO_i^{NP}$ and $\widehat{PORTFOLIO}_{k_i}^{NP}$. Matched pairs are finally checked manually with respect to the partial de-anonymization provided by Affidavits of Robert McCullough, as discussed at the beginning of this section. The procedure results in 93.5% of total capacity matched to GENIDs. The remaining plants are difficult to identify because they belong to firms with only a few plants of small sizes. Descriptive statistics of the final results are given in Tables C.1, C.2 and C.3 in appendix C.3.

Plant-level marginal costs

This section describes the construction of plant-level marginal costs, based on a standard engineering approach used in the economics literature (Mansur, 2007). I account for fuel prices, polluting emission prices and variable operation and maintenance costs.[58] The marginal production costs $mc_{p,f,t}$ (expressed in USD/MWh) of a thermal power plant p at date t using fuel f depends on its heat rate ($HR_{p,f}$ in $MMbtu/MWH$),[59] its CO2 and SO2 emission rates ($CO2R_{p,f}$ and $SO2R_{p,f}$ in ton/MWh) when regulated by environmental policies,[60] its variable operation and maintenance costs ($VOM_{p,f}$

[58] The New York State is a member of the Regional Greenhouse Gases Initiative (RGGI), which aims at pricing carbon dioxide (CO2) emissions in the north-eastern U.S. using a cap-and-trade system. Most polluting generation facilities emitting carbon dioxide (CO2) as a by-product of electricity production in NYISO must buy compliance certificates for each ton of emissions. Under the U.S. Environmental Protection Agency (EPA)'s Acid Rain Program (ARP), power plants in New York are also mandated to purchase compliance certificates for their emissions of sulfur dioxide (SO2). Additionally, the EPA's Clean Air Interstate Rule (CAIR) mandates plants to acquire allowances for their nitrogen oxide (NOx) emissions.

[59] The power plant's heat rate corresponds to the efficiency factor of converting energy from the fuel source into power. MMbtu denotes million british thermal units, a well-used measure of energy content.

[60] I neglect regulations on NOx since those mainly concern coal plants which are largely left idle in New York, and because recent data on NOx allowance certificates are difficult

in USD/MWh) and naturally, fuel f prices ($FP_{f,t}$ in $USD/MMbtu$) and $CO2$ and $SO2$ emission prices ($CO2P_t$ and $SO2P_t$ in USD/ton). It is constructed using a formula given by

$$mc_{p,f,t} = VOM_{p,f} + HR_{p,f}\left(FP_{f,t} + CO2P_t \times CO2R_{p,f} + SO2P_t \times SO2R_{p,f}\right). \quad (3.69)$$

I collect all necessary data from the EIA forms 860 and 923,[61] various EIA reports, and the EPA Continuous Emission Monitoring Systems (CEMS).[62] A preliminary first step consists in matching the firm-level information from the "Gold Book" to the EIA forms 860 and 923 and the EPA CEMS to obtain cost information from regulatory data and identify plants regulated by environmental policies. I use a matching algorithm based on text strings, such as the facility name and its fuel types, and numerical values, such as the facility's capacity and commissioning date. This results in 99.5% of total capacity in the Gold Book accurately matched to EIA data. I finally use plant names to match this data to the EPA CEMS. I am able identify 95% of New York state power plants under RGGI and 88% of plants under ARP in the EPA data.

The EIA form 923 is used to calculate the median heat rate for each pair of primer mover-fuel at the unit-level in the state of New York.[63] The prime mover corresponds to the primary source of power, or technology, used to produce electricity using a given fuel. Plants may have multiple prime mover-fuel pairs. In that case, I calculate the plant-level heat rate as an average by fuel type across prime mover by weighting each unit by its net output in MWh.[64] Heat rates by primemover-fuel are given in column 5 of Table C.4. I compute average unit-level emission rates for CO2 and SO2 for each prime mover-fuel pair in EPA CEMS. The plant-level emission rates correspond to averages across units weighted by net output in MWh and whether the unit is regulated for that emission type. Emission rates

to obtain.

[61] https://www.eia.gov/electricity/data/eia860/ and https://www.eia.gov/electricity/data/eia923/

[62] https://ampd.epa.gov/ampd/

[63] It could have been possible to use plant-level heat rate but the presence of outliers would have required additional data manipulations. Outliers are the main reason to the median value rather than the mean.

[64] Dual-fired plants are sometimes misrepresented by these weights. I adjust the weights of Roseton 1 and 2 and Astoria 3 and 5 to 50% oil and 50% natural gas.

by primemover-fuel are given in columns 6 and 7 of Table C.4. Finally, I use variable operation and maintenance costs from EIA (2013),[65] and from NREL (2012) for hydro plants. I then compute plant-level counterparts by weighting each prime mover-fuel pair with its net capacity in EIA form 860. They are given in column 8 of Table C.4.

I also use the EIA form 923's fuel receipts and costs from regulated power plants to construct time series of fuel prices. There are many missing values for plants located in the state of New York. Therefore, I take data for all states covered by the New York ISO, New England ISO and PJM.[66] It yields fuel prices in $USD/MMbtu$ for bituminous coal (BIT), sub-bituminous coal (SUB), distillate fuel oil (DFO), residual fuel oil (RFO), other petroleum products (PET). I complement this data with monthly state-level fuel prices from the EIA website for kerosene (KER).[67] The price of biomass-type fuels (BIO) is set annually using values from EIA reports.[68] Finally, I extract daily natural gas prices traded at the Henry Hub from the EIA website. The Henry Hub consists in the main reference natural gas trading hub in the U.S. The emission prices are extracted from two sources. The weekly RGGI CO2 allowance prices in USD/ton are extracted from RGGI (2014; 2015; 2016).[69] The SO2 allowance price is set at the annual auction clearing price available on the EPA website.[70] Table C.5 collects descriptive statistics for fuel and emisssion prices.

Firm-level data

The hourly supply and marginal cost functions are constructed for each firm identified in the NYISO's Gold Book 2014 by aggregating bids and marginal costs across plants as obtained from Sections 3.6.1 and 3.6.1. As discussed

[65] http://www.eia.gov/forecasts/capitalcost/pdf/updated_capcost.pdf

[66] The states include New York, Maine, New Hampshire, Vermont,Massachusetts, Rhode Island, Connecticut,Delaware,Maryland, District of Columbia, Illinois, Indiana, Kentucky, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee and Virginia.

[67] The rare missing values are interpolated using cubic spline smoothing.

[68] I use data from Table E7. Electric Power Sector Energy Price Estimates of the EIA's State Energy price and Expenditure Estimates (2013, p.9; 2014 p.11; 2015 p.11).

[69] Specifically, I take the realized auction price at auction dates, the future price in-between auction dates, and the physical delivery in COATS when the future price is not available. Missing data are set to midpoints.

[70] https://www.epa.gov/airmarkets/so2-allowance-auctions

earlier, bids and marginal costs are stable across hours of any given day. Thus, I focus on a daily sample by considering only curves at 6pm in the day-ahead market.

Each energy offer bid may be of four possible types: i) ISO-Committed bids are treated as a bid curve; ii) ISO-Committed Flexible bids are similar but may also serve to signal willingness to supply ancillary services; iii) Self-Committed Fixed bids are treated as price-taking capacity; iv) Self-Committed Flexible bids are treated as price-taking capacity up to some quantity and then as a bid curve, and can be used to offer ancillary services. I transform all price-taking bids into bids with a constant price equal to the minimum price of -1000 USD/MWh. All bids can hence be aggregated into a bid curve.[71]

For all firms, I recover contract positions using the identification result discussed in Section 3.4. Computational details and descriptive statistics are to be found in Table 3.5 below for the 10 largest firms. Similar statistics for the other firms are in Table C.6 in appendix C.3. All firms with average contracted quantity QC close to 1 can be considered as price-takers in the day-ahead market. This is the case of most small producers, wind, hydro and nuclear power-dominated firms.[72] The other firms will be referred to as strategic, in the sense that they do not act as price takers in the auction.

Examples of bids and marginal costs. For illustration purposes, I will present results for two firms throughout the application: Astoria Energy and Long Island Power Authority (LIPA). Figure 3.4 and Figure 3.5 show illustrative inverse supply functions and marginal cost functions for three days at 6pm for Astoria Energy and LIPA, respectively. The quantities are normalized by the maximum quantity bid in that auction, shown at the top-right of each plot.

Both firms are among the largest but have vastly different profiles. Asto-

[71] In the rare instances where a firm has no energy offer bids in the day-ahead market but some bids in the real-time market, it means the firm has scheduled a bilateral transaction that cannot be observed. Therefore they are treated as price-takers in the DAM with capacity equal to their maximum real-time bid.

[72] Some hydro-dominated firms have smaller forward positions, which may be due to the absence of the shadow value of water in hydro reservoirs in the marginal cost. This shadow value is crucial to calculate the marginal cost of hydro power plants with reservoirs. It is the solution of a dynamic management problem given water inflows as modelled in Debia et al. (2018). This approach is however difficult to pursue here.

Firm	Capacity	Max Bid	Mean QC	Mean PC	# GENIDs	# obs
New York Power Authority	6730.60	5771.90	0.42	6.00	25	1095
Long Island Power Authority	5216.10	4910.70	0.51	65.33	67	1095
NRG Power Marketing	4117.60	3621.00	0.32	57.02	28	1095
Entergy Nuclear Power Marketing	3193.00	2927.20	1.00	2.14	3	1095
Consolidated Edison	2593.80	1706.20	0.50	75.74	6	1095
TC Ravenswood	2557.10	2308.70	0.66	59.21	19	1095
Exelon Generation Company	2546.70	2538.40	1.00	38.69	10	1095
Astoria Generating Company	1771.00	1829.80	0.34	57.44	51	1095
Athens Generating Company	1323.00	1198.80	0.83	47.77	3	1081
Astoria Energy	1300.00	1262.40	0.85	75.18	4	1094

NOTES. This table reports descriptive statistics for firm-level data. A firm's max capacity denotes the maximum capacity this firm has bid in the DAM. Mean QC is the average contract quantity as a share of total quantity bid. Mean PC is the average contract price. # GENIDs denote the number of Masked Gen IDs associated to the firm. # obs is the number of days with day-ahead market bids over the sample period (2013-2015).

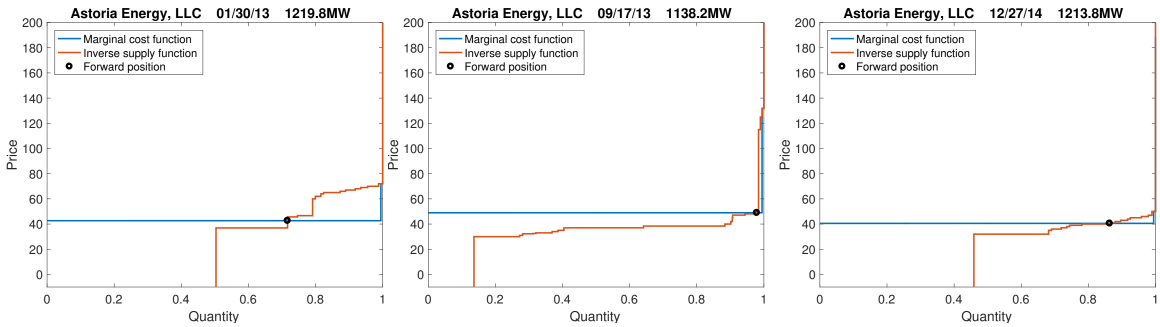
Table 3.5: Descriptive statistics of firm-level data

ria Energy owns 2 natural gas-fired combined cycle power plants of respective nameplate capacities 640 MW and 660 MW both located in Queens, New York. Each plant has 2 generating units which use distillate fuel oil as a secondary fuel. Each of the 4 units is associated to a single Masked Gen ID. Astoria Energy is a lightly regulated power producer with the ability to earn revenues from its profitable location in New York City, where electricity prices are high due to a tight supply.

On the other hand, LIPA is a large utility consisting of 76 generating units for a total of 5,216 MW of capacity in Long Island. The utility is vertically integrated from production to distribution and subject to cost-of-service regulation, though it does not own the generating assets. It has been operated by National Grid USA up until the end of 2013. The Office of the New York State Comptroller, the entity in charge of overseeing State authorities, had "identified areas requiring improvement, including adequacy of regulatory oversight, rate relief, [...]". The Governor's Program Bill #20 of 2013 has been enacted in an effort to address some of the concerns regarding LIPA through significant restructuring. In January 2014, Public Service Enterprise Group (PSEG) took over the management and operations under a 10-year contract with LIPA.[73] It is possible that the restructuring resulted

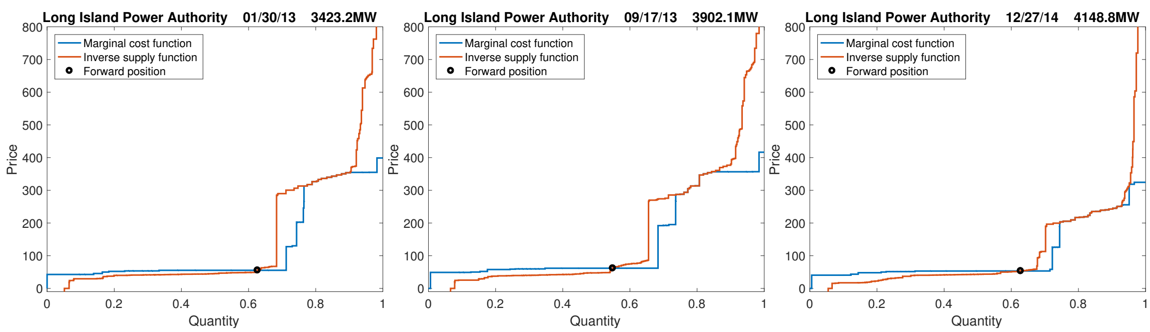
[73] See the Office of the State Comptroller's reports, Public Authorities by the Numbers: Long Island Power Authority (October 2012), and Long Island Power Authority by the Numbers: A Public Authority in Transition (July 2015) available at https://www.osc.

in a different bidding behavior. However, there is no reason to expect this large utility to bid at marginal costs since several of its assets are allowed to offer energy at wholesale prices.[74]



(a) January 30th, 2013 (b) September 17th, 2013 (c) December 27th, 2014

Figure 3.4: Astoria Energy



(a) January 30th, 2013 (b) September 17th, 2013 (c) December 27th, 2014

Figure 3.5: Long Island Power Authority

Function smoothing. The functional analysis uses smoothed versions of the supply and residual demand functions defined over a range of prices. I now focus on $S_{id}(p)$ rather than $S_{id}^{-1}(p)$. I use penalized spline smoothing to

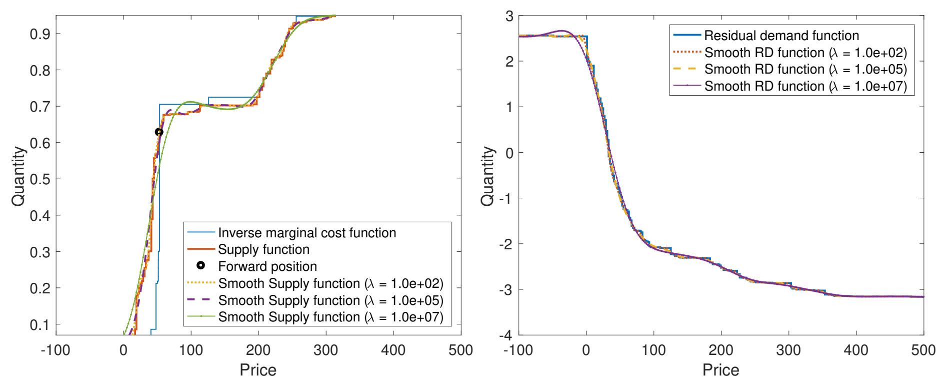
---

`state.ny.us/`.

[74]There is evidence of strategic behavior of the former managing firm (National Grid). In 2008, the firm paid USD 12 millions in disgorgement following allegations of misconduct in the capacity market (Ledgerwood and Carpenter, 2012).

construct functional data objects and estimate derivative functions. Supply and residual demand functions are smoothed using 30 splines of order 4 with a penalty on the 3rd derivative. The smoothing is performed on the rescaled functions. The rescaled residual demand function is defined as $RD^{scaled}_{i,d}(p) = \frac{Demand_d - \sum_{j \neq i} S_{j,d}(p)}{K_{i,d}}$.

In the empirical analysis, I use smoothed functions obtained with the regularization parameter $\lambda_S = 10^5$ and $\lambda_{RD} = 10^7$, respectively for supply and residual demand. The resulting dataset of smoothed functions will be denoted $\{\tilde{S}_{i,d}, \tilde{RD}_{i,d}, \tilde{S}'_{i,d}, \tilde{RD}'_{i,d}\}_{d=1}^{D}$. Figure 3.6 show illustrative examples for LIPA on December 27th, 2014 with various degrees of smoothing.[75]



(a) Smoothed Supply Function

(b) Smoothed Residual Demand Function

Figure 3.6: Long Island Power Authority

Estimation of price distribution functions

I follow the estimation method laid out in Section 3.5.4 to estimate the daily price distribution functions. Table 3.6 show descriptive statistics of hourly zonal prices. The zonal prices span over relatively similar ranges from 0 USD to 600 USD although 98% of the distribution is concentrated on prices from 10 USD to 250 USD.

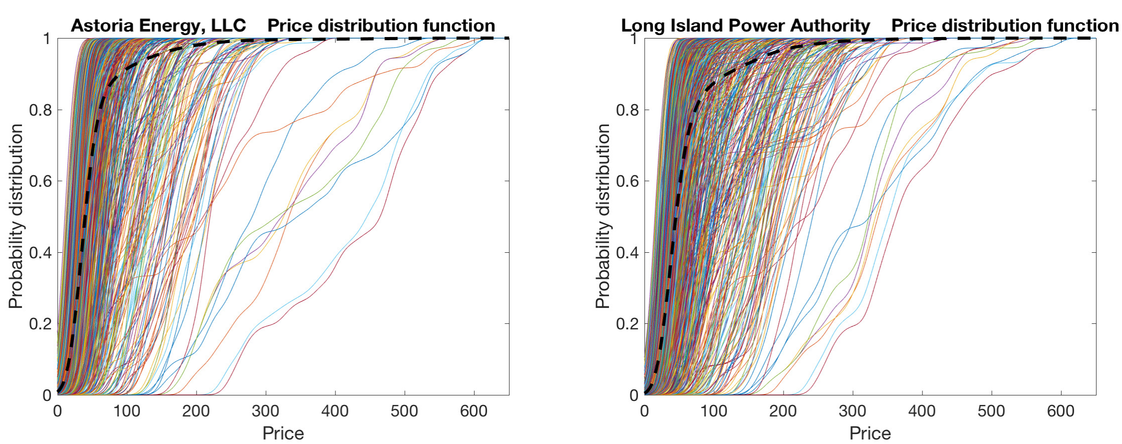
[75] Remark that the smoothed functions may not be monotone, which is not problematic for this analysis. It is possible to force monotonicity using a transformation approach (Ramsay, 1998).

Zone	Name	Mean P	Std.	Min	P01	P99	Max
A	West	38.28	31.52	1.03	6.09	189.06	500.00
B	Genesee	37.05	32.70	0.93	5.91	194.10	522.16
C	Central	38.75	34.34	1.04	6.18	202.60	542.63
D	North	34.80	34.77	0.34	2.14	201.90	553.70
E	Mohawk Valley	39.68	36.31	1.01	6.42	213.31	573.07
F	Capital	49.88	47.65	2.18	10.49	247.60	593.02
G	Hudson Valley	49.26	44.15	2.00	10.47	236.26	599.11
H	Millwood	49.62	44.54	2.02	10.52	238.26	602.43
I	Dunwoodie	49.54	44.36	2.01	10.53	237.27	600.74
J	New York City	50.83	46.72	2.02	10.58	241.26	604.64
K	Long Island	59.99	52.54	2.57	13.59	276.22	601.38
Total		47.48	41.27	1.77	10.00	225.41	577.54

NOTES. This table reports descriptive statistics of hourly zonal prices including the mean, standard deviation, minimum, 1st percentile, 99th percentile, and maximum prices. The dataset contains 26,280 observations. The "total" is computed as the load-weighted average series.

Table 3.6: Descriptive statistics of zonal prices (2013-2015)

Figure 3.7 presents the estimated CDFs and its mean (thick dashed line) for 1095 days from January 1, 2013 to December 31, 2015 for Astoria and LIPA. Each of those two firms are located within a single zone, hence their associated price simply corresponds to the zonal prices J and K, respectively.



(a) Astoria Energy (2013-2015)

(b) Long Island Power Authority (2013-2015)

Figure 3.7: Estimated price distribution functions

The data having been detailed, I present the empirical results in the next section.

3.6.2 Empirical results

In the New York's day-ahead electricity market, the late disclosure of bids and private information may limit the ability to act as if they knew their residual demand function. I consider that firms form their bids on the basis of past realizations of the hourly day-ahead prices. Consequently, they can only be expected to behave optimally around price values that occur with non-zero probabilities. I study each firm's bid function on the price range that correspond to 98% of the probability distribution of that firm's prevailing price, as explained in Section 3.6.1. The comparison of optimal bids to actual bids is then conducted on the basis of the range of prices that is relevant for each given day: $\mathcal{P}_t$, as characterized in Section 3.5.4.

Testing optimal bidding under additive separability

I use the testing procedures developed in Section 3.5.2 to evaluate whether the optimality condition under the AS restriction, defined in (3.7), holds in expectations for any of the 10 largest firms. For each firm i, the null hypothesis is

$$H_0 : E[p - C'_{i,d}(S_{i,d}(p)) - (S_{i,d}(p) - QC_{i,d})\frac{1}{-\tilde{RD}'_{i,d}(p)}] = 0, \quad \forall p \in [\underline{P}, \overline{P}]. \tag{3.70}$$

Table 3.7 reports the p-values of the tests CvM and J_2, which performed best in the simulations, separately for each year 2013, 2014 and 2015. The table also shows the price range for each firm, where $\underline{\text{P}}$ and $\overline{P}$ denote the lower and upper bounds respectively. Both tests consistently reject the optimality condition defined in (3.7) at the 5% level for all firms. This can be due to a suboptimal bidding behavior, a failure of the AS restriction, or a more general shortcoming of the static profit-maximization model that neglects the effect of binding transmission constraints on the residual demand functions.

	2013		2014		2015			
Firm	CvM	J_2	CvM	J_2	CvM	J_2	$\underline{P}$	$\overline{P}$
New York Power Authority	0.00	0.00	0.00	0.00	0.00	0.00	9.0	205.2
Long Island Power Authority	0.00	0.00	0.00	0.00	0.00	0.00	15.0	275.3
NRG Power Marketing	0.00	0.00	0.00	0.00	0.00	0.00	9.0	213.2
Entergy Nuclear Power Marketing	0.00	0.00	0.00	0.00	0.00	0.00	11.0	225.2
Consolidated Edison	0.00	0.00	0.00	0.00	0.00	0.00	11.0	235.2
TC Ravenswood	0.01	0.00	0.00	0.00	0.00	0.00	11.0	241.2
Exelon Generation Company	0.00	0.00	0.00	0.00	0.00	0.00	7.0	199.2
Astoria Generating Company	0.00	0.00	0.00	0.00	0.00	0.00	11.0	241.2
Athens Generating Company	0.00	0.00	0.00	0.00	0.00	0.00	11.0	247.2
Astoria Energy	0.00	0.00	0.00	0.00	0.00	0.00	11.0	241.2

NOTES. This table reports the p-values of the test statistics CvM and J_2 for the 10 largest firms.

Table 3.7: P-values of the tests of optimal bidding

Testing the additive separability restriction

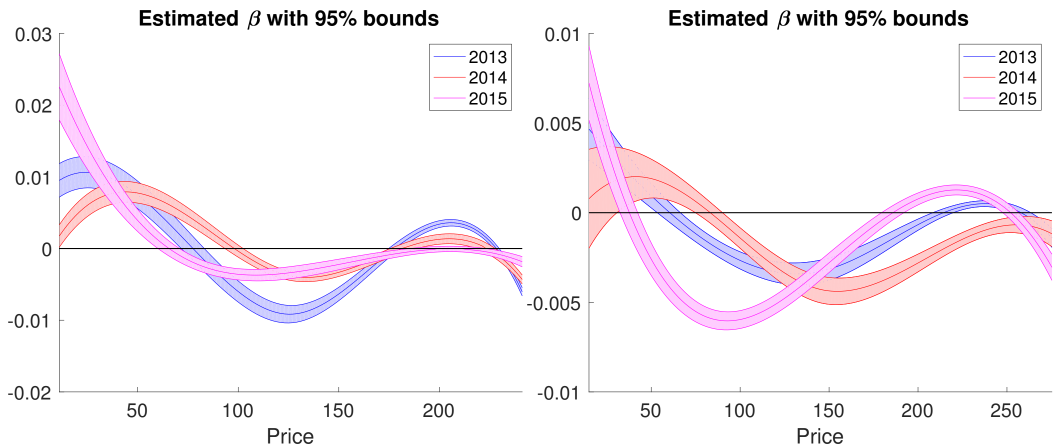
The testing procedures developed in Section 3.5.3 are now used to evaluate the validity of the AS restriction for the same group of firms. Let us specify for each firm i the functional linear model

$$\tilde{S}'_{i,d}(p) = \alpha_i(p)+\beta_i(p)QC_{i,d}+\gamma_{i,RD}(p)\tilde{RD}'_{i,d}(p)+\gamma_{i,NGP}(p)NGP_d+\gamma_{i,trend}(p)d+U_{i,d}(p), \tag{3.71}$$

for $p \in [\underline{P}, \overline{P}]$, and where the control variables include: a constant functional parameter α_i, the derivative of the residual demand function $\tilde{RD}'_{i,d}$, the daily price of natural gas NGP_d and a time trend d. $U_{i,d}$ is a functional error term. The AS restriction is evaluated separately for each firm i using the null hypothesis

$$H_0 : \beta_i(p) = 0, \quad \forall p \in [\underline{P}, \overline{P}]. \tag{3.72}$$

I estimate the functional parameters in (3.71) using 5 cubic spline basis and $\lambda = 0$. The estimated functional parameters associated to $QC_{i,d}$ along with 95% (pointwise) confidence bounds are shown in Figure 3.8a for Astoria Energy and 3.8b for LIPA. The shape of the parameters seems consistent between firms and across years. The slope of the supply functions are found positively correlated with contracted quantities for low prices, and negatively so for larger prices. This effect can be observed in Figures 3.4a and 3.4b. As QC increases, the inverse supply function S^{-1} is flatter for low prices and steeper for larger prices. Since I study S instead of S^{-1} here, the interpretation is simply reversed. This pattern suggests that greater contracted sales are associated with less aggressive bids for small prices and more aggressive bids for larger ones.



(a) Astoria Energy (2013-2015) (b) Long Island Power Authority (2013-2015)

Figure 3.8: Estimated β with confidence bands

On the other hand, I perform the HP test based on the estimates of β from the analogous linear model

$$\overline{S}'_{i,d} = \alpha_i + \beta_i QC_{i,d} + \gamma_{i,RD}\overline{RD}'_{i,d} + \gamma_{i,NGP} NGP_d + \gamma_{i,trend} d + U_{i,d}, \quad (3.73)$$

where $\overline{S}'_{i,d}$ and $\overline{RD}'_{i,d}$ denote the slope of the linearized functions, obtained as described in Section 3.5.3.[76] Table 3.8 reports the p-values of the test statistics CvM, J_2 and t_{HP}. The tests do not reject the null for Entergy and Exelon. Those firms operate nuclear power plants and thus act as price-takers with supply functions kept constant over time. There is simply not enough variation to estimate β. Besides those two cases, the functional tests overwhelmingly reject the AS restriction for all firms at the 5% level. The HP

[76]Remark that those control variables are similar to those used in Hortaçsu and Puller (2008). The only difference is that I do not use the marginal cost functions as it is itself a function of the bids, instead I use natural gas prices which explain most of the variations in cost functions across days.

test accepts the null hypothesis in 20% of the cases. Following the previous simulations based on a parameter of similar shape, the results obtained with the functional tests should be more reliable. For example, the functional parameter for Astoria in 2014 appears to integrate to zero over the range of prices in Figure 3.8a, and the corresponding p-value of HP test in Table 3.8 suggests not to reject, as predicted by the simulations. As a conclusion, the AS restriction can be rejected with confidence in this dataset. The market power parameters are estimated in the next section.

	2013			2014			2015		
Firm	CvM	J_2	t_{HP}	CvM	J_2	t_{HP}	CvM	J_2	t_{HP}
New York Power Authority	0.02	0.00	0.00	0.00	0.00	0.73	0.00	0.00	0.00
Long Island Power Authority	0.00	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.00
NRG Power Marketing	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Entergy Nuclear Power Marketing	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Consolidated Edison	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
TC Ravenswood	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Exelon Generation Company	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Astoria Generating Company	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Athens Generating Company	0.00	0.00	0.00	0.00	0.00	0.47	0.00	0.00	0.03
Astoria Energy	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00

NOTES. This table reports p-values of the test statistics CvM, J_2 and t_{HP} for the null hypothesis of additive separability.

Table 3.8: P-values of AS tests

Measures of market power

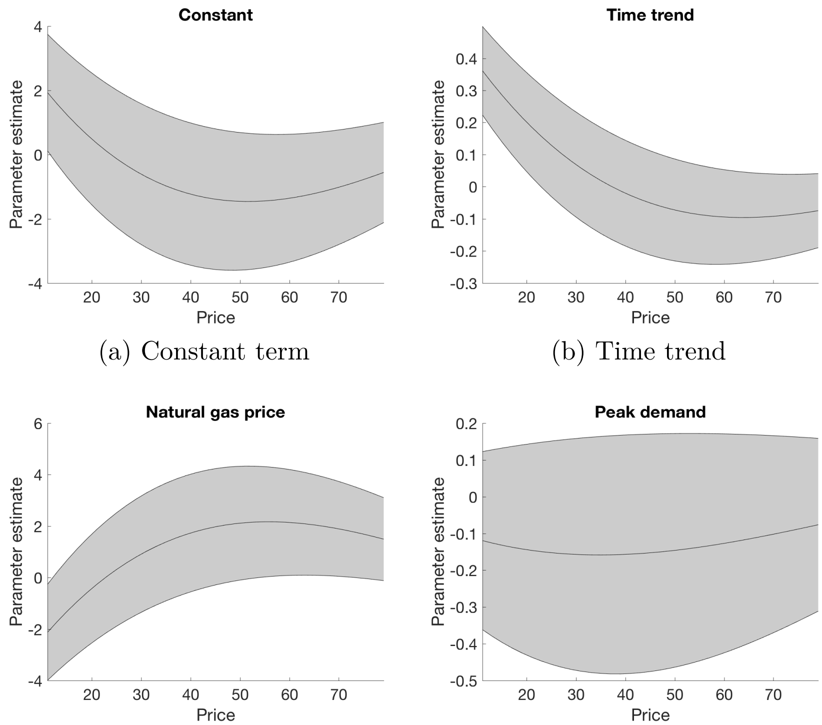
For each firm i, I estimate the functional linear model given by

$$\hat{F}_{i,d}(p) = \overline{F}_{i,t}(p) + \tilde{S}_{i,d}(p)\pi_{i,t}(p) + \boldsymbol{Z}_{i,d}\boldsymbol{\gamma}_{i,t}(p) + V_{i,d}(p), \tag{3.74}$$

separately for each sample $\{\hat{F}_{i,d}, \tilde{S}_{i,d}, \boldsymbol{Z}_{i,d}\}_{d=t-15}^{d=t+15}$ with $t \in \{16, ..., T-16\}$. $\overline{F}_{i,t}$ is a constant function, $\boldsymbol{Z}_{i,d}$ is a vector of controls including a time trend, the price of natural gas and the daily peak demand. $V_{i,d}$ denotes a functional error term. As discussed in Section 3.5.4, the model is also separately estimated using the forward quantity $QC_{i,d}$ and the lagged supply function $\tilde{S}_{i,d-1}$ as instrumental variables for the endogenous variable $\tilde{S}_{i,d}$. The basis expansions of each functional parameter are performed onto 5 cubic splines and the hold-out validation criterion is used to choose the regularization parameter for each sample.

Illustrative estimates along with 95% pointwise confidence bounds are shown in Figures 3.9 and 3.10 for Astoria Energy, and in Figures 3.11 and 3.12 for LIPA. The parameters associated to control variables are displayed only for the least-squares regression without instrument, whereas the main parameter of interest π is plotted for all three cases. In general, the estimates suggest that natural gas prices and peak demand are associated with lower probabilities of small prices.[77] Unsurprisingly, the higher natural gas prices and peak demand are, the higher the realizations of day-ahead electricity prices will be. Additionally, the parameter of interest π is found to be positive, and larger in the IV regressions.

[77]Remark that, although the pointwise bands may contain the zero parameter, it does not necessarily imply that the parameter is not statistically different from zero. This also depends on the correlations along the function. Uniform bands can be computed in this setting following Babii (2016).



(a) Constant term (b) Time trend

(c) Natural gas price (d) Peak demand

Figure 3.9: Estimated parameters (Astoria Energy: May 7th, 2014)

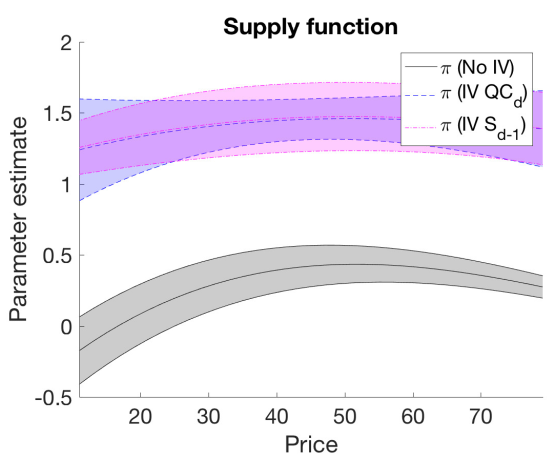
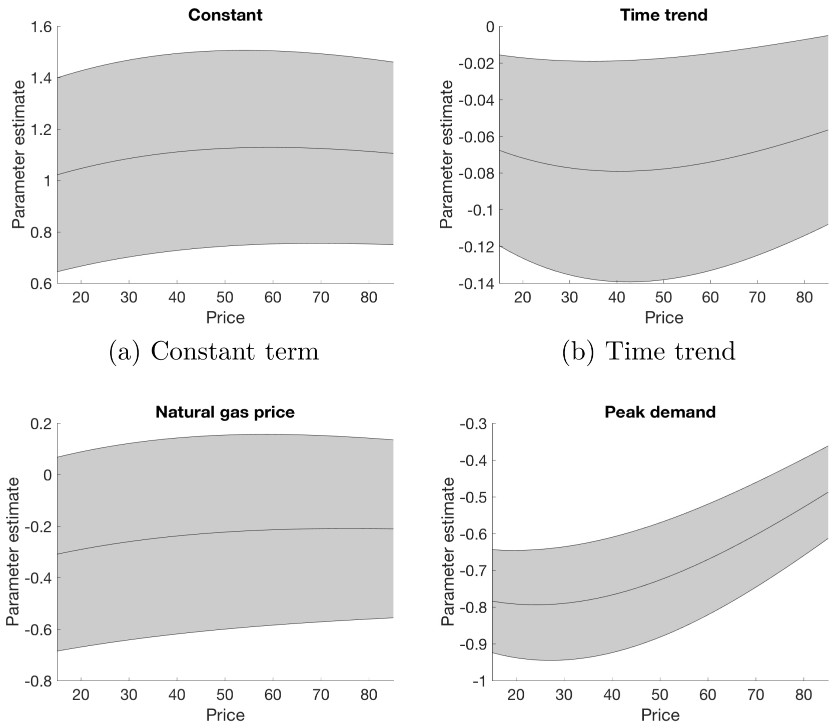


Figure 3.10: $\hat{\pi}$ (Astoria Energy: May 7th, 2014)



(a) Constant term

(b) Time trend

(c) Natural gas price

(d) Peak demand

Figure 3.11: Estimated parameters (LIPA: August 20th, 2015)

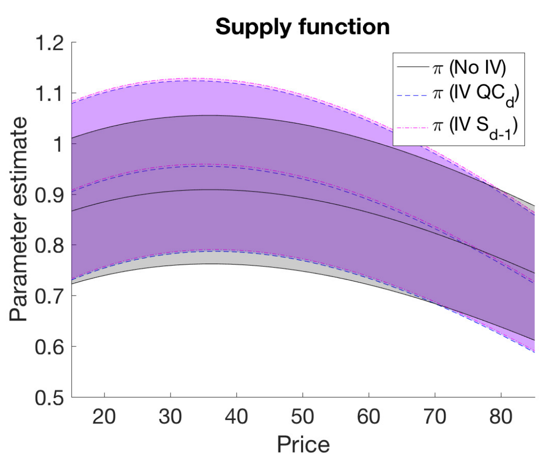


Figure 3.12: $\hat{\pi}$ (LIPA: August 20th, 2015)

The parameter π measures the effect of a marginal increase of firm i's supplied quantity at price p onto the probability of observing that price p. In that sense, it corresponds to a measure of market power. I propose to use the average integrated standardized estimate (AISE) to compare firm-level values. It is defined as

$$AISE = T^{-1} \sum_{t=1}^{T} \frac{1}{\overline{p}_t - \underline{p}_t} \int_{\underline{p}_t}^{\overline{p}_t} \hat{\pi}_t(p) \frac{K_t}{K_{max}} dp, \tag{3.75}$$

where $\frac{K_t}{K_{max}}$ is used to standardize $\hat{\pi}_t$ so that it corresponds to a change in quantity in percent of available capacity K_t.[78] It gives the mean value of the parameter once standardized by capacity. Table 3.9 shows the AISE of $\hat{\pi}$ for estimates obtained with and without instruments for each of the top ten firms across years. The table also reports the rejection rates of $H0 : \pi = 0$ based on the functional test CvM for every case.

[78] Recall that $\hat{\pi}_t$ is obtained from $S_t^{scaled}(p) = \frac{S_t(p)}{K_{max}}$, hence $\hat{\pi}_t S_t^{scaled}(p) = \left(\hat{\pi}_t \frac{K_t}{K_{max}}\right) \frac{S_t(p)}{K_t}$ where $\frac{S_t(p)}{K_t} \in [0, 1]$.

Firm	2013 $\hat{\pi}$	2013 $\hat{\pi}^{IV1}$	2013 $\hat{\pi}^{IV2}$	2014 $\hat{\pi}$	2014 $\hat{\pi}^{IV1}$	2014 $\hat{\pi}^{IV2}$	2015 $\hat{\pi}$	2015 $\hat{\pi}^{IV1}$	2015 $\hat{\pi}^{IV2}$
New York Power Authority	0.1	-1.8	-1.9	0.3	-0.9	-1.0	0.1	-1.2	-1.2
	(0.9)	(0.9)	(0.8)	(0.5)	(0.8)	(0.9)	(0.4)	(0.9)	(0.9)
Long Island Power Authority	0.8	1.2	1.2	0.9	1.0	1.0	0.8	0.9	0.9
	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(0.9)	(0.9)
NRG Power Marketing	1.0	1.8	1.8	0.9	1.6	1.6	0.4	1.1	1.2
	(0.9)	(0.9)	(0.9)	(1.0)	(1.0)	(0.9)	(0.8)	(0.9)	(0.9)
Entergy Nuclear Power Marketing	-0.3	0.8	0.2	0.3	0.8	-0.1	0.1	-0.5	-0.0
	(0.2)	(0.8)	(0.8)	(0.3)	(0.9)	(0.9)	(0.3)	(0.8)	(0.8)
Consolidated Edison	0.8	1.6	1.7	0.6	0.9	0.9	0.3	0.8	0.9
	(1.0)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.8)	(0.9)	(0.9)
TC Ravenswood	0.6	0.6	0.6	0.4	0.7	0.7	0.4	0.6	0.7
	(1.0)	(1.0)	(1.0)	(0.9)	(1.0)	(1.0)	(1.0)	(1.0)	(0.9)
Exelon Generation Company	-0.0	-6.1	-8.6	-0.1	-4.6	-8.3	-0.3	-1.3	-0.8
	(0.3)	(1.0)	(1.0)	(0.4)	(0.9)	(0.9)	(0.3)	(0.8)	(0.8)
Astoria Generating Company	1.1	1.3	1.3	1.0	1.3	1.2	0.5	1.0	1.1
	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(0.9)	(0.9)	(1.0)
Athens Generating Company	0.6	0.6	0.7	0.5	0.8	0.8	0.5	0.8	0.8
	(1.0)	(0.9)	(0.9)	(1.0)	(0.9)	(0.9)	(1.0)	(0.8)	(0.9)
Astoria Energy	0.4	0.8	0.8	0.4	1.1	1.2	0.3	0.8	0.9
	(0.9)	(1.0)	(1.0)	(0.8)	(0.9)	(0.9)	(0.8)	(0.8)	(0.8)

NOTES. This table reports the average integrated standardized estimate (AISE) of π defined in (3.75) and rejection rates (in parentheses) of $H0 : \pi = 0$ based on the test CvM for the 10 largest firms. $\hat{\pi}$ denotes the least-square estimate, $\hat{\pi}^{IV1}$ denotes the IV estimate using the forward position and $\hat{\pi}^{IV2}$ that with the lagged supply function.

Table 3.9: AISE of $\hat{\pi}$ and rejection rate of $H0 : \pi = 0$

Consider the values of Astoria Energy and LIPA. On average, in 2013, Astoria Energy is found to increase the probability of occurrence of some price p by 0.4 percentage point when increasing its quantity offered at p by 1 percentage point of its total capacity. IV estimates give a two-fold increase of that effect. LIPA is found to have a slightly greater effect in comparison. That effect appears larger for 2013 than 2014 and 2015, implying that the firm possibly chose smaller markups during the second period. This might be a consequence of its restructuring in 2014.

Those summary statistics deliver several important patterns. From the first columns of each year, only three firms: NYPA, Entergy and Exelon, seem to have nearly no impact on realized prices. Their AISE and rejection rates are small, which suggest that their supply functions have no statistically significant effects on the price distribution. The IV estimates for those firms are either negatively signed and rather large in absolute terms or close to zero. The findings that the supply functions of those firms do not affect the price distribution should not be surprising. Entergy and Exelon operate nuclear power plants and consistently act as price-takers. That is, their bids are kept constant so no identification is possible. Besides, NYPA mostly operate hydro plants. Those two technologies are infra-marginal, and thus almost never set the market price. Using instruments yield smaller estimates for NYPA, unlike for all other firms.

All other firms are mainly endowed with gas-fired and oil-fired power plants. For those firms, all three estimates are positive, take value around 1 and have high rejection rates of the null. Their value also seem relatively consistent across years, and the IV estimates are found larger and close together. It implies that the supply functions are negatively correlated with the error term. I interpret this result as evidence that those firms react to anticipations of smaller prices by decreasing their supplied quantity, so as to attenuate the price decrease, and reversely.

Mean elasticities of the implicit residual demand functions

Elasticities are widely used as a measure of unilateral market power. In the context of multi-unit auctions, the elasticity that matters is that of the residual demand function faced by the firm, as shown in (3.8). However, in the presence of uncertainty, it is possible to compute the elasticity of the implicit residual demand function induced by the conditional distribution of prices as defined in (3.9). The optimality condition also permits to calculate the elasticities induced by observed supply functions as

$$\varepsilon_0 = \frac{p}{p - C'(S(p))} \frac{S(p) - QC}{S(p)}. \tag{3.76}$$

I compare the expected elasticities averaged across the sample, given by

$$\overline{\varepsilon} = T^{-1} \sum_{t=1}^{T} \int \varepsilon_t(p) \hat{f}_t(p) dp, \tag{3.77}$$

for the elasticities defined in (3.8), (3.9) and (3.76). The values are presented in the Table 3.10. The elasticities $\overline{\varepsilon}_F^{IV1}$ and $\overline{\varepsilon}_F^{IV2}$ corresponding to the IV estimates are found to be reasonably close to the observed ones $\overline{\varepsilon}_0$.

Firm	$\overline{\varepsilon}_0$	$\overline{\varepsilon}_F$	$\overline{\varepsilon}_F^{IV1}$	$\overline{\varepsilon}_F^{IV2}$	$\overline{\varepsilon}_{RD}$
New York Power Authority	0.29	3.71	0.63	0.58	2.04
Long Island Power Authority	1.98	3.34	3.22	3.32	7.15
NRG Power Marketing	8.85	13.43	6.66	6.88	25.21
Entergy Nuclear Power Marketing	0.00	4.41	1.73	1.91	3.10
Consolidated Edison	2.93	6.43	3.07	3.20	18.75
TC Ravenswood	4.80	9.02	7.29	7.11	19.84
Exelon Generation Company	0.00	3.57	0.37	0.34	3.26
Astoria Generating Company	5.48	11.72	7.52	7.62	36.06
Athens Generating Company	1.15	3.10	2.79	2.52	13.89
Astoria Energy	1.43	5.57	2.20	2.01	12.57

NOTES. This table reports the average expected elasticites for the 10 largest firms. $\overline{\varepsilon}_0$, $\overline{\varepsilon}_F$, $\overline{\varepsilon}_F^{IV1}$, $\overline{\varepsilon}_F^{IV2}$, and $\overline{\varepsilon}_{RD}$ respectively denote the elasticity induced by observed bids, induced by the least-squares estimate, the IV estimate with QC, the IV estimate with the lagged supply function, and the elasticity of the observed residual demand function.

Table 3.10: Mean elasticities of residual demand functions

Optimal bidding

I now use the estimates to compute the optimal supply bids and compare them to the observed bids. The comparison is based on two statistics. First, the integrated mean squared error ($IMSE$) defined by

$$IMSE = T^{-1}\sum_{t=1}^{T}\left(\frac{1}{\overline{p}_t - \underline{p}_t}\int_{\underline{p}_t}^{\overline{p}_t}\Big(S_t^\star(p) - S_t^0(p)\Big)^2 dp\right) \tag{3.78}$$

provides a standardized metric to evaluate how close the optimal bid function $S_t^\star$ and the observed bid function S_t^0 are on average. However, the IMSE criterion capturtes both the bias and variance. Second, I consider the integrated mean error (IME) as defined in

$$IME = T^{-1}\sum_{t=1}^{T}\left(\frac{1}{\overline{p}_t - \underline{p}_t}\left[\int_{\underline{p}_t}^{PC_t}\Big(S_t^\star(p) - S_t^0(p)\Big)dp\right) - \int_{PC_t}^{\overline{p}_t}\Big(S_t^0(p) - S_t^\star(p)\Big)dp\right]\right), \tag{3.79}$$

as a means to evaluate whether $S_t^\star$ tend to be above or below S_t^0 on average. In that sense, this statistic measures the sign and magnitude of the bias between the two functions. This aims at understanding whether observed bids are more or less aggressive that optimal bids.

The $IMSE$ are given in Table 3.11, and the IME are shown in Table 3.12 for the 10 largest firms. Consider first the results for Astoria Energy and LIPA for illustration. Their IMSE are small and the IV estimates deliver a substantially better prediction than the HP benchmark. The IME indicate that Astoria Energy's observed supply functions are only 3.5% below the optimal functions derived from the IV estimates and 12% below the HP benchmark. Within the same order of magnitude, the observed bids of LIPA are found only 3% below the IV benchmark and about 6% below the HP optimum.

I show illustrative plots of the optimal and actual supply functions for Astoria Energy in Figure 3.13a and the optimal bids based on the IV estimates with 95% confidence bounds in Figure 3.13b. Similar plots for LIPA are shown in Figure 3.14a and Figure 3.14b. In order to give a sense of the magnitudes, the examples for both firms are chosen so that the daily IMSE is close to the firm-level IMSE reported in Table 3.11. It somehow corresponds to the average prediction in terms of IMSE. Note that the density is small at the tails, which explains the larger deviations.

Firm	$IMSE$	$IMSE^{IV1}$	$IMSE^{IV2}$	$IMSE^{HP}$
New York Power Authority	0.073	0.087	0.087	0.063
Long Island Power Authority	0.016	0.016	0.016	0.039
NRG Power Marketing	0.017	0.020	0.020	0.021
Entergy Nuclear Power Marketing	0.000	0.000	0.000	0.000
Consolidated Edison	0.033	0.028	0.029	0.058
TC Ravenswood	0.077	0.064	0.065	0.138
Exelon Generation Company	0.000	0.000	0.000	0.000
Astoria Generating Company	0.054	0.056	0.053	0.125
Athens Generating Company	0.074	0.044	0.042	0.196
Astoria Energy	0.121	0.034	0.035	0.196

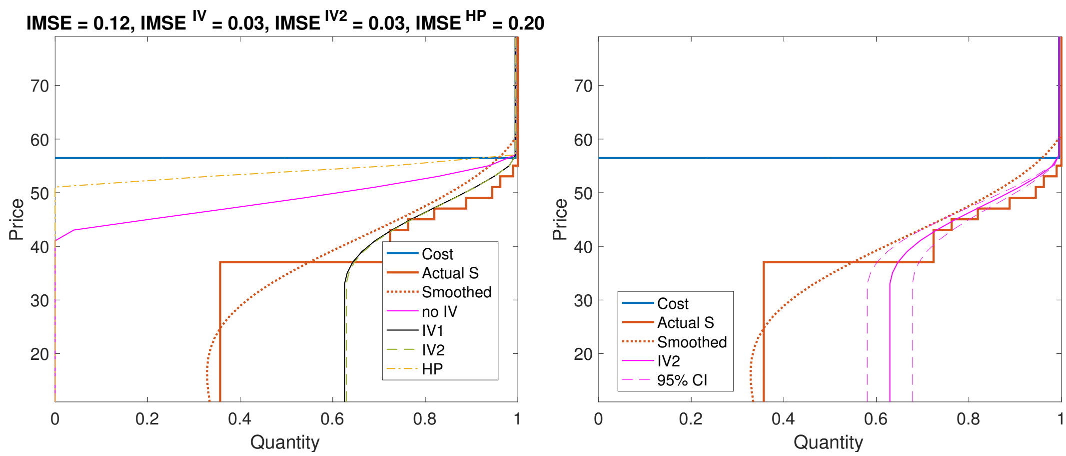
NOTES. This table reports the IMSE between observed and optimal bid functions derived from the standard estimates, both IV estimates for QC (IV1) and lagged supply (IV2), and the HP benchmark, for the 10 largest firms.

Table 3.11: IMSE between observed and optimal supply functions

Firm	IME	IME^{IV1}	IME^{IV2}	IME^{HP}
New York Power Authority	-0.187	-0.202	-0.201	-0.152
Long Island Power Authority	-0.030	-0.030	-0.032	-0.063
NRG Power Marketing	-0.080	-0.053	-0.052	-0.026
Entergy Nuclear Power Marketing	0.000	0.000	0.000	0.000
Consolidated Edison	-0.063	-0.036	-0.035	-0.053
TC Ravenswood	-0.062	-0.051	-0.050	-0.117
Exelon Generation Company	-0.002	-0.002	-0.002	-0.002
Astoria Generating Company	-0.140	-0.129	-0.126	-0.205
Athens Generating Company	-0.046	-0.036	-0.032	-0.112
Astoria Energy	-0.079	-0.035	-0.034	-0.119

NOTES. This table reports the IME between observed and optimal bid functions derived from the standard estimates, IV estimates for QC (IV1) and lagged supply (IV2), and the HP benchmark, for the 10 largest firms.

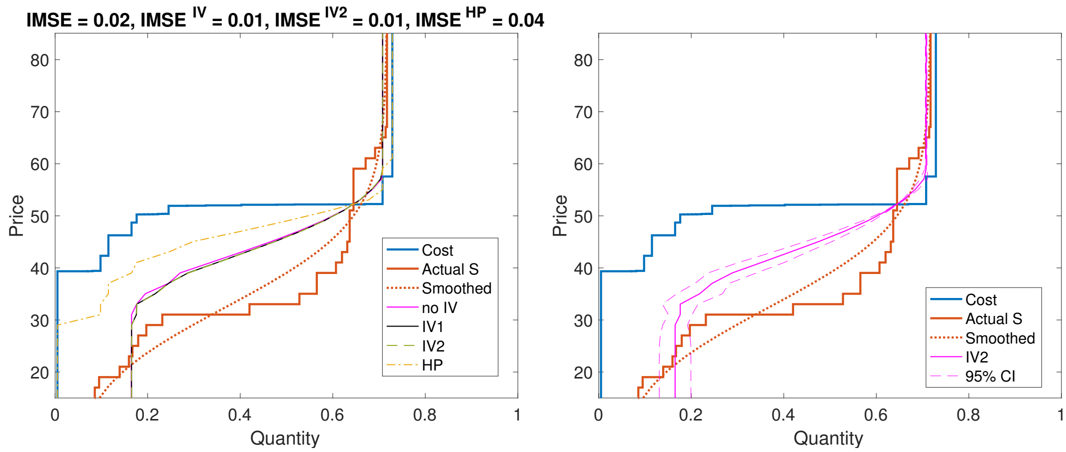
Table 3.12: IME between observed and optimal supply functions



(a) Inverse supply functions

(b) $S^{\star IV2}$ with 95% CI

Figure 3.13: Optimal supply function (Astoria Energy: May 7th, 2014)



(a) Inverse supply functions (b) $S^{\star IV2}$ with 95% CI

Figure 3.14: Optimal supply function (LIPA: August 20th, 2015)

In the previous section, I found that the strategies of Entergy, Exelon and NYPA have almost no effect onto the price distribution. The HP benchmark performs similarly than the functional approach for these firms. Entergy and Exelon being price-takers with nearly zero marginal cost, they always offer their full capacity over the relevant price range and their behavior is hence trivial to predict. The optimal functions based on the functional estimates and the HP benchmark suggest that the three firms should bid close to their marginal cost. Yet, I observe from the IME of NYPA that observed bid functions are much less aggressive than predicted. The firm tends to submit supply functions with quantity offers around 20% lower than the optimum. I believe this is due to the absence of the shadow cost of water in the marginal cost of NYPA, which leads to greatly underestimate both the firm's cost function and forward position.

For most firms, the optimal bids derived from the functional IV estimates appear to be substantially closer to actual bids than both the bids from the least-squares estimate and the HP benchmark. The IME show that on average, the IV benchmark delivers supply functions less than 5% below the observed functions for all firms except NYPA and Astoria Generating Company.[79] Those results suggest that firms behave relatively close to optimum despite their limited available information about rivals' bids. The actual

[79] It must be noted that Astoria Generating Company operates dual-fired capacity using oil and gas interchangeably, thus its cost function may not be accurate as the shares of

supply functions appear to be less aggressive than at the optimum, i.e. firms tend to choose actual profit margins that are slightly too large on average than predicted by the model.

Remark that the IMSE for IV estimates is not always smaller than for least-squares estimates. This is because the IMSE captures both the bias and variance, and the IV estimates reduce the bias at the cost of an increased variance. The IMEs clearly show that IV estimates lead to a reduced bias.

In the next section, counterfactual expected profits are compared to actual ones as a measure of firm's performance.

Expected profits

This section proposes an evaluation of the extent to which each firm's bidding strategy performs in terms of maximum expected profit achieved. Under the first-order approximation, it is possible to compute the counterfactual price distribution $F_t^\star$ resulting from the optimal supply function as

$$F_t^\star(p) = \hat{F}_t(p) + \hat{\pi}(p)\Big(S_t^\star(p) - \tilde{S}_t(p)\Big), \tag{3.80}$$

which admits the counterfactual density function $f_t^\star$.[80] Optimal bids appear to result in slightly larger prices in expectations.

The counterfactual expected profit is given by

$$E[\Pi_t\Big(S_t^\star\Big)] = \int\limits_{\mathcal{P}} \Bigg(pS_t^\star(p) - c_t(S_t^\star(p)) - (p - PC_t)QC_t\Bigg) f_t^\star(p)dp, \tag{3.81}$$

and is used to compute a measure of maximum profit achieved, as in Hortaçsu and Puller (2008) and Sioshansi and Oren (2007). Let us consider the ratio of the sums of expected profits[81] over the sample defined by

$$\Pi_{\%} = 100 \times \frac{\sum_{t=1}^T E[\Pi_t\Big(S_t^0\Big)]}{\sum_{t=1}^T E[\Pi_t\Big(S_t^\star\Big)]}. \tag{3.82}$$

each fuel is kept fixed over time whereas, in reality, it may change from one day to the next.

[80] I ensure that the counterfactual distribution is well-defined by forcing values to be between 0 and 1. The density function is obtained using penalized spline smoothing. It is forced to take positive values and rescaled to ensure that $\int f_t^\star(p)dp = 1$.

[81] This measure is preferred to the mean of ratios because of the total firm's capacity varies across periods. The ratio of sums weights each period's percent achieved by the profit size.

Firm	Performance $\Pi_{\%}$
New York Power Authority	41%
Long Island Power Authority	95%
NRG Power Marketing	75%
Entergy Nuclear Power Marketing	100%
Consolidated Edison	92%
TC Ravenswood	84%
Exelon Generation Company	100%
Astoria Generating Company	67%
Athens Generating Company	82%
Astoria Energy	94%

NOTES. This table reports the performance measure in terms of expected profit achieved for the 10 largest firms. Counterfactual profits for NYIPA, Entergy and Exelon are based on the least-squares estimates, whereas all other firm's measures are obtained using the IV estimates.

Table 3.13: Percent achieved of maximum expected profits

Table 3.13 presents the percent achieved for the 10 largest firms, based on IV estimates. I find that most firms achieve more than 75% of maximum counterfactual profits over the sample period. Only NYPA and Astoria Generating Company achieve smaller performance.

It appears that the method proposed in this paper captures reasonably well the observed behavior of firms in the New York electricity market. I consider an explanation for the observed deviations from optimal bidding in the next section.

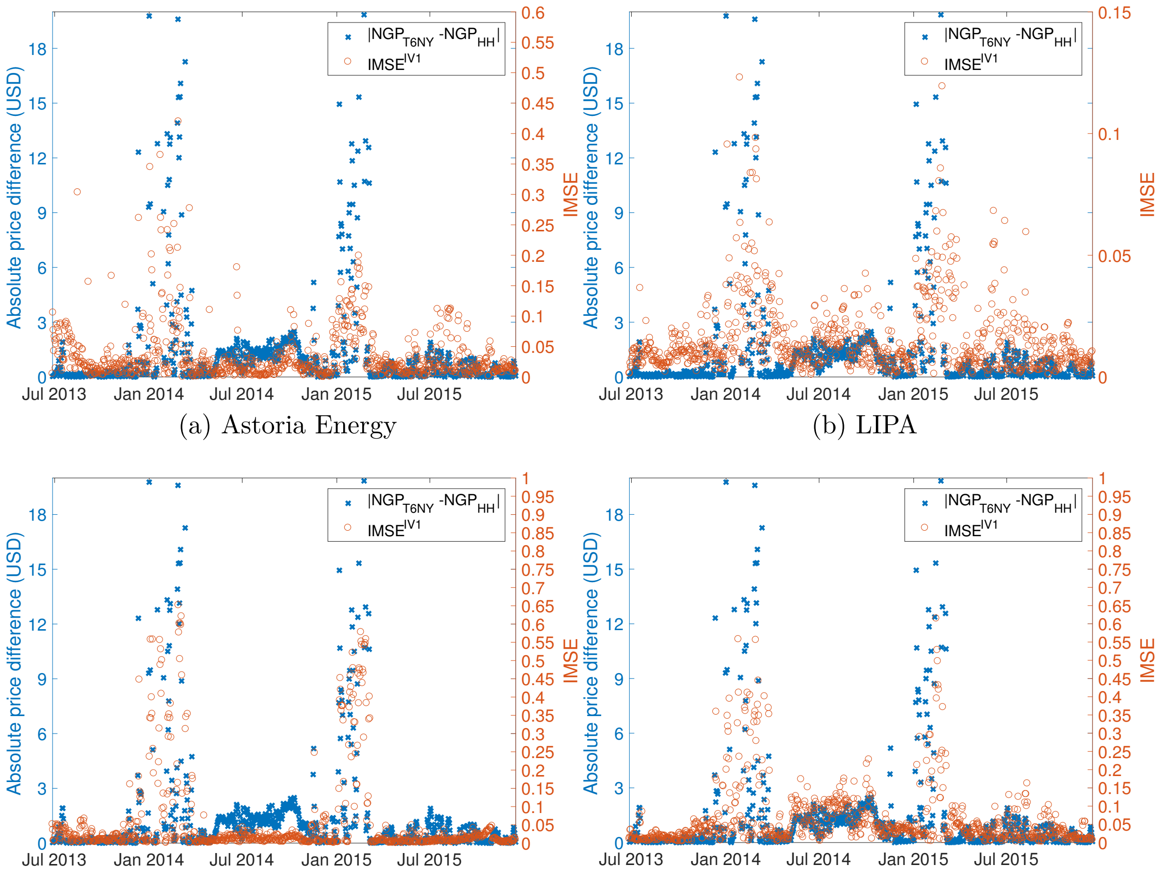
Deviations from optimal bidding

I consider the deviations from optimal bidding as relatively small, and most likely due to errors in the construction of the marginal cost functions. Most importantly, the price of natural gas paid by firms in the New York State may differ from the spot price at the reference hub. That price paid varies also across cities, depending on the delivery hub location and existing congestions. It may as well vary across firms as they hold fuel inventories and long-term contracts.

The effect of measurement errors in natural gas prices on supply function estimates is illustrated in Figure 3.15 for Astoria Energy, LIPA, Astoria Generating Company and TC Ravenswood. The figures plot the $IMSE^{IV1}$ and the absolute difference between the daily spot price of natural gas at Henry Hub located in Louisiana, that is used to construct marginal costs, and at Transco Zone 6 in New York State, from July 2013 to December 2015.[82] During the winter storms of 2014 and 2015, the price spread becomes large – up to USD 120 in January 2014 – and the IMSE increases substantially. These price spreads are explained by delivery issues caused by congested gas transportation infrastructures.[83]

[82]The data on the spot price of natural gas at the Transco Zone 6 New York state is available for July 2013 onwards from the Wall Street Journal at http://www.wsj.com/mdc/public/page/2_3023-cashprices.html.

[83]See https://www.eia.gov/todayinenergy/detail.php?id=14651.



(a) Astoria Energy (b) LIPA

(c) Astoria Generating Company (d) TC Ravenswood

Figure 3.15: IMSE and natural gas price spread between Henry Hub and Transco 6 NY

However, the price at Transco Zone 6 only provides yet another proxy of the true price paid by firms because of hedging contracts. Better data on natural gas prices could thus further improve the results in terms of IMSE and other statistics.

If one is convinced that the method delivers accurate estimates of market power, then marginal costs may be directly estimated from the optimality condition. I discuss this estimation in the next section.

Estimation of marginal costs under optimal bidding

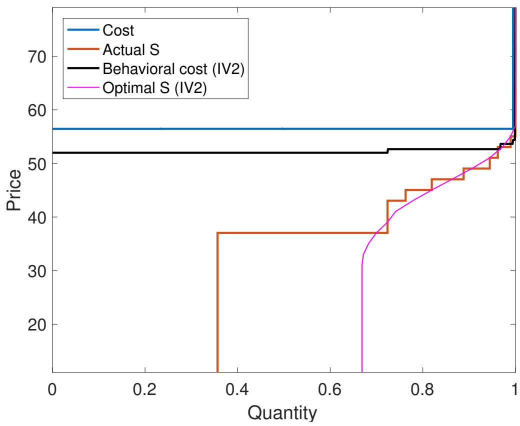
The previous results provide an empirical validation of the IV estimation method. It may hence be used to estimate marginal costs under the assumption that the observed bidding behavior is optimal, by using the IV estimates of market power. This approach yields daily estimates of marginal cost functions. In a sense, they consist of non-parametric versions of the *behavioral* marginal cost estimates of Wolak (2003) and related works.

Non-parametric estimates are obtained using Algorithm 4 presented in appendix C.4. The algorithm works as follows for any given day in the sample. For each value of the forward quantity QC in a given grid $\mathcal{QC}$, I solve recursively for the monotone function that satisfies the optimality condition given $\hat{\pi}$, $\hat{f}$ and the observed supply function S^0. This yields a set of candidate marginal cost functions C_{QC}, each associated to a single forward position $QC \in \mathcal{QC}$. Then, I choose the cost function that corresponds to the solution of the minimization problem

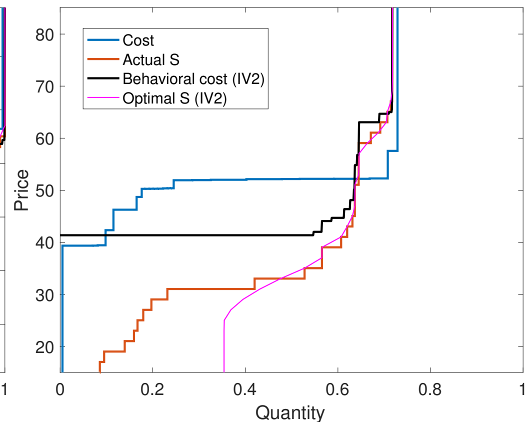
$$\min_{QC \in \mathcal{QC}} \int_{\mathcal{P}} \left(p - C_{QC}(p) - \left(S^0(p) - QC\right) \frac{\hat{\pi}(p)}{\hat{f}(p)} \right)^2 \hat{f}(p) dp, \qquad (3.83)$$

where $\hat{f}$ is used as a weighting function. Confidence intervals can be constructed like for optimal bids.

For illustration purposes, I compute estimates of marginal cost functions for Astoria Energy and LIPA. Resulting estimates using the lagged supply function as IV are presented in Figure 3.16a and Figure 3.16b. The optimal bid functions computed from these marginal costs are also shown.



(a) Astoria Energy (May 7th, 2014)



(b) LIPA (August 20th, 2015)

Figure 3.16: Estimated behavioral marginal costs

3.6.3 Conclusion of the application

This empirical analysis of the New York day-ahead electricity market delivers several important conclusions with regards to the bidding behavior of participants.

The first step of the analysis has shown evidence that the observed bidding behavior of market participants does not correspond to the standard model of strategic behavior in the absence of uncertainty. This model, referred to as the HP benchmark in the analysis, is widely rejected by the data on the basis of a three-fold analysis. First, the functional testing procedures suggest that the optimality condition resulting from that model is not satisfied for any firm. Second, the additive separability restriction of supply strategies in the firms' private information, which somehow allows to neglect the effect of uncertainty on strategies, is largely rejected by the data for most firms. Finally, the observed bids appear to diverge systematically from the optimal bids predicted by that model. This result suggests that private information should be accounted for in this market. It is however possible that some deviations between the HP benchmark and the observed behavior arise from binding transmission constraints instead of the failure of the AS restriction. In effect, conducting the analysis on segmented markets – as proposed by Mercadal (2016) – may provide further insights on this aspect. It would notably allow to identify whether deviations between optimal bids obtained under the AS restriction and from the functional estimates differ because of private information per se or the complex market environment.

In a second stage, I proposed an alternative benchmark based on the estimated firm-level ability to affect the market-clearing price probability distribution with its supply bids. This ability to affect the price distribution corresponds to a measure of a firm's unilateral market power. This approach avoids the exercise of modelling the market-clearing mechanism of the day-ahead auctions, while still being able to provide insights on bidding behavior. I estimate that 7 out of the 10 largest firms do affect the price distribution with their bid functions. These firms mainly operate gas-fired capacity, whereas the 3 remaining firms operate infra-marginal technologies. Furthermore, the optimal supply functions resulting from these estimates are found to be reasonably close to the observed offer bids. It suggests that firms are well aware of their ability to exercise market power when forming their bids.

Finally, I compare the expected profits computed from the observed price distribution to the counterfactual expected profits under the optimal strategy and find that firms are able to earn more than 70% of maximum profit over 2013-2015. I consider these results as evidence of strategic bidding for the largest firms in NYISO, and show that the largest deviations between observed and optimal bid functions can be attributed to measurement errors in the firm-level marginal cost functions.

I believe that the repeated nature of these auctions may facilitate the collection of strategic information, in spite of the limitations imposed by the regulatory framework. Consequently, a dynamic model of strategic bidding may provide additional insights on how firms collect and use information in repeated multi-unit auctions.

3.7 Conclusion

This paper presents a novel approach for the empirical analysis of multi-unit auctions based on functional econometric methods. Functional estimation and inference results are developed. Notably, the method aims at estimating the optimal bid function in a multi-unit auction from a given marginal cost function in a private information context, without having to replicate the market mechanism.

An application to the New York state's day-ahead electricity market is proposed. The transmission limitations in this market generate uncertainty surrounding the realizations of market-clearing location-based prices. Firms also hold contract positions that consist of private information. Furthermore, firms in this market cannot observe any recent rival bids. In this context, one may suspect that forecasting residual demand functions is difficult. I construct a dataset for the New York state's day-ahead electricity auctions by de-anonymizing the bid data and calibrate firm-level marginal cost functions using an engineering method. The observed supply functions are found to be relatively close to the optimal ones estimated with my method. This suggests that firms are well-aware of the influence of their bidding strategy on the distribution of equilibrium prices, and are able to exert market power accordingly.

The analysis provides evidence in support of the economic intuition that firms do not need to observe past realizations of rival bids to develop their strategies in multi-unit auctions. Instead, they can simply use their own past bids or private information, own marginal cost estimation and the realizations of equilibrium prices. This is not to say that late bid disclosure is not helpful to prevent the abuse of market power. Late bid disclosure can be effective in deterring collusion and preventing firms to evaluate their market power in rare events. Yet, it cannot be considered as sufficient to prevent firms to collect strategic information about their ability to influence market prices – at least for prices that occur regularly.

On a side note, I find that the anonymity of bidders is not guaranteed. Firms with substantial stake cannot be expected to be fooled by the apparent anonymity of bids. However, it may deter independent research on the bidding behavior of firms.

There are several possible extensions of this research. As for econometric methods, it would be interesting to have a functional approach for settings where the market mechanism is modelled. Moment conditions could then be evaluated not only around the observed equilibrium price, but the range of prices with non-zero probabilities. Non-parametric functional regression methods extended to the IV setting would consist of an interesting alternative approach. Concerning the empirical studies of multi-unit auctions, an economic model of oligopoly learning may provide a useful framework for the study of dynamic bidding strategies.

Conclusion

This thesis proposed an analysis of energy markets and the development of econometric methods in three chapters. A micro-economic framework was developed in Chapter 1 as a means to study market power exertion and market manipulations in sequential markets under imperfect competition. The new theoretical results for functional regression models derived in Chapter 2 were partly motivated by the functional nature of supply strategies in electricity markets. Finally, Chapter 3 presented a new empirical approach to analyze firm-level market power in multi-unit procurement auctions based on functional econometric methods. The method was then applied in an empirical study of strategic behaviors in the New York electricity market.

Market power exertion and misconducts threaten the very own objectives of electricity markets: an improved economic efficiency of the energy sector and a fair distribution of the efficiency gains. This thesis underlines the importance of sound regulation, market design and economic policies to control market power and curb anti-competitive conducts in energy markets. Results suggest that firms are well-aware of their market power and act accordingly, in spite of the regulatory tools used for market power mitigations. In addition, the regulatory instruments designed to mitigating market power may sometimes give birth to undesirable incentives for manipulative conducts, which can be particularly concerning.

This research also makes the argument that functional data analysis provides a natural framework for the empirical analysis of market behaviors in electricity markets because of the functional nature of strategies. The methodology nonetheless has a more general scope.

Finally, this thesis provides a ground for future research. An important extension would be to consider an endogenous forward market. Another promising research avenue lies in the development of functional methods in structural econometrics.

Bibliographie

Bibliography

Ait-Sahalia, Y. and A. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *The Journal of Finance*, LIII, 499–547.

Akerlof, G. A. and R. J. Shiller (2015): *Phishing for phools: The economics of manipulation and deception*, Princeton University Press.

Allaz, B. and J.-L. Vila (1993): "Cournot competition, forward markets and efficiency," *Journal of Economic theory*, 59, 1–16.

Andersson, J. and J. Lillestol (2010): "Modeling and Forecasting Electricity Consumption by Functional Data Analysis," *The Journal of Energy Markets*, 3–15.

Aneiros, G., J. M. Vilar, R. Cao, and A. M. San Roque (2013): "Functional Prediction for the Residual Demand in Electricity Spot Markets," *IEEE Transactions on Power Systems*, 28, 4201–4208.

Antoch, J., L. Prchal, M. De Rosa, and P. Sarda (2010): "Electricity Consumption Prediction with Functional Linear Regression using Spline Estimators," *Journal of Applied Statistics*, 2027–2041.

Armantier, O., J.-P. Florens, and J.-F. Richard (2008): "Approximation of Nash eEquilibria in Bayesian Games," *Journal of Applied Econometrics*, 23, 965–981.

Armantier, O. and E. Sbaï (2006): "Estimation and Comparison of Treasury Auction Formats when Bidders are Asymmetric," *Journal of Applied Econometrics*, 21, 745–779.

AUE, A., D. NORINHO, AND S. HORMANN (2015): "On the Prediction of Stationary Functional Time Series," *Journal of American Statistical Association*, 378–392.

AUSUBEL, L. M., P. CRAMTON, M. PYCIA, M. ROSTEK, AND M. WERETKA (2014): "Demand Reduction and Inefficiency in Multi-unit Auctions," *The Review of Economic Studies*, rdu023.

BABII, A. (2016): "Honest Confidence Sets in Nonparametric IV Regression and Other Ill-Posed Models," *Working Paper*.

BACK, K. AND J. F. ZENDER (1993): "Auctions of Divisible Foods: On the Rationale for the Treasury Experiment," *The Review of Financial Studies*, 6, 733–764.

BAGNOLI, M., S. W. SALANT, AND J. E. SWIERZBINSKI (1989): "Durable-goods monopoly with discrete demand," *Journal of Political Economy*, 97, 1459–1478.

BAUTISTA, G., M. F. ANJOS, AND A. VANNELLI (2007): "Modeling Market Power in Electricity Markets: Is the Devil Only in the Details?" *The Electricity Journal*, 20, 82 – 92.

BELLMAN, R. E. (1961): *Adaptive Control Processes*, Princeton University Press.

BENATIA, D., E. BILLETTE DE VILLEMEUR, AND P.-O. PINEAU (2018): "Strategic Reneging in Sequential Markets under Imperfect Competition," *Working Paper*.

BENATIA, D., M. CARRASCO, AND J.-P. FLORENS (2017): "Functional Linear Regression with Functional Response," *Journal of Econometrics*, 201, 269–291.

BENATIA, D. AND T. KOŹLUK (2016): "The Determinants of Entry in The Electricity Generation Sector in OECD Countries: A Focus on Renewable Energy," *OECD Environment Working Papers, no. 111, OECD Publishing, Paris.*

BERGLER, J., S. HEIM, AND K. HÜSCHELRATH (2017): "Strategic Capacity Withholding through Failures in the German-Austrian Electricity Market," *Energy Policy*, 102, 210–221.

Bernhardt, D. and D. Scoones (1994): “A Note on Sequential Auctions,” *The American economic review*, 84, 653–657.

Birge, J., A. Hortaçsu, I. Mercadal, and M. Pavlin (2014): “Limits to Arbitrage in Electricity Markets,” *Working Paper*.

Bondarenko, O. (2003): “Estimation of Risk-Neutral Densities using Positive Convolution Approximation,” *Journal of Econometrics*, 85–112.

Borenstein, S., J. B. Bushnell, and F. A. Wolak (2002): “Measuring Market Inefficiencies in California’s Restructured Wholesale Electricity Market,” *American Economic Review*, 1376–1405.

Bosco, B., L. Parisio, and M. Pelagatti (2012): “Strategic Bidding in Vertically Integrated Power Markets with an Application to the Italian Electricity Auctions,” *Energy economics*, 34, 2046–2057.

——— (2013): “Price-Capping in Partially Monopolistic Electricity Markets with an Application to Italy,” *Energy Policy*, 54, 257–266.

Bosq, D. (2000): *Linear Processes in Function Spaces: Theory and Linear Processes in Function Spaces: Theory and applications*, Springer-Verlag, New York.

Brown, D. P. and D. E. Olmstead (2017): “Measuring Market Power and the Efficiency of Alberta’s Restructured Electricity Market: An Energy-only Market Design,” *Canadian Journal of Economics/Revue canadienne d’économique*, 50, 838–870.

Cardot, H., F. Ferraty, and P. Sarda (2003): “Spline Estimators for the Functional Linear Model,” *Statistica Sinica*, 571–591.

Carlton, D. W. and K. Heyer (2008): “Appropriate Antitrust Policy towards Single-Firm Conduct,” *Economic Analysis Group Discussion Paper No. EAG*, 08–2.

Carrasco, M. and J.-P. Florens (2000): “Generalization of GMM to a Continuum of Moment Conditions,” *Econometric Theory*, 16, 797–834.

——— (2011): “A Spectral Method for Deconvolving a Density,” *Econometric Theory*, 27, 546–581.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," *Handbook of econometrics*, 6, 5633–5751.

——— (2014): "Asymptotic Normal Inference in Linear Inverse Problems," *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 73, 140.

Centorrino, S. (2016): "Data-Driven Selection of the Regularization Parameter in Additive Nonparametric Instrumental Regressions," Mimeo, Stony Brook University.

Centorrino, S., F. Fève, and J. P. Florens (2017): "Additive Nonparametric Instrumental Regressions: A Guide to Implementation," *Journal of Econometric Methods*, 6, 571–591.

Centorrino, S. and J. Florens (2015): "Nonparametric Instrumental Variable Estimation of Binary Response Models," Mimeo, Stony Brook University.

Chen, X. (2011): *Penalized Sieve Estimation and Inference of Semi-Nonparametric Dynamic Models: a Selective Review*, Cambridge University Press.

Coase, R. H. (1972): "Durability and Monopoly," *The Journal of Law and Economics*, 15, 143–149.

Comte, F. and J. Johannes (2012): "Adaptive Functional Linear Regression," *Annals of Statistics*, 40, 2765–2797.

Conley, T., L. Hansen, E. Luttmer, and J. Scheinkman (1997): "Short-Term Interest Rates as Subordinated Diffusions," *The Review of Financial Studies*, 10, 2765–2797.

Crambes, C., A. Kneip, and P. Sarda (2009): "Smoothing Splines Estimators for Functional Linear Regression," *Ann. Statist.*, 37, 35–72.

Crambes, C. and A. Mas (2013): "Asymptotics of Prediction in Functional Linear Regression with Functional Outputs," *Bernoulli*, 19, 2627–2651.

Cuevas, A., M. Febrero, and R. Fraiman (2002): "Linear Functional Regression: The Case of Fixed Design and Functional Response," *Canadian Journal of Statistics*, 30, 285–300.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

Dauxois, J., A. Pousse, and Y. Romain (1982): "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Spplications to Statistical Inference," *Journal of Multivariate Analysis*, 12, 136 – 154.

Debia, S., D. Benatia, and P.-O. Pineau (2018): "Evaluating an Interconnection Project: Do Strategic Interactions Matter?" *Working Paper*.

Dufour, J.-M. and P. Valéry (2016): "Rank-Robust Wald-type Tests: a Regularization Approach," Working paper.

Elyakime, B., J. J. Laffont, P. Loisel, and Q. Vuong (1994): "First-Price Sealed-Bid Auctions with Secret Reservation Prices," *Annales d'Economie et de Statistique*, 115–141.

Engl, H., M. Hanke, and A. Neubauer (1996): *Regularization of Inverse Problems*, Mathematics and Its Applications, Springer Netherlands.

Farebrother, R. (1990): "Algorithm AS 256: The Distribution of a Quadratic Form in Normal Variables," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39, 294–309.

FERC (2012): "Constellation Energy Commodities Group, Inc., 138 FERC Â¶ 61,168. March 9, 2012," Tech. rep.

——— (2013a): "Barclays Bank PLC, et al., 144 FERC Â¶ 61,041. July 16, 2013," Tech. rep.

——— (2013b): "Deutsche Bank Energy Trading, LLC, 142 FERC Â¶ 61,056. January 22, 2013," Tech. rep.

——— (2013c): "In re Make-Whole Payments and Related Bidding Strategies (JP Morgan Ventures Energy Corporation), 144 FERC Â¶ 61,068. July 30, 2013," Tech. rep.

Ferraty, F. and P. Vieu (2006): *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Science & Business Media.

Florens, J.-P. (2003): "Inverse Problems and Structural Econometrics," *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, 2, 46–85.

Florens, J.-P., J. Johannes, and S. Van Bellegem (2011): "Identification and Estimation by Penalization in Nonparametric Instrumental Regression," *Econometric Theory*, 27, 472–496.

Florens, J.-P. and S. Van Bellegem (2015): "Instrumental Variable Estimation in Functional Linear Models," *Journal of Econometrics*, 186, 465–476.

Fogelberg, S. and E. Lazarczyk (2014): "Strategic Withholding through Production Failures," *IFN Working Paper*.

Gans, J. S. and F. A. Wolak (2008): "A Comparison of Ex-Ante versus Ex-Post Vertical Market Supply: Evidence from the Electricity Supply Industry," *Working Paper*.

Garcia, R. and R. Gençay (2000): "Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint," *Journal of Econometrics*, 94, 93–115.

Garcia, R., E. Ghysels, and E. Renault (2010): "The Econometrics of Option Pricing," *Handbook of Financial Econometrics*, 1, 479–552.

Genc, T. S. and S. S. Reynolds (2011): "Supply Function Equilibria with Capacity Constraints and Pivotal Suppliers," *International Journal of Industrial Organization*, 29, 432–442.

Goldenshluger, A., O. Lepski, et al. (2011): "Bandwidth Selection in Kernel Density Estimation: Oracle Inequalities and Adaptive Minimax Optimality," *The Annals of Statistics*, 39, 1608–1632.

Gourieroux, C. and J. Jasiak (2001): "Financial Econometrics," .

Green, R. and D. Newberry (1992): "Competition in the British Electricity Spot Market," *B Journal of Political Economy*, 100.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574.

HALL, P., J. L. HOROWITZ, ET AL. (2007): "Methodology and Convergence Rates for Functional Linear Regression," *The Annals of Statistics*, 35, 70–91.

HANSEN, L. P. AND T. J. SARGENT (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and control*, 2, 7–46.

HARVEY, S. M. AND W. W. HOGAN (2001): "Market Power and Withholding," *Harvard Univ., Cambridge, MA*.

HILL, P. D. (1985): "Kernel Estimation of a Distribution Function," *Communications in Statistics-Theory and Methods*, 14, 605–620.

HOGAN, W. W. (1997): "A Market Power Model with Strategic Interaction in Electricity Networks," *The Energy Journal*, 107–141.

HOLMBERG, P., D. NEWBERY, AND D. RALPH (2013): "Supply Function Equilibria: Step Functions and Continuous Representations," *Journal of Economic Theory*, 148, 1509–1551.

HOROWITZ, J. L. AND S. LEE (2007): "Nonparametric instrumental variables estimation of a quantile regression model," *Econometrica*, 75, 1191–1208.

HORTAÇSU, A. (2002): "Mechanism Choice and Strategic Bidding in Divisible Good Auctions: An Empirical Analysis of the Turkish Treasury Auction Market," .

HORTAÇSU, A. AND D. MCADAMS (2010): "Mechanism Choice and Strategic Bidding in Divisible Good Auctions: An Empirical Analysis of the Turkish Treasury Auction Market," *Journal of Political Economy*, 118, 833–865.

——— (2017): "Empirical Work on Auctions of Multiple Objects," *Journal of Economic Literature*, Forthcoming.

Hortaçsu, A. and S. 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.

Horváth, L. and P. Kokoszka (2012): *Inference for functional data with applications*, vol. 200, Springer Science & Business Media.

Imhof, J. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.

Ito, K. and M. Reguant (2016): "Sequential Markets, Market Power, and Arbitrage," *American Economic Review*, 106, 1921–57.

Jackwerth, J. C. and M. Rubinstein (1996): "Recovering probability distributions from option prices," *The Journal of Finance*, 51, 1611–1631.

Jha, A. and F. A. Wolak (2015): "Testing for Market Efficiency with Transaction Costs: An Application to Financial Trading in Wholesale Electricity Markets," *Working Paper*.

Joskow, P. L. (2008a): "Capacity Payments in Imperfect Electricity Markets: Need and Design," *Utilities Policy*, 16, 159 – 170, capacity Mechanisms in Imperfect Electricity Markets.

——— (2008b): "Lessons Learned from Electricity Market Liberalization," *The Energy Journal*, 29, 9–42.

Joskow, P. L. and E. Kahn (2002): "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000," *The Energy Journal*, 0, 1–35.

Kargin, V. and A. Onatski (2008): "Curve Forecasting by Functional Autoregression," *Journal of Multivariate Analysis*, 99, 2508–2526.

Kastl, J. (2012): "On the Properties of Equilibria in Private Value Divisible Good Auctions with Constrained Bidding," *Journal of Mathematical Economics*, 48, 339–352.

Kennan, J. (1979): "The Estimation of Partial Adjustment Models with Rational Expectations," *Econometrica: Journal of the Econometric Society*, 1441–1455.

KLEMPERER, P. D. AND M. A. MEYER (1989): “Supply Function Equilibria in Oligopoly under Uncertainty,” *Econometrica*, 57, 1243–1277.

KRESS, R. (1999): “Linear integral equations, volume 82 of Applied Mathematical Sciences,” .

LEDGERWOOD, S. D. AND P. R. CARPENTER (2012): “A Framework for the Analysis of Market Manipulation,” *Review of Law & Economics*, 8, 253–295.

LEDGERWOOD, S. D. AND J. P. PFEIFENBERGER (2013): “Using Virtual Bids to Manipulate the Value of Financial Transmission Rights,” *The Electricity Journal*, 26, 9 – 25.

LEHMANN, E. L. AND J. P. ROMANO (2007): *Testing Statistical Hypotheses, 3rd edn*, Springer.

LIEBL, D. ET AL. (2013): “Modeling and Forecasting Electricity Spot Prices: A Functional Data Perspective,” *The Annals of Applied Statistics*, 7, 1562–1592.

MANRESA, E. (2013): “Estimating the Structure of Social Interactions using Panel Data,” *Unpublished Manuscript. CEMFI, Madrid.*

MANSUR, E. T. (2007): “Do Oligopolists Pollute Less? Evidence from a Restructured Electricity Market,” *The Journal of Industrial Economics*, 55, 661–689.

MCAFEE, R. P. AND D. VINCENT (1993): “The Declining Price Anomaly,” *Journal of Economic Theory*, 60, 191–212.

MCLAUGHLIN, L., L. JIA, Z. YU, AND L. TONG (2011): “Thermal Dynamic for Home Energy Management: A Case Study,” *Cornell Univ., Ithaca, NY, USA, Tech. Rep. ACSP-TR-10-11-01.*

MERCADAL, I. (2016): “Dynamic Competition and Arbitrage in Electricity Markets: The Role of Financial Players,” *Working paper.*

MSA (2012): “Application for Approval of a Settlement Agreement between the Market Surveillance Administrator and TransAlta Energy Marketing Corp, Decision 2012-182, July 3, 2012,” .

——— (2015): "Market Surveillance Administrator allegations against TransAlta Corporation et al., Mr. Nathan Kaiser and Mr. Scott Connelly, Phase 1, Decision 3110-D01-2015, July 27, 2015," *Alberta Utilities Commission*.

Muth, J. F. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica: Journal of the Econometric Society*, 315–335.

NYISO (2015): "Market Participants User's Guide," Tech. rep., New York Independent System Operator.

——— (2016): "Day-Ahead Scheduling Manual," Tech. rep., New York Independent System Operator.

Olmstead, D. E. and M. J. Ayres (2014): "Notes from a Small Market: The Energy-only Market in Alberta," *The Electricity Journal*, 27, 102–111.

Paarsch, H. J. (1992): "Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of econometrics*, 51, 191–215.

Park, J. Y. and J. Qian (2012): "Functional Regression of Continuous State Distributions," *Journal of Econometrics*, 167, 397–412.

Patrick, R. H. and F. A. Wolak (2001): "Estimating the Customer-level Demand for Electricity under Real-time Market Prices," Tech. rep., National Bureau of Economic Research.

Pedersen, G. K. (1989): *Analysis now*, vol. 118, Springer Science & Business Media.

Pelagatti, M. (2013): "Supply Function Prediction in Electricity Auctions," in *Complex Models and Computational Methods in Statistics*, Springer, 203–213.

Prüggler, N., W. Prüggler, and F. Wirl (2011): "Storage and Demand Side Management as power generator's strategic instruments to influence demand and prices," *Energy*, 36, 6308–6317.

Ramsay, J. and B. Silverman (2005): "Springer series in statistics," .

Ramsay, J. O. (1998): "Estimating Smooth Monotone Functions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60, 365–375.

Reguant, M. (2014): "Complementary bidding mechanisms and startup costs in electricity markets," *The Review of Economic Studies*, 81, 1708–1742.

Reiss, P. C. and F. A. Wolak (2007): "Structural Econometric Modeling: Rationales and Examples from Industrial Organization," *Handbook of econometrics*, 6, 4277–4415.

Reiss, R.-D. (1981): "Nonparametric Estimation of Smooth Distribution Functions," *Scandinavian Journal of Statistics*, 116–119.

Rudin, W. (1991): "Functional analysis. International series in pure and applied mathematics," .

Ryan, N. (2013): "The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market," *Working paper.*

Schill, W.-P. and C. Kemfert (2011): "Modeling Strategic Electricity Storage: the Case of Pumped Hydro Storage in Germany," *Energy Journal*, 32, 59–87.

Schöne, S. (2009): *Auctions in the Electricity Market: Bidding when Production Capacity is Constrained*, vol. 617, Springer Science & Business Media.

Shorack, G. and J. Wellner (1986): *Empirical Processes with Applications to Statistics*, John Wiley&Sons.

Silverman, B. W. (1982): "On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method," *The Annals of Statistics*, 795–810.

Sioshansi, R. and S. Oren (2007): "How Good are Supply Function Equilibrium Models: an Empirical Analysis of the ERCOT balancing market," *Journal of Regulatory Economics*, 31, 1–35.

Tirole, J. (2014): "Prize Lecture: Market Failures and Public Policy," *Nobel Media AB.*

Vilenkin, N. J. (1968): *Special Functions and the Theory of Group Representation*, American Mathematical Society.

von der Fehr, N.-H. M. and D. Harbord (1993): "Spot Market Competition in the UK Electricity Industry," *The Economic Journal*, 103, 531–546.

Wahba, G. (1990): *Spline Models for Observational Data*, SIAM.

Weber, R. J. et al. (1981): *Multiple-Object Auctions*, Northwestern University.

Wilson, R. (1979): "Auctions of Shares," *The Quarterly Journal of Economics*, 675–689.

Wolak, F. (2009): "An Assessment of the Performance of the New Zealand Wholesale Electricity Market," *Report for the New Zealand Commerce Commission.*

Wolak, F. A. (2000): "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market," *International Economic Journal*, 14, 1–39.

——— (2003): "Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets," *Advances in Economics and Econometrics - Theory and Applications*, 115–149.

——— (2007): "Quantifying the Supply-Side Benefits from Forward Contracting in Wholesale Electricity Markets," *Journal of Applied Econometrics*, 22, 1179–1209.

Y., A.-S., Y. Wang, and F. Yared (2001): "Do Option Markets Correctly Price the Probabilities of Movement of the Underlying Asset?" *Journal of Econometrics*, 67–110.

Yao, F., H.-G. Müller, J.-L. Wang, et al. (2005): "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903.

Yu, Z., L. Jia, M. C. Murphy-Hoye, A. Pratt, and L. Tong (2013): "Modeling and Stochastic Control for Home Energy Management," *IEEE Transactions on Smart Grid*, 4, 2244–2255.

Annexes

Appendices

A Chapter 1

Proof of Lemma 1. For $\theta \in [\underset{\sim}{\theta}, \tilde{\theta}]$, the strategic firm faces two competing strategies, either playing as a residual monopolist to earn

$$\pi_m^\star(\theta) = \frac{b(\theta - K_e - \rho_f K_f)^2}{4} + rK_e,$$

or as a pivotal supplier to obtain

$$\pi_m^\dagger(\theta) = (\bar{P} - c)(\theta - K_f - K_e) + rK_e.$$

It is thus profitable for the monopolist to favor the pivotal strategy if

$$\pi_m^\dagger(\theta) - \pi_m^\star(\theta) = (\bar{P} - c)(\theta - K_f - K_e) - \frac{b(\theta - K_e - \rho_f K_f)^2}{4} \geq 0,$$

which is satisfied as soon as $\theta \geq \theta_1 = \inf\{\theta \in [\underset{\sim}{\theta}, \tilde{\theta}] : \pi_m^\dagger(\theta) - \pi_m^\star(\theta) \geq 0\}$, given that $\pi_m^\dagger(\theta) - \pi_m^\star(\theta)$ is monotone increasing. This is assured by Assumption 1 since

$$\pi_m^{\dagger\prime}(\theta) - \pi_m^{\star\prime}(\theta) = (\bar{P} - c) - \frac{b(\theta - K_e - \rho_f K_f)}{2} > 0, \forall \theta \in [\underset{\sim}{\theta}, \tilde{\theta}] \leftrightarrow \bar{P} > c + b(1 - \rho_f)K_f$$

Consequently, finding θ_1 consists in solving for the roots of the 2nd-order polynomial

$$(\bar{P} - c)q - \frac{b(q + (1 - \rho_f)K_f)^2}{4}, \tag{A.1}$$

where $q = \theta - K_e - K_f$. Under Assumption 1, this polynomial admits two roots since $\Delta_q = 16\frac{\bar{P}-c}{b}((\frac{\bar{P}-c}{b} - (1 - \rho_f)K_f) > 0$, although only the smaller one being feasible given that $2\frac{\bar{P}-c}{b} - (1-\rho_f)K_f) > (1-\rho_f)K_f$. The equilibrium solution is given by

$$q_1^\star = 2\frac{\bar{P}-c}{b} - (1-\rho_f)K_f - 2\sqrt{\frac{\bar{P}-c}{b}((\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)}, \qquad \text{(A.2)}$$

which yields the unique demand threshold $\theta_1 = K_f + K_e + q_1^\star$, completing the proof of (a).

With regards to the properties of this solution, we have

$$\frac{\partial q_1^\star}{\partial \bar{P}} = \frac{2}{b}\left(1 - \frac{1}{2}(\frac{\bar{P}-c}{b}((\frac{\bar{P}-c}{b} - (1-\rho_f)K_f))^{-\frac{1}{2}}(2\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)\right) < 0,$$

since $(1-\rho_f K_f) > 0$ implies $2(\frac{\bar{P}-c}{b}(\frac{\bar{P}-c}{b} - (1-\rho_f)K_f))^{\frac{1}{2}} < 2\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)$. From the definition of θ_1, one directly obtains $\frac{\partial \theta_1}{\partial \bar{P}} < 0$.

Taking the limits of (A.2) respectively for $\bar{P} \to c + b(1-\rho_f)K_f$ and $\bar{P} \to +\infty$ yields

$$\lim_{\bar{P}\to c+b(1-\rho_f)K_f} q_1^\star = (1-\rho_f)K_f,$$

$$\lim_{\bar{P}\to+\infty} q_1^\star = 2\lim_{\bar{P}\to+\infty}\left(\frac{\bar{P}-c}{b}\right) - 2\lim_{\bar{P}\to+\infty}\left(\frac{\bar{P}-c}{b}\right) = 0.$$

Finally, we have that $q_m^\star(\theta_1) = \frac{\theta_1 - K_e - \rho_f K_f}{2} = \frac{q_1^\star + (1-\rho_f)K_f}{2} \geq q_1^\star$, since $q_1^\star \leq (1-\rho_f)K_f$.

□

Proof of Proposition 1. We first prove the following lemma.

Lemma 4 (Naive Equilibria)**.** *Under Assumptions 1-5 and assuming that the strategic player does not anticipate the effects of its own strategy on its rival's capacity constraints, all equilibria are such that:*

(a) Competitive Regime ($\theta \in [\underline{\theta}, \theta_c]$): The fringe produces with the constant marginal cost capacity and both players make zero market profit.

(b) Monopoly Regime ($\theta \in (\theta_c, \tilde{\theta}]$): The fringe produces with his increasing marginal cost capacity and the strategic firm acts as a monopolist on its residual demand.

(c) Pivotal Regime ($\theta \in [\tilde{\theta}, \overline{\theta}]$): The fringe is capacity constrained and the strategic firm has a pivotal position as she supplies only the inelastic residual demand at the price cap.

witse $\tilde{\theta} = K_e + K_f + (1 - \rho_f)K_f$. *The corresponding equilibrium supply schedule strategies are such that:*

$$S_0^\star(p) = \begin{cases} 0 & \text{if } p \in [0, c) \\ q_m \in [0, K_m] & \text{if } p = c \\ \frac{p-c}{b} & \text{if } p \in (c, c + bq_m^\star(\tilde{\theta})] \\ K_m & \text{if } p = \bar{P} \end{cases}$$

The first-order condition of the maximization problem defined in (1.4) is given by

$$P'(q_m)q_m + P(q_m) - c = \overline{\lambda}_m - \underline{\lambda}_m. \tag{A.3}$$

For $\theta \in [\underline{\theta}, \theta_c]$, we have $q_e(\theta) = K_e$ yielding net market demand $\theta - K_e > 0$. The fringe production is $q_f = \{\theta - q_m - q_e(\theta), 0\}^+ \leq \rho_f K_f$. From (1.2), $P(q_m) = c$ and $P'(q_m) = 0$ for $q_f \leq \rho_f K_f$, thus (A.3) becomes $0 = \overline{\lambda}_m - \underline{\lambda}_m$. The solution is hence such that $q_m \in [0, K_m]$ with $q_m + q_f = \theta - K_e$, proving *(a)*.

For $\theta \in (\theta_c, \tilde{\theta}]$, we have an interior solution defined by the FOC

$$-bq_m + b(\theta - K_e - \rho_f K_f - q_m) = 0, \tag{A.4}$$

which yields the explicit solution $q_m = \frac{(\theta - K_e - \rho_f K_f)}{2}$. Substitution in (1.2) gives the market-clearing price and q_f is obtained from the equilibrium condition, which completes the proof of *(b)*.

By definition, $\tilde{\theta}$ corresponds the demand level at which the fringe's capacity constraint is binding when the monopolist follows the strategy above. Formally,

$$\tilde{\theta} = \inf\{\theta : q_f = \theta - K_e - \frac{(\theta - K_e - \rho_f K_f)}{2} \geq K_f\} \tag{A.5}$$

admits the closed-form expression is $\tilde{\theta} = K_e + K_f + (1 - \rho_f K_f)$. For $\theta \in (\tilde{\theta}, \overline{\theta}]$, by construction $q_e(\theta) = K_e$ and $q_f(\theta) = K_f$ and from the equilibrium condition we obtain $q_m(\theta) = \theta - K_f - K_e \geq 0$. Any price strictly below the price cap cannot be an equilibrium since the monopolist always increase its payoff by setting a larger price, up to the maximum price. Hence, we have $p = \bar{P}$ in equilibrium, which completes the proof of the above proposition. □

Let us now turn to the proof of Proposition 1. Lemma 4 implies *(a)* and, $q_m(\theta) = q_m^\star(\theta)$ for $\theta \in [\theta_c, \underset{\sim}{\theta}]$. Lemma 1 implies $q_m(\theta) = q_m^\star$ for $\theta \in [\underset{\sim}{\theta}, \theta_1]$ and $q_m(\theta) = q_m^\dagger(\theta)$ for $\theta \in [\theta_1, \tilde{\theta}]$. The rest of the proof follows directly from Lemma 4.

Proof of Proposition 2. Proposition 1 implies *(a)* and, $q_m(\theta) = q_m^\star(\theta)$ for $\theta \in [\theta_c, \underset{\sim}{\theta}]$. Lemma 1 implies $q_m(\theta) = q_m^\star$ for $\theta \in [\underset{\sim}{\theta}, \theta_1]$ and $q_m(\theta) = q_m^\dagger(\theta)$ for $\theta \in [\theta_1, \tilde{\theta}]$. The rest of the proof follows directly from Proposition 1. □

Proof of Lemma 2. In order to characterize optimal manipulations we consider all five cases.

★ (Within 1) As already proved, manipulations are never profitable if $r > 0$ for $\theta \in [\underline{\theta}, \rho_f K_f)$.

★ (Within 2) For $\theta \in [\theta_c, \theta_1 - K_e)$, RD manipulations are profitable if $\pi_m^\star(\theta, 0) - \pi_m^\star(\theta, K_e) \geq 0$ which is equivalent to

$$\begin{aligned} & \frac{b(\theta - \rho_f K_f)^2}{4} - \left(\frac{b(\theta - K_e - \rho_f K_f)^2}{4} + rK_e \right) \geq 0 \\ \leftrightarrow \quad & \frac{b(\theta - \rho_f K_f)K_e}{2} - \frac{bK_e^2}{4} - rK_e \geq 0 \\ \leftrightarrow \quad & \theta \geq \theta^M := \rho_f K_f + \frac{K_e}{2} + \frac{2r}{b}, \end{aligned}$$

where θ^M is feasible if it belongs to $[\theta_c, \theta_1 - K_e)$, that is for $r \in [\frac{bK_e}{4}, bq_m^\star(\theta_1) - \frac{bK_e}{4})$. Hence, it is profitable to withhold K_e over the whole interval if $r < \frac{bK_e}{4}$, whereas as r goes to the upper bound, it is never profitable in this interval.

★ (Within 3) As already proved, for $\theta \in [\theta_1, \tilde{\theta}]$ it is profitable to set $q_e^\star = 0$ as soon as $r \leq \bar{P} - c$. For $r > \bar{P} - c$, we have $q_e^\star = K_e$.

★ (Shift 1) For $\theta \in [\theta_c - K_e, \theta_c)$, the profitability condition $\pi_m^\star(\theta, 0) - \pi_m^c(\theta, K_e)$ is equivalent to

$$\begin{aligned} & \frac{b(\theta - \rho_f K_f)^2}{4} - rK_e \geq 0 \\ \leftrightarrow \quad & \theta \geq \theta^M := \rho_f K_f + 2\sqrt{\frac{rK_e}{b}}, \end{aligned}$$

which is feasible as long as $r \in [0, \frac{bK_e}{4})$.

★ (Shift 2) For $\theta \in [\theta_1 - K_e, \theta_1)$, if it is profitable to withhold the whole exogenous capacity in both the monopoly and pivotal regimes then $r \in$

$[b(q^\star_m(\theta_1) - \frac{bK_e}{4}, \{b(q^\star_m(\theta_1) + \frac{bK_e}{4}, \bar{P} - c\}^+]$. The optimal $q^\star$ is determined by $\pi^\dagger_m(\theta, 0) - \pi^\star_m(\theta, K_e)$ which is equivalent to

$$(\bar{P} - c)(q^\star) - \frac{b(q^\star + (1-\rho_f)K_f)^2}{4} \geq 0, \tag{A.6}$$

where the left-hand side (LHS) corresponds to the polynomial in (A.1). Thus the solution $q^\star_2$ coincides with $q^\star_1$.

If it is profitable to withhold the exogenous capacity only in the pivotal regime then $\frac{b(q^\star_1 + (1-\rho_f)K_f)}{2} + \frac{bK_e}{4} \leq r \leq \bar{P} - c$. The optimal $q^\star$ is determined by the condition $\pi^\dagger_m(\theta, 0) - \pi^c_m(\theta, 0)$ which is equivalent to

$$\begin{aligned} & (\bar{P} - c)q^\star - \frac{b(q^\star + (1-\rho_f)K_f - K_e)^2}{4} - rK_e \geq 0 \\ \leftrightarrow \quad & \left((\bar{P} - c)q^\star - \frac{b(q^\star + (1-\rho_f)K_f)^2}{4}\right) + K_e\left(b(q^\star_m(\theta_1) - \frac{bK_e}{4} - r\right) \geq 0. \end{aligned}$$

The second term on the LHS is non-positive because $r \geq bq^\star_m(\theta_1) - \frac{bK_e}{4}$. Hence the solution will be such that $q^\star_2 \geq q^\star_1$ given that the first term of the LHS is increasing in $q^\star$ and coincides with the problem without manipulations. The closed-form solution corresponds to the smaller root of the above second-order polynomial, and is given by

$$q^\star_2 = 2\frac{\bar{P} - c}{b} - ((1-\rho_f)K_f - K_e) - 2\sqrt{\frac{\bar{P} - c}{b}(\frac{\bar{P} - c}{b} - ((1-\rho_f)K_f - K_e)) - \frac{rK_e}{b}}.$$

When $r \to \bar{P} - c$, we have

$$\lim_{r \to \bar{P} - c} q^\star_2 = q^\star_1 + K_e,$$

and thus $\lim_{r \to \bar{P} - c} \theta_2 = K_f + \lim_{r \to \bar{P} - c} q^\star_2 = \theta_1$. Conversely, when r goes to its lower bound, $q^\star_2 = q^\star_1$, and thus $\theta_2 = \theta_1 - K_e$.

Finally, if r is such that withholding within both regimes is not profitable, full withholding of K_e is never profitable. A peculiar strategy in this situation might consist in shifting the residual demand in order to exactly trigger the pivotal regime, i.e. setting $q_e(\theta) = \theta - K_f - q^\star \leq K_e$. Such a manipulation is profitable for θ such that

$$\pi^\dagger_m(\theta, \theta - K_f - q^\star) - \pi^\star_m(\theta, K_e) \geq 0, \tag{A.7}$$

which strict inequality never holds for feasible demand levels since the optimal transition across profit functions is continuous in θ. This can be proved formally using the following argument. If $r > \bar{P} - c$, setting $q_e(\theta) = \theta - K_f - q^\star$ is profitable if (A.7) holds or equivalently

$$(\bar{P}-c)(q^\star) - \frac{b(q^\star + (1-\rho_f)K_f)^2}{4} + (K_e - q_e(\theta))(\frac{b(q^\star + (1-\rho_f)K_f))}{2} - \frac{b(K_e - q_e(\theta)}{4} - r) \geq 0,$$

from which substitution of $q_e(\theta)$ and straightforward manipulations yield the condition

$$q^\star \leq \frac{r(\theta - K_f - K_e) - \frac{b(\theta - Ke - \rho_f K_f)^2}{4}}{r - (\bar{P} - c)}.$$

On the other hand, feasibility imposes $q_e = \theta - K_f - q^\star \leq K_e \leftrightarrow q^\star \geq \theta - K_f - K_e$, hence partial withholding is profitable for $\theta \in [\theta_1 - K_e, \theta_1)$ satisfying

$$\theta - K_f - K_e \leq \frac{r(\theta - K_f - K_e) - \frac{b(\theta + (1-\rho_f)K_f - Ke - K_f)^2}{4}}{r - (\bar{P} - c)},$$

from which rearrangements yield (A.6). Therefore, the above condition is satisfied for $\theta \geq \theta_1$ which is not feasible unless $\theta = \theta_1$, hence residual demand manipulations are never profitable when $r > \bar{P} - c$.

□

Proof of Proposition 3. For $r > \bar{P} - c$, Lemma 2 establishes that no manipulation is profitable. The problem simplifies to that of Section 1 and the equilibria are characterized in Proposition 1 . For $r < \bar{P} - c$, two cases exist:

- ★ For $r \in [0, bq_m^\star(\theta_1) - \frac{bK_e}{4}]$, from Lemma 2 the monopoly regime prevails for $\theta \in (\rho_f K_f, \theta_1 - K_e)$, and the pivotal regime does so for $\theta \in [\theta_1 - K_e, \overline{\theta}]$. Therefore, the transition is exactly shifted by K_e. For $r \in [0, \frac{bK_e}{4}]$, the monopoly regime prevails for lower demand levels than in absence of manipulation opportunities. For $r \in (\frac{bK_e}{4}, \frac{b}{2}(q_1^\star + (1 - \rho_f)K_f) - \frac{bK_e}{4}]$, the competitive regime coincides with that of the no manipulation case, hence only the transition from $q_m^\star$ and $q_m^\dagger$ occurs for a lower demand level.

- ★ For $r \in (q_m^\star(\theta_1) - \frac{bK_e}{4}, \bar{P} - c]$, from Lemma 2 the demand interval over which the competitive regime prevails is unchanged. However, the pivotal regime exists for lower demand levels since $\theta_2 < \theta_1$, although with a larger market position $q_2^\star > q_1^\star$.

□

The results of Lemma 2 and Proposition 4 are summarized in Figure A.1. The gray, white and red area represents the pairs (θ, r) where the competitive regime, monopoly regime and pivotal regime respectively prevail. Reneging is profitable below the thick black line marked by θ^M and not profitable above it. The red dashed line shows the demand level at which the pivotal regime is reached. Finally, the blue dotted line represents the demand level at which the monopolist's capacity constraint K_m is binding.

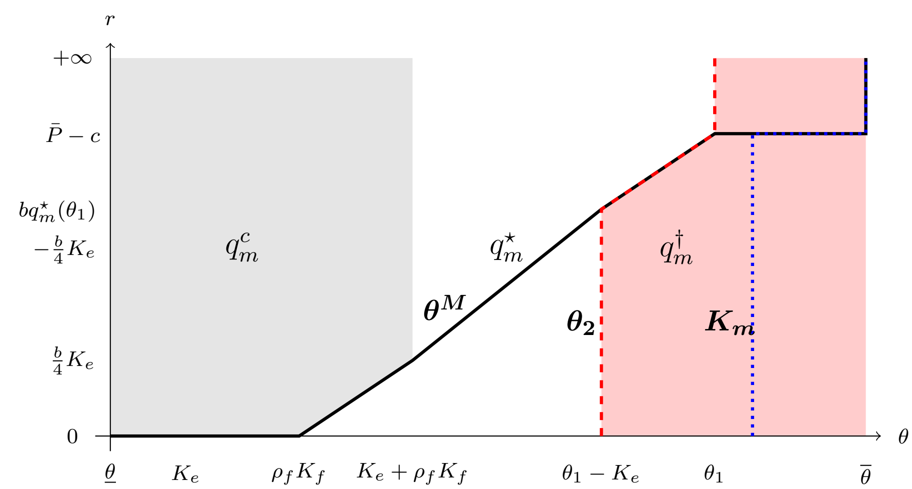


Figure A.1: Illustration of Proposition 3

Proof of Proposition 4. Under the uniformity assumption, the FOC (1.10) becomes

$$\frac{b}{2}\Big(\theta^+ - \theta^-\Big)^2 - \Big(\bar{P} - c - b(1-\rho_f)K_f\Big)q, \tag{A.8}$$

which simplifies to the second-order polynomial

$$q^2 - 2\frac{\bar{P} - c}{b}q + ((1-\rho_f)K_f)^2. \tag{A.9}$$

Differentiation with respect to q yields $-2(\frac{\bar{P}-c}{b} - q) \leq 0$ by assumption 1, proving the SOC is satisfied. The smaller root of (A.9) characterizes the optimal q given by

$$q_3^\star = \frac{\bar{P}-c}{b} - \sqrt{(\frac{\bar{P}-c}{b})^2 - ((1-\rho_f)K_f)^2} \tag{A.10}$$

(i) Let us now show that $q_3^\star \geq q_1^\star$.

$$\begin{aligned}
&\frac{\bar{P}-c}{b} - \sqrt{(\frac{\bar{P}-c}{b})^2 - ((1-\rho_f)K_f)^2} \geq 2\frac{\bar{P}-c}{b} - (1-\rho_f)K_f - 2\sqrt{\frac{\bar{P}-c}{b}(\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)} \\
\leftrightarrow \quad & 2\sqrt{\frac{\bar{P}-c}{b}(\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)} - \sqrt{(\frac{\bar{P}-c}{b})^2 - ((1-\rho_f)K_f)^2} \geq \frac{\bar{P}-c}{b} - (1-\rho_f)K_f \\
\leftrightarrow \quad & 2(\frac{\bar{P}-c}{b})^{1/2} - (\frac{\bar{P}-c}{b} + (1-\rho_f)K_f)^{1/2} \geq (\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)^{1/2} \\
\leftrightarrow \quad & 4(\frac{\bar{P}-c}{b}) \geq ((\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)^{1/2} + (\frac{\bar{P}-c}{b} + (1-\rho_f)K_f)^{1/2})^2 \\
\leftrightarrow \quad & (\frac{\bar{P}-c}{b}) \geq (\frac{\bar{P}-c}{b} - (1-\rho_f)K_f)^{1/2}(\frac{\bar{P}-c}{b} + (1-\rho_f)K_f)^{1/2} \\
\leftrightarrow \quad & (\frac{\bar{P}-c}{b})^2 \geq (\frac{\bar{P}-c}{b})^2 - ((1-\rho_f)K_f)^2 \\
\leftrightarrow \quad & ((1-\rho_f)K_f)^2 \geq 0.
\end{aligned}$$

(ii) We now show that $\theta_1 \in [\theta^-(q_3^\star), \theta^+(q_3^\star)]$. Since $q_3^\star \geq q_1^\star$, we have $\theta^+(q_3^\star) \geq \theta_1$. Furthermore, $\theta^-(q_3^\star) \leq \theta_1 \Leftrightarrow q_1^\star - 2q_3^\star \geq (1-\rho_f)K_f \Leftrightarrow 2(\frac{\bar{P}-c}{d} - (1-\rho_f)K_f)^{1/2}[(\frac{\bar{P}-c}{d} + (1-\rho_f)K_f)^{1/2} - (\frac{\bar{P}-c}{d})^{1/2}] \geq 0$, which is always true.

(iii) We prove the failure of ex-post optimality over the withholding interval. Let us consider the cost of uncertainty – which corresponds to that of economic withholding. It is defined as the difference between the ex-post optimal profit and the expected profit in (1.9) as

$$CU(q^\star) = \int_{\theta^-(q^\star)}^{\theta_1} \Big(\pi_m^\star - \pi_m^w\Big) dF(\theta) + \int_{\theta_1}^{\theta^+(q^\star)} \Big(\pi_m^\dagger - \pi_m^w\Big) dF(\theta). \tag{A.11}$$

By construction, $\pi_m^\star(\theta^-) = \pi_m^w(\theta^-)$ and $\frac{d(\pi_m^\star(\theta) - \pi_m^w(\theta))}{d\theta} > 0 \; \forall \theta \in (\theta^-, \theta_1]$ since $\frac{\bar{P}-c}{d} > (1-\rho_f)K_f$, hence the first term in (A.11) is positive. Also, from

Lemma 1, we have $\pi^\star_m(\theta_1) = \pi^\dagger_m(\theta_1)$ and it is easily shown that $\frac{d(\pi^\dagger_m(\theta)-\pi^w_m(\theta))}{d\theta} > 0$ if $\frac{\bar{P}-c}{b} > (1-\rho_f)K_f$. Hence, the second term of (A.11) is also positive. Therefore $\frac{\bar{P}-c}{b} > (1-\rho_f)K_f \rightarrow \int_{\theta^-(q^\star)}^{\theta_1} \left(\pi^\star_m - \pi^w_m\right) dF(\theta) > 0$ and $\int_{\theta_1}^{\theta^+(q^\star)} \left(\pi^\dagger_m - \pi^w_m\right) dF(\theta) > 0$, i.e. $CU(q^\star) > 0$.

(iv) Finally, if $\theta \sim U[\theta_1, \overline{\theta}]$ Lemma 1 implies that it is optimal to choose the pivotal strategy for every demand realizations. Choosing $q^\star_3 = q^\star_1$ or below is hence best strategy. The reverse holds for $\theta \sim U[\underline{\theta}, \theta_1]$ with $q^\star_3 = q^\star_m(\theta_1)$ or above.

□

Proof of Proposition 5. There are only two cases where manipulations affect the choice of $q^\star$, which ultimately depends on r.

(✓ Intermediate case) Suppose it is not profitable to renege before it can be used to trigger the withholding regime or within that regime to exacerbate market power. That is, we have $\pi^\dagger_m(\theta, 0) \geq \pi^\dagger_m(\theta, K_e)$ and $\pi^\dagger_m(\theta, 0) \geq \pi^w_m(\theta, K_e)$. Those conditions are equivalent to $r \geq dq^\star$ and $r \leq \bar{P} - c$, respectively. The last condition holds for all $\theta \geq \theta^M_I(q) = K_f + \frac{r-bq}{\bar{P}-c-bq}K_e + \frac{b((1-\rho_f)K_f - q)}{\bar{P}-c-bq}q$, where feasibility requires $\theta^M_I(q) \in [\theta^+(q,0), \theta^+(q, K_e)]$ as depicted in Figure 1.5. The optimal $q^\star$ is hence solution of

$$\max_{q^\star} \int_{\rho_f K_f + K_e}^{\theta^-(q^\star, K_e)} \pi^\star_m(\theta, K_e) dF(\theta) + \int_{\theta^-(q^\star, K_e)}^{\theta^M_I(q^\star)} \pi^w_m(\theta, K_e) dF(\theta) + \int_{\theta^M_I(q^\star)}^{K_m + K_f} \pi^\dagger_m(\theta, 0) dF(\theta)$$
$$+ \int_{K_m + K_f}^{\bar{\theta}} \pi^\dagger_m(\theta, \theta - K_m - K_f) dF(\theta) \quad \text{s.t.} \quad \theta^M_I(q^\star) \geq \theta^+(q^\star, 0) \perp \mu. \tag{A.12}$$

The FOC is given by

$$\int_{\theta^-(q^\star, K_e)}^{\theta^M_I(q^\star)} \frac{d\pi^w_m(\theta, K_e)}{dq} dF(\theta) = \mu(1 - \theta^{M'}_I(q^\star)), \tag{A.13}$$

where $\frac{d\pi_w(\theta, K_e)}{dq} = 0$ at $\theta^-(q^\star, K_e)$ and > 0 elsewhere in the interval. Remark that

$$\theta_I^{M'}(q^\star) = \frac{b(r-(\bar{P}-c))K_e}{(\bar{P}-c-bq)^2} + \frac{b((1-\rho_f)K_f - q)(\bar{P}-c-bq) - b(\bar{P}-c-d(1-\rho_f)K_f)q}{(\bar{P}-c-bq)^2} \leq 1, \tag{A.14}$$

since $b(r-\bar{P}-c)K_e - (\bar{P}-c)(\bar{P}-c-b(1-\rho_f)K_f) \leq 0$. Hence the right-hand-side of (A.13) is non-negative.

Suppose $\theta_I^M > \theta^+(q^\star, 0)$ then $\mu = 0$ from the complementary slackness condition. From (A.13) it must be that $\theta_I^M(q^\star) = \theta^-(q^\star, K_e)$, a contradiction with $\theta_I^M(q^\star) > \theta^+(q^\star, 0)$. Therefore we have $\theta^M(q^\star) = \theta^+(q^\star, 0)$ with $\mu > 0$. It is straightforward to obtain the analytical solution $q_4^\star = \frac{rK_e}{\bar{P}-c-b(1-\rho_f)K_f+bK_e}$.[84] This case prevails when $r \in [bq_3^\star, \bar{P}-c]$.

In order to ensure that this is the only case to study in this situation, we assume $q_4^\star \leq (1-\rho_f)K_f - K_e$ so that $\theta^+(q_4^\star, 0) \geq \theta^-(q_4^\star, K_e)$ for all r satisfying the above condition. Since $\frac{dq_4^\star}{dr} > 0$, it is sufficient to assume $q_4^\star|_{r=\bar{P}-c} \leq (1-\rho_f)K_f - K_e \leftrightarrow \bar{P}-c \geq \frac{b((1-\rho_f)K_f-K_e)^2}{(1-\rho_f)K_f-2K_e}$ with $(1-\rho_f)K_f > 2K_e$. From the expressions of $q_4^\star$ and $q_3^\star$, it is easy to show that $q_4^\star \leq q_3^\star$, for all $r \in [bq_3^\star, \frac{\bar{P}-c-b(1-\rho_f)K_f+bK_e}{K_e}q_3^\star]$. Assuming K_e to be small enough, i.e. $K_e \leq \frac{(\frac{\bar{P}-c}{b}-(1-\rho_f)K_f)(\frac{\bar{P}-c}{b}-\sqrt{(\frac{\bar{P}-c}{b})^2-(1-\rho_f)^2K_f^2})}{\sqrt{(\frac{\bar{P}-c}{b})^2-(1-\rho_f)^2K_f^2}}$, ensures that this inequality always holds. This condition is satisfied for $\bar{P}$ large enough. Finally, we must verify the second-order condition to ensure that this candidate is a global maximum. The derivative of the FOC gives

$$\theta_I^{M'}(q) b\left(\theta_I^M(q) - K_e - \rho_f K_f - 2q\right) f(\theta_I^M(q)) - 2b \int_{\theta^-(q,K_e)}^{\theta_I^M(q)} dF(\theta),$$

which under the uniformity assumption simplifies to

$$b\left(\theta_I^{M'}(q) - 2\right)\left(\theta_I^M(q) - \theta^-(q, K_e)\right).$$

Hence the SOC is satisfied if $\theta_I^{M'}(q) \leq 2$, which holds since $\theta_I^{M'}(q) \leq 1$ as proved above.

[84]Remark that the feasibility constraint is always binding since in its absence the solution would be characterized by $\theta^-(q^\star, K_e) = \theta_I^M(q^\star)$ which rewrites as the polynomial equation $q^2 - \frac{2(\bar{P}-c)}{d}q + \frac{\bar{P}-c}{d}(1-\rho_f)K_f - \frac{\bar{P}-c-r}{d}K_e = 0$ admitting the closed-form solution as lower root $\frac{\bar{P}-c}{d} - \sqrt{(\frac{\bar{P}-c}{d})^2 - \frac{(\bar{P}-c)(1-\rho_f)K_f-(\bar{P}-c-r)K_e}{d}} \in [q_3^\star, q_m(\theta_1)]$.

(✓ Partial case) Suppose now it is not profitable to renege before it can be used to trigger the withholding regime, but it not profitable to use it within the pivotal regime either. Formally, $\pi_m^w(\theta, K_e) \geq \pi_m^w(\theta, 0)$ and $\pi_m^\dagger(\theta, 0) \leq \pi_m^\dagger(\theta, K_e)$ but now $\pi_m^\dagger(\theta, q_e^\star) \geq \pi_m^w(\theta, K_e)$. The first and second conditions are equivalent to $r \geq dq^\star$ and $r \geq \bar{P} - c$, respectively. The last condition states that reneging is used only to trigger the pivotal regime, but not beyond, i.e. $q_e^\star = \theta - K_f - q$. This condition holds $\forall \theta \geq \theta_P^M(q) = K_f + K_e + q\left(1 - \frac{\bar{P}-c-d(1-\rho_f)K_f}{r-dq}\right) \in [\theta^+(0, K_e), \theta^+(q, K_e))$. The lower bound holds for $r = \bar{P} - c$ and the upper one is attained at the limit $r \to \infty$. By construction we have $\theta_P^M|_{r \geq \bar{P}-c} \geq \theta_I^M|_{r=\bar{P}-c}$. The optimal $q^\star$ is solution of

$$\max_{q^\star} \int_{\rho_f K_f + K_e}^{\theta^-(q^\star, K_e)} \pi_m^\star(\theta, K_e) dF(\theta) + \int_{\theta^-(q^\star, K_e)}^{\theta_P^M(q^\star)} \pi_w(\theta, K_e) dF(\theta) + \int_{\theta_P^M(q^\star)}^{\theta^+(q, K_e)} \pi_m^\dagger(\theta, \theta - K_f - q) dF(\theta)$$
$$+ \int_{\theta^+(q, K_e)}^{\bar{\theta}} \pi_m^\dagger(\theta, K_e) dF(\theta) \quad \text{s.t.} \quad \theta_P^M(q^\star) \geq \theta^+(q^\star, 0) \perp \mu. \tag{A.15}$$

The FOC writes

$$\int_{\theta^-(q^\star, K_e)}^{\theta_P^M(q^\star)} \frac{d\pi_w(\theta, K_e)}{dq} dF(\theta) = (r - (\bar{P} - c)) \int_{\theta_P^M(q^\star)}^{\theta^+(q^\star, K_e)} dF(\theta) + \mu(1 - \theta_P^{M'}(q^\star)). \tag{A.16}$$

When $r = \bar{P} - c$, the problem coincides to the previous one and so does the solution. On the other hand, $\theta_P^M \to \theta^+(q, K_e)$ as $r \to \infty$, i.e. reneging is never profitable only if r is infinite: the maximization problem converges to that of Proposition 4 and the solution converges to $q_3^\star$.

Under uniformity the SOC is satisfied only if

$$\theta_P^{M'}(q)\left(b(\theta_P^M(q) - K_e - \rho_f K_f - 2q) + (r - (\bar{P} - c))\right) - 2b \int_{\theta^-(q, K_e)}^{\theta_P^M(q)} dF(\theta) - (r - (\bar{P} - c)) \leq 0$$
$$b\left(\theta_P^{M'}(q) - 2\right)\left(\theta_P^{M'}(q) - \theta^-(q, K_e)\right) + \left(\theta_P^{M'}(q) - 1\right)(r - (\bar{P} - c)) \leq 0,$$

where both terms are negative since $\theta_P^{M'}(q) = 1 - \frac{r(\bar{P}-c-d(1-\rho_f)K_f)}{(r-dq)^2} \leq 1$.

(✓ Early case) For $r \leq dq$. It is profitable to renege on commitments not only to trigger the pivotal regime but also within the withholding regime and possibly before. It is easy to show that $\pi_m^w(\theta, 0) \geq \pi_m^w(\theta, K_e) \leftrightarrow r \leq dq$, thereby $\theta^M(q) \leq \theta^-(q, K_e)$. For simplicity, we consider the case where manipulations become profitable in the monopoly regime, that is $\theta^M(q)$ is such that $\pi_m^\star(\theta^M(q), K_e) = \pi_m^\star(\theta^M(q), 0)$. The profit-maximization problem takes the form

$$\max_{q^\star} \int_{\theta_c}^{\theta^M(q)} \pi_m^\star(\theta, K_e) dF(\theta) + \int_{\theta^M(q)}^{\theta^-(q,0)} \pi_m^\star(\theta, 0) dF(\theta) + \int_{\theta^-(q,0)}^{\theta^+(q,0)} \pi_m^w(\theta, 0) dF(\theta)$$
$$+ \int_{\theta^+(q,0)}^{K_m+K_e} \pi_m^\dagger(\theta, 0) dF(\theta) + \int_{K_m+K_e}^{\bar{\theta}} \pi_m^\dagger(\theta, \theta - K_m - K_f) dF(\theta), \tag{A.17}$$

which yields the FOC

$$\theta^{M\prime}(q) \left(\pi_m^\star(\theta^M(q), K_e) - \pi_m^\star(\theta^M(q), 0)\right) + 2 \left(\pi_m^\star(\theta^-(q, 0), 0) - \pi_m^w(\theta^-(q, 0), 0)\right)$$
$$+ \left(\pi_m^w(\theta^+(q, 0), 0) - \pi_m^\dagger(\theta^+(q, 0), 0)\right) + \int_{\theta^-(q,0)}^{\theta^+(q,0)} \frac{d\pi_m^\dagger(\theta, 0)}{dq} dF(\theta) = 0, \tag{A.18}$$

where the first two terms are zero by definition of θ^M and because both the monopoly and withholding profit functions coincide when the withholding regime begins at $\theta^-(q, 0)$. Under the uniformity assumption, this FOC coincides with that of no manipulations in (1.10).

□

B Chapter 2

Proof of Proposition 1.

$$\begin{aligned}\left\|\hat{\Pi}_{\alpha}\right\|_{HS}^{2} &= \left\|\hat{C}_{YZ}\left(\alpha I+\hat{V}_{Z}\right)^{-1}\right\|_{HS}^{2}\\ &\leq \left\|\hat{C}_{YZ}\right\|_{HS}^{2}\left\|\left(\alpha I+\hat{V}_{Z}\right)^{-1}\right\|_{op}\end{aligned}$$

using the fact that, if A is a HS operator and B is a bounded operator, $\|AB\|_{HS} \leq \|A\|_{HS}\|B\|_{op}$ where $\|B\|_{op} \equiv \sup_{\|\phi\|\leq 1}\|B\phi\|$ is the operator norm. Then, we have

$$\left\|\hat{\Pi}_{\alpha}\right\|_{HS}^{2} \leq \frac{1}{\alpha}\left\|\hat{C}_{YZ}\right\|_{HS}^{2}.$$

It remains to show that $\hat{C}_{YZ}$ is a HS operator. $\hat{C}_{YZ}$ is an integral operator with degenerate kernel $\frac{1}{n}\sum_{i=1}^{n} y_i(s)\, z_i(t)$. A sufficient condition for $\hat{C}_{YZ}$ to be HS is that its kernel is square integrable which is true because Y_i and Z_i are elements of Hilbert spaces. The result of Proposition 1 follows.

□

Proof of Proposition 2. To prove Proposition 2, we need three preliminary lemmas.

Lemma 5. *Let $A = B + C$ where B is a zero mean random operator and C is a nonrandom operator. Then,*

$$E\left(\|A\|_{HS}^{2}\right) = E\left(\|B\|_{HS}^{2}\right) + \|C\|_{HS}^{2}.$$

Proof of Lemma 5.

$$\begin{aligned} E\left(\|A\|_{HS}^2\right) &= E\left(\sum_j \langle A\phi_j, A\phi_j\rangle\right) \\ &= E\left(\sum_j \langle A^*A\phi_j, \phi_j\rangle\right) \\ &= E\left(\sum_j \langle (B+C)^*(B+C)\phi_j, \phi_j\rangle\right) \\ &= E\left(\sum_j \langle B^*B\phi_j, \phi_j\rangle\right) \\ &\quad +E\left(\sum_j \langle C^*B\phi_j, \phi_j\rangle\right) \\ &\quad +E\left(\sum_j \langle B^*C\phi_j, \phi_j\rangle\right) \\ &\quad +E\left(\sum_j \langle C^*C\phi_j, \phi_j\rangle\right). \end{aligned}$$

The second and third terms on the r.h.s are equal to zero because $E(B) = 0$ and C is deterministic. We obtain $E\left(\|A\|_{HS}^2\right) = E\left(\|B\|_{HS}^2\right) + \|C\|_{HS}^2$. □

Lemma 6. *Let A be a random Hilbert-Schmidt operator from $\mathcal{E}$ to $\mathcal{F}$.*

$$E\left(\|A\|_{HS}^2\right) = trE\left(A^*A\right) = trE\left(AA^*\right).$$

Proof of Lemma 6. We have

$$\begin{aligned} E\left(\|A\|_{HS}^2\right) &= E\left(tr(A^*A)\right) \\ &= E\left(tr(AA^*)\right) \\ &= E\sum_j \langle AA^*\phi_j, \phi_j\rangle \\ &= trE\left(AA^*\right) \end{aligned}$$

where the second equality follows from the fact that $tr(AB) = tr(BA)$ when A and B are Hilbert-Schmidt operators (see Pedersen, 1989). □

Lemma 7. $\left\|\alpha\left(\alpha I+V_Z\right)^{-1} V_Z^{\beta/2} R\right\|_{HS}^2=O\left(\alpha^{\beta\wedge 2}\right).$

Proof of Lemma 7. Let $\{\lambda_j,\varphi_j\}$ be the eigenvalues and orthonormal eigenfunctions of V_Z.

$$\begin{aligned}
&\left\|\alpha\left(\alpha I+V_Z\right)^{-1} V_Z^{\beta/2} R\right\|_{HS}^2 \\
= \;& \alpha^2 \sum_j \left\langle \left(\alpha I+V_Z\right)^{-1} V_Z^{\beta/2} R\varphi_j, \left(\alpha I+V_Z\right)^{-1} V_Z^{\beta/2} R\varphi_j \right\rangle \\
= \;& \alpha^2 \sum_j \frac{\lambda_j^\beta}{\left(\lambda_j+\alpha\right)^2} \left\langle R\varphi_j, R\varphi_j\right\rangle^2 \\
\leq \;& \alpha^2 \sup_\lambda \frac{\lambda^\beta}{\left(\lambda+\alpha\right)^2} \sum_j \left\langle R\varphi_j, R\varphi_j\right\rangle^2 \\
= \;& O\left(\alpha^{\beta\wedge 2}\right).
\end{aligned}$$

The last equality follows from the fact that $\sum_j \left\langle R\varphi_j, R\varphi_j\right\rangle^2=\|R\|_{HS}^2<\infty$ and, using the notation $\lambda=\mu^2$, we have

$$\sup_\lambda \frac{\alpha^2\lambda^\beta}{\left(\lambda+\alpha\right)^2}=\sup_\mu \frac{\alpha^2\mu^{2\beta}}{\left(\mu^2+\alpha\right)^2}=O\left(\alpha^{\beta\wedge 2}\right)$$

by Carrasco, Florens, and Renault (2007, Proposition 3.11). Consequently,

$$\left\|\alpha\left(\alpha I+V_Z\right)^{-1} V_Z^{\beta/2} R\right\|_{HS}^2=O\left(\alpha^{\beta\wedge 2}\right).$$

This ends the proof of Lemma 7. □

We turn to the proof of Proposition 2.

Replacing y_i by Πz_i+u_i in the expression of $\hat{C}_{ZY}$, we obtain

$$\begin{aligned}
\hat{C}_{ZY} &= \frac{1}{n}\sum_i z_i\left\langle y_i, .\right\rangle \\
&= \frac{1}{n}\sum_i z_i\left\langle u_i, .\right\rangle+\frac{1}{n}\sum_i z_i\left\langle \Pi z_i, .\right\rangle \\
&= \hat{C}_{ZU}+\hat{V}_Z\Pi^*.
\end{aligned}$$

We decompose $\hat{\Pi}^*_\alpha - \Pi^*$ in the following manner:

$$\begin{aligned}
\hat{\Pi}^*_\alpha - \Pi^* &= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZY} - \Pi^* \\
&= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZU} && \text{(B.1)} \\
&\quad + \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z \Pi^* - \left(\alpha I + V_Z\right)^{-1} V_Z \Pi^* && \text{(B.2)} \\
&\quad + \left(\alpha I + V_Z\right)^{-1} V_Z \Pi^* - \Pi^*. && \text{(B.3)}
\end{aligned}$$

To study the rate of convergence of the MSE, we will study the rates of the three terms (B.1), (B.2), and (B.3).

Applying Lemma 5 on the decomposition (B.1), (B.2), and (B.3), we have

$$\begin{aligned}
& E\left(\left\|\hat{\Pi}_\alpha - \Pi\right\|^2_{HS} |Z_1, Z_2, ..., Z_n\right) \\
= \; & E\left(\|(B.1)\|^2_{HS} |Z_1, Z_2, ..., Z_n\right) + \|(B.2) + (B.3)\|^2_{HS} \\
\leq \; & E\left(\|(B.1)\|^2_{HS} |Z_1, Z_2, ..., Z_n\right) + 2\,\|(B.2)\|^2_{HS} + 2\,\|(B.3)\|^2_{HS}.
\end{aligned}$$

We study the first term of the r.h.s.. By Lemma 6,

$$\begin{aligned}
& E\left(\|(B.1)\|^2_{HS} |Z_1, Z_2, ..., Z_n\right) \\
= \; & E\left(\left\|\left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZU}\right\|^2_{HS} |Z_1, Z_2, ..., Z_n\right) \\
= \; & trE\left(\left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZU}\hat{C}^*_{ZU} \left(\alpha I + \hat{V}_Z\right)^{-1} |Z_1, Z_2, ..., Z_n\right) \\
= \; & tr\left\{\left(\alpha I + \hat{V}_Z\right)^{-1} E\left(\hat{C}_{ZU}\hat{C}^*_{ZU}|Z_1, Z_2, ..., Z_n\right)\left(\alpha I + \hat{V}_Z\right)^{-1}\right\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\hat{C}_{ZU}\hat{C}^*_{ZU}\varphi &= \frac{1}{n^2}\sum_{i,j} z_i \langle z_j, \varphi\rangle \langle u_i, u_j\rangle, \\
E\left(\hat{C}_{ZU}\hat{C}^*_{ZU}\varphi|Z_1, Z_2, ..., Z_n\right) &= \frac{1}{n}\sum_i z_i \langle z_i, \varphi\rangle E\left[\langle u_i, u_i\rangle |Z_1, Z_2, ..., Z_n\right] \\
&= \frac{1}{n}\sum_i z_i \langle z_i, \varphi\rangle tr\left(V_U\right) \\
&= \frac{1}{n} tr\left(V_U\right) \hat{V}_Z \varphi
\end{aligned}$$

because the u_i are uncorrelated. To see that $E[\langle u,u\rangle] = trV_U$, decompose u on the basis formed by the eigenfunctions ψ_j of V_U so that $u = \sum_j \langle u, \psi_j\rangle \psi_j$. It follows that $\langle u,u\rangle = \sum_j \langle u,\psi_j\rangle^2$ and $E\langle u,u\rangle = \sum_j \langle V_U\psi_j,\psi_j\rangle = tr(V_U)$. Hence,

$$\begin{aligned} E\left(\|(B.1)\|_{HS}^2 |Z_1, Z_2, ..., Z_n\right) &= \frac{1}{n} tr(V_U)\, tr\left(\left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z \left(\alpha I + \hat{V}_Z\right)^{-1}\right) \\ &\leq \frac{C}{n\alpha} \end{aligned}$$

where C is a generic constant. It follows that $E\left(\|(B.1)\|_{HS}^2\right) \leq \frac{C}{n\alpha}$.

Now, we turn toward the term (B.2). We have

$$\begin{aligned} &\left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z \Pi^* - (\alpha I + V_Z)^{-1} V_Z \Pi^* \\ = &\left[-\left(I - \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z\right) + \left(I - (\alpha I + V_Z)^{-1} V_Z\right)\right] \Pi^*. \end{aligned}$$

Using $I = \left(\alpha I + \hat{V}_Z\right)^{-1}\left(\alpha I + \hat{V}_Z\right)$, we obtain

$$I - \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z = \alpha \left(\alpha I + \hat{V}_Z\right)^{-1}.$$

Hence,

$$\begin{aligned} (B.2) &= \left[-\alpha\left(\alpha I + \hat{V}_Z\right)^{-1} + \alpha (\alpha I + V_Z)^{-1}\right] \Pi^* \\ &= -\left(\alpha I + \hat{V}_Z\right)^{-1} \left(V_Z - \hat{V}_Z\right) \alpha (\alpha I + V_Z)^{-1} \Pi^* \end{aligned}$$

where the last equality follows from $A^{-1} - B^{-1} = A^{-1}(B - A) B^{-1}$.

Now, we have

$$\begin{aligned} &\left\|\left(\alpha I + \hat{V}_Z\right)^{-1}\left(V_Z - \hat{V}_Z\right) \alpha (\alpha I + V_Z)^{-1} \Pi^*\right\|_{HS}^2 \\ \leq &\left\|\left(\alpha I + \hat{V}_Z\right)^{-1}\right\|_{op}^2 \left\|\left(V_Z - \hat{V}_Z\right)\right\|_{op}^2 \left\|\alpha (\alpha I + V_Z)^{-1} \Pi^*\right\|_{HS}^2 \end{aligned}$$

where $\left\|\left(\alpha I + \hat{V}_Z\right)^{-1}\right\|_{op}^2 \leq 1/\alpha^2$, $\left\|\left(V_Z - \hat{V}_Z\right)\right\|_{op}^2 = O_p(1/n)$ by Assumption 2 and $\left\|\alpha (\alpha I + V_Z)^{-1} \Pi^*\right\|_{HS}^2 = O\left(\alpha^{\beta\wedge 2}\right)$ by Lemma 7.

If $\beta > 1$ then the term corresponding to (B.2) is negligible with respect to (B.1). If $\beta < 1$, then (B.1) is negligible with respect to (B.2).

Now, we turn our attention toward the term (B.3). We have

$$\begin{aligned} & (\alpha I + V_Z)^{-1} V_Z \Pi^* - \Pi^* \\ = \ & (\alpha I + V_Z)^{-1} (V_Z - \alpha I - V_Z) \Pi^* \\ = \ & \alpha (\alpha I + V_Z)^{-1} \Pi^* \\ = \ & \alpha (\alpha I + V_Z)^{-1} V_Z^{\beta/2} R \end{aligned}$$

by Assumption 4. The rate of this term follows from Lemma 7.

This concludes the proof of Proposition 2. □

Proof of Proposition 3. We decompose $\hat{\Pi}_\alpha^* - \Pi_{N^\perp}^*$ in the following manner:

$$\begin{aligned} \hat{\Pi}_\alpha^* - \Pi_{N^\perp}^* &= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZY} - \Pi_{N^\perp}^* \\ &= \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{C}_{ZU} && \text{(B.4)} \\ &\quad + \left(\alpha I + \hat{V}_Z\right)^{-1} \hat{V}_Z \Pi^* - (\alpha I + V_Z)^{-1} V_Z \Pi^* && \text{(B.5)} \\ &\quad + (\alpha I + V_Z)^{-1} V_Z \Pi_{N^\perp}^* - \Pi_{N^\perp}^*. && \text{(B.6)} \end{aligned}$$

where (B.6) comes from the fact that $V_Z \Pi^* = V_Z \Pi_{N^\perp}^*$. Following the proof of Proposition 2, we can establish that the rates of (B.4) and (B.6) are the same as those of (B.1) and (B.3). The rate of (B.5) will be however different from that of (B.2). The reason is that whereas $V_Z \Pi^* = V_Z \Pi_{N^\perp}^*$, $\hat{V}_Z \Pi^* \neq \hat{V}_Z \Pi_{N^\perp}^*$. Using the steps of the proof of Proposition 2, we have

$$\begin{aligned} & \|(B.5)\|_{HS}^2 \\ = \ & \left\| \left(\alpha I + \hat{V}_Z\right)^{-1} \left(V_Z - \hat{V}_Z\right) \alpha (\alpha I + V_Z)^{-1} \Pi^* \right\|_{HS}^2 \\ \leq \ & \left\| \left(\alpha I + \hat{V}_Z\right)^{-1} \right\|_{op}^2 \left\| \left(V_Z - \hat{V}_Z\right) \right\|_{op}^2 \left\| \alpha (\alpha I + V_Z)^{-1} \Pi^* \right\|_{HS}^2 \end{aligned}$$

where $\left\| \left(\alpha I + \hat{V}_Z\right)^{-1} \right\|_{op}^2 \leq 1/\alpha^2$, $\left\| \left(V_Z - \hat{V}_Z\right) \right\|_{op}^2 = O_p(1/n)$ by Assumption 2 and $\left\| \alpha (\alpha I + V_Z)^{-1} \Pi^* \right\|_{HS}^2 = O(1)$. We do not get the rate $\left\| \alpha (\alpha I + V_Z)^{-1} \Pi^* \right\|_{HS}^2 = O\left(\alpha^{\beta \wedge 2}\right)$ because Lemma 7 does not apply here (it applies with Π replaced by $\Pi_{N^\perp}$). Hence the rate of $\|(B.5)\|_{HS}^2$ is $\mathrm{O}_p\left(1/\left(n\alpha^2\right)\right)$. □

Proof of Proposition 4. Under the assumptions (U_i, Z_i) i.i.d., $\mathrm{E}\|Z_i\|^4 < \infty$, and $\mathrm{E}\|U_i\|^2\|Z_i\|^2 < \infty$, Theorem 1 ensures the root-n asymptotic normality of $\frac{1}{n}\sum_i \begin{pmatrix} U_i \otimes Z_i \\ Z_i \otimes Z_i \end{pmatrix}$. By the continuous mapping theorem, one has (2.23) using the continuous transformation $\begin{pmatrix} A \\ B \end{pmatrix} \mapsto (\alpha I + V_Z)^{-1}A + \alpha(\alpha I + V_Z)^{-1}B(\alpha I + V_Z)^{-1}\Pi^\star$. The covariance operator of $\sqrt{n}(\hat{\Pi}^\star_\alpha - \Pi^\star_\alpha)$ may be written using (2.22) as

$$\begin{aligned}\Omega_\alpha =& \mathrm{E}\left[\left(\frac{1}{\sqrt{n}}\sum_i\Big((u_i + \alpha\Pi\tilde{z}_i)\otimes\tilde{z}_i - \alpha C_{\tilde{Z}\Pi\tilde{Z}}\Big)\right)\tilde{\otimes}\left(\frac{1}{\sqrt{n}}\sum_i\Big((u_i + \alpha\Pi\tilde{z}_i)\otimes\tilde{z}_i - \alpha C_{\tilde{Z}\Pi\tilde{Z}}\Big)\right)\right] \\ =& \mathrm{E}\left[\left((U + \alpha\Pi\tilde{Z})\otimes\tilde{Z} - \alpha C_{\tilde{Z}\Pi\tilde{Z}}\right)\tilde{\otimes}\left((U + \alpha\Pi\tilde{Z})\otimes\tilde{Z} - \alpha C_{\tilde{Z}\Pi\tilde{Z}}\right)\right],\end{aligned}$$

where the second line is obtained from the i.i.d. assumption. Straightforward developments yield (2.24). Now letting $\alpha \to 0$ gives (2.25). □

Proof of Proposition 5. We have

$$\hat{\Pi}^{m*}_\alpha - \Pi^* = \hat{\Pi}^{m*}_\alpha - \hat{\Pi}^*_\alpha + \hat{\Pi}^*_\alpha - \Pi^*.$$

We focus on the term $\hat{\Pi}^{m*}_\alpha - \hat{\Pi}^*_\alpha$.

$$\begin{aligned}\hat{\Pi}^{m*}_\alpha - \hat{\Pi}^*_\alpha &= \left(\alpha I + \hat{V}^m_Z\right)^{-1}\hat{C}^m_{ZY} - \left(\alpha I + \hat{V}_Z\right)^{-1}\hat{C}_{ZY} \\ &= \left(\alpha I + \hat{V}^m_Z\right)^{-1}\left(\hat{C}^m_{ZY} - \hat{C}_{ZY}\right) + \left[\left(\alpha I + \hat{V}^m_Z\right)^{-1} - \left(\alpha I + \hat{V}_Z\right)^{-1}\right]\hat{C}_{ZY}.\end{aligned}$$

$$\begin{aligned}\left\|\hat{C}^m_{ZY} - \hat{C}_{ZY}\right\|^2_{HS} &= \left\|\frac{1}{n}\sum_{i=1}^n z_i^m\langle y_i^m, .\rangle - \frac{1}{n}\sum_{i=1}^n z_i\langle y_i, .\rangle\right\|^2_{HS} \\ &= \left\|\frac{1}{n}\sum_{i=1}^n\{(z_i^m - z_i)\langle y_i, .\rangle + z_i^m\langle y_i^m - y_i, .\rangle\}\right\|^2_{HS} \\ &\le \frac{2}{n^2}\sum_{i=1}^n\left\{\|(z_i^m - z_i)\langle y_i, .\rangle\|^2_{HS} + \|z_i^m\langle y_i^m - y_i, .\rangle\|^2_{HS}\right\}.\end{aligned}$$

$$\begin{aligned}
\left\|(z_i^m - z_i)\langle y_i, .\rangle\right\|_{HS}^2 &= \sum_j \left\langle (z_i^m - z_i)\langle y_i, \phi_j\rangle, (z_i^m - z_i)\langle y_i, \phi_j\rangle\right\rangle \\
&= \left\|z_i^m - z_i\right\|^2 \sum_j \langle y_i, \phi_j\rangle^2 \\
&= O_p\left(f(m)^2\right).
\end{aligned}$$

$$\begin{aligned}
\left\|z_i^m \langle y_i^m - y_i, .\rangle\right\|_{HS}^2 &= \sum_j \left\langle z_i^m \langle y_i^m - y_i, \phi_j\rangle, z_i^m \langle y_i^m - y_i, \phi_j\rangle\right\rangle \\
&= \left\|z_i^m\right\|^2 \sum_j \langle y_i^m - y_i, \phi_j\rangle^2 \\
&= \left\|z_i^m\right\|^2 \left\|y_i^m - y_i\right\|^2 \\
&= O_p\left(f(m)^2\right).
\end{aligned}$$

Hence,

$$\left\|\left(\alpha I + \hat{V}_Z^m\right)^{-1}\left(\hat{C}_{ZY}^m - \hat{C}_{ZY}\right)\right\|_{HS}^2 = O_p\left(\frac{f(m)^2}{\alpha^2 n^2}\right).$$

$$\begin{aligned}
&\left[\left(\alpha I + \hat{V}_Z^m\right)^{-1} - \left(\alpha I + \hat{V}_Z\right)^{-1}\right]\hat{C}_{ZY} \\
&= \left(\alpha I + \hat{V}_Z^m\right)^{-1}\left(\hat{V}_Z - \hat{V}_Z^m\right)\left(\alpha I + \hat{V}_Z\right)^{-1}\hat{C}_{ZY} \\
&= \left(\alpha I + \hat{V}_Z^m\right)^{-1}\left(\hat{V}_Z - \hat{V}_Z^m\right)\hat{\Pi}_\alpha^*.
\end{aligned}$$

Hence,

$$\left\|\left(\alpha I + \hat{V}_Z^m\right)^{-1}\left(\hat{V}_Z - \hat{V}_Z^m\right)\hat{\Pi}_\alpha^*\right\|_{HS}^2 = O_p\left(\frac{f(m)^2}{\alpha^2 n^2}\right).$$

This concludes the proof of Proposition 5. □

Proof of Proposition 6. Using the fact that $\|a+b+c\|_{HS}^2 \leq 3\left(\|a\|_{HS}^2 + \|b\|_{HS}^2 + \|c\|_{HS}^2\right)$, we can evaluate the terms (2.35), (2.36), and (2.37) separately. The proof follows closely that of Proposition 2. Let $\mathbf{Z}$ and $\mathbf{W}$ be the sets $(Z_1, Z_2, ..., Z_n)$ and $(W_1, W_2, ..., W_n)$.

$$\begin{aligned}
&E\left(\|(2.35)\|_{HS}^2 \,|\mathbf{Z}, \mathbf{W}\right) \\
&= tr\left\{\left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\hat{C}_{ZW} E\left(\hat{C}_{WU}\hat{C}_{UW}|\mathbf{Z}, \mathbf{W}\right)\hat{C}_{WZ}\left(\alpha I + \hat{C}_{ZW}\hat{C}_{WZ}\right)^{-1}\right\}.
\end{aligned}$$

Using

$$E\left(\hat{C}_{WU}\hat{C}_{UW}|\mathbf{Z},\mathbf{W}\right)=\frac{1}{n}tr\left(V_U\right)\hat{V}_W,$$

we obtain

$$E\left(\|(2.35)\|_{HS}^2\right)\leq\frac{C}{n\alpha}$$

for some constant C.

The proof regarding the rates of convergence of (2.36) and (2.37) is similar to that of Proposition 2 and is not repeated here. □

C Chapter 3

C.1 Proofs

Proof of Lemma 3 (Asymptotic normality of expansion coefficients). Let us rewrite and rearrange (3.22) using the introduced notations as

$$
\begin{aligned}
\hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}} - \boldsymbol{b}_{\boldsymbol{K},\boldsymbol{\lambda}} =& \left[\int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p) \sum_{t=1}^{T} z_t(p)^2 \boldsymbol{\Phi}_{\boldsymbol{K}}(p)^\top dp + \lambda \boldsymbol{R}_{\boldsymbol{K}} \right]^{-1} \int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p) \sum_{t=1}^{T} z_t(p) y_t(p) dp \\
& - \left[\int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p) E[Z(p)^2] \boldsymbol{\Phi}_{\boldsymbol{K}}(p)^\top dp + \lambda \boldsymbol{R}_{\boldsymbol{K}} \right]^{-1} \left[\int_{\mathcal{P}} \boldsymbol{\Phi}_{\boldsymbol{K}}(p) E[Z(p)^2] \boldsymbol{\Phi}_{\boldsymbol{K}}(p)^\top dp \right] \boldsymbol{b}_{\boldsymbol{K}} \\
=& (\hat{V}_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \langle \boldsymbol{\Phi}_{\boldsymbol{K}}, T^{-1} \sum_{t=1}^{T} z_t \cdot y_t \rangle + (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \langle \boldsymbol{\Phi}_{\boldsymbol{K}}, E[Z \cdot Z] \cdot \boldsymbol{\Phi}_{\boldsymbol{K}}^\top \rangle \boldsymbol{b}_{\boldsymbol{K}} \\
=& (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \langle \boldsymbol{\Phi}_{\boldsymbol{K}}, T^{-1} \sum_{t=1}^{T} z_t \cdot u_t \rangle \\
& + \lambda (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \langle \boldsymbol{\Phi}_{\boldsymbol{K}}, \Big(T^{-1} \sum_{t=1}^{T} z_t \cdot z_t - E[Z \cdot Z] \Big) \cdot \boldsymbol{\Phi}_{\boldsymbol{K}}^\top \rangle (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \boldsymbol{R}_{\boldsymbol{K}} \boldsymbol{b}_{\boldsymbol{K}} \\
& + \Big[(\hat{V}_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} - (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \Big] \langle \boldsymbol{\Phi}_{\boldsymbol{K}}, T^{-1} \sum_{t=1}^{T} z_t \cdot u_t \rangle \\
& + \lambda \Big[(\hat{V}_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} - (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \Big] \langle \boldsymbol{\Phi}_{\boldsymbol{K}}, \Big(T^{-1} \sum_{t=1}^{T} z_t \cdot z_t - E[Z \cdot Z] \Big) \cdot \boldsymbol{\Phi}_{\boldsymbol{K}}^\top \rangle \\
& \times (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \boldsymbol{R}_{\boldsymbol{K}} \boldsymbol{b}_{\boldsymbol{K}}
\end{aligned}
\tag{C.1}
$$

where the first two terms are $O_p(1/\sqrt{T})$ and the last two terms are $O_p(1/T)$. Neglecting the $O_p(1/T)$ terms yields

$$
\begin{aligned}
\hat{\boldsymbol{b}}_{\boldsymbol{K},\boldsymbol{\lambda}} - \boldsymbol{b}_{\boldsymbol{K},\boldsymbol{\lambda}} =& T^{-1} \sum_{t=1}^{T} \langle (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \boldsymbol{\Phi}_{\boldsymbol{K}}, z_t \cdot u_t \\
& + \lambda \Big(z_t \cdot z_t - E[Z \cdot Z] \Big) \cdot \boldsymbol{\Phi}_{\boldsymbol{K}}^\top (V_Z + \lambda \boldsymbol{R}_{\boldsymbol{K}})^{-1} \boldsymbol{R}_{\boldsymbol{K}} \boldsymbol{b}_{\boldsymbol{K}} \rangle.
\end{aligned}
\tag{C.2}
$$

The asymptotic normality is shown in Proposition 2. □

Proof of Proposition 2 (Asymptotic normality). Under the assumptions (U, Z) *i.i.d.*, $\mathcal{B}_{K,\lambda} < \infty$, $E||Z||^4 < \infty$, $E||U||^2||Z||^2 < \infty$, Theorem 2.7 in Bosq (2000) ensures the root-n asymptotic normality of $T^{-1} \sum_t \begin{pmatrix} u_t \cdot z_t \\ z_t \cdot z_t \end{pmatrix}$. The continuous mapping theorem yields the asymptotic normality result using the continuous transformation $\begin{pmatrix} A \\ B \end{pmatrix} \mapsto (V_Z + \lambda \boldsymbol{R_K})^{-1} \langle \boldsymbol{\Phi_K}, A \rangle + \lambda (V_Z + \lambda \boldsymbol{R_K})^{-1} \langle \boldsymbol{\Phi_K}, (B - E[Z \cdot Z]) \boldsymbol{\Phi_K^\top} \rangle (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{R_K} \boldsymbol{b_K}$. Making use of (C.2) gives the asymptotic covariance matrix of expansion coefficients $B_{K,\lambda}$ defined as

$$\begin{aligned}
B_{K,\lambda} =& E\Big[T^{-1} \sum_{t=1}^{T} \sum_{t'=1}^{T} \langle (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K}, z_t \cdot u_t + \lambda \Big(z_t \cdot z_t - E[Z \cdot Z] \Big) \cdot \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{R_K} \boldsymbol{b_K} \rangle \\
& \times \langle \lambda \boldsymbol{b_K^\top} \boldsymbol{R_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K} \cdot \lambda \Big(z_{t'} \cdot z_{t'} - E[Z \cdot Z] \Big) + u_{t'} \cdot z_{t'}, \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \rangle \Big] \\
=& E\Big[T^{-1} \sum_{t=1}^{T} \langle (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K}, z_t \cdot u_t + \lambda \Big(z_t \cdot z_t - E[Z \cdot Z] \Big) \cdot \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{R_K} \boldsymbol{b_K} \rangle \\
& \times \langle \lambda \boldsymbol{b_K^\top} \boldsymbol{R_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K} \cdot \lambda \Big(z_t \cdot z_t - E[Z \cdot Z] \Big) + u_t \cdot z_t, \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \rangle \Big] \\
=& E\Big[\langle (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K}, z_t \cdot u_t \rangle \langle u_t \cdot z_t, \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \rangle \\
& + \lambda^2 \langle (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K}, \Big(z_t \cdot z_t - E[Z \cdot Z] \Big) \cdot \boldsymbol{\Phi_K^\top} (V_Z \boldsymbol{R_K})^{-1} \boldsymbol{R_K} \boldsymbol{b_K} \rangle \\
& \times \langle \lambda \boldsymbol{b_K^\top} \boldsymbol{R_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K} \cdot \lambda \Big(z_t \cdot z_t - E[Z \cdot Z] \Big), \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \rangle \\
& + \lambda \langle (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K}, z_t \cdot u_t \rangle \langle \lambda \boldsymbol{b_K^\top} \boldsymbol{R_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K} \\
& \cdot \lambda \Big(z_t \cdot z_t - E[Z \cdot Z] \Big), \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \rangle \\
& + \lambda \langle (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{\Phi_K}, \Big(z_t \cdot z_t - E[Z \cdot Z] \Big) \cdot \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \boldsymbol{R_K} \boldsymbol{b_K} \rangle \\
& \times \langle u_t \cdot z_t, \boldsymbol{\Phi_K^\top} (V_Z + \lambda \boldsymbol{R_K})^{-1} \rangle \Big].
\end{aligned} \tag{C.3}$$

□

Proof of Proposition 3 (Tests distribution). Under the assumptions (Y, X) *i.i.d.*, $K < \infty$, $E||Y - X||^2 < \infty$, $E(Y - X) = 0$, Theorem 2.7 in Bosq (2000) ensures that

$$T^{-1/2} \sum_t (Y_t - X_t) \xrightarrow{d} N(0, C), \tag{C.4}$$

where the asymptotic covariance operator C admits the eigendecomposition

$C = \sum_{l=1}^{\infty} \lambda_l \langle \cdot, \phi_l \rangle \phi_l$ since it is a compact linear operator,[85] where λ_l denotes an eigenvalue and ϕ_l its associated orthonormal eigenfunction. By definition, I have that for any $\delta \in \mathbb{L}^2$

$$\langle T^{-1/2} \sum_t (Y_t - X_t), \delta \rangle \xrightarrow{d} N(0, \langle K\delta, \delta \rangle), \tag{C.5}$$

from which $ii)$ and $iii)$ follow. Furthermore, I obtain

$$\frac{\langle T^{-1/2} \sum_t (Y_t - X_t), \phi_l \rangle}{\lambda_l} \xrightarrow{d} N(0, 1) \tag{C.6}$$

since $\langle K\phi_l, \phi_l \rangle = \lambda_l$ which implies

$$\frac{\langle T^{-1/2} \sum_t (Y_t - X_t), \phi_l \rangle^2}{\lambda_l} \xrightarrow{d} \chi^2(1). \tag{C.7}$$

Given that $\{\phi_l\}_{l=1}^{\infty}$ defines a complete orthonormal sequence, I can write $CvM_T = \sum_{l=1}^{+\infty} \langle T^{-1/2} \sum_t (Y_t - X_t), \phi_l \rangle^2$ by Parseval's formula. From (C.7), the asymptotic distribution in $i)$ follows

$$CvM_T = \sum_{l=1}^{+\infty} \lambda_l \frac{\langle T^{-1/2} \sum_t (Y_t - X_t), \phi_l \rangle^2}{\lambda_l} \xrightarrow{d} \sum_{l=1}^{+\infty} \lambda_l \chi_l^2(1) \tag{C.8}$$

where χ_l^2 are independent chi-square random variables. A more general proof of these results is provided by Shorack and Wellner (1986). □

Proof of Proposition 4 (Tests consistency). The test statistic CvM_T writes

$$CvM_T = T||T^{-1} \sum_{t=1}^{T} (y_t - x_t)||^2. \tag{C.9}$$

By the L.L.N for i.i.d sequences, stated as Theorem 2.4 in Bosq (2000), we have

$$T^{-1} \sum_{t=1}^{T} (y_t - x_t) \xrightarrow{p} E[Y - X], \tag{C.10}$$

as $T \to \infty$. Given that the $\mathbb{L}^2$-norm is continuous, applying the continuous mapping theorem yields

$$||T^{-1} \sum_{t=1}^{T} (y_t - x_t)||^2 \xrightarrow{p} ||E[Y - X]||^2, \tag{C.11}$$

[85] Theorem 2.41 in Carrasco et al. (2007).

where $0 < ||E[Y-X]||^2 < \infty$ since $Y, X \in \mathbb{L}^2$ and $E[Y-X] \neq 0$. Therefore, as $T \to \infty$, we have $CvM_T \xrightarrow{p} +\infty$ under any alternative $H_1 : E[Y-X] \neq 0$. This proves i).

Fix any $\psi \in \mathbb{L}^2$, and consider an alternative such that $E[Y(p)-X(p)] \neq 0$, for some $p \in \mathcal{P}$. Let us rewrite $J_{1,T}$ as

$$J_{1,T} = \sqrt{T}\frac{\langle T^{-1}\sum_t (y_t - x_t), \psi\rangle}{\langle \hat{C}\psi, \psi\rangle^{1/2}}, \tag{C.12}$$

and assume that $\hat{C}$ is a consistent estimator of C such that by the L.L.N we have $\langle \hat{C}\psi, \psi\rangle^{1/2} \xrightarrow{p} \langle C\psi, \psi\rangle^{1/2} < \infty$. Applying the L.L.N to the numerator of (C.12) yields

$$\langle T^{-1}\sum_t (y_t - x_t), \psi\rangle \xrightarrow{p} E[\langle Y, \psi\rangle] - E[\langle X, \psi\rangle], \tag{C.13}$$

where the interchange of the inner product and the expectation operator follows from linearity. Therefore, we have $J_{1,T} \xrightarrow{p} +\infty$ as $T \to \infty$ if and only if $E[\langle Y, \psi\rangle] \neq E[\langle X, \psi\rangle]$. This proves ii).

Similarly, fix any $\psi_1, ..., \psi_Q \in \mathbb{L}^2$. Let us consider an alternative such that $E[Y(p) - X(p)] \neq 0$, for some $p \in \mathcal{P}$, and rewrite $J_{Q,T}$ as

$$J_{Q,T} = \sum_{q=1}^{Q} T\frac{\langle T^{-1}\sum_t (y_t - x_t), \psi_q\rangle^2}{\langle \hat{C}\psi_q, \psi_q\rangle}, \tag{C.14}$$

and assume that $\langle \hat{C}\psi, \psi\rangle \xrightarrow{p} \langle C\psi, \psi\rangle < \infty$. For each q, applying the L.L.N to the numerator of (C.14) yields

$$\langle T^{-1}\sum_t (y_t - x_t), \psi_q\rangle \xrightarrow{p} E[\langle Y, \psi_q\rangle] - E[\langle X, \psi_q\rangle]. \tag{C.15}$$

Therefore, we have $J_{Q,T} \xrightarrow{p} +\infty$ as $T \to \infty$ if and only if $E[\langle Y, \psi_q\rangle] \neq E[\langle X, \psi_q\rangle]$ for some q. This proves ii).

□

C.2 Optimal bids and standard errors

Optimal bids

Algorithm 1: Optimal supply function

```
initialize q_0 = QC and p_0 = PC;
while p_0 ≤ max{p ∈ P} do
    Set p = min{p ∈ P | p > p_0} and q_ = q_0;
    Find q = argmin_{q ∈ [q_, K]} ( p − c'_t(q) − (q − QC_t) π̂_t(p) / f̂_t(p) )^2 ;
    Set Ŝ*(p) = q, q_0 = q and p_0 = p;
end
initialize q_0 = QC and p_0 = PC;
while p_0 ≥ min{p ∈ P} do
    Set p = max{p ∈ P | p < p_0} and q_ = q_0;
    Find q = argmin_{q ∈ [0, q_]} ( p − c'_t(q) − (q − QC_t) π̂_t(p) / f̂_t(p) )^2 ;
    Set Ŝ*(p) = q, q_0 = q and p_0 = p;
end
return Ŝ*;
```

In the listing, q_ stands for the printed underlined q, $\underline{q}$. The two argmin lines print as:

$$q=\operatorname{argmin}_{q\in[\underline{q},K]}\left(p-c_t'(q)-(q-QC_t)\frac{\hat{\pi}_t(p)}{\hat{f}_t(p)}\right)^2;$$

$$q=\operatorname{argmin}_{q\in[0,\underline{q}]}\left(p-c_t'(q)-(q-QC_t)\frac{\hat{\pi}_t(p)}{\hat{f}_t(p)}\right)^2;$$

NOTES. The algorithm solves for the optimal function recursively over a grid of prices starting at the contract price.

Calculations of standard errors

The algorithm used to obtain $\hat{S}^\star$ imposes a monotonicity restriction. For that reason, it is not trivial to compute the analytical expression of the covariance operator of this estimated function. I circumvent this issue by using a simulation approach. Pointwise confidence bands for $\hat{S}^\star$ are obtained from this estimated covariance operator, that can eventually be used to test hypothesis. The method is as follows. For every t:

1. Draw a random function π^b from the distribution $N(\hat{\pi},\hat{\mathcal{B}})$, where $\hat{\mathcal{B}}$ is the estimated covariance operator of $\hat{\pi}$ discussed in Proposition 2.

2. Compute the optimal supply function $S_t^{\star b}$ using (3.61) from π^b.

3. Perform 1000 replications of steps 1 and 2.
4. Estimate the covariance operator of $S_t^{\star}$ as in (3.30) after recentering.

C.3 Data

Algorithms

Algorithm 2: Computing the pseudometric between portfolios of possibly different cardinalities

---

function d_3^H (i, k);

Input : $PORTFOLIO_i^H$, $\widehat{PORTFOLIO}_k^H$, and $\mathcal{L} = \{1, ..., L_k^H\}$ the indices of the ordered set $\widehat{PORTFOLIO}_k^H$

Output: $d_3^H(i, k)$

initialize $j = 1$; $\mathcal{L}_1^i = \mathcal{L}$; $\mathcal{L}_1^k = \emptyset$;

while $j \leq \min(J_i, L_k)$ **do**

$\quad l_j = \text{argmin}_{l \in \mathcal{L}_j^i} |CAP_{i,j}^H - \widehat{CAP}_{k,l}^H|$;

$\quad \mathcal{L}_{j+1}^k = \mathcal{L}_j^k \cup l_j$;

$\quad \mathcal{L}_{j+1}^i = \mathcal{L}_j^i \setminus l_j$;

$\quad j = j + 1$;

end

$PF_i^H = \{PORTFOLIO_i^H, \mathbf{0}_{\min(L_k - J_i, 0)}\}$;

$PF_k^H = \{\widehat{PORTFOLIO}_k^H(\mathcal{L}_j^k), \widehat{PORTFOLIO}_k^H(\mathcal{L} \setminus \mathcal{L}_j^k), \mathbf{0}_{\min(J_i - L_k, 0)}\}$;

$d_3^H = 0.5 \times \frac{\sum_{s=1}^{max(J_i, L_k)} (PF_i^H(s) - PF_k^H(s))^2}{\sum_{s=1}^{max(J_i, L_k)} PF_i^H}$;

return d_3^H;

NOTES. The algorithm first reorders elements in bidder k's portfolio so that they match to those of firm i's. Second, it adds zeros so that the two set cardinalities match. Third it computes the distance metric as a normalized sum of squared differences. $\widehat{PORTFOLIO}_k^H(\mathcal{L}_j^k)$ denotes the set $\widehat{PORTFOLIO}_k^H$ rearranged according to indices in $\mathcal{L}_j^k$, and $\mathbf{0}_{\min(J_i - L_k, 0)}$ denotes a set containing $\min(J_i - L_k, 0)$ zeros.

Algorithm 3: Stage 1: Matching firms to bidders

function *match1* $(\mathcal{I}, \mathcal{K})$;
Input : Firms $\mathcal{I} = \{1, ..., I\}$, bidders $\mathcal{K} = \{1, ..., K\}$, the number of plants by firm $\{J_1^H, ..., J_I^H\}$ for all H, and the constraints $\mathcal{P}_s, \forall s$.
Output: The set of firms sorted in decreasing order by number of plants $\mathcal{I}_I$ and the corresponding set of bidder $\mathcal{K}_I$.

```
initialize t = 1; I_0 = ∅;
while t ≤ I do
    i_t = argmax_i J_i^{NP} ;
    I_t = I_{t-1} ∪ i_t;
    t = t + 1;
end
initialize r = 1; K_0 = ∅;
for i ∈ I_I do
    for k ∈ K do
        for s = 1 : 3 do
            if d_s^H(i,k) > 0.01 then
                p_{i,s}^H(k) = 2 exp(1/d_s^H(i,k)) / (1+exp(1/d_s^H(i,k))) − 1;
            else
                p_{i,s}^H(k) = 1;
            end
        end
        matchprob_i(k) = (1/9) Σ_H Σ_{s=1}^3 p_{i,s}^H(k);
    end
    P_1 = ∪_{∀H} {p_1^H > 0.5, p_2^H > 0.5, (1/3) Σ_{s=1}^3 p_s^H > 0.7};
    P_2 = {(1/9) Σ_H Σ_{s=1}^3 p_s^H > 0.35} ∩ {(1/3) Σ_H p_1^H > 0.5} ∩ {(1/3) Σ_H p_3^H > 0.5};
    P_3 = ∪_{∀H} {(1/3) Σ_{s=1}^3 p_s^H = 1};
    K̂ = {k ∈ K \ K_{r-1} | {p_{i,s}^H(k)}_{∀s,H} ∈ (P_1 ∩ P_2) ∪ P_3} ;
    k_i = argmax_{k∈K̂} matchprob_i(k);
    K_r = K_{r-1} ∪ k_i;
    r = r + 1;
end
```

NOTES. The algorithm first sorts the firms in decreasing order according to their number of plants. Second, each firm is matched to its closest bidder if the matching probabilities satisfies a set of constraints. Once a bidder is matched it is removed from the set of potential candidates. Note that I have $K > I$. The set of constraints is chosen by trial and error.

Matching results

Owner	Status	#PTIDS	Capacity (MW)	RMSE	MCL
AES ES Westover	Matched	0	0.0	N/A	0
	Not matched	1	8.0		
Albany Energy	Matched	1	5.6	1.8	0
	Not matched	0	0.0		
Astoria Energy II	Matched	2	660.0	13.8	2
	Not matched	0	0.0		
Astoria Energy	Matched	2	640.0	6.0	2
	Not matched	0	0.0		
Astoria Generating Company	Matched	51	1771.0	1.8	2
	Not matched	0	0.0		
Athens Generating Company	Matched	3	1323.0	43.0	3
	Not matched	0	0.0		
Bayonne Energy Center	Matched	8	512.0	0.0	0
	Not matched	0	0.0		
Boralex Hydro Operations Inc	Matched	4	20.7	0.0	0
	Not matched	0	0.0		
CHI Energy Inc	Matched	1	2.0	0.0	0
	Not matched	0	0.0		
Caine Energy Service	Matched	4	264.8	14.0	0
	Not matched	0	0.0		
CaN/Adaigua Power Partners	Matched	1	125.0	0.0	1
	Not matched	0	0.0		
Canastota Windpower	Matched	1	30.0	0.0	0
	Not matched	0	0.0		
Carr Street Generating Station	Matched	1	122.6	19.6	0
	Not matched	0	0.0		
Castleton Power	Matched	1	72.0	0.0	1
	Not matched	0	0.0		
Cayuga Operating Company	Matched	3	328.1	4.3	2
	Not matched	0	0.0		
Central Hudson Gas & Elec.	Matched	0	0.0	N/A	0
	Not matched	10	78.9		
Commerce Energy	Matched	1	20.0	0.0	0
	Not matched	0	0.0		
Consolidated Edison	Matched	6	1764.6	21.0	4
	Not matched	10	829.2		
Consolidated Hydro New York	Matched	0	0.0	N/A	0
	Not matched	2	4.4		
Covanta Niagara	Matched	1	50.0	11.7	0
	Not matched	0	0.0		
Delaware County	Matched	0	0.0	N/A	0
	Not matched	1	2.0		
Dynegy Marketing and Trade	Matched	1	1254.0	166.0	1
	Not matched	0	0.0		
Eagle Creek Hydro Power	Matched	1	21.8	0.7	0
	Not matched	0	0.0		
East Coast Power	Matched	0	0.0	N/A	0
	Not matched	1	1034.9		
Empire Generating Co	Matched	2	670.0	7.0	2
	Not matched	0	0.0		
Entergy Nuclear Power Marketing	Matched	3	3193.0	147.3	3
	Not matched	0	0.0		
Erie Blvd.	Matched	16	659.5	6.3	0
	Not matched	1	4.7		
Exelon Generation Company	Matched	7	31.6	1.6	0
	Not matched	0	0.0		

NOTES. This table reports descriptive results of the matching algorithm. RMSE denotes the root mean squared errors across plants within portfolios and MCL denotes the number of PTIDs given/confirmed in McCullough's Affidavits.

Table C.1: Matching PTIDs to MaskedGenIDs

Owner	Status	#PTIDS	Capacity (MW)	RMSE	MCL
Flat Rock Windpower II	Matched	1	90.7	0.1	0
	Not matched	0	0.0		
Flat Rock Windpower	Matched	1	231.0	0.0	0
	Not matched	0	0.0		
Freeport Electric	Matched	2	91.8	9.2	0
	Not matched	0	0.0		
GenOn Energy Management	Matched	2	1242.0	311.3	0
	Not matched	0	0.0		
Hampshire Paper Co.	Matched	1	3.4	0.1	0
	Not matched	0	0.0		
Hardscrabble Wind Power	Matched	1	74.0	0.0	0
	Not matched	0	0.0		
Howard Wind	Matched	1	57.4	4.1	0
	Not matched	0	0.0		
Indeck Energy Services of Silver Springs	Matched	1	56.6	1.1	0
	Not matched	0	0.0		
Indeck-Corinth	Matched	1	147.0	16.0	0
	Not matched	0	0.0		
Indeck-Olean	Matched	1	90.6	9.6	1
	Not matched	0	0.0		
Indeck-Oswego	Matched	1	57.4	2.8	0
	Not matched	0	0.0		
Indeck-Yerkes	Matched	1	59.9	5.9	0
	Not matched	0	0.0		
Innovative Energy Systems	Matched	7	38.4	1.7	0
	Not matched	0	0.0		
International Paper Company	Matched	1	42.1	32.5	0
	Not matched	0	0.0		
Jamestown Board of Public Utilities	Matched	0	0.0	N/A	0
	Not matched	2	101.0		
Lakeside New York	Matched	2	210.5	11.4	0
	Not matched	0	0.0		
Long Island Power Authority	Matched	67	5209.0	5.6	4
	Not matched	5	7.1		
Lyonsdale BioMass	Matched	2	76.6	8.1	0
	Not matched	0	0.0		
Madison Windpower	Matched	1	11.6	0.4	0
	Not matched	0	0.0		
Marble River	Matched	1	215.5	0.0	0
	Not matched	0	0.0		
Model City Energy	Matched	1	5.6	1.0	0
	Not matched	0	0.0		
Modern Innovative Energy	Matched	1	6.4	1.1	0
	Not matched	0	0.0		
NRG Power Marketing	Matched	28	4114.1	23.3	4
	Not matched	5	3.5		
New York Power Authority	Matched	25	6553.8	65.0	0
	Not matched	2	176.8		
New York State Elec. & Gas	Matched	5	109.0	1.5	0
	Not matched	13	20.9		
Niagara Generation	Matched	1	56.0	14.5	0
	Not matched	0	0.0		
Niagara Mohawk Power	Matched	3	138.8	5.1	0
	Not matched	14	334.4		
Nine Mile Point Nuclear Station	Matched	2	1901.1	29.3	2
	Not matched	0	0.0		

NOTES. This table reports descriptive results of the matching algorithm. RMSE denotes the root mean squared errors across plants within portfolios and MCL denotes the number of PTIDs given/confirmed in McCullough's Affidavits.

Table C.2: Matching PTIDs to MaskedGenIDs (continued)

Owner	Status	#PTIDS	Capacity (MW)	RMSE	MCL
Noble Altona Windpark	Matched	1	97.5	0.0	1
	Not matched	0	0.0		
Noble Bliss Windpark	Matched	1	100.5	0.0	1
	Not matched	0	0.0		
Noble Chateaugay Windpark	Matched	1	106.5	0.0	0
	Not matched	0	0.0		
Noble Clinton Windpark 1	Matched	1	100.5	0.0	1
	Not matched	0	0.0		
Noble Ellenburg Windpark	Matched	1	81.0	0.0	1
	Not matched	0	0.0		
Noble Wethersfield Windpark	Matched	1	126.0	0.0	1
	Not matched	0	0.0		
Northbrook Lyons Falls	Matched	0	0.0	N/A	0
	Not matched	1	8.0		
Orange and Rockland Utilities	Matched	0	0.0	N/A	0
	Not matched	4	8.6		
PSEG Energy Resource & Trade	Matched	1	893.1	321.1	1
	Not matched	0	0.0		
R.E. Ginna Nuclear Power Plant	Matched	1	614.0	32.0	1
	Not matched	0	0.0		
ReEnergy Chateaugay	Matched	0	0.0	N/A	0
	Not matched	1	19.7		
Rochester Gas and Electric	Matched	1	57.1	1.7	0
	Not matched	5	20.6		
Rockville Centre Village of	Matched	0	0.0	N/A	0
	Not matched	1	33.8		
Selkirk Cogen Partners	Matched	2	446.0	8.2	0
	Not matched	0	0.0		
Seneca Energy II	Matched	0	0.0	N/A	0
	Not matched	2	24.0		
Seneca Power Partners	Matched	6	350.9	3.9	0
	Not matched	0	0.0		
Sheldon Energy	Matched	1	112.5	0.0	0
	Not matched	0	0.0		
Shell Energy North America	Matched	3	288.4	5.3	0
	Not matched	0	-0.0		
Somerset Operating Company	Matched	1	655.1	30.9	1
	Not matched	0	0.0		
Stephentown Spindle	Matched	0	0.0	N/A	0
	Not matched	1	20.0		
Stony Creek Energy	Matched	1	93.9	0.1	0
	Not matched	0	0.0		
Syracuse Energy Corporation	Matched	0	0.0	N/A	0
	Not matched	2	101.6		
TC Ravenswood	Matched	19	2557.1	17.5	0
	Not matched	0	0.0		
TransAlta Energy Marketing	Matched	1	285.6	5.6	1
	Not matched	0	0.0		
Triton Power Company	Matched	0	0.0	N/A	0
	Not matched	1	3.0		
Western New York Wind	Matched	0	0.0	N/A	0
	Not matched	1	6.6		
Wheelabrator Hudson Falls	Matched	1	14.4	2.8	0
	Not matched	0	0.0		
Wheelabrator Westchester LP	Matched	1	59.7	10.1	0
	Not matched	0	0.0		

NOTES. This table reports descriptive results of the matching algorithm. RMSE denotes the root mean squared errors across plants within portfolios and MCL denotes the number of PTIDs given/confirmed in McCullough's Affidavits.

Table C.3: Matching PTIDs to MaskedGenIDs (continued)

Plant-level costs

Fuel	Fuel Type	P-M	Capacity (MW)	Heat rate	CO2 rate	SO2 rate	Var. O& M
LFG	BIO	IC	108.2	11.16	9.35e-02	1.20e-04	17.5
OBG	BIO	IC	2.0	6.62	9.35e-02	1.20e-04	10.4
BLQ	BIO	ST	25.0	5.90	9.35e-02	1.20e-04	15.4
MSB	BIO	ST	155.0	18.35	9.35e-02	1.20e-04	15.4
MSN	BIO	ST	148.9	18.35	9.35e-02	1.20e-04	15.4
TDF	BIO	ST	28.9	17.23	9.35e-02	1.20e-04	15.4
WDS	BIO	ST	103.8	8.27	9.35e-02	1.20e-04	15.4
BIT	BIT	ST	987.2	10.89	9.36e-02	4.20e-04	4.5
SUB	SUB	ST	634.7	11.67	9.43e-02	2.54e-04	4.5
DFO	DFO	CC	12.0	11.53	7.35e-02	1.95e-04	3.6
DFO	DFO	GT	1213.0	14.85	7.35e-02	1.95e-04	15.4
DFO	DFO	IC	62.0	10.20	7.35e-02	1.95e-04	10.4
DFO	DFO	ST	8.8	11.37	7.35e-02	1.95e-04	15.4
RFO	RFO	ST	2076.5	10.71	6.28e-02	1.20e-04	15.4
KER	KER	CC	6.9	11.55	N/A	N/A	3.6
KER	KER	GT	375.3	14.12	N/A	N/A	15.4
WO	PET	CC	0.2	7.57	N/A	N/A	3.6
WO	PET	ST	0.0	14.95	N/A	N/A	15.4
NG	NG	CC	8175.3	11.26	5.44e-02	6.75e-06	3.6
NG	NG	GT	4109.1	10.97	5.44e-02	6.75e-06	15.4
NG	NG	IC	18.2	9.65	5.44e-02	6.75e-06	10.4
NG	NG	ST	7974.9	11.49	5.44e-02	6.75e-06	15.4
PG	NG	ST	0.1	17.92	5.44e-02	6.75e-06	0
NUC	NUC	ST	5708.1	10.45	N/A	N/A	2.1
WAT	REN	HY	5329.1	0.00	N/A	N/A	6.0
SUN	REN	PV	31.5	0.00	N/A	N/A	0.0
WND	REN	WT	1729.7	0.00	N/A	N/A	0.0

NOTES. This table reports heat rates (MMbtu/MWh), emission rates (ton/MMbtu) and variable operation and maintenance costs (USD/MWh) at the prime mover-fuel level. P-M denote prime mover and include internal combustion (IC) engine, steam turbine (ST), combined cycle (CC), gas turbine (GT), hydro turbine (HY), photovoltaic panels (PV) and wind turbine (WT). Biomass (BIO) includes landfill gas (LFG), other biomass gases (OBG), black liquor (BLQ), Municipal Solid Waste - Biogenic component (MSB), Municipal Solid Waste - Non-biogenic components (MSN) , tire-derived fuels (TDF), and wood waste solids (WDS). Other fuels include bituminous coal (BIT), sub-bituminous coal (SUB), distillate fuel oil (DFO), residual fuel oil (RFO), kerosene (KER), other petroleum-bases fuels (PET) such as waste oil (WO), gaseous propane (PG), natural gas (NG), nuclear fuel (NUC). Renewable energy sources include water (WAT), sun (SUN) and wind (WND). N/A denotes a missing value due to the absence of that unit type regulated by the RGGI or ARP.

Table C.4: Cost parameters by prime mover-fuel

Fuel/Emission Type	Mean	Std.	Min	Max	Frequency
BIO	2.54	0.21	2.25	2.70	yearly
BIT	2.59	0.08	2.44	2.73	monthly
SUB	2.27	0.09	2.12	2.50	monthly
DFO	19.17	4.51	9.09	24.07	monthly
RFO	13.41	4.45	6.45	20.47	monthly
KER	17.50	4.90	7.95	23.65	monthly
PET	17.65	5.21	7.24	22.95	monthly
NG	3.54	0.87	1.54	7.78	daily
NUC	0.00	0.00	0.00	0.00	constant
REN	0.00	0.00	0.00	0.00	constant
CO2	5.12	1.48	1.86	8.39	weekly
SO2	0.21	0.10	0.11	0.35	yearly

NOTES. This table reports descriptive statistics for the fuel and emission types that are used to construct plant-level marginal cost functions. The acronyms are given in the previous table.

Table C.5: Descriptive statistics of fuel and emission prices

Forward positions

As explained in section 3.4, the intersection of a firm's supply and marginal cost functions identifies its forward position.

However, curves may cross multiple times due to the measurement errors in the constructed cost functions. I use the following algorithm to identify the most plausible candidate.

Consider firm i in day d. First, the quantity bids are normalized by the maximum capacity bid $K_{i,d}$ in order to obtain inverse supply functions $S_{i,d}^{-1}(q)$ and cost functions $C_{i,d}(q)$ defined over the fixed quantity interval $[0,1]$. The set of candidates includes all quantity values at which the two curves intersect,[86] it is defined formally as $\mathcal{QC}_{i,d} = \{q \in [0,1] | S_{i,d}^{-1}(q) = C_{i,d}(q)\}$. If $|\mathcal{QC}_{i,d}| = 1$, then there is a unique candidate. If $|\mathcal{QC}_{i,d}| = \emptyset$, the curves never cross and thus the forward quantity is $QC_{id} = 0$ if $\int_0^1 \left(S_{i,d}^{-1}(q) - C_{i,d}(q)\right) dq < 0$ and $QC_{id} = 1$ otherwise. If $|\mathcal{QC}_{i,d}| = J > 1$, there are multiple candidates to choose from. Denote $\mathcal{QC}_{i,d}(j)$ the j^{th} element of the set of candidates.[87]

[86] When the curves are confounded over some non-degenerate intervals, i.e. $S_{i,d}^{-1}(q) = C_{i,d}(q)$ for $\forall q \in [a,b]$ with $b > a$, then a single candidate $a + (b-a)/2$ is selected for the whole interval.

[87] Remark that any candidate j such that $\lim S_{i,d}^{-1}(\mathcal{QC}_{i,d}(j)+\varepsilon) = C_{i,d}(\mathcal{QC}_{i,d}(j)+\varepsilon) < 0$ as $\varepsilon > 0$ approaches 0 can be discarded because it means that the supply curve goes below the cost curve at that point.

I consider that a plausible forward quantity QC_{id} must be such that for the inverse supply is clearly above the cost for larger quantities and below it for smaller ones. Thus, I choose $QC_{id} = \mathcal{QC}_{i,d}(k)$ where

$$k = \text{argmax}_j \int_0^{\mathcal{QC}_{i,d}(j)} \Big(C_{i,d}(q) - S_{i,d}^{-1}(q)\Big)dq + \int_{\mathcal{QC}_{i,d}(j)}^1 \Big(S_{i,d}^{-1}(q) - C_{i,d}(q)\Big)dq. \tag{C.16}$$

Once the forward quantity $QC_{i,d}$ is selected, I correct the cost function so that the curves cross only at that quantity value and I compute the corresponding contracted price $PC_{i,d}$.[88]

Table 3.5 shows firm-level descriptive statistics of estimated forward prices and quantities, the number of day-ahead bids and of Masked Gen IDs.[89]

[88] Additionally, I consider a contract price above 150 USD as an outlier, therefore if there are other candidates I choose to neglect intersection points leading to such large values.

[89] Firms have been re-arranged to reflect the actual ownership structure during the sample period. This is because the *owner* category in the NYISO's Gold Book does not offer a correct picture of the ownership structure of New York 's power plants. I aggregate owners into firms as follows: Astoria Energy, LLC represents Astoria Energy, LLC and Astoria Energy II, LLC, Upstate New York Power Producers, Inc. represents Cayuga Operating Company, LLC and Somerset Operating Company, LLC, Flat Rock Windpower, LLC represents Flat Rock Windpower, LLC and Flat Rock Windpower II, LLC, Indeck Energy Services, Inc. includes all Indeck plants, Noble Energy Inc. includes all Noble wind parks, and Exelon Generation Company, LLC also contains R.E. Ginna Nuclear Power Plant, LLC and Nine Mile Point Nuclear Station, LLC, and Wheelabrator Technologies contains both Wheelabrator facilities.

Firm	Capacity	Max Bid	Mean QC	Mean PC	# GENIDs	# obs
Dynegy Marketing and Trade	1254.00	279.00	0.95	47.71	1	980
GenOn Energy Management	1242.00	1136.00	0.78	97.76	2	987
Upstate New York Power Producers	983.20	995.40	0.83	38.33	3	1090
PSEG Energy Resource & Trade	893.10	290.10	0.97	47.65	1	997
Empire Generating Co	670.00	682.00	0.81	124.12	2	1057
Erie Blvd.	659.50	581.70	0.78	5.09	16	1095
Noble Energy	612.00	612.00	1.00	0.00	6	1095
Bayonne Energy Center	512.00	512.00	0.80	233.41	8	1053
Niagara Mohawk Power	473.20	136.60	1.00	57.39	3	1095
Selkirk Cogen Partners	446.00	436.00	0.86	47.86	2	1051
Indeck Energy Services	411.50	380.30	0.80	49.69	5	1095
Seneca Power Partners	350.90	330.70	0.23	47.79	6	1095
Flat Rock Windpower	321.70	321.80	1.00	0.00	2	1066
Shell Energy North America	288.40	302.90	0.90	47.73	3	1095
TransAlta Energy Marketing	285.60	280.00	1.00	47.73	1	1087
Caine Energy Service	264.80	218.30	0.87	50.15	4	1095
Marble River	215.50	215.50	1.00	0.00	1	1088
Lakeside New York	210.50	188.40	0.58	47.69	2	1072
New York State Elec. & Gas	129.90	100.50	1.00	55.39	5	1095
Canandaigua Power Partners	125.00	125.00	1.00	0.00	1	1085
Carr Street Generating Station	122.60	103.00	0.87	193.83	1	1059
Sheldon Energy	112.50	118.10	0.96	0.00	1	1079
Stony Creek Energy	93.90	94.00	0.93	0.00	1	769
Freeport Electric	91.80	74.00	1.00	212.98	2	1069
Rochester Gas and Electric	77.70	55.40	1.00	6.00	1	1073
Lyonsdale BioMass	76.60	72.00	0.43	37.65	2	262
Wheelabrator Technologies	74.10	61.40	0.96	55.30	2	1095
Hardscrabble Wind Power	74.00	74.00	1.00	0.00	1	1054
Castleton Power	72.00	72.00	0.94	173.64	1	1027
Howard Wind	57.40	61.50	0.88	0.00	1	1064
Niagara Generation	56.00	41.50	1.00	276.47	1	1051
Covanta Niagara	50.00	34.00	1.00	61.21	1	1095
International Paper Company	42.10	9.70	1.00	158.84	1	1095
Innovative Energy Systems	38.40	28.50	1.00	38.69	7	1095
Canastota Windpower	30.00	30.00	1.00	0.00	1	1093
Eagle Creek Hydro Power	21.80	21.10	0.35	6.00	1	1095
Boralex Hydro Operations Inc	20.70	20.80	0.36	6.00	4	1095
Commerce Energy	20.00	20.00	1.00	0.00	1	1094

NOTES. This table reports descriptive statistics for firm-level data. A firm's max capacity denotes the maximum capacity this firm has bid in the DAM. Mean QC is the average contract quantity as a share of total quantity bid. Mean PC is the average contract price. # GENIDs denote the number of Masked Gen IDs associated to the firm. # obs is the number of days with day-ahead market bids over the sample period (2013-2015).

Table C.6: Descriptive statistics of firm-level data

C.4 Estimation of marginal costs

Algorithm 4: Estimation of the marginal cost function

for $QC \in \mathcal{QC}$ **do**
  Find $PC_{QC} = \text{argmin}_{PC \in \mathcal{P}} |QC - S^0(PC)|$;
  Initialize $c_0 = PC_{QC}$ and $p_0 = PC_{QC}$;
  while $p_0 \leq \max\{p \in \mathcal{P}\}$ **do**
    Set $p = \min\{p \in \mathcal{P} | p > p_0\}$ and $\underline{c} = c_0$;
    Find $c = \text{argmin}_{c \in [\underline{c}, \overline{P}]} \left(p - c - (S_t^0(p) - QC) \frac{\hat{\pi}_t(p)}{\hat{f}_t(p)} \right)^2$;
    Set $C_{QC}(p) = c$, $c_0 = c$ and $p_0 = p$;
  end
  Initialize $c_0 = PC_{QC}$ and $p_0 = PC_{QC}$;
  while $p_0 \geq \min\{p \in \mathcal{P}\}$ **do**
    Set $p = \max\{p \in \mathcal{P} | p < p_0\}$ and $\underline{c} = c_0$;
    Find $c = \text{argmin}_{c \in [0, \underline{c}]} \left(p - c - (S_t^0(p) - QC) \frac{\hat{\pi}_t(p)}{\hat{f}_t(p)} \right)^2$;
    Set $C_{QC}(p) = c$, $c_0 = c$ and $p_0 = p$;
  end
end
Find $\widehat{QC}_t =$
$\text{argmin}_{QC \in \mathcal{QC}} \int_{\mathcal{P}} \left(p - C_{QC}(p) - \left(S^0(p) - QC \right) \frac{\hat{\pi}(p)}{\hat{f}(p)} \right)^2 \hat{f}(p) dp$;
Set $\hat{c}_t' = C_{\widehat{QC}_t}$;
return $\widehat{QC}_t, \hat{c}_t'$;

NOTES. The algorithm solves for the cost function recursively over a grid of prices.