2021-07-15

How Should Information Technology be Regulated

Vijairaghavan, Vaarun

UNIVERSITY OF CALGARY

How Should Information Technology be Regulated?

by

Vaarun Vijairaghavan

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MANAGEMENT

CALGARY, ALBERTA

JULY, 2021

Abstract

With Information Technology (IT) playing a more central role in the modern economy, governments the world over are displaying an active interest in regulating IT firms with the goal to increase competition, foster local firms or increase social surplus. My research informs such policymakers in two critical and relevant areas: the effect of regulations on IT investments and societal wellbeing (measured as total surplus), and the net productivity effects of increased IT investments.

In my first research stream, I use analytical modelling methodologies to evaluate the impact of specific regulatory mechanisms, thereby creating a theoretical template for how such regulations can be evaluated. In my first chapter in this stream, I analyze how firm investments can be incentivized and coordinated through a platform, and whether the platform as an institutional mechanism requires regulatory intervention. In my second chapter, I analyze Data Portability Regulations which require that platforms allow users to download their data and port it to competing firms. Such laws have been passed by the E.U. and the state of California, and are being considered by the U.S. Congress. I study whether this regulation accomplishes the law's goal of encouraging competition in the data economy.

In a second research stream, I use structural econometric models to measure the effect of IT investments on energy productivity. I execute this by mathematically deriving a structural econometric model to estimate the impact of IT on the output elasticity of energy. Thus, if a policy decision to regulate platforms causes a decrease in firm IT investments, my empirical work measures the impact of such decreases, both in terms of its direct effect on marginal product as well its indirect effect through the change in the productivity of energy.

Epigraph

The obstacles, trials and tribulations in the path ... can be removed to a large extent with the help of a Guru. The syllable gu means darkness and ru means light. He alone is a Guru who removes darkness and brings enlightenment. The conception of a Guru is deep and significant. He is not an ordinary guide. He is a spiritual teacher who teaches a way of life, and not merely how to earn a livelihood. He transmits knowledge of the Spirit and one who receives such knowledge is a sisya, a disciple.

- B.K.S. Iyengar, *Light on Yoga*

Acknowledgements

I would like to express my gratitude to Dr. Barrie Nault who inspired me to take this path, who beyond being my supervisor, is also my guru.

I would also like to thank my wife, Rakhee, for providing me with the sense of calm, and the sense of purpose that is crucial in an endeavour like this. Her unwavering support, encouragement and understanding made this journey possible; she is my source of strength.

I also want to thank my parents and grandparents for their love and encouragement, my parents-in-law for their help and support, my daughter Meera for all the joy she brought me, and my sister Mridula for her 1000 watt smile and booming laughter!

I want to thank my Ph.D. committee for their time, encouragement and direction over these past years; FGS, GSA, Haskayne's Ph.D. office, Dr. Nault and Dr. Hidaji for funding support; and Jeanette Burman for excellent editing assistance.

I dedicate this work to my late grandfather, Krishnaswamy Rajagopalan. For when a man of his stature believes in you, you begin to believe in yourself.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The impact of information technology (IT) on firms, the economy, and society is a crucial debate in the current era. IT represents a larger proportion of consumer and industry budgets today than ever before - for instance, IT capital in U.S. manufacturing firms is growing faster than other inputs by an annual growth rate of 3.46% over 1998-2018. On the supply side, IT firms have dominated - the top five firms in the S&P500 are IT firms and represent 19% of the S&P500 market capitalization (Mackintosch 2018; Rajkumar 2020). In addition, IT firms are continuing to grow so that they occupy a central place in our economies. Such developments have garnered attention from policy-makers because there are significant implications for market power, competition, innovation, and taxation. As such, there is deep interest in regulating IT and

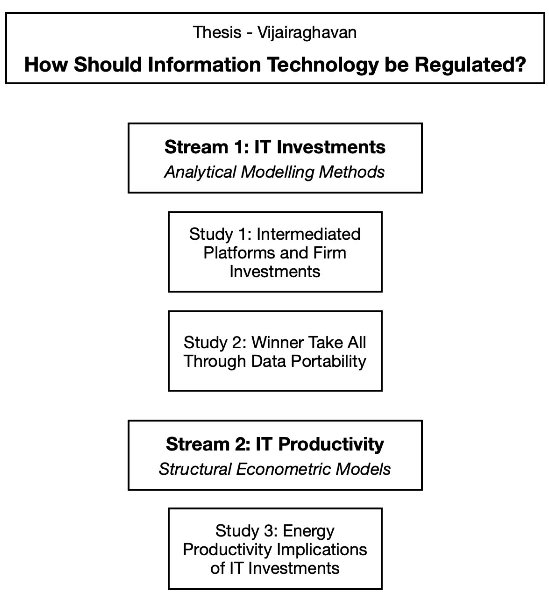


Figure 1.1: Thesis Structure

varied perspectives on whether IT should be regulated, and if so how this is best accomplished.

My thesis informs this debate on two fronts. Firstly, we use general form analytical models based in the Economics of Information Systems to study the impact of IT regulation on IT investments. The goal of this stream of research is to inform the policy-maker's decision about whether to regulate IT industries related to data analytics, the internet of things, and platforms, and if so, how to regulate them. In our second stream of research, we measure the impact of IT investments on productivity by using structural econometric models viewed through the lens of IT productivity. Thus, the intent of the research program is to theoretically evaluate the effects of regulation on IT investments, and to empirically evaluate the effects of IT investments on productivity, as seen in Figure 1.1. We briefly introduce each of our research streams and studies below.

Research Stream 1. IT Industry Regulation - Analytical Modelling Methods

Study 1 – Platform Intermediation and Firm Investments: Platforms such as Microsoft Windows have the scale and reach to influence the participation, investments, and profits of hundreds of firms that produce Windows compatible hardware and software. Although such intermediated platform structures are widely prevalent in the modern economy, they have not garnered much attention from IS researchers. Intermediated platforms are different from other two sided platform models in that a separate set of decisions are made by firms operating on the platform, and these decisions are based on each firm's objectives. In addition, while pricing strategies for platform markets have been widely studied, investments in platform-based markets have received little attention. In the context of compatible products being sold on an intermediated platform, the coordination of investments by firms operating on the platform becomes a pertinent question.

We analyze the structure of intermediated platforms, where firms operating on a platform sell their platform-compatible wares to end-customers. In particular, we evaluate how investments by such firms can be coordinated by the platform. We operationalize this study using the example of the Smart Grid (SG), which improves energy productivity through large multi-party firm investments in IoT, data analytics, and real-time control systems. In this setup, the platform charges a per-transaction fee from firms for its services. In addition, platforms enable a per-transaction transfer from one side of the market to another, which is a mechanism to coordinate prices and investments on each side of the market. Firms operating on the platform observe the fees and transfers set by the platform, and then set the investments and prices for end-customers. Thus, in this two-stage game the platform sets the fees and transfers in the first stage, then the firms set their investments and prices in the second stage.

Our main finding is that the platform's choice of transfers maximizes platform profits, while also maximizing the policy-maker's goal of end-customer adoption of the SG. While the platform's choice of fees decreases firm profits and participation, its effect on end-customer adoption depends on the shape of the demand curve facing each firm. Given certain conditions such as convex, less responsive to own price, or more responsive to cross price, demand, prices set by firms can increase faster than fees, a phenomenon we term P-Convexity. When there is P-Convexity, we show that firm investments increase in fees charged to firms. We term these conditions P-Investments. For the policy-maker, our study explains the mechanism through which platforms influence firm participation, investments, pricing, and profits, and it explains how platform choices affect end-customer adoption. Most crucially, for policy-makers interested in customer adoption such as in the case of the SG, the decision to regulate platforms depends on the characteristics of demand facing each firm, and on the presence or absence of P-Investments or P-Convexity.

Study 2 - Winner Takes All With Data Portability: Data Portability Regulations (DPR) are a key aspect of the European Union's General Data Protection Guidelines (Council of European Union 2016). With the objective of increasing innovation and competition in the data economy, DPR legislates that Data Controllers (DCs) such as Amazon and Expedia allow their users to download their personal data in a commonly used format so that users can port their data to other competing DCs. Platforms that do not comply with DPR risk being fined. Similar laws have been passed by the state of California (California State Legislature 2018) and are being mooted federally by the U.S. Congress (United States House of Representatives 2020).

In this second study, we evaluate the effect of fines to enforce DPR on the structure of the DC industry. We model a setup in which DCs differ in capability, where capability is defined as the ability to generate revenue per unit of output. The sequence of events is modeled in two stages, where the policy-maker first sets fines to enforce DPR, and in the second stage DCs choose whether to participate in a given industry, and if so, whether to comply with DPR.

We find that fines to enforce DPR divides DCs into three categories according to capability. The most capable DCs participate and comply with DPR, less capable DCs participate but do not comply with DPR, while the least capable DCs do not participate. Further, variable fines that depend on output have the additional effect of decreasing consumer surplus generated by DCs that do not comply with DPR, while also increasing the consumer surplus generated by compliant DCs. If the gains in consumer surplus from more capable DCs are larger in magnitude than the lost surplus from those less capable, then increasing fines has a positive effect on consumer surplus. However, because compliant DCs are also the most capable DCs, they gain at the expense of less capable DCs. On the whole, while fines can increase consumer surplus, and producer surplus, this can come at the expense of the least capable DCs exiting the industry, the mid-capability DCs experiencing decreases in output and profits, and the most capable DCs gaining in terms of output and profits. Thus, fines to enforce DPR can run counter to one of the purposes of the regulation -

decreasing market concentration.

General Form Analytical Models - A Note

We use the general form to model each of the above research questions, which differentiates our work from functional form models. In general form models, assumptions are made about the curvature conditions of the demand or profit function, thus the theory developed is generalizable to situations where curvature conditions assumed are satisfied. Therefore, general form models can lead to more generalizeable theories.

Research Stream 2. Impact of IT Investments - Structural Econometric Models

Study 3 – Energy Productivity Implications of IT Investments: "Advancing American Energy" is at the heart of the U.S. energy policy, with strategic investments being made in order to increase energy productivity. Given the well documented evidence on the variety of ways that IT improves operational effectiveness and the productivity of other inputs like capital and labor, we posit that IT improves the output elasticity of energy. Further, given the large scale IT investments to modernize and computerize the electric grid, an interesting question is whether these investments have disproportionately improved the output elasticity of electric energy.

In this study, we assess the impact of IT on energy productivity. We execute this study by developing a structural econometric model based on the Cobb-Douglas functional form to assess the effect of IT on energy productivity. Additionally, we estimate the effect of IT on the productivity of electric energy and non-electric energy. Here, we use the varying coefficient model for the base specification, and use U.S. industry level data collected from the U.S. Bureau of Economic Analysis, the Bureau of Labor Statistics, the the Energy Information Administration.

We find that IT has a significant effect on the output elasticity of energy, so that a 1% increase in IT is associated with a 0.13% increase in the output elasticity of energy. In addition, we find that the effect of IT on the productivity of electric energy is an order of magnitude larger than its effect on the productivity of non-electric energy. Specifically, a 1% increase in IT is associated with a 1.52% increase in the output elasticity of electric energy, whereas IT increases the output elasticity of non-electric energy by 0.12%.

Structural Econometric Models - A Note

We execute our empirical work through the use of structural econometric models. Thus, our work formally builds theory prior to econometric testing, and lays a foundation for a theoretical explanation of the phenomena being observed and measured. Such an approach can add insight to empirical findings.

Chapter 2

Platform Intermediation and Firm Investments

We study here, how firm investments can be coordinated in a setting where firms offering two kinds of complementary technology operate on a platform that intermediates the relationship between firms. We model this coordination problem by abstracting from the case of Demand Response (DR) on the Smart Grid, where the platform enables interoperability between two distinct but complementary types of DR products, firms invest to develop DR products, customers purchase DR products from firms, and the policymaker favors greater customer adoption of DR. In our setting, demand generated by a firm depends on own investments and prices which they choose, as well as cross-side investments and prices by other firms. The platform sets fees and transfers between firms. We set up our analysis as a two-stage model. In the first stage, the DR platform chooses both fees and transfers between firms. In the second stage, firms choose investments and prices. Importantly, we find that the platform and the policymaker goals are aligned: the platform's choice of transfer coordinates firm prices and investments, and maximizes aggregate platform demand. Although fees decrease firm profitability and firm participation, its effect on customer adoption depends on the shape of demand facing firms. If a firm faces demand that is convex, less responsive to own price, or more responsive to cross-prices – then we can observe a platform phenomenon we term P-Convexity, where prices increase faster than fees. Further, we demonstrate how firm investments and prices are inter-related – if a firm faces P-Convexity, then it is inevitable that firm investment increases with fees, conditions we call P-Investment. On the other hand, if investments decrease in fees, then prices can also decrease in fees resulting in a negative investment externality.

2.1 Introduction

Over the last two decades, the behavior of two-sided markets has been extensively studied - particularly from the perspectives of pricing and subsidies. For instance, the circumstances under which monopolist ownership is beneficial to a platform have been made evident, how a platform can cross-subsidize prices on each side of the market has been clarified, and it has been shown how a platform with a longer term purview can benefit from initial subsidization (Eisenmann 2007; Parker and Van Alstyne 2005; Rochet and Tirole 2006). However, the role of investments in two-sided markets has been relatively under-studied - often, studies re-categorize investments as subsidies, and provide results about the effect of such subsidies. For instance, Parker and Van Alstyne (2005) view Microsoft's operating system as a platform that charges for complementary applications on the consumer side of the market, but subsidizes system developer toolkits for the developer side of the market. However in such a setup, the investments in system developer toolkits are abstracted away in favor of pricing an already existing toolkit.

Although investments and price subsidies can each drive demand, investments are fundamentally different from price subsidies. Firstly, price subsidies vary with quantity demanded, whereas investments are usually set to affect demand. Secondly, investments can have the effect of changing demand, so that a different quantity is realized at each price point; a price subsidy merely involves going down a given demand curve so that a higher quantity is realized. Thirdly, prices and investments can interact so that there are particular combinations of prices and investments that maximize profit. To illustrate this third point, consider a firm's product investment decision - it estimates the price that each of its potential products will command in the market, and chooses to invest to bring a particular product to market.

Separately, the subject of platform intermediation has not received much attention (Ba and Nault 2017). Platforms intermediate firms operating on the platform, so that the platform intermediates the relationship between the firms, and between firms and end-customers. For example, Microsoft Windows is a platform that is shared by several hardware and software firms, and these firms sell their own Windows-compatible wares to end-customers. Such platform intermediation is structurally different from standard models of two-sided markets, where a platform sells directly to end-customers on each side of the market. Further, compared to simple price intermediaries, we study a more active form of intermediation where firms choose both prices as well as investments to increase the value of platform-compatible products. Thus, compared to simple 2-sided markets, intermediated platforms result in a separate set of pricing and investment decisions made by a set of firms operating on the platform, leading to a model that embeds important network externalities. In the case of Windows, end-customers secure hardware and software which work together as "systems of compatible products", necessitating an approach that is different from the standard two-sided model (pp 1496, Parker and Van Alstyne 2005). Several prior works have modeled intermediated settings such as Microsoft Windows platform, Google's Android platform and video game consoles as simple two sided platforms, thereby assuming that any subsidies offered by the platform are passed on fully to customers by firms operating on the platform. However, this may not be incentive compatible for firms. In our study, we focus on how firm behavior (price and investment) changes with platform choices of fees, and transfers among firms.

If customers gain value only from acquiring a system of compatible products, then this requires the coordination of investments (and prices) across the two sides of the market. Individual investments by firms on one side of the market are not as effective as coordinated investments on both sides of the market. In other words, investments on one side of the market may not deliver the expected benefits if complementary investments are not made. Although platforms can potentially play a role in coordinating investments of each side of the market, it is an unclear how this mechanism functions, and specifically how firm investments, profits and participation are affected by a platform's choices.

Further, the modern regulatory environment has evolved so that regulators and policy-makers are taking an active interest in platforms. In terms of size, Information Technology firms have clearly dominated - the largest five firms in the S&P500 are technology firms with an average valuation exceeding $1 Trillion, accounting for 19% of S&P500 market value (Mackintosch 2018, Rajkumar 2020). The risks emanating from economic concentration of this scale has attracted the attention of policymakers the world over, and in Washington, D.C., the U.S. Congress and anti-trust enforcement agencies are actively updating anti-trust regulation to "keep pace with technological change" (Kendall and McKinnon 2019). Such regulators may have objectives that are different than profit maximization (United States House of Representatives 2020), or even maximizing total surplus - which has been the focus of several prior works in the stream of literature on platforms. For instance, platforms clearly have an impact on the profitability of firms operating on them - an interesting question is then the effect that platform choices have on firm investments, profitability and participation, in addition to the effect on end-customer adoption of the platform.

We model a setting where firms operate on a platform that intermediates the relationship between the firms, and between the firms and end-customers. We focus on how the platform intermediates the relationship between firms that offer two types of strictly complementary products. The platform charges a fee from firms for its services, and uses the mechanism of transfers to coordinate the two sides of the market. Firms operating on the platform invest to bring products to market on either one or both sides of the platform, and set prices for their products on the end-customer market. We study how the platform uses fees and transfers to profit maximize, and coordinate firm investments. We also study the effect of the platform's choice of fees and transfers on platform profits, firm participation, firm profits, customer adoption of the platform, and social surplus.

We find that the platform's choice of transfer is aligned with the objective of the policymaker: the transfer coordinates firm investments and prices in order to maximize aggregate platform demand. However, the effect of the platform's choice of fees is both interesting and multi-faceted. Although fees decrease firm profitability and firm participation, its effect on customer adoption depends on the shape of demand facing firms. If a firm faces demand that is convex, less responsive to own price, or more responsive to cross-prices - then we can observe a platform phenomenon we term P-Convexity, where the rate at which firm prices increase with platform fees is greater than unity. Further, we demonstrate how firm investments and prices are inter-related - if a firm faces P-Convexity, then it is inevitable that firm investment increases with fees, conditions we call P-Investment. On the other hand, if investments decrease in fees, then prices can also decrease in fees resulting in a negative investment externality. For policymakers interested in increasing investments, total surplus, or customer adoption, the absence or presence of P-effects is an important determinant of whether the platform should be regulated. As well, our results are informative to platforms that are interested in firm investments or customer adoption.

2.1.1 Literature Review

Rochet and Tirole (2006) define a market as two-sided *"if the platform can affect the volume of transactions by charging more to one side of the market and reducing the price paid by the other side by an equal amount"* (pp 648). Thus, the very presence of one customer type creates value for the other type. This is exposited with clarity by Katz and Shapiro (1985) *"Consumption externalities give rise to demand-side economies of scale, which will vary with customer expectations. As a result, multiple fulfilled expectations equilibria may exist for a given set of cost and utility functions."* Further, they state, *"if customers expect a seller to be dominant, then customers will be willing to pay more for the firm's product, and it will, in fact be dominant"* (P 425).

In a two-sided market, it is possible that customers on one side do not internalize the welfare impact of their joining the network on customers on the other side. This provides an opportunity for the platform to subsidize one side, thereby creating positive externalities for the other side. The side that generates the greater externality benefit is subsidized, and this pricing strategy does not harm welfare (Parker and Van Alstyne 2005). Several studies have shown that a monopoly platform is beneficial to customer adoption (Katz and Shapiro 1985, Rochet and Tirole 2006). First, monopolies can reduce uncertainty through initial subsidization of user entry (Eisenmann 2007). Second, a monopoly platform can enable cross-subsidization, which can lead to superior outcomes (Emch and Thompson 2006).

In two-sided markets, demand on each side depends on prices on both sides of the market. Similarly, demand on each side depends on the value of products available (and therefore, investments) on both sides of the market. Anderson et al. (2014) analyze how platforms choose investment when investment can have a negative effect on demand on one side of the market. Although their work concerns a negative effect of platform investment on demand, we are primarily concerned about firm investments to increase demand. In an intermediated platform setting, Nault and Dexter (2006) show that cross-subsidization by the platform weakly increases investments by firms operating on the platform and that firms are rewarded for focusing on one side of the platform. They model demand on each side of the market as being dependent on investment choices made by firms on both sides of the market, and exogenously determined prices. Our work is also related to the agency model of intermediation, where the firm sets its price, and the platform takes a fee from each transaction (Geng et al. 2018; Tan and Carrillo 2014). While we confirm their findings that the fee decreases firm profits, and observe the *loss-sharing effect* when the firms sets their prices (pp612, Geng et al. 2018), our focus is on the effect of fees on firm investments, the interaction between prices and investments, and how the platform coordinates firm investments.

Given the important role that Information Technology plays to "manage and coordinate the supply and demand of energy" (Gupta 2017, "Road for the Future" Section), we use the example of Demand Response on the Smart Grid to elucidate the mechanics of an intermediated platform.

2.1.2 Our example

The Smart Grid (SG) is characterized by the use of digital information and control technology to make the electric grid reliable, efficient, and secure. The technology consists of sensors collecting ambient information, data storage solutions, real time analytics, and control systems to effect changes in the SG (Litos Communication 2010). We focus on a specific SG initiative termed Demand Response (DR), the purpose of which is to shift electricity consumption from a peak (high demand) hour to an off-peak hour through real-time pricing. The most critical advantage to DR is reduced infrastructure costs and increased capacity utilization - generation, transmission, and distribution infrastructure can then be planned for a lower level of peak demand. These form compelling reasons for the policymaker's interest in ensuring wide adoption of the SG in general and DR in particular. Their intent is evidenced by the large investments by governments worldwide (See Appendix A.1 for a description of the SG and DR).

Customer adoption of DR requires two kinds of technology: Energy Management Systems (EMSs), and DR-apps. EMSs are decision processing systems owned by customers, and their function is to make decisions about electricity consuming device usage on behalf of the customer. They control electric devices based on customer instructions and current electricity prices. The Electric Power Research Institute (EPRI) defines EMS as *"the controller, making decisions based on exigent conditions viewed in the light of a predefined instructions set"* (p 6-15, Gellings 2011). DR-apps are embedded into customer owned electricity consuming devices such as heating or cooling systems, lighting, manufacturing machinery, and electric vehicles (EVs), for example. In its simplest form, a DR-app communicates with the EMS, and can turn on and turn off the energy consuming device. We describe the mechanics of DR adoption using the example of EV fleets.

In 2017, the global stock of EVs stood at over 3.3 million, and it is projected to rise to 130 million by 2030 (International Energy Agency 2018). EV owners are important DR customers because they are substantial users of electric energy and have the ability to draw electric energy at off-peak hours and store it for future use. These electricity customers decide whether to participate in DR by purchasing an EMS and DR-apps. The combined ownership of an EMS and DR-apps gives customers the ability to respond to shocks in electricity prices - we consider a customer to be a DR adopter upon ownership of an EMS and at least one DR-app. Our example applies to the DR retrofit industry for EVs. The retrofit consists of a DR-app embedded into each EV, and an EMS that can be located at and operated from a central office. EVs begin to charge their batteries upon plug-in, but the retrofit enables vehicles to charge during off-peak hours. We focus on owners or operators of fleets of EVs as end-customers like city public transit operators and private rental fleet operators.

EMS Providers can be automobile original equipment manufacturers (OEMs), aftermarket auto manufacturers, or technology firms. They develop EMSs to enable customer programming of energy use to charge vehicle batteries, and to communicate with each of the customer's DR-apps. EMSs can link with other enterprise systems like production planning and order management systems. For instance, an EMS for an electric bus fleet can incorporate the scheduled start time for bus operations, or an EMS for rental car operators can incorporate the scheduled assignment of a car to a rental customer. Further, EMSs can differ in performance characteristics such as speed, reliability, and ease of use. Thus, the quality and functionality of EMSs can be heterogenous.

DR-app providers can also be OEMs, aftermarket auto manufacturers, or technology firms. They develop and make available DR-apps that can be retrofitted into EVs. The DR-app is bundled with an electric relay that can turn on or turn off the flow of electricity to the battery. The purpose of the DR-app is to enable communication with EMSs. DR-apps can also differ in performance and value-added features; as such their quality and functionality can also be heterogenous.

EMS and DR-app providers are firms that invest in the development of either (or both) EMSs or DR-apps. These investments increase customer demand for their products. Firms also set prices for their EMSs and DR-apps so as to maximize profits. They adhere to the standards laid out by the DR platform and make their products available on the platform.

The DR platform is the point source of sales for DR customers. It provides directory and matching services, so that customers can find and purchase compatible EMSs and DR-apps. It also provides real-time electricity pricing information, sets standards, and enables interoperability between EMSs and DR-apps, thereby ensuring a wide selection of compatible DR technology.

Most U.S. states are regulated electricity markets (Electric Choice 2018). Regulated markets are vertically integrated, and customers buy electricity from a monopoly provider.[1] Electricity providers have a relationship with customers, with ongoing communications about electricity usage, billing, and payments, putting them in a superior position to market, incentivize, and sell DR.

Customers buy a DR solution consisting of EMS and DR-apps if it is individually rational to do so. The DR system enables customers to respond to real-time prices, thereby providing cost savings - these potential cost savings equate to the value that the customer can derive from the DR system, and form the basis for their individual rationality constraint.

We take that each EMS and DR-app provider knows the shape of its demand function. Analytically, the demand for each EMS and DR-app provider can be determined by aggregating the value functions across customers. Several works econometrically estimate demand (e.g., Ata et al. 2018; Ghose and Han 2014), and in practice, firms routinely engage in demand estimation as an input to marketing decisions. In other words, key demand characteristics such as price responsiveness and concavity are known by the firm. We model an intermediated two-sided platform setting for DR where demand depends on own investments and prices as well as cross-side investments and prices. Through this model, we identify the relationship between firm prices and firm investments, and how the platform's choices impact firm prices, investments, participation and profits, customer adoption in DR, and social surplus

We set up our analysis as a two-stage model. In the first stage, the DR platform chooses both fees and transfers between EMS and DR-app providers. In the second stage, firms choose investments and prices. These prices and investments by firms determine customer demand. In this set-up, investments and prices are a function of the platform's choice of fees and transfers. Figure 2.1 depicts the structure of the setting that we analyze.

Our work provides insights that helps policymakers deliver better returns on their SG investments. Firstly, the platform's choice of transfer maximizes aggregate platform demand for DR. Thus, monopolistic ownership of the DR platform aids in coordinating prices and investments across the components of DR. Secondly, an increase in platform fees towards profit maximizing levels can lead to an increase or decrease

[1] In unregulated electricity markets, multiple providers can compete for customers. Their costs include the purchase of electricity in bulk from generators, and a fee to transmitters and distributors that is regulated. They compete with other providers on price and quality of service, and charge customers for electricity.

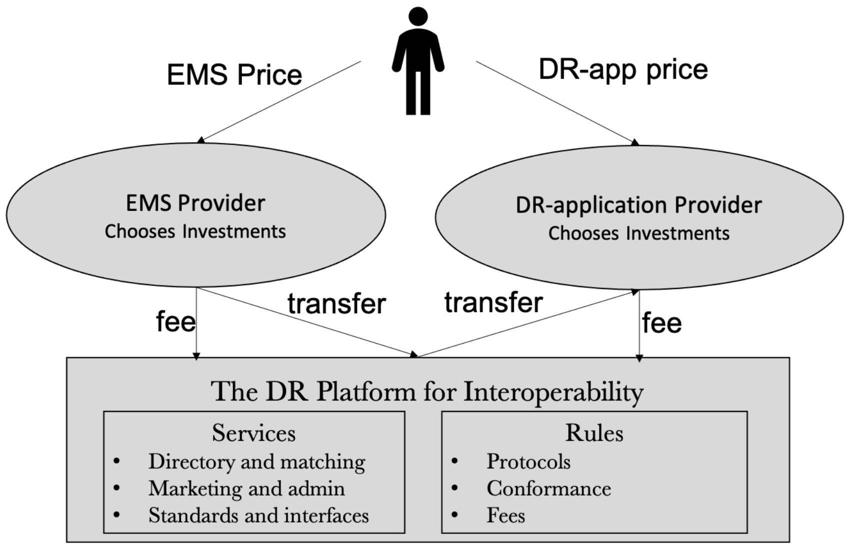


Figure 2.1: DR Platform - Structure of the Setting

in prices. If prices increase faster than fees, then it is inevitable that investments increase in fees. On the other hand, if investments decrease rapidly in fees, then it is possible that prices decrease in fees. Thus, it is unclear ex-ante if the policymaker should use Ramsey-Boiteux pricing (or mechanisms) to regulate the monopolist platform (Wikipedia 2020); our model provides insights to this question of regulation.

2.2 Structure of the Setting, Notation, and Assumptions

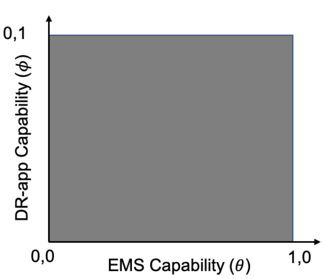


Figure 2.2: Firm Distribution

Firms vary in their capability to develop EMSs, θ, and in their capability to develop DR-apps, ϕ. Each are normalized to be in $[0, 1]$, and are uniformly distributed, $\theta, \phi \sim U[0, 1]$. A higher θ identifies a more capable EMS provider and a higher ϕ identifies a more capable DR-app provider. Thus, the tuple (θ, ϕ) identifies a particular firm in a 2-dimensional capability continuum (denoted by $\theta\phi$ in the following sections) and $\backslash(\theta\phi)$ represents other firms (Figure 2.2).

Using the superscript "E" for EMS and "A" for DR-apps, each firm sets its *own price* for its EMS, $P^E_{\theta\phi} \in [0, \bar{P^E}]$, and its *own price* for its DR-app $P^A_{\theta\phi} \in [0, \bar{P^A}]$. Because each firm $\theta\phi$ has its own choice of $P^E_{\theta\phi}$ and $P^A_{\theta\phi}$, then $P^E, P^A \in \mathbb{R}^2$ exist as matrices in the 2-dimensional space, where each $\theta\phi$ tuple indexes the price chosen by a firm. The space of prices of EMS and DR-apps provided by other firms is then $P^E_{\backslash(\theta\phi)}$ and $P^A_{\backslash(\theta\phi)}$. The platform provider collects the same fee, $f \in \mathbb{R}_{>0}$ for a unit of demand each from the EMS provider and the DR-app provider. Further, the platform enables cross-subsidization by setting a transfer from the EMS provider to the DR-app provider, $t \in \mathbb{R}$ for each unit of demand. Thus, for each unit of demand the EMS provider keeps $P^E_{\theta\phi} - f - t$ and the DR-app provider keeps $P^A_{\theta\phi} - f + t$. As is the case with most information goods, we assume that marginal costs are zero.

Each firm chooses *own investment* in EMSs, $\omega^E_{\theta\phi} \in [0, \bar{\omega^E}]$, and *own investment* in DR-apps, $\omega^A_{\theta\phi} \in [0, \bar{\omega^A}]$. Firm investments, $\omega^E_{\theta\phi}$ and $\omega^A_{\theta\phi}$, increase the value of the firm's EMS and DR-app, respectively, to all of its customers. Because each firm $\theta\phi$ has its own choice of $\omega^E_{\theta\phi}$ and $\omega^A_{\theta\phi}$, then $\omega^E, \omega^A \in \mathbb{R}^2$ exists, where each $\theta\phi$ tuple indexes the investment choice of a firm. Note that $\omega^E_{\backslash(\theta\phi)} = \int_\phi \int_\theta \omega^E_{\theta\phi} d\theta d\phi - \omega^E_{\theta\phi}$ is the cumulative investment in EMSs by the other firms participating in the platform and $\omega^A_{\backslash(\theta\phi)} = \int_\theta \int_\phi \omega^A_{\theta\phi} d\phi d\theta - \omega^A_{\theta\phi}$ is the cumulative investment in DR-apps by the other firms participating in the platform. $\Upsilon \in [0, 1]$ is the proportion of firms participating in the platform. We show how Υ is calculated later.

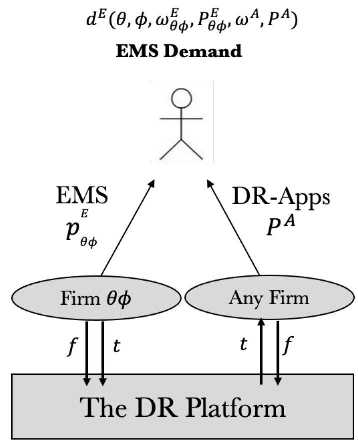


Figure 2.3: EMS Demand for Firm $\theta\phi$

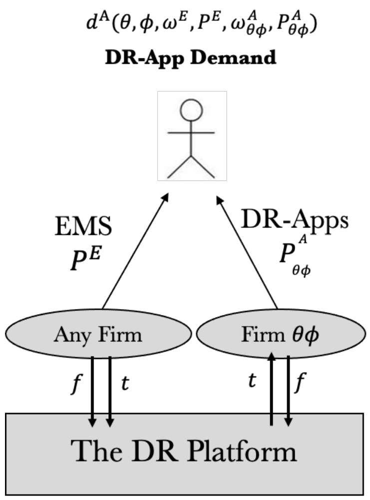


Figure 2.4: DR-App Demand for firm $\theta\phi$

Our set-up results in a two-sided platform mediated network with EMS providers on one side and DR-app providers on the other. Given each firm's EMS capability and DR-app capability, each firm generates its EMS demand and DR-app demand. We follow the standard terminology whereby for a given EMS provider, same-side refers to other EMS providers and cross-side refers to DR-app providers (Eisenmann 2007). Similarly for a given DR-app provider, where other DR-app providers are same-side and EMS providers are cross-side.

When a firm $\theta\phi$ generates EMS demand through its *own price* $P^E_{\theta\phi}$ and *own investment* $\omega^E_{\theta\phi}$, given *cross-side prices* P^A and *cross-side investments* ω^A, we denote this *EMS demand* as $d^E(\theta, \phi, P^E_{\theta\phi}, \omega^E_{\theta\phi}, P^A, \omega^A)$, or more simply as $d^E(\cdot)$ (Figure 2.3). The terms, θ, ϕ, index the firm to which the demand applies. As with DR-app demand below, the arguments to the demand generated by a firm will become clear with our assumptions.

Similarly, when a firm $\theta\phi$ generates demand for DR-apps through its *own price* $P^A_{\theta\phi}$, *own investment* $\omega^A_{\theta\phi}$, *cross-side prices* P^E, and *cross-side investments* ω^E, we denote this *DR-app demand* as $d^A(\theta, \phi, P^A_{\theta\phi}, \omega^A_{\theta\phi}, P^E, \omega^E)$, or simply $d^A(\cdot)$. Figure 2.4 provides a visual representation of a firm's DR-app demand. In Appendix A.2, we illustrate the demands using a simplified example where the market consists of two firms.

Assumption 2.1 (Capability and demand) *EMS Demand is increasing in own EMS capability. DR-app demand is increasing in own DR-app capability. EMS Demand is unaffected by own DR-app capability, and DR-app demand is unaffected by own EMS capability.*

$$\frac{\partial d^E(\cdot)}{\partial \theta}, \frac{\partial d^A(\cdot)}{\partial \phi} > 0; \qquad \frac{\partial d^E(\cdot)}{\partial \phi}, \frac{\partial d^A(\cdot)}{\partial \theta} = 0.$$

Greater firm capability implies a greater level of expertise in the technologies used or a higher level of exposure to the markets that they intend to sell to. A firm with expertise in web applications, communication protocols over networks, and human interfaces is at an advantage when manufacturing an EMS. Similarly, a firm with experience developing embedded applications and instrumentation design is able to better penetrate the DR-app market. Using our SG example, capability enables firms developing EMS and DR-apps to increase their value to fleet operators of EVs, thereby increasing demand.

Assumption 2.2 (Price and demand) *(a) Each demand decreases in own and cross-side prices:*

$$\frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}}, \frac{\partial d^E(\cdot)}{\partial P^A}, \frac{\partial d^A(\cdot)}{\partial P^A_{\theta\phi}}, \frac{\partial d^A(\cdot)}{\partial P^E} < 0.$$

(b) The firm profit function, $\pi(\theta, \phi, P^E, P^A, \omega^E, \omega^A, f, t) = \pi(\cdot)$, is concave in own price:

$$\frac{\partial^2 \pi(\cdot)}{\partial {P^E_{\theta\phi}}^2}, \frac{\partial^2 \pi(\cdot)}{\partial {P^A_{\theta\phi}}^2} < 0$$

EV fleet operators participate in DR if their willingness to pay is higher than the price. An increase in price of a given EMS or DR-app decreases demand, leading to Assumption 2.2(a).

Assumption 2.2(b) ensures a single optimum to profit maximization. This means that the demand can be concave, linear or convex, but if the demand is convex, then the extent of demand convexity is limited (see Aguirre et al. 2010). We describe the firm profit function in more detail later.

Assumption 2.3 (Investment and demand) *(a) Demand is increasing and concave in own and cross-side investment.*

$$\frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}}, \frac{\partial d^E(\cdot)}{\partial \omega^A}, \frac{\partial d^A(\cdot)}{\partial \omega^A_{\theta\phi}}, \frac{\partial d^A(\cdot)}{\partial \omega^E} > 0.$$
$$\frac{\partial^2 d^E(\cdot)}{\partial {\omega^E_{\theta\phi}}^2}, \frac{\partial^2 d^E(\cdot)}{\partial {\omega^A}^2}, \frac{\partial^2 d^A(\cdot)}{\partial {\omega^A_{\theta\phi}}^2}, \frac{\partial^2 d^A(\cdot)}{\partial {\omega^E}^2} < 0.$$

(b) Demand is zero when own or cross-side investment is zero:

$$d^E(\cdot) = 0 \;\; if \;\; [\omega^E_{\theta\phi} = 0 \; or \; \omega^A = 0], \quad d^A(\cdot) = 0 \;\; if \;\; [\omega^A_{\theta\phi} = 0 \; or \; \omega^E = 0].$$

Higher levels of investment generate higher demand. For example, investments by firms in the development of EMS and DR-apps increase their value to fleet operators of EVs, thereby increasing demand. However, the incremental impact of investment at high levels of investment is lower, when compared with the incremental impact of investment at lower levels of investment. This is because the investment is first applied to opportunities that afford the greatest return on investment. Further, no demand is generated if own or cross-side investments are zero.

Assumption 2.4 (Competition silos) *Own demand is unaffected by same-side prices and investments.*

$$\frac{\partial d^E(\cdot)}{\partial P^E_{\backslash\theta\phi}}, \; \frac{\partial d^E(\cdot)}{\partial \omega^E_{\backslash\theta\phi}}, \; \frac{\partial d^A(\cdot)}{\partial P^A_{\backslash\theta\phi}}, \; \frac{\partial d^A(\cdot)}{\partial \omega^A_{\backslash\theta\phi}} = 0.$$

Specifically, a firm's EMS demand is not affected by another firm's prices of or investment in EMS, and a firm's DR-app demand is not affected by another firm's prices of or investment in DR-apps.

In the presence of differentiated products (the absence of perfect substitutes), own effects on own demand are strictly larger in magnitude than same-side effects. Thus, the impact of same-side prices and investments on own demand is weakly smaller in magnitude than the impact of own prices and investments. Given the nature of differentiated providers in our setting, we abstract away these weak same-side effects. This assumption is easily satisfied in the early days of a platform when there are few competitors and several geographies and industries, each of which forms a localized market. The effect of this assumption is to simplify our analysis, but there is no loss of generality compared to assuming weak same-side effects.

Assumption 2.5 (Independence of prices and investments) *Prices and investments have independent effects on demand.*

$$\frac{\partial^2 d^q(\cdot)}{\partial P^r_{ij} \partial \omega^s_{kl}} = 0 \quad for \quad ij, kl \in [\theta\phi], \quad and \quad q, r, s \in [E, A].$$

Investments in products have two effects (Johnson and Myatt 2006, Dixit and Norman 1978). When investments improve the quality of a product for all customers, the demand curve shifts outward, and is represented in the case of the effect of own price and investments on EMS demand as $\partial d_E(\cdot)/\partial \omega^E_{\theta\phi} > 0$ (as in Assumption 2.3). When investments have a differing impact on customers, the demand curve rotates, and the extent of the rotation is $\partial d^2_E(\cdot)/[\partial \omega^E_{\theta\phi} \partial P^E_{\theta\phi}]$. We assume that investments do not have a differing impact on customers.

2.3 Our Model

As we described in the Introduction, we model this setting as a 2-stage game. In Stage 1 the platform sets the fees and transfer. In Stage 2 firms set their investments and prices for EMSs and DR-apps. Finally, the platform and the firms realize their profits as customers adopt. We work backwards through the stages to obtain our results by first solving the firm's problem and then solving the platform's problem. The stages of the game are depicted in Figure 2.5.

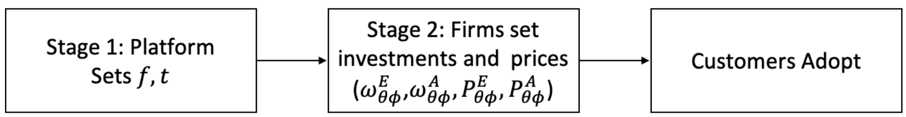


Figure 2.5: Stages of the Game

Firms profit maximize by choice of their investments and prices,

$$\max_{\omega^E_{\theta\phi}, \omega^A_{\theta\phi}, P^E_{\theta\phi}, P^A_{\theta\phi}} \pi(\cdot) = [P^E_{\theta\phi} - f - t]d^E(\cdot) + [P^A_{\theta\phi} - f + t]d^A(\cdot) - \omega^E_{\theta\phi} - \omega^A_{\theta\phi}. \tag{2.1}$$

The first term is profits generated from EMS demand from which the platform and the DR-app provider take f and t, respectively. The second term is profits generated from DR-app demand, from which the platform takes f and gives t. The third and fourth terms are the costs of the firm's investments in EMS and DR-apps.

2.3.1 Stage 2: The Firm's Choices - Investments and Prices

When a firm maximizes profits by choosing EMS investment, its first-order condition is

$$\frac{\partial \pi(\cdot)}{\partial \omega^E_{\theta\phi}} = [P^E_{\theta\phi} - f - t]\frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}} + [P^A_{\theta\phi} - f + t]\frac{\partial d^A(\cdot)}{\partial \omega^E_{\theta\phi}} - 1 = 0 = \psi(\cdot). \tag{2.2}$$

The first and second terms in (2.2) are positive from Assumption 2.3(a) and represents incremental profits from EMS demand and DR-app demand due to increased investment. Setting the first-order condition in (2.2) to zero, $\psi(\cdot) = 0$ implicitly defines an optimal value of EMS investments as a function of the other variables. The second derivative of profits with respect to EMS investment is

$$\frac{\partial^2 \pi(\cdot)}{\partial {\omega^E_{\theta\phi}}^2} = [P^E_{\theta\phi} - f - t]\frac{\partial^2 d^E(\cdot)}{\partial {\omega^E_{\theta\phi}}^2} + [P^A_{\theta\phi} - f + t]\frac{\partial^2 d^A(\cdot)}{\partial {\omega^E_{\theta\phi}}^2} = \frac{\partial \psi(\cdot)}{\partial \omega^E_{\theta\phi}}. \tag{2.3}$$

Each of the terms above are negative, and thus, $\partial^2\pi(\cdot)/\partial{\omega^E_{\theta\phi}}^2 = \partial\psi(\cdot)/\partial\omega^E_{\theta\phi} < 0$. By symmetry, a similar set of conditions obtain when each firm chooses its DR-app investment, $\omega^A_{\theta\phi}$.

Next, the first-order condition when choosing $P^E_{\theta\phi}$ is chosen by setting the first derivative of (2.1) with respect to prices equal to zero,

$$\frac{\partial\pi(\cdot)}{\partial P^E_{\theta\phi}} = [P^E_{\theta\phi} - f - t]\frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}} + d_E(\cdot) + [P^A_{\theta\phi} - f + t]\frac{\partial d^A(\cdot)}{\partial P^E_{\theta\phi}} = 0 = \Lambda(\cdot), \tag{2.4}$$

where $\Lambda(\cdot) = 0$ implicitly defines price as a function of the other variables. The first and third terms in (2.4) represent the loss on the margin from falling demand due to increased prices. These marginal losses are attenuated by the externalities - fees and transfer. The second term is the gain on the infra-margin due to increased prices. The profit maximizing price equates these inframarginal gains with the attenuated marginal losses in the first and third terms.

The second-order condition choosing EMS prices is

$$\frac{\partial^2\pi(\cdot)}{\partial {P^E_{\theta\phi}}^2} = [P^E_{\theta\phi} - f - t]\frac{\partial^2 d^E(\cdot)}{\partial {P^E_{\theta\phi}}^2} + \frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}} + \frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}} + [P^A_{\theta\phi} - f + t]\frac{\partial^2 d^A(\cdot)}{\partial {P^E_{\theta\phi}}^2}. \tag{2.5}$$

For a downward sloping demand curve, a price increase has three effects on marginal profit. The first is the marginal loss effect and is represented by the second term. This effect occurs because marginal losses increase at a higher price, and is signed negative by Assumption 2.2(a). If demand is not linear, then the marginal loss effect needs to be adjusted for the extent of demand convexity or concavity. This non-linearity effect is captured by the first and fourth terms - these terms are positive if demand is convex and negative if demand is concave. The third is the inframarginal effect and is captured by the third term. This effect is negative and occurs because at a higher price, the infra-marginal gain from a unit increase in price is lower because the demand curve is sloping downward. Note that the marginal loss effect and the inframarginal effect by choice of $P^E_{\theta\phi}$ do not apply to $d^A(\cdot)$, the cross side demand. Our Assumption 2.2(b) effectively means that we take the sum of each of the above effects as negative, and the second order condition is satisfied so that $\partial^2\pi(\cdot)/\partial {P^E_{\theta\phi}}^2 < 0$. Thus, at the optimum, $\partial\Lambda(\cdot)/\partial P^E_{\theta\phi} = \partial^2\pi(\cdot)/{P^E_{\theta\phi}}^2 < 0$. A similar set of conditions is obtained for firms setting DR-app prices, $P^A_{\theta\phi}$.

With investments and prices compact, and profits continuous in both investments and prices together with concavity, the fulfillment of the first-order conditions for every firm participating in the platform represents a Nash equilibrium. In other words, the simultaneous solution of the first order conditions for EMS and DR-app investments and prices $\forall\ \theta\phi$ gives a Nash equilibrium in investments and prices, $\omega^E(f,t)$, $\omega^A(f,t)$, $P^E(f,t)$ and $P^A(f,t)$. These are each two-dimensional spaces of EMS investments, DR-app investments, EMS prices and DR-app prices choicen by the firms. We denote individual investments made by each firm more simply as $\omega^E_{\theta\phi}(\cdot)$ and $\omega^A_{\theta\phi}(\cdot)$, and the individual prices set by each firm as $P^E_{\theta\phi}(\cdot)$ and $P^A_{\theta\phi}(\cdot)$. If more than one Nash equilibrium exists, then we assume that the one with the highest investment levels is obtained.

2.3.2 Stage 1: Platform transfer and fee

In order to address the platform's problem of choosing the transfer and fee, we define the firm that is indifferent between participating in the platform and not, the proportion of firms that participate, and aggregate demands.

Firm profits in the platform is (2.1) but now with prices as well as investments as functions. Note that the 2-dimentional space of equilibrium prices, $P^E(f,t)$ and $P^A(f,t)$ contains $P^E_{\theta\phi}(\cdot)$ and $P^A_{\theta\phi}(\cdot)$:

$$\pi(\cdot) = [P^E_{\theta\phi}(\cdot) - f - t]d^E(\cdot) + [P^A_{\theta\phi}(\cdot) - f + t]d^A(\cdot) - \omega^E_{\theta\phi}(\cdot) - \omega^A_{\theta\phi}(\cdot). \tag{2.6}$$

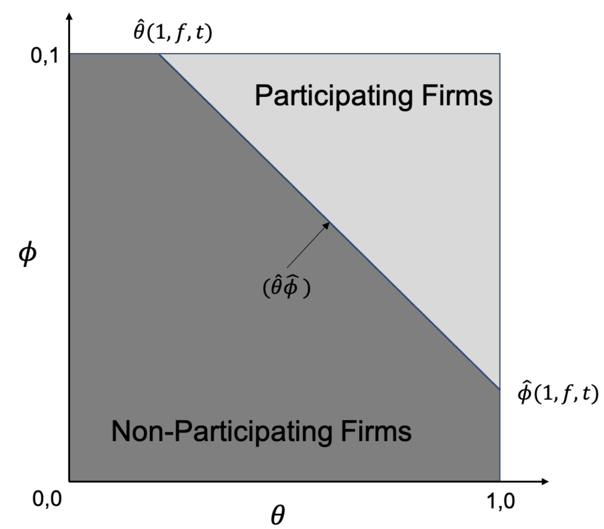


Figure 2.6: Firm participation and non-participation: An example

A proportion of firms are profitable and hence participate in the DR market, and the rest are not profitable and do not participate. When firm capability is distributed along a line, the indifferent, or marginal, firm is a point. When firm capability is distributed along two dimensions, the set of marginal firms is represented by the marginal firm curve $\hat{\theta}\hat{\phi}$ (Figure 2.6). This curve is derived by identifying the set of marginal firms that generate zero profits. Therefore, $\pi(\hat{\theta},\hat{\phi},f,t) = 0$ in (2.6) implicitly defines $\hat{\theta}(\hat{\phi},f,t)$ and $\hat{\phi}(\hat{\theta},f,t)$. In other words, for a given f and t, each of the functions $\hat{\theta}(\hat{\phi},f,t)$ and $\hat{\phi}(\hat{\theta},f,t)$ represent the same set of marginal firms.

An example of a marginal firm curve is shown in Figure 2.6. Here, $\hat{\phi}(1,f,t)$ represents the DR-app capability of the marginal firm with an EMS capability value of unity and given a f and t. Similarly, $\hat{\theta}(1, f, t)$ represents the EMS capability of the marginal firm with a DR-app capability value of unity and the given f and t.

Lemma 2.1 *Participating firms are more capable than the marginal firm.*

Proof: The partial derivative of the profit function of the marginal firm defined in (2.6) with respect to $\hat{\theta}$, after applying the envelope theorem is

$$\frac{\partial \pi(\cdot)}{\partial \hat{\theta}} = [P^E_{\theta\phi}(\cdot) - f - t]\frac{\partial d^E(\cdot)}{\partial \hat{\theta}}.$$

The above equation represents the change in profits of a firm that has a slightly higher EMS capability than the marginal firm, $\partial\pi(\cdot)/\partial\hat{\theta}$. Because EMS demand increases in EMS capability, the above equation is positive, and $\partial\pi(\cdot)/\partial\hat{\theta} > 0$. Similarly, $\partial\pi(\cdot)/\partial\hat{\phi} > 0$. □

From this Lemma, any increase in capability along the θ axis or the ϕ axis moves the firm from a marginal firm to a participating firm, the marginal firm curve is negatively sloped so that $\partial\hat{\theta}(\hat{\phi}, f, t)/\partial\hat{\phi} < 0$. Thus, the marginal firm defined by $\pi(\hat{\theta}, \hat{\phi}, f, t)$ makes zero profits, and more capable firms make positive profits. Firms that are less capable than a marginal firm either in EMS capability or DR-app capability can at best make negative profits if they participate, and hence choose not to participate.

We can now define the proportion of participating firms. This proportion, $\Upsilon(f, t)$, is calculated as

$$\Upsilon(f, t) = \int_{\hat{\phi}(1,f,t)}^{1} \int_{\hat{\theta}(\phi,f,t)}^{1} d\theta d\phi. \tag{2.7}$$

The above equation identifies the case where the marginal firm curve intersects the 2-dimensional firm capability continuum at $\theta = 1$ and $\phi = 1$, as shown in Figure 2.6.[2] At the platform level, the aggregate EMS and DR-app demands are, respectively,

$$D^i(\cdot) = \int_{\hat{\phi}(1,f,t)}^{1} \int_{\hat{\theta}(\phi,f,t)}^{1} d^i(\cdot) d\theta d\phi, \quad i \in [E, A]. \tag{2.8}$$

Platform profits are fees collected from both sides of the market times aggregate platform-matched

[2]There are three other cases based on where the marginal firm curve intersects the 2-dimensional firm capability continuum and these are described in Appendix A.3. For each of the cases, the proportion of participating firms is given by the area to the top and right of the marginal firm curve $\theta\phi$ (Figure 2.6). The terms under the integration sign differ by case, but share the common property that the proportion of participating firms decreases in the marginal firm capability, $\partial\Upsilon(f, t)/\partial\hat{\theta} < 0$ and $\partial\Upsilon(f, t)/\partial\hat{\phi} < 0$, so our results are identical across the cases.

demand, $D^E(\cdot) + D^A(\cdot)$. The platform maximizes profits by choosing transfers and fees,

$$\max_{f,t} \Pi(f,t) = \max_{f,t} f[D^E(\cdot) + D^A(\cdot)]. \tag{2.9}$$

The necessary first-order condition for an interior profit maximum by choice of fee is

$$\frac{\partial \Pi(\cdot)}{\partial f} = f[\frac{\partial D^E(\cdot)}{\partial f} + \frac{\partial D^A(\cdot)}{\partial f}] + D^E(\cdot) + D^A(\cdot) = 0, \tag{2.10}$$

and by choice of transfer is

$$\frac{\partial \Pi(\cdot)}{\partial t} = f[\frac{\partial D^E(\cdot)}{\partial t} + \frac{\partial D^A(\cdot)}{\partial t}] = 0, \tag{2.11}$$

where for both the fee and transfer first-order conditions the envelope theorem cancels the indirect effects through investments and prices. The second-order conditions require that $f[\partial^2 D^E(\cdot)/\partial f^2 + \partial^2 D^A(\cdot)/\partial f^2] \leqslant -2\partial D^E(\cdot)/\partial f - 2\partial D^A(\cdot)/\partial f$, and $\partial^2 D^E(\cdot)/\partial t^2 + \partial^2 D^A(\cdot)/\partial t^2 \leqslant 0$, which are reasonable and we take to hold.

2.4 Our Main Results

2.4.1 Prices

We examine the impact of fees on firm prices, and isolate conditions that determine the rate of price increase with fees. We begin by differentiating (2.4) with respect to fees, and using the envelope theorem,

$$\frac{\partial \Lambda(\cdot)}{\partial f} = -\frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}} - \frac{\partial d^A(\cdot)}{\partial P^E_{\theta\phi}} + \frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}}\frac{\partial \omega^E_{\theta\phi}(\cdot)}{\partial f} + \frac{\partial d^E(\cdot)}{\partial \omega^A}\frac{\partial \omega^A(\cdot)}{\partial f} +$$

$$[P^E_{\theta\phi} - f - t][\frac{\partial^2 d^E(\cdot)}{\partial \omega^E_{\theta\phi}\partial P^E_{\theta\phi}}\frac{\partial \omega^E_{\theta\phi}(\cdot)}{\partial f} + \frac{\partial^2 d^E(\cdot)}{\partial \omega^A \partial P^E_{\theta\phi}}\frac{\partial \omega^A(\cdot)}{\partial f}] +$$

$$[P^A_{\theta\phi} - f + t][\frac{\partial^2 d^A(\cdot)}{\partial \omega^E_{\theta\phi}\partial P^E_{\theta\phi}}\frac{\partial \omega^E_{\theta\phi}(\cdot)}{\partial f} + \frac{\partial^2 d^A(\cdot)}{\partial \omega^A \partial P^E_{\theta\phi}}\frac{\partial \omega^A(\cdot)}{\partial f}].$$

The fourth and fifth terms (second and third lines in the above equation) drop out when we assume the independence of prices and investments (Assumption 2.5). This leaves us with

$$\frac{\partial \Lambda(\cdot)}{\partial f} = -\frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}} - \frac{\partial d^A(\cdot)}{\partial P^E_{\theta\phi}} + \frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}}\frac{\partial \omega^E_{\theta\phi}(\cdot)}{\partial f} + \frac{\partial d^E(\cdot)}{\partial \omega^A}\frac{\partial \omega^A(\cdot)}{\partial f}. \tag{2.12}$$

The first two terms are positive by Assumption 2.2(a). At a higher level of fees, the marginal losses from an increase in price are attenuated. This is because price net of fees represents the proportion of marginal losses that are internalized, and with higher fees, a smaller proportion of the marginal losses are internalized. The first term captures this effect as it relates to own prices, while the second term captures this effect as it pertains to cross-side prices. The third and fourth terms represent the indirect effect of fees through investments. The third term represents this effect as it pertains to own investments, $\omega^E_{\theta\phi}$, and the fourth term represents this effect as it pertains to cross side investments, ω^A. This leads to our first proposition where we obtain a counter-intuitive result whereby prices can decrease in fees if there is a strong negative investment externality.

Proposition 2.1 *In the presence of a negative investment externality, prices can decrease with fees.*

Observing (2.12), for $\partial\Lambda(\cdot)/\partial f < 0$, we require that

$$\frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}}\frac{\partial \omega^E_{\theta\phi}}{\partial f} + \frac{\partial d^E(\cdot)}{\partial \omega^A}\frac{\partial \omega^A}{\partial f} < \frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}} + \frac{\partial d^A(\cdot)}{\partial P^E_{\theta\phi}}. \tag{2.13}$$

From the implicit function theorem we have

$$\frac{\partial P^E_{\theta\phi}(\cdot)}{\partial f} = -\frac{\partial\Lambda(\cdot)/\partial f}{\partial\Lambda(\cdot)/\partial P^E_{\theta\phi}} \tag{2.14}$$

The numerator is negative if the condition in (2.13) is satisfied. The denominator is negative from the second-order condition in (2.5) and its subsequent discussion, and $\partial P^E_{\theta\phi}/\partial f < 0$. By following a similar set of steps, the condition for $\partial P^A_{\theta\phi}/\partial f < 0$ can be derived. □

We define P-Prices as the phenomenon where prices decrease in fees. Fees represent a constant marginal cost for firms that engage in DR-app and EMS provision while operating on the platform, and although an increase in fees can be expected to lead to an increase in prices, the investment externality alters the impact of fees on prices. In particular, a strong negative investment externality can negate a rise in price that is expected with an increase in fees.

Even though several prior works assume that price subsidies offered by platforms are passed on directly to customers, Proposition 2.1 shows that firms can face demand conditions under which they do not pass the price subsidies to end-customers. Our next proposition explores conditions under which prices increase faster than an increase in fees.

Proposition 2.2 *The rate of price increase with fees can be greater than unity.*

Proof: Substituting (2.5) and (2.12) into (2.14), we get

$$\frac{\partial P^E_{\theta\phi}}{\partial f} = -\frac{-\partial d^E(\cdot)/\partial P^E_{\theta\phi} - \partial d^A(\cdot)/\partial P^E_{\theta\phi} + [\partial d^E(\cdot)/\partial \omega^E_{\theta\phi}][\partial \omega^E_{\theta\phi}(\cdot)\partial f] + [\partial d^E(\cdot)\partial \omega^A][\partial \omega^A(\cdot)/\partial f]}{[P^E_{\theta\phi} - f - t]\partial^2 d^E(\cdot)/\partial {P^E_{\theta\phi}}^2 + 2\partial d^E(\cdot)/\partial P^E_{\theta\phi} + [P^A_{\theta\phi} - f + t]\partial^2 d^A(\cdot)/\partial {P^E_{\theta\phi}}^2}.$$

The rate of price increase with fees depends on the response of each demand to prices, the indirect effect of fees through investments, and the concavity of demand faced by a firm. By setting the RHS of the above equation to be greater than unity, the condition for the rate of price increase with fees to be greater than unity can be expressed as

$$-\frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}} + \frac{\partial d^A(\cdot)}{\partial P^E_{\theta\phi}} - \frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}}\frac{\partial \omega^E_{\theta\phi}(\cdot)}{\partial f} - \frac{\partial d^E(\cdot)}{\partial \omega^A}\frac{\partial \omega^A(\cdot)}{\partial f} < [P^E_{\theta\phi} - f - t]\frac{\partial^2 d^E(\cdot)}{\partial {P^E_{\theta\phi}}^2} + [P^E_{\theta\phi} - f + t]\frac{\partial^2 d^A(\cdot)}{\partial {P^E_{\theta\phi}}^2}.$$

If the relationship described in the above equation holds, then the rate of price increase with fees is greater than unity. □

We define *P-Convexity* as the region along a demand curve where the rate of price increase with fees is greater than unity, and P-Convexity is observed when the inequality in Proposition 2.2 is satisfied. This inequality defines a relationship between the price and investment responsiveness of demand, and its convexity. Observing the inequality in Proposition 2.2, when a firm's demand is not responsive to own-price, highly responsive to cross-price, does not have strong negative indirect effects through investments, or is convex in prices, then it can be faced with demand that is P-Convex.

Proposition 2.2 depends on the relationship between the price responsiveness of demand and demand convexity. Demand responsiveness to own-price refers to the term $\partial d^E(\cdot)/\partial P^E_{\theta\phi}$ in the inequality in Proposition 2.2. Demand is non-responsive to own price when customers continue to purchase despite a price increase. In the case of EMS or DR-apps, this can happen if the cost savings from engaging in DR are large for most customers. Demand responsiveness to cross-price refers to the term $\partial d^E(\cdot)/\partial P^A_{\theta\phi}$ in the inequality in Proposition 2.2, the specific cross-price responsiveness of EMS demand to own DR-app prices. This cross-price responsiveness of demand can increase as the number of compatible DR-apps decrease. In an extreme case, if the focal firm $\theta\phi$ is the only producer of a DR-app that is compatible with its own EMS, then the cross-price responsiveness $\partial d^E(\cdot)/\partial P^A_{\theta\phi}$ can be higher. Note that we qualify demand as convex if the second derivative of demand with respect to price is positive, and as well that indirect effects through investments occur as a result of the effect of fees on investments, which in turn impact demand.

The intuition behind P-Convexity can be discerned by observing how marginal losses and infra-marginal gains are equated through the choice of EMS price in a setting where demand is linear in price. An increase in EMS price leads to a marginal loss and an infra-marginal gain on EMS demand. With downward sloping demand, this marginal loss increases and the infra-marginal gain decreases with price. On the other hand, raising EMS price produces no infra-marginal gains on DR-app demand because the gains from the DR-app sales are given as $[P^A_{\theta\phi}(\cdot) - f + t]d^A(\cdot)$. Further, marginal losses on DR-app demand with changes in EMS prices remain constant if demand is linear. Thus, as EMS prices increase, marginal losses on DR-app demand are constant, marginal losses on EMS demand increase, and infra-marginal gains on EMS demand decrease. Noting that other effects are added if demand is not linear, EMS prices are increased until these gains and losses are equal.

An increase in fees decreases the internalized portion of marginal losses on each of EMS demand and DR-app demand from an EMS price increase, and the equilibrium price increases so that the infra-marginal gains equal the internalized portion of the marginal losses. Because the internalized portion of the marginal losses are smaller than the total losses, this phenomenon is termed the *loss-sharing effect* by Geng et al. 2018. Infra-marginal gains decrease (increase) in the presence of negative (positive) investment externalities, and the price increase is tempered (accelerated).

2.4.2 Investment

To assess the impact of an increase in fees on firm investment, we differentiate the firm's first-order condition when it chooses its EMS investment in (2.2) with respect to fees, and apply the envelope theorem to get

$$\begin{aligned}\frac{\partial\psi(\cdot)}{\partial f} = [P^E_{\theta\phi} - f - t][\frac{\partial^2 d^E(\cdot)}{\partial\omega^E_{\theta\phi}\partial P^E_{\theta\phi}}.\frac{\partial P^E_{\theta\phi}}{\partial f} + \frac{\partial^2 d^E(\cdot)}{\partial\omega^E_{\theta\phi}\partial P^A}.\frac{\partial P^A}{\partial f}] + \frac{\partial d^E(\cdot)}{\partial\omega^E_{\theta\phi}}[\frac{\partial P^E_{\theta\phi}}{\partial f} - 1] \\ +[P^E_{\theta\phi} - f + t][\frac{\partial^2 d^A(\cdot)}{\partial\omega^E_{\theta\phi}\partial P^E}.\frac{\partial P^E}{\partial f} + \frac{\partial^2 d^A(\cdot)}{\partial\omega^E_{\theta\phi}\partial P^A_{\theta\phi}}.\frac{\partial P^A_{\theta\phi}}{\partial f}] + \frac{\partial d^A(\cdot)}{\partial\omega^E_{\theta\phi}}[\frac{\partial P^A_{\theta\phi}}{\partial f} - 1].\end{aligned}$$

The above equation enumerates each of the effects of an increase in fees on the firm's EMS investment first order condition. The first line in the above equation enumerates these effects as it relates to the firm's EMS demand, and the second line enumerates these effects as it relates to the firm's DR-app demand. The first term in the first line and the first term in the second line drop out when we assume that own investments do not change the effects of own price on demand, our Assumption 2.5. What we are left with then is

$$\frac{\partial\psi(\cdot)}{\partial f} = \frac{\partial d^E(\cdot)}{\partial\omega^E_{\theta\phi}}[\frac{\partial P^E_{\theta\phi}}{\partial f} - 1] + \frac{\partial d^A(\cdot)}{\partial\omega^E_{\theta\phi}}[\frac{\partial P^A_{\theta\phi}}{\partial f} - 1]. \tag{2.15}$$

The first term here specifies how fees increase EMS prices, thereby improving the return on marginal EMS demand generated through EMS investments, net of the fee externality. The second term specifies how fees increase DR-app prices, thereby improving the return on marginal DR-app demand generated through

EMS investments (a cross-side effect), again net of the fee externality. Each of the two terms in the above equation can be positive or negative depending on the condition in Proposition 2.2. Our first theorem provides a counter-intuitive case of the relationship between investments and fees.

Theorem 2.1 *If EMS and DR-app demand are P-Convex, then investments increase in fees.*

Proof: In (2.15), demand increases in investment from Assumption 2.3. From Proposition 2.2, if both EMS demand and DR-app demand are P-Convex, then the rate of price increase with fees is greater than unity, making $\partial\psi(\cdot)/\partial f > 0$. By the implicit function theorem,

$$\frac{\partial \omega^E_{\theta\phi}}{\partial f} = -\frac{\partial\psi(\cdot)/\partial f}{\partial\psi(\cdot)/\partial\omega^E_{\theta\phi}} > 0.$$

The denominator is negative from (2.3), and as we saw from the discussion above, the numerator is positive, giving us that $\partial\omega^E_{\theta\phi}/\partial f > 0$. Through a similar set of steps it can be shown that P-Convexity implies that DR-app investment increases in fees, $\partial\omega^A_{\theta\phi}/\partial f > 0$. $\square$

We term the phenomenon of increasing investments with increases in fees as *P-Investments*. The intuition for Theorem 2.1 is that when prices increase at a faster rate than fees, marginal demand generated through investments by DR-app and EMS providers are better rewarded. It is worth observing that if EMS demand is not P-Convex, then $0 < \partial P^E_{\theta\phi}/\partial f \leqslant 1$, and if DR-app demand is not P-Convex, then $0 < \partial P^A_{\theta\phi}/\partial f \leqslant 1$. Thus, in the RHS of (2.15), either or both terms can be negative, and fees can have a positive, negative, or no effect on investments depending on the relative magnitudes of the first and second terms. In the special case where the first and second terms are equal and opposite in magnitude, we have the following corollary.

Corollary 2.1 *If EMS or DR-app demand is P-Convex, then investments can be unaffected by fees.*

Proof: By the implicit function theorem, the condition for lack of change in investments with fees is

$$\frac{\partial \omega^E_{\theta\phi}}{\partial f} = -\frac{\partial\psi(\cdot)/\partial f}{\partial\psi(\cdot)/\partial\omega^E_{\theta\phi}} = 0.$$

The numerator is zero if (2.15) is zero, which occurs when the first and second terms in (2.15) are equal and opposite in magnitude. The denominator is negative from (2.3), making $\partial\omega^E_{\theta\phi}/\partial f = 0$. $\square$

Holding other terms constant, if $\partial P^E_{\theta\phi}/\partial f$ is smaller than its value in setting (2.15) to zero, then firm EMS investments decrease with fees. If $\partial P^E_{\theta\phi}/\partial f$ is greater than its value in setting (2.15) to zero, then EMS

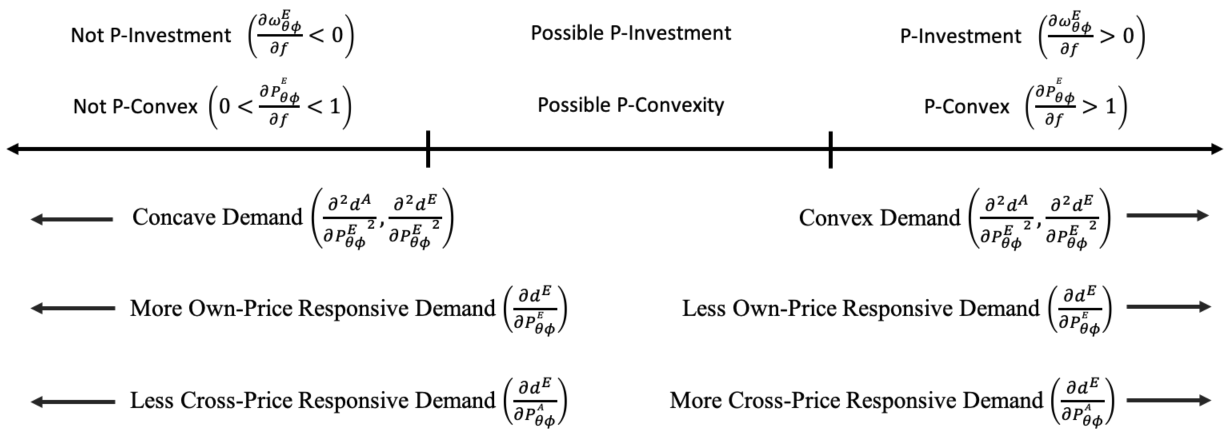


Figure 2.7: Demand convexity and price response, and the effect of fees on firm prices and investments

investments increase with fees. For EMS, P-investments occur when

$$\frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}}[\frac{\partial P^E_{\theta\phi}}{\partial f} - 1] + \frac{\partial d^A(\cdot)}{\partial \omega^E_{\theta\phi}}[\frac{\partial P^A_{\theta\phi}}{\partial f} - 1] > 0. \tag{2.16}$$

To summarize, fees are an externality because they decrease the return on marginal demand generated through investments. However, if demand is P-Convex, then with a fee increase a firm experiences a higher return on marginal demand generated through investments despite the fee externality. This causes an increase in firm investments. If demand is not P-Convex, then with a fee increase, a firm experiences a lower return on marginal demand generated through investments net of the fee externality. This places downward pressure on firm investments. If the rate of increase in EMS and DR-app prices with an increase in fees equals unity, then firm investments are indifferent to fee increases.

Figure 2.7 encapsulates these conditions. The right-hand side of the figure represents firm demand that is more convex, less responsive in own prices and more responsive in cross-prices. The left-hand side of the figure represents firm demand that is more concave, more responsive in own-price and less responsive in cross-price. The right side represents conditions that are favourable for the presence of P-Convexity and P-Investments. If demand for both EMS and DR-apps is P-Convex, then P-Investments (increasing investments from an increase in fees) are inevitable as seen in the right-hand side of the figure. P-investments can also occur if one of EMS demand or DR-app demand is P-Convex - a necessary condition for P-investment is that one of EMS or DR-app demand is P-Convex. On the left, conditions are not favorable for either of these phenomena.

2.4.3 Fees and Transfers

We now evaluate the effect of platform fees and transfers on demand and aggregate demand.

Lemma 2.2 *Fees decrease firm profits and firm participation.*

Proof: The profit function $\pi(\cdot)$ is defined in (2.6). Consider the effect of an increase in fees on the profits of the firm holding investments and prices constant,

$$\frac{\partial \pi(\cdot)}{\partial f} = -d^E(\cdot) - d^A(\cdot). \tag{2.17}$$

The firm first order conditions when choosing investments and prices eliminates the indirect effect of fees through investments and prices, and we are left with the two terms signifying a reduction in profits from current sales because of an increase in fees. Signing the above two terms confirms that $\partial\pi(\cdot)/\partial f < 0$. □

The above holds for the marginal firm as well, whose profit dips below zero with an increase in fees, causing this firm to exit the platform. This reduces the proportion of participating firms in the platform. The new marginal firm is a more capable one within the firm capability continuum, as more capable firms are more profitable than the marginal firm (Lemma 2.1). In other words, the minimum capability required in order to be profitable increases with an increase in fees. This implies that the proportion of firms participating in the platform decreases as fees increase because the proportion of participating firms decreases in the capability of the marginal firm $(\partial\Upsilon(f,t)/\partial\hat{\theta}) < 0$.

Proposition 2.3 *In the absence of P-Prices and P-Investments, (a) firm demand decreases in fees, and (b) aggregate demand decreases in fees.*

Proof: (a) The derivative of EMS demand with respect to fees is

$$\frac{\partial d^E(\cdot)}{\partial f} = \overset{(-)}{\frac{\partial d^E(\cdot)}{\partial P^E_{\theta\phi}}}\overset{(+)}{\frac{\partial P^E_{\theta\phi}}{\partial f}} + \overset{(+)}{\frac{\partial d^E(\cdot)}{\partial \omega^E_{\theta\phi}}}\overset{(-)}{\frac{\partial \omega^E_{\theta\phi}}{\partial f}} + \overset{(-)}{\frac{\partial d^E(\cdot)}{\partial P^A}}\overset{(+)}{\frac{\partial P^A}{\partial f}} + \overset{(+)}{\frac{\partial d^E(\cdot)}{\partial \omega^A}}\overset{(-)}{\frac{\partial \omega^A}{\partial f}}$$

In the absence of P-Prices and P-Investments, prices increase and investments decrease in fees. Because demand decreases in price and increases in investments (Assumptions 2.2 and 2.3), all the terms above can be signed negative, and $\partial d^E(\cdot)/\partial f < 0$. By symmetry, $\partial d^A(\cdot)/\partial f < 0$.

(b) To evaluate the effect of fees on aggregate EMS demand, we differentiate (2.8) with respect to fees.

We use Leibnitz's rule to expand this to get

$$\frac{\partial D^E(\cdot)}{\partial f} = \int_{\hat{\phi}(1,f,t)}^{1} \int_{\hat{\theta}(\phi,f,t)}^{1} \frac{\partial d^E(\cdot)}{\partial f} d\theta d\phi - \int_{\hat{\phi}(1,f,t)}^{1} d^E(\cdot) \frac{\partial \hat{\theta}(\cdot)}{\partial f} d\phi,$$

where $d^E(\cdot)$ in the second term represents the demand of the marginal firm with EMS capability $\hat{\theta}(\phi, f, t)$. The second term captures a marginal effect - fees decrease firm participation, leading to a loss of demand generated by firms that cease to participate due to the fee increase. The first term captures an inframarginal effect - the net change in demand generated by participating firms, which is negative in the absence of P-Investments from Proposition 2.3(a), and $\partial D^E(\cdot)/\partial f < 0$. Similarly, $\partial D^A(\cdot)/\partial f < 0$. $\square$

Without platform-specific information about the presence of P-effects, it is apriori unclear how fees affect aggregate demand when investments are considered: a result that affirms the importance of determining the specifics of P-effects within each platform before regulating it. In addition, because customer surplus increases in aggregate demand, the effect of fees on customer surplus is also indeterminate apriori. We now evaluate the impact of the platform's choice of transfer on aggregate demand in our second theorem.

Theorem 2.2 *The profit-maximizing choice of transfer also maximizes aggregate demand.*

Proof: The value of t at which aggregate demand is maximized is

$$\arg\max_t \, [D^E(\cdot) + D^A(\cdot)] = \arg\max_t \int_{\hat{\phi}(\cdot)}^{1} \int_{\hat{\theta}(\cdot)}^{1} [d^E(\cdot) + d^A(\cdot)] d\theta d\phi.$$

Platform profits are fees times aggregate demand. The value of t at which platform profits are maximized is

$$\arg\max_t \; f[D^E(\cdot) + D^A(\cdot)] = f \int_{\hat{\phi}(\cdot)}^{1} \int_{\hat{\theta}(\cdot)}^{1} [d^E(\cdot) + d^A(\cdot)] d\theta d\phi.$$

Because the transfer does not depend on the fine, the choice of t is the same in each of the above equations. Thus, the profit maximizing choice of transfer also maximizes aggregate demand. $\square$

This important result confirms that in the context of transfers, the platform and the policymaker are aligned - the choice of transfer that maximizes platform profits also maximizes aggregate demand. In this way the policymaker's objective of maximizing the penetration of DR technology is achieved.

We now explore the impact of the platform's choice of fees on firm surplus (FS),

$$FS = \int_{\hat{\phi}(1,f,t)}^{1} \int_{\hat{\theta}(\phi,f,t)}^{1} \pi(\cdot) d\theta d\phi. \tag{2.18}$$

Proposition 2.4 *Firm surplus decreases in fees.*

Proof: To find the effect of fees on firm surplus, we differentiate (2.18) with respect to fees by applying Leibnitz's rule twice, so that

$$\frac{\partial FS(\cdot)}{\partial f} = \int_{\hat{\phi}(1,f,t)}^{1} \int_{\hat{\theta}(\phi,f,t)}^{1} \frac{\partial \pi(\cdot)}{\partial f} d\theta d\phi$$

FS decreases because the profits of firms that continue to participate decrease in fees (Lemma 2.2), and the above term can be signed negative. Thus, fees cause a decrease in firm surplus. □

Thus, fees cause a decrease in the profits of firms that continue to participate, firm participation, and aggregate firm surplus. To be profitable the platform sets fees at a positive value. However, any increase in fees from its lower bound leads to a decrease in firm surplus (Proposition 2.4). Thus, the platform does not maximize firm surplus through its choice of fees, but its effect on customer surplus is indeterminate. The presence of P-effects can alter customer surplus, thereby reducing the extent of the misalignment between the platform and the policymaker. This is because P-investments increase firm investments, and P-Prices decrease prices, thereby driving demand. The strength of the P-effects, and the proportion of firms that face them determines the extent of the misalignment.

2.5 Conclusion

We develop a general form model that can be used to analyze platform intermediation, where a platform intermediates between firms operating on the platform, and the relationship between firms and end-customers. We describe how firms operating on the platform set prices and investments in response to the platform's choices of fees and transfer.

We find that the platform's choice of transfer concomitantly coordinates prices and investments across the two sides of the market, and maximizes customer adoption of the platform. Here, the platform and the policymaker are aligned. However, the platform's choice of fees decreases firm profits and firm participation. As the platform increases its fees towards profit maximizing levels, the rate of price increase with fees depends on the nature of demand facing each firm. If a firm has strong negative investment externalities, then prices can decrease in fees, a platform phenomenon we term P-Prices. If firm demand is convex, less responsive to own price, or highly responsive in cross-prices, then prices can increase faster than fees, demand conditions we term P-Convexity. Such price increases decrease demand. If firm demands are P-Convex, then it is inevitable that firm investments increase in fees, which is a novel positive investment externality. Such investments increase demand. Given each of these opposing effects on firm demand, it is ex-ante indeterminate if the platform's profit maximizing choice of fees decreases customer adoption of the platform. In the absence of P-effects, fees decrease firm demand, investments, profits, and participation, in addition to decreasing customer adoption and total surplus.

As we increasingly observe governments becoming involved in regulating monopolistic platforms, we can imagine that a policymaker may consider forms of Ramsey-Boiteux pricing or other mechanisms as a way to regulate and reward monopolist platforms. This is particularly true if there are weak or non-existing P-effects. However, if P-effects are strong and prevalent, then the benefits of regulating a monopoly platform depend on the platform's specific characteristics.

Generalizability and Limitations Our model is formulated in a general form whereby demands are defined by monotonicity, curvature, and inequality conditions rather than specific functional forms. As such we can determine conclusively situations when certain results can be obtained rather than special cases that depend on specific parameterization or numerical values. To show the generality of our formulation we provide an interpretation of our model in the context of Microsoft Windows in our Appendix A.4.

There are, of course, limitations of choosing a more general formulation whereby more detailed results could be obtained by choosing specific functional forms, but results from the latter approach cannot be extended beyond those specific forms.

A substantial limitation in our model is the assumption of competition silos whereby we take there to be sufficient differentiation within each kind of technology that same-side competitive effects are suppressed. Our results do generalize to settings where such competition is at a small enough scale that it does not dominate our other effects.

Chapter 3

Winner Take All Through Data Portability

Many jurisdictions have enacted regulations that require Data Controllers (DCs) such as Expedia and Amazon to enable their Data Subjects (DSs), like travel and shopping customers, to download the personal data gathered on them so that they can port their data to competing DCs. The intention of this regulation, that we call data portability regulation (DPR), is to improve DS choice by reducing the switching costs of moving between DCs. The goal is to increase the number of DCs that can be profitable and participate in the market, and to encourage DC innovation more generally. To achieve this, DPR often imposes a fine on DCs that do not allow DSs to port their data. We model the interaction between the policy-maker and DCs as a two-stage game where the policy-maker sets the fine, and then DCs choose whether to participate in a given industry, and if so, whether to comply with DPR. Modelling DCs as differing in capability, the ability to generate more revenue per unit of output, we find that more capable DCs comply, less capable DCs either do not comply or exit, and fines increase compliance and decrease participation. In some cases welfare-maximizing fines ensure all participating DCs comply, leaving only more capable DCs in the market. We also find that fines increase the aggregate output (that is, consumer surplus) and aggregate DC surplus from DCs that comply. The conclusion is that the policy-maker can maintain or improve social welfare through the use of fines to enforce DPR.

3.1 Introduction

Data Controllers (DCs) such as Expedia and Amazon are able to leverage the personal data, including user generated content, of their Data Subjects (DSs) like travel and shopping customers to provide personalized value-added features. This personal data provides a competitive advantage for DCs. Notable examples include the use of personal data by Expedia to provide a customized welcome page, travel suggestions, and pre-loaded booking preferences; and Amazon's recommendation of new products and simplified transactions. A DS moving from Expedia or Amazon to another DC does not typically get these personalized features from the other DC. Some policy-makers view this as a DS lock-in problem, where high switching costs hinder DSs from moving to a DC that they would otherwise prefer. This barrier to switching is thought to have negative implications for competition, DS surplus, and welfare. In order to alleviate these issues, policy-makers, including those in the European Union and the State of California, have intervened by requiring that DCs allow DSs to download their personal data and easily move it to a competing DC. We refer to such regulatory interventions requiring DCs to enable data portability as *data portability regulations* (DPR). DCs that violate DPR are fined under European Union's General Data Protection Regulation (GDPR, Council of European Union 2016), and State of California's Consumer Privacy Act (CCPA, California State Legislature 2018).[1]

Data portability as a right was first defined in the General Data Protection Regulation passed in the European Parliament in 2016 (Council of European Union 2016). The regulation stipulates that DSs have the right to easily receive their personal data in a commonly used format and freely transmit or port such data to competing DCs. DCs that do not comply with this regulation are subject to fines of up to 20M EUR, or variable fines of up to 4% of the total worldwide revenue. The CCPA reserves a similar right to data portability, while imposing fines of up to $7,500 per violation.

The purpose of our work is to understand the implications of a fine used to enforce data portability on the structure of the DC industry. First, we model the impact of fines on DC participation and examine the forces that lead to entry to and exit from the DC industry, thereby characterizing the types of DCs that continue to participate. Second, we determine the costs and benefits of compliance with DPR and characterize the types of DCs that comply (compliant DCs) with DPR or choose not to comply (non-compliant DCs) and pay the fine. Finally, we study the impact of fines on consumer surplus, producer (DC) surplus, and social welfare.

By requiring dominant DCs to make their data easily portable to less-dominant DCs, DPR is expected to diminish lock-in. Through this lens, DPR is seen as an integral part of restoring competition in the

[1] See Data Portability Requirements, Article 20 (which is the actual article in the EU framework), ACCESS (the US act), or the Data Portability Act.

digital economy by giving a measure of control over personal data to DSs. The United States House of Representatives report on the big tech (large and dominant DCs) states

> *"... A user may upload a variety of data ... including photos and personal information, but may not be able to easily download that data and move it to another social media site; instead, the user would have to start from scratch, re-uploading her photos and re-entering her personal information to the new platform"* (p42, United States House of Representatives 2020).

In this regard, the accumulation of data at DCs is seen as a powerful barrier to entry, giving dominant DCs a significant competitive advantage. Policy-makers believe that these issues can be resolved through imposition of DPR and fines for at-fault DCs. The basis of DPR is that data portability hinders the market power of large incumbent platforms and improves competition (Article 4, Council of European Union 2016, p20 and p40-44, United States House of Representatives 2020, Cyphers and O'Brien 2018; Seamans and Bytes 2018).

We use a general model to study the implications of DPR where DCs differ in their capability and decide on their output, where we define capability as the ability to generate more revenue per unit of output. We consider four effects on DCs from the imposition of DPR, of which the first two pertain to non-compliant DCs and the next two apply to compliant DCs. The first effect is fixed and variable (as proportion of revenue) fines for non-compliant DCs. The second effect is the loss in revenue resulting from a reduction in appeal of non-compliant DCs due to the lack of data portability as a feature. The third effect is the gain or loss in revenue due to the ability of DSs to port their data between compliant DCs. The fourth effect is the cost of compliance for compliant DCs. We find that the output of both compliant and non-compliant DCs increases with their capability, and that the imposition of fines sorts DCs into three groups based on their capability: the least capable DCs exit; the moderately capable DCs participate but do not comply with DPR and pay the fines; and the most capable DCs participate and comply with DPR. Intuitively, both fixed and variable fines increase the proportion of compliant DCs but also decrease the participation of DCs. These findings have important implications for competition law. Even though DPR increases the proportions of DCs that enable data portability, this comes at the cost of fewer DCs in the market overall, thus reducing the choices of DCs for DSs.

3.2 Literature Review

A common concern expressed by researchers in law is that the intersection between DPR and other laws (competition law, intellectual property law and consumer protection law) remains unclear (De Hert et al.

2018, Engels 2016). Swire and Lagos (2012) describe the EU's rationale behind DPR as reducing *"monopoly power in the market, so that new services can innovate and attract customers away from the original service"*, while noting that this law applies to *"start-up software companies in a garage just as it does to a monopolist"* (pp 338-339). Thus, data portability decreases switching costs for DSs by providing a competing DC with the same DS-specific data as the incumbent. We abstract away from the details of data portability implementation and focus on data portability as a construct that allows for the easy movement of DSs between DC platforms.

On the technical front, data portability has progressed over decades with the emergence of enterprise systems, inter-organizational systems, and the Internet. Modern interoperability standards for data portability between DCs in the contexts of clouds and social networks have been well researched (e.g., Alomari et al. 2014, Razmerita et al. 2009, Bojārs et al. 2008, Shirazi et al. 2012), and widely accepted standards for interoperability and portability between DCs already exist, making data portability between DCs technically feasible (ISO Central Secretary 2017, Cloud Standards Customer Council 2017). Additionally, there are several initiatives for simplifying transfer of personal data among DCs, including the Data Transfer Project[2] and the Universal Digital Profile[3]. We take the DC's decision to enable data portability to be a logical rather than a technical one and focus on the implications of DPR.

Our work is related to the literature on law enforcement and the behavior of firms responding to fines. Shavell (1987) and Polinsky and Shavell (2000) analyze the behavior of potential violators of law in a setting where courts have imperfect information. They incorporate the probability of detection into their economic model of law enforcement and suggest appropriate sanctions (e.g., fines) for offenders. In our model of data portability, the courts enforcing DPR have perfect information as consumer groups and regulators reporting to the courts can make it clear if DC compliance to DPR is followed. Our focus is on how fines for violations of DPR impact the participation and compliance of DCs.

Our analysis is also related to the literature on asset ownership, which we extend to include data as an asset. Hart and Moore (1990) show that in the presence of incomplete contracts asset ownership should reside with the party that is most able to generate value from it. If this is not the case, then there is a risk of under-investment and the first-best solution cannot be reached. Brynjolfsson (1994) and Bakos and Nault (1997) apply the Hart-Moore framework to information assets and electronic networks, respectively. Given that Hart and Moore (1990) define asset ownership as the right to exclude others from its use, DPR is a partial transfer of data asset ownership as the DC that originally gathered the data remains able to use it. Nevertheless, losing exclusive use of the data may impact DC investments. In a similar vein, the actions of

[2] `https://datatransferproject.dev`

[3] `https://techcrunch.com/2018/05/22/the-birth-of-the-universal-digital-profile/`

DCs can be evaluated through the lens of the principal-agent model (Hölmstrom 1979, Lambert 2001). The principal (DS) assigns an agent (DC) to create value from their data - this value is then shared between the DS and DC. With DPR, the nature of the contract between the two parties changes - the DS has the ability to port their data to a different DC, thereby potentially changing the actions of the DC.

Using arguments in competition law, Graef et al. (2013) and Swire and Lagos (2012) suggest that DPR could stifle competition, innovation, and investments, and go on to suggest that DPR can lead to a decrease in consumer welfare. Further, Lam and Liu (2020) find that if DPR encourages DSs to share more data with their current DC because of the prospect of easier switching, then this could strengthen the incumbent and raise the entry barrier for new DCs. Christensen et al. (2013) estimate the financial impact of data protection regulation, including DPR, on firms within the EU. They consider the costs of compliance and lack of access to data and estimate a significant negative impact on firms and the economy as a result of enforcement of DPR. Wohlfarth (2019) considers the impact of DPR on the amount of data that DCs collect on DSs, finding that in some situations the increase in data collection due to DPR enforcement may harm users. We extend this stream of literature and uncover the mechanisms by which fines influence a DC's decisions about output levels, participation in the market and DPR compliance, as well as competition, DS surplus, and DC surplus. We find that even in absence of data collection decisions, fines on DCs for violations of DPR can negatively impact DCs and DSs.

We restrict our analysis to an industry consisting of a set of competing DCs that differ in capability. Although our analysis applies to any set of competing DCs that derive value from personal data, we illustrate our setup using the travel booking industry which consists of well-known DCs such as Expedia, Travelocity, and Kayak, as well as lesser known local (European) DCs such as hrs.de. The structure of our assumptions is related to Nault and Zimmermann (2019), where DC (edge provider in their model) costs and profits change with decisions made in earlier stages. Our use of reduced form profit functions is related to a long stream of research that abstracts from the details of a DC's production choices and competition so as to help the study focus on the interactions between external parties and the DC (see pp 147-150 Jehle and Reny 2011, pp 224-226 Tirole 1988, Nault 1996, Levi and Nault 2004, Nault and Zimmermann 2019). Our theoretical contribution to this literature is the inclusion of port functions which measure the gains or losses in revenue of both compliant and non-compliant DCs when DPR is implemented and enable us to discuss the intricacies of the impact of DPR on DCs.

Our study contributes to the established streams of literature on regulation of market entry and incumbency (Tirole 1988), and switching costs and user lock-in (Klemperer 1987, Farrell and Shapiro 1988), specifically in the domain of information technology (IT). Nault and Vandenbosch (2000) consider the timing of entry from an entrant with disruptive technologies and discuss how the incumbent and entrant may act to increase barriers to entry. We extend this literature by discussing the impact of DPR and fines on the entry and exit of DCs, as well as on their compliance decisions. The main finding of the analytical studies on switching costs is that such costs hinder competition through higher prices for users. Our work demonstrates that in the context of DPR, imposing fines to enforce DPR compliance as a way to decrease switching costs may not be an effective regulatory mechanism.

Finally, our work contributes to the emerging literature about the impact of data on competition. Braulin and Valletti (2016) consider the exclusivity of data sales by a monopolist data broker to two competing retailers and find that the data broker sells data exclusively to either the high-quality or the low-quality retailer. Kim et al. (2019) study the impact of data on horizontal mergers and its implications for consumer surplus. de Corniere and Taylor (2020) propose a competition-in-utilities framework for studying the impact of data on competition, specifically in the contexts of data-based algorithms, targeted advertising, price discrimination, and data-driven mergers. This stream focuses on the competitive effects from exclusive access to data, for example, through buying data from data brokers, and does not reserve a right for customers to access or port their data. We extend this literature by considering the implications of allowing users to port their data through DPR and its implications on competition and welfare.

3.2.1 Our Setting

Following the Council of European Union (2016), we broadly define DC as the body which "determines the purposes and means of processing of personal data", and DS as an identifiable natural person who shares their personal data with a DC. The DS might be an end customer of a DC and receive more valuable products and services from the DC in return for sharing their personal data (pp 2-5, Council of European Union 2016). As a context, one can take Expedia to be a DC, and their users to be the DSs. We analyze an industry that consists of competing heterogenous DCs and we use the term *capability* which we define to be the ability to generate more revenue per unit of output to describe the dimension along which DCs are distributed. Capability is a representation of a DC's production technology, and can include expertise in applications development, database, data mining, analytics, and user preference estimation. More capable DCs may provide greater ease of use, quality of recommendations, availability through multiple channels or platforms, reliability, or speed.

In part, DSs hand over their data in lieu of prices, which is in turn monetized by DCs. However, DSs typically do not know the relative level of privacy intrusion that they face with each DC. For instance, although DSs are concerned about their privacy, they are not aware of how their personal data is collected and used (Graeff and Harmon 2002). In practice, the details of how personal data can be used by a DC are often embedded within a multi-page contract with the DS, whereby DSs often *"... view policies as a nuisance, ignoring them to pursue the ends of digital production, without being inhibited by the means"* (p1 Graeff and Harmon 2002). Thus, we assume that DSs do not internalize the differences in privacy costs when choosing between competing DCs.

We study the effect of fines on DC participation, compliance, and profits, given pre-existing investments and an exogenously determined capability. Each DC maximizes profit by deciding if it should comply with DPR or not comply and pay the fine, and by choosing its level of output. Complying with DPR brings with it gains or losses in revenue from DSs switching from or to other DCs. The policy-maker's interest is in maximizing social welfare. To this end the policy-maker decides the level of fines to levy on DCs that do not comply with DPR. Although we do not explicitly model individual DS decisions, we take DSs as rational utility maximizers - they choose a DC that offers them the greatest benefit so long as it satisfies their individual rationality constraint.

In the following section we describe the notation and assumptions of our model. Then, we use backward induction to solve the DC problem, followed by the policy-maker's problem. Finally, we use comparative statics to explain the effects of fines on the DCs' decisions to comply with DPR and participate in the market, as well as on consumer surplus, producer surplus and social welfare, before concluding with our findings and recommendations to the policy-maker.

3.3 Notation and Assumptions

We take DCs to be heterogeneous and differ in capability. We denote DC capability by θ, which is normalized to be in $[0, 1]$, and is uniformly distributed, $g(\theta) \sim U[0, 1]$, so that the density is positive over its support, $g(\theta) > 0 \ \forall \ \theta \in [0, 1], G(0) = 0$ and $G(1) = 1$. We treat DC capability as increasing in θ so that a more capable DC has a higher θ. An example of the prior use of capability as a construct is in Gal-Or et al. (2018), where they model platforms that have different targeting capabilities. DC capability can be visualized to be distributed along a line. We refer to this visual representation of DC distribution in the following sections as the *capability line*. Every point θ on this capability line represents a DC. The policy-maker knows the distribution of θ but cannot infer the capability of a specific DC.

DCs provide products and services to DSs. We denote a generic unit of revenue-generating output, that we hereafter refer to as *output*, by $x \in [0, \bar{x}]$. An example of output for Expedia is the number of travel bookings or reservations that are made using the platform. Each DC produces output for which it receives revenue, and we denote that output with the subscript θ so that x_θ represents output from the DC at θ on the capability line. We refer to the vector of DC outputs over the support of θ as $\vec{x} = (x_\theta, \vec{x}_{\backslash\theta})$ where $\vec{x}_{\backslash\theta}$ is a vector of outputs from DCs other than θ.

For each DC we define a reduced form revenue function $R(\theta, x_\theta, \vec{x}_{\setminus\theta})$, allowing for revenue to depend on capability, own output, and output from any or all other DCs. We also define a DC cost function that depends on own output only, denoted as $C(x_\theta)$. We take the revenues of a DC with zero capability as zero, $R(\theta = 0, \vec{x}) = 0$, and both revenues and costs with zero output as zero, $R(\theta, x_\theta = 0, \vec{x}_{\setminus\theta}) = 0,\ C(x_\theta = 0) = 0$. We note that x can be characterized as resulting from an effort to output function, $x(e)$, where $e \in \mathbb{R}_{>0}$ and $x'(e) > 0$. There is no loss of generality from working directly with x in our analysis. Our first assumption defines reasonable properties of $R(\theta, x_\theta, \vec{x}_{\setminus\theta})$ and $C(x_\theta)$.

Assumption 3.1 (DC revenues and costs) *(a) DC revenues are increasing in capability, increasing and concave in output, and decreasing in other DC output, (b) DC costs are increasing and convex in output, and (c) marginal revenue is increasing in capability.*

Mathematically the parts of Assumption 3.1 are

$$(a) \quad \frac{\partial R(\theta, \vec{x})}{\partial \theta} > 0, \quad \frac{\partial R(\theta, \vec{x})}{\partial x_\theta} > 0, \quad \frac{\partial^2 R(\theta, \vec{x})}{\partial x_\theta^2} \leqslant 0, \quad \frac{\partial R(\theta, \vec{x})}{\partial x_{\setminus\theta}} \leqslant 0,$$
$$(b) \quad \frac{\partial C(x_\theta)}{\partial x_\theta} > 0, \quad \frac{\partial^2 C(x_\theta)}{\partial x_\theta^2} \geqslant 0,$$
$$(c) \quad \frac{\partial^2 R(\theta, \vec{x})}{\partial \theta \partial x_\theta} > 0.$$

Assumption 3.1(a) defines the effect of DC capability in our model: more capable DCs generate more revenue per unit of output. Further, we take revenues to be increasing and concave in output. In addition, allowing DC revenues to depend on output from other DCs means that the reduced form revenue function and its properties in Assumption 3.1(a) can accommodate almost any form of DC competition.

Assumption 3.1(b) takes costs to be increasing and convex in output. These costs can include new product and service development, as well as marketing and advertising. We could also define the cost function to depend on capability with properties that are complementary to the revenue function, and this would simply reinforce our results without additional insights. Although the marginal costs of technology can be concave, the costs of revenue as well as the costs of selling and marketing are convex. A large proportion of DC costs are typically apportioned to convex costs. For example, Expedia Group reports that in 2019, their costs of revenue, selling, and marketing were \$8,298M, while the accounting allowances related to their technology assets were \$537M.

Finally, in Assumption 3.1(c) we assume that more capable DCs generate greater revenue per unit of output and thus more capable DCs are more profitable per unit of output.

DC profits can then be written as

$$PR(\theta, \vec{x}) = R(\theta, \vec{x}) - C(x_\theta). \tag{3.1}$$

From Assumption 3.1 DC profits are increasing in capability, $\partial PR(\theta, \vec{x})/\partial\theta > 0$, concave in output, $\partial^2 PR(\theta, \vec{x})/\partial x_\theta^2 < 0$, and marginal profits are increasing in capability, $\partial^2 PR(\theta, \vec{x})/\partial\theta\partial x_\theta > 0$. DSs play a passive role in our model - they influence the profit function of each DC through their demand preferences. This is captured without loss of generality within the revenue functions of DCs.

Imposition of DPR: fines and porting The imposition of DPR has four effects. The first is that DCs that do not comply with DPR are fined. Following the EU model, we allow for both fixed fines and variable fines as a proportion of revenues. The fixed fine on non-compliant DCs is $F \in \mathbb{R}_{>0}$. The variable fine on non-compliant DCs is $f \in [0, 1]$, and represents the proportion of DC revenue that is transferred as a fine. Thus, the variable fine paid by a non-compliant DC is f times their revenues. The policy-maker chooses the level of fines. There is no fine for compliant DCs.

The second and third effects concern DC revenues. The second effect is that DCs which do not comply with DPR are less attractive to DSs relative to DCs which do, because they do not allow the porting of DS data. Everything else equal, all DSs prefer a DC that allows them to port their data because DPR is considered an additional feature. This leads to a potential loss in revenues for non-compliant DCs as DSs may move some or all of their business to compliant DCs in order to gain the added feature of portability. The underlying cause for the change in revenues is a shift in demand. When DPR is imposed, the choice of a DC to not enable portability weakens demand (shifts its demand down and to the left), resulting in decreased revenues (a downward shift of its revenue function). In other words, for a particular level of output, revenues for a non-compliant DC are weakly lower.

The third effect concerns DCs that comply with DPR. By complying with DPR, DSs can port their data from one compliant DC to another. Thus, for DCs that comply with DPR, there is the potential of gains or losses in revenues as a result of DSs porting their data along with some or all of their business to or from other DCs.

The fourth effect is the cost of compliance for DCs that comply. The DC decision to allow DSs to download their data may be a logical one, or may require either new equipment or substantial system development. There are standards that simplify data portability such as the Data Transfer Project and the Universal Digital Profile. The result is often in form a button on the website to enable the download of DS data in a common and easy-to-port format using a simple query.

We model such effects by gains and losses in revenue from compliance as a result of the imposition of DPR using a *compliant port* function that is specific to a given compliant DC. The compliant port function captures the change in revenues from compliance as well as the cost of compliance, if any. This compliant port function depends on the proportion of DCs that comply, where for the moment we use ρ to denote the proportion of the population of DCs that comply, and the output of all DCs, $\zeta(\theta, \rho, \vec{x}) \in \mathbb{R}$. The compliant port function $\zeta(\theta, \rho, \vec{x})$ can be positive or negative for a given DC, as the third effect above can cause some compliant DCs to lose business to other compliant DCs and there may exist the additional cost of compliance. Thus, the net DC revenues from compliance is $R(\theta, \vec{x}) + \zeta(\theta, \rho, \vec{x})$.

Similarly, the loss in revenue resulting from non-compliance with DPR is modeled using a *non-compliant port* function, $\omega(\theta, \rho, \vec{x}) \in \mathbb{R}_{>0}$. The non-compliant port function captures the losses in revenues from non-compliance. Thus, the non-compliant port function, $\omega(\theta, \rho, \vec{x})$ represents lost revenue from non-compliance, which when subtracted from DC revenues gives the net DC revenues from non-compliance, $R(\theta, \vec{x}) - \omega(\theta, \rho, \vec{x})$.

These port functions are the novel innovation in our formulation, and we examine how the port functions change with DPR. We use the port functions to account for all the changes in DC revenues that result from the imposition of DPR. An exception is an indirect effect whereby the impact of DPR on DC outputs affects the value of the DC reduced form profit function, $PR(\theta, \vec{x})$ in (3.1). As distinctions between output of compliant versus non-compliant DCs matters later on, we use the superscript c to denote compliance so x_θ^c is the compliant DC's choice of output and we use the superscript nc to denote non-compliance so that x_θ^{nc} is the non-compliant DC's choice of output. We make general assumptions regarding the behavior of each port function below.

Assumption 3.2 (Compliant Port Function) *The compliant port function is (a) increasing in DC capability, (b) concave in own output, (c) increasing in the proportion of compliant DCs, (d) at the margin is increasing in capability, and (e) at the margin is increasing in the proportion of compliant DCs.*

Using weak inequalities, mathematically the parts of Assumption 3.2 are

$$(a)\ \frac{\partial \zeta(\theta, \rho, \vec{x})}{\partial \theta} \geqslant 0, \quad (b)\ \frac{\partial^2 \zeta(\theta, \rho, \vec{x})}{\partial [x_\theta^c]^2} \leqslant 0, \quad (c)\ \frac{\partial \zeta(\theta, \rho, \vec{x})}{\partial \rho} \geqslant 0,$$
$$(d)\ \frac{\partial^2 \zeta(\theta, \rho, \vec{x})}{\partial \theta \partial x_\theta^c} \geqslant 0, \quad and \quad (e)\ \frac{\partial^2 \zeta(\theta, \rho, \vec{x})}{\partial \rho \partial x_\theta^c} \geqslant 0.$$

The port function for DCs that comply has similar properties to the DC profit function in (3.1). In Assumption 3.2(a) DSs widely agree on their preference for more capable DCs, and we model gains or losses in revenue from data porting by DSs as a result of DPR as a possible migration from a lower capability DC to a higher capability DC. In the compliant port function for a given DC, the gains or losses in revenue depend in part on the DC's current level of output, x^c_θ. In Assumption 3.2(b) we take such gains as concave in output as may be the case if there were capacity constraints. As we point out later, Assumption 3.2(b) is sufficient but not necessary for our results.

In Assumption 3.2(c), the change in revenue from compliance increases with the proportion of DCs that comply, ρ. This is because for a particular compliant DC demand increases as the proportion of compliant DCs increases. In Assumption 3.2(d), the marginal change in revenue from compliance increases in capability as more capable DCs are able to generate more gains in revenues with an additional unit of output – the underlying cause being superior technological capability – in part due to DSs preferring more capable DCs. For the same reason as in Assumption 3.2(c), in Assumption 3.2(e), the marginal change in revenue from compliance increases in the proportion of DCs that comply. We more fully explain Assumptions 3.2(c) and (e) after Theorem 3.1 when ρ can be characterized and before they are used in our main results.

Assumption 3.3 (Non-Compliant Port Function) *DC losses in revenue from non-compliance (a) are decreasing in DC capability, (b) are increasing in output, (c) are convex in output, (d) are increasing in the proportion of compliant DCs, (e) at the margin are decreasing in capability, and (f) at the margin are increasing in the proportion of compliant DCs.*

Using weak inequalities, mathematically the parts of Assumption 3.3 are

$$(a)\ \frac{\partial \omega(\theta,\rho,\vec{x})}{\partial \theta} \leqslant 0, \quad (b)\ \frac{\partial \omega(\theta,\rho,\vec{x})}{\partial x^{nc}_\theta} \geqslant 0, \quad (c)\ \frac{\partial^2 \omega(\theta,\rho,\vec{x})}{\partial [x^{nc}_\theta]^2} \geqslant 0, \quad (d)\ \frac{\partial \omega(\theta,\rho,\vec{x})}{\partial \rho} \geqslant 0,$$
$$(e)\ \frac{\partial^2 \omega(\theta,\rho,\vec{x})}{\partial \theta \partial x^{nc}_\theta} \leqslant 0, \quad and\ (f)\ \frac{\partial^2 \omega(\theta,\rho,\vec{x})}{\partial \rho \partial x^{nc}_\theta} \geqslant 0.$$

In Assumption 3.3(a) DC capability mitigates the loss from non-compliance as more capable DCs offer more value to DSs and better resist losing DSs to compliant DCs. Assumption 3.3(b) means that DC losses in revenue from non-compliance are increasing in output, although not enough to reverse Assumption 3.3(a). Such would be the case if non-compliant DCs lose a constant proportion of revenue from their output to compliant DCs. In Assumption 3.3(c), the losses from non-compliance are convex in output. Similar to Assumption 3.2(b), Assumption 3.3(c) is sufficient but not necessary for our results as it simply ensures that the net DC revenue from non-compliance is concave. Assumption 3.3(d) implies that DC losses in revenue from non-compliance are increasing in the proportion of compliant DCs, because there are more alternative compliant DCs for DSs to choose from. Assumption 3.3(e) means that more capable non-compliant DCs lose less at the margin for the same reason as Assumption 3.3(a). For the same reason as 3.3(d), the marginal losses increase in the proportion of compliant DCs, our Assumption 3.3(f). Similar to our Assumptions 3.2(c) and (e), we more fully explain Assumptions 3.3(d) and (f) after Theorem 3.1 when ρ can be characterized and before they are used in our main results.

It is reasonable to expect that for non-compliant DCs not all profits are lost, $PR(\theta, \vec{x}) \geqslant \omega(\theta, \rho, \vec{x}) \geqslant 0$. This implies that revenues are no less than the loss of revenue from non-compliance, $R(\theta, \vec{x}) \geqslant \omega(\theta, \rho, \vec{x}) \geqslant 0$. In addition, the marginal revenue from additional output is no less than the marginal loss of revenue from non-compliance so that $\partial R(\theta, \vec{x})/\partial x_\theta^{nc} - \partial \omega(\theta, \rho, \vec{x})/\partial x_\theta^{nc} \geqslant 0$. In other words, loss of revenue from non-compliance can at worst equal gains in revenue from the additional output, so that net revenue from non-compliance weakly increases in output. This is distinct from convex increases in cost resulting from additional output.

Thus, changes in revenue from compliance are captured by $\zeta(\theta, \rho, \vec{x})$, and losses in revenues from non-compliance are captured by $\omega(\theta, \rho, \vec{x})$. Further, $\zeta(\theta, \rho, \vec{x}) \in \mathbb{R} \ \ \forall \ \ \theta \in [0, 1]$ can be positive or negative (and increasing in capability), but $\omega(\theta, \rho, \vec{x}) \in \mathbb{R}_{\geqslant 0} \ \ \forall \ \ \theta \in [0, 1]$ is always positive, thereby decreasing net revenues when it is subtracted from the revenue function $R(\theta, \vec{x})$. Our next assumption compares the effects of output and capability on the two port functions.

Assumption 3.4 (Relative Effects) *(a) A DC always gains more of its marginal revenue by complying with DPR, and (b) for a given DC increases in capability have larger effects on porting if they comply with DPR.*

Mathematically, Assumption 3.4 is

$$(a)\ \frac{\partial \zeta(\theta, \rho, \vec{x})}{\partial x_\theta^c} + \frac{\partial \omega(\theta, \rho, \vec{x})}{\partial x_\theta^{nc}} > 0, \quad (b)\ \frac{\partial \zeta(\theta, \rho, \vec{x})}{\partial \theta} + \frac{\partial \omega(\theta, \rho, \vec{x})}{\partial \theta} > 0.$$

The first term of the inequality describing Assumption 3.4(a), the change in revenue from compliance, can be negative when a compliant DC loses some of its marginal revenue to higher capability compliant DCs. However, this cannot be larger than the second term which is the marginal losses in revenue from non-compliance. Given that DSs with non-compliant DCs face friction in acquiring their data to move to a compliant DC, our Assumption 3.4(a) implicitly assumes that for such DSs the benefit of portability as an additional feature outweighs the friction in acquiring their data. In Assumption 3.4(b), with an increase in capability, gains from portability for compliant DCs are larger than the reduction in losses for non-compliant DCs. That is, the first term is positive from Assumption 3.2 and the second term is negative, although smaller in magnitude than the first given Assumption 3.3.

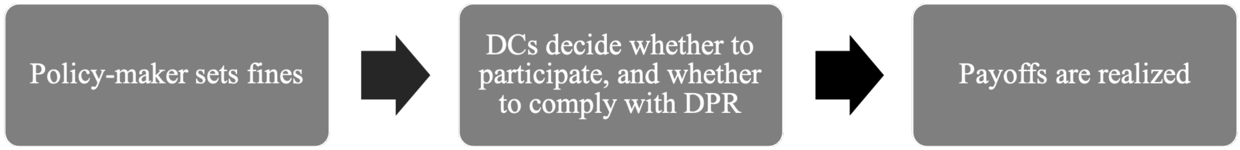


Figure 3.1: Stages of the game.

3.4 The Effects of Data Portability

We structure our analysis with the goal to explain the effect of fines to enforce DPR on the structure of the DC industry. We analyze a setting where DCs have complete information and consistent with the above, we use the terms *compliant* and *non-compliant* to denote a DC's compliance or non-compliance with DPR. In contrast, we use the term *participating* to denote DCs that choose to operate because they make positive profits, while *non-participating* DCs choose zero output and effectively exit the industry.

We model the process of policy-maker fines and DC compliance as a two-stage game where the latter stage has two decisions. In Stage 1 the policy-maker sets fines for non-compliance with DPR. In Stage 2 DCs decide their output and whether to comply with DPR. Finally, payoffs are realized (Figure 3.1). To analyze the impact of fines, we work backwards by first solving the DC production (output) decisions when complying with DPR and not, and next choosing whether to comply - this determines which DCs comply. Then we solve the policy-maker's decision on fines and analyze the effect of fines on consumer surplus (DS surplus), producer surplus (DC surplus), and social welfare.

Given that the purpose of DPR is to enable DSs to port their data to a different DC by decreasing switching costs, we first consider the status quo as an equilibrium prior to the imposition of DPR. We then consider a setting where the imposition of DPR causes some DSs to port their data from non-compliant DCs to compliant DCs, and from lower capability compliant DCs to higher capability ones, as captured by our port functions.

Using the reduced form profit function in (3.1), prior to the imposition of DPR each DC maximizes profits by choice of own output, where we use the superscript *pre* to denote output pre-DPR:

$$\max_{x_\theta^{pre}} PR(\theta, x_\theta^{pre}, \vec{x}_{\setminus\theta}),$$

and the resulting set of first-order conditions are

$$\partial PR(\theta, x_\theta^{pre}, \vec{x}_{\setminus\theta})/\partial x_\theta^{pre} = 0, \;\; \forall \;\; \theta \in [0,1].$$

With concavity of the profit function and output defined above (continuous over a compact set), the first-order conditions lead to pre-DPR equilibrium output for all DCs, $x_\theta^{pre}(\vec{x}_{\setminus\theta})$.

3.4.1 DC Production Decisions

DCs that Comply with DPR

Payoffs for DCs that comply are profits that would have been attained pre-DPR plus changes in revenue from the compliant port function. Compliant DCs are not fined. Using (3.1) the payoff maximization problem for a compliant DC is

$$\max_{x_\theta^c} \Pi^c(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta}) = PR(\theta, x_\theta^c, \vec{x}_{\setminus\theta}) + \zeta(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta}). \tag{3.2}$$

The first-order condition by choice of output is

$$\frac{\partial \Pi^c(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial x_\theta^c} = \frac{\partial PR(\theta, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial x_\theta^c} + \frac{\partial \zeta(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial x_\theta^c} = 0 = \psi_1(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta}), \tag{3.3}$$

where $\psi_1(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta}) = 0$ implicitly defines $x_\theta^c(\rho, \vec{x}_{\setminus\theta})$. Thus, $x_\theta^c(\rho, \vec{x}_{\setminus\theta})$ is an optimal value function. Economizing on notation so that $\psi_1(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta}) = \psi_1(\cdot)$, and differentiating (3.3) with respect to output, using Assumptions 3.1(a) and (b), and 3.2(b) we get

$$\frac{\partial^2 \Pi^c(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial [x_\theta^c]^2} = \frac{\partial \psi_1(\cdot)}{\partial x_\theta^c} = \frac{\partial^2 PR(\theta, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial [x_\theta^c]^2} + \frac{\partial^2 \zeta(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial [x_\theta^c]^2} < 0, \tag{3.4}$$

noting that our Assumption 3.2(b) on concavity of the compliant port function is not necessary so long as it is outweighed by the concavity of the reduced form profit function.

Our first lemma illustrates the effect of capability on output for DCs that comply with DPR.

Lemma 3.1 *For compliant DCs, output increases in capability and in the proportion of compliant DCs.*

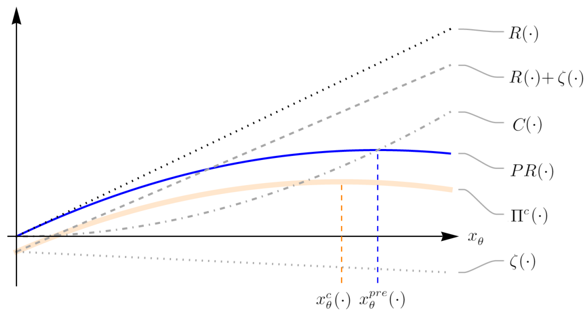
Proof: Differentiating (3.3) with respect to capability, we get

$$\frac{\partial \psi_1(\cdot)}{\partial \theta} = \frac{\partial^2 PR(\theta, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial x_\theta^c \partial \theta} + \frac{\partial^2 \zeta(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial x_\theta^c \partial \theta} > 0,$$

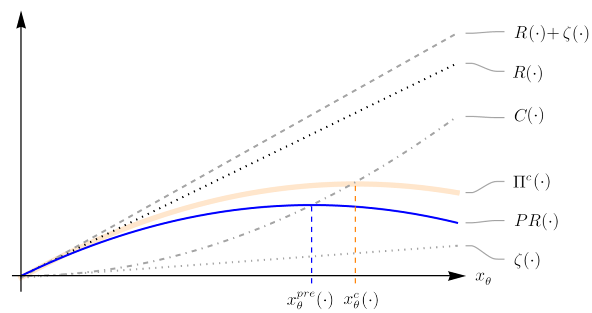
which is positive by Assumptions 3.1(c) and 3.2(d). From (3.4) we have $\partial \psi_1(\cdot)/x_\theta^c < 0$. Using the implicit function theorem,

$$\frac{\partial x_\theta^c(\rho, \vec{x}_{\setminus\theta})}{\partial \theta} = -\frac{\partial \psi_1(\cdot)/\partial \theta}{\partial \psi_1(\cdot)/\partial x_\theta^c} > 0.$$

Differentiating (3.3) with respect to the proportion of DCs that comply with DPR and using the envelope



(a) Loss of revenue from compliance



(b) Gain in revenue from compliance

Figure 3.2: Payoffs from compliance: An illustrative example.

theorem to eliminate the effect of the proportion of compliant DCs on output, we get

$$\frac{\partial \psi_1(\cdot)}{\partial \rho} = \frac{\partial^2 \zeta(\theta, \rho, x_\theta^c, \vec{x}_{\setminus\theta})}{\partial x_\theta^c \partial \rho} \geqslant 0,$$

which is positive from Assumption 3.2(e). Using the implicit function theorem,

$$\frac{\partial x_\theta^c(\rho, \vec{x}_{\setminus\theta})}{\partial \rho} = -\frac{\partial \psi_1(\cdot)/\partial \rho}{\partial \psi_1(\cdot)/\partial x_\theta^c} \geqslant 0. \qquad \square$$

Figure 3.2 provides an illustration of the payoffs from compliance. Pre-DPR profit is revenue less cost, which is optimized at $x_\theta^{pre}(\cdot)$. After imposition of DPR, depending on the functional forms or a DC's capability, the change in revenues from compliance, $\zeta(\cdot)$, can be either negative (Figure 3.2a), or positive (Figure 3.2b), which results in a consequently lower or higher payoff from compliance, $\Pi^c(\cdot)$, respectively.

DCs that Do Not Comply with DPR

Payoffs for DCs that do not comply with DPR are revenues that would have been attained pre-DPR less the non-compliant port function adjusted by the variable fine on revenues, less costs and the fixed fine. In the payoff maximization problem for a non-compliant DC below, f is the proportion of net DC revenues transferred to the policy-maker as variable fines, and F is the fixed fine transferred to the policy-maker. The payoff maximization is

$$\max_{x_\theta^{nc}} \Pi^{nc}(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta}, F, f) = [1-f][R(\theta, x_\theta^{nc}, \vec{x}_{\setminus\theta}) - \omega(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta})] - C(x_\theta^{nc}) - F. \tag{3.5}$$

Economizing on notation by using $\Pi^{nc}(\cdot)$ to capture the arguments in the left-hand side of (3.5), for non-compliant DCs the first-order condition by choice of output is

$$\frac{\partial \Pi^{nc}(\cdot)}{\partial x_\theta^{nc}} = [1-f][\frac{\partial R(\theta, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial x_\theta^{nc}} - \frac{\partial \omega(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial x_\theta^{nc}}] - \frac{\partial C(x_\theta^{nc})}{\partial x_\theta^{nc}} = 0 = \psi_2(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta}, f), \tag{3.6}$$

where $\psi_2(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta}, f) = 0$ implicitly defines $x_\theta^{nc}(\rho, \vec{x}_{\setminus\theta}, f)$. Using the description following Assumption 3.3, positive output is only possible if marginal revenue is greater than the marginal loss from porting, which is straightforward, as a DC cannot lose revenue from porting what it does not have. Further economizing on notation so that $\psi_2(\theta, x_\theta^{nc}, \vec{x}_{\setminus\theta}, f) = \psi_2(\cdot)$, differentiating (3.6) with respect to output we get

$$\frac{\partial \psi_2(\cdot)}{\partial x_\theta^{nc}} = \frac{\partial^2 \Pi^{nc}(\cdot)}{\partial [x_\theta^{nc}]^2} = [1-f][\frac{\partial^2 R(\theta, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial [x_\theta^{nc}]^2} - \frac{\partial^2 \omega(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial [x_\theta^{nc}]^2}] - \frac{\partial^2 C(x_\theta^{nc})}{\partial [x_\theta^{nc}]^2} < 0, \tag{3.7}$$

which is negative from Assumptions 3.1(a) and (b), and 3.3(c). Similar to the second-order conditions in (3.4), our Assumption 3.3(c) is not necessary so long as the effects on revenues and costs outweigh those from the non-compliant port function.

Our second lemma characterizes the effect of fines, capability, and the proportion of compliant DCs on output of non-compliant DCs.

Lemma 3.2 *For non-compliant DCs, output increases in capability, decreases in the proportion of complaint DCs and variable fines, and is unaffected by fixed fines.*

Proof: Differentiating (3.6) with respect to capability we get

$$\frac{\partial \psi_2(\cdot)}{\partial \theta} = \frac{\partial^2 \Pi^{nc}(\cdot)}{\partial \theta \partial x_\theta^{nc}} = [1-f][\frac{\partial^2 R(\theta, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial \theta \partial x_\theta^{nc}} - \frac{\partial^2 \omega(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial \theta \partial x_\theta^{nc}}] > 0, \tag{3.8}$$

which is positive by Assumptions 3.1(c) and 3.3(e). Using the implicit function theorem, we have

$$\frac{\partial x_\theta^{nc}(\rho, \vec{x}_{\setminus\theta}, f)}{\partial \theta} = -\frac{\partial \psi_2(\cdot)/\partial \theta}{\partial \psi_2(\cdot)/\partial x_\theta^{nc}}.$$

The numerator is positive from (3.8), and the denominator is negative from (3.7), so that $\partial x_\theta^{nc}(\rho, \vec{x}_{\setminus\theta}, f)/\partial \theta > 0$.

Differentiating (3.6) with respect to the proportion of compliant DCs and applying the envelope theorem to eliminate the effect of the proportion of compliant DCs on output, we get

$$\frac{\partial \psi_2(\cdot)}{\partial \rho} = [1-f]\Big[- \frac{\partial^2 \omega(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial x_\theta^{nc} \partial \rho}\Big] \leqslant 0,$$

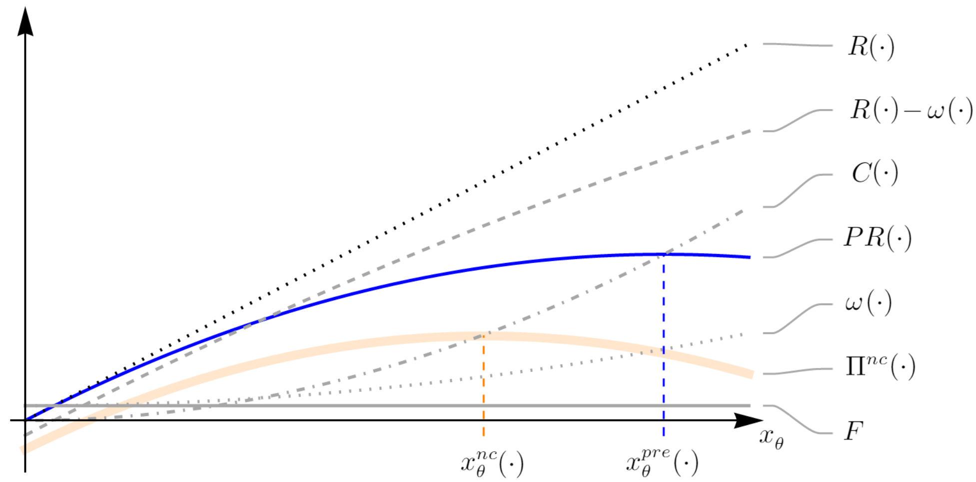


Figure 3.3: Output and payoffs from non-compliance: An illustrative example.

which is negative by Assumption 3.3(f). Then, using (3.7), by the implicit function theorem, we have

$$\frac{\partial x_\theta^{nc}(\rho, \vec{x}_{\setminus\theta}, f)}{\partial \rho} = -\frac{\partial \psi_2(\cdot)/\partial \rho}{\partial \psi_2(\cdot)/\partial x_\theta^{nc}} \leqslant 0.$$

Now differentiating (3.6) with respect to variable fines and applying the envelope theorem to eliminate the effect of fines on output, we get

$$\frac{\partial \psi_2(\cdot)}{\partial f} = -\Big[\frac{\partial R(\theta, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial x_\theta^{nc}} - \frac{\partial \omega(\theta, \rho, x_\theta^{nc}, \vec{x}_{\setminus\theta})}{\partial x_\theta^{nc}}\Big] < 0,$$

which is negative by Assumptions 3.1(a) and 3.3(a). Again, using (3.7), and by the implicit function theorem,

$$\frac{\partial x_\theta^{nc}(\rho, \vec{x}_{\setminus\theta}, f)}{\partial f} = -\frac{\partial \psi_2(\cdot)/\partial f}{\partial \psi_2(\cdot)/\partial x_\theta^{nc}} < 0.$$

Finally, $\psi_2(\cdot)$ is not a function of fixed fines, therefore, output is unaffected by fixed fines. □

Figure 3.3 illustrates the impact of imposition of DPR on DC payoffs from non-compliance. The net revenues from non-compliance, $R(\cdot) - \omega(\cdot)$, and the fixed fine, F, result in lower payoffs, $\Pi^{nc}(\cdot)$, and optimal output, $x_\theta^{nc}(\cdot)$. Non-zero variable fines cause a further decrease in the net revenues from non-compliance, payoffs, and optimal output.

3.4.2 DC Industry Response

We begin by comparing DC output for compliant and non-compliant DCs. Examining the first-order conditions in (3.3) and (3.6) for a DC with a given capability θ, and taking the variable fine as being set to zero, the difference between complying and not,

$$\frac{\partial\zeta(\theta,\rho,x^c_\theta,\vec{x}_{\setminus\theta})}{\partial x^c_\theta}+\frac{\partial\omega(\theta,\rho,x^{nc}_\theta,\vec{x}_{\setminus\theta})}{\partial x^{nc}_\theta},$$

is positive from our Assumption 3.4(a). For a given DC, output is higher if it complies. This is reinforced with a positive variable fine if the DC does not comply because the effect of such a fine is greater on marginal revenue than on the non-compliant port function as per the explanation of (3.6). Consequently,

$$x^c_\theta(\rho,\vec{x}_{\setminus\theta})>x^{nc}_\theta(\rho,\vec{x}_{\setminus\theta},f). \tag{3.9}$$

Given a known level of fines and DC participation, DCs choose whether to comply with DPR. The payoffs for DCs that comply are as in (3.2) where outputs are stated as optimal value functions, $x^c_\theta(\rho,\vec{x}_{\setminus\theta})=x^c_\theta(\cdot)$ and $x^{nc}_\theta(\rho,\vec{x}_{\setminus\theta},f)=x^{nc}_\theta(\cdot)$, and where we write out the elements of the reduced form profit function from (3.1) as

$$R(\theta,x^c_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))+\zeta(\theta,\rho,x^c_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))-C(x^c_\theta(\cdot)). \tag{3.10}$$

The payoffs for DCs that do not comply are as in (3.5) but again with outputs stated as optimal value functions,

$$[1-f][R(\theta,x^{nc}_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))-\omega(\theta,\rho,x^{nc}_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))]-C(x^{nc}_\theta(\cdot))-F. \tag{3.11}$$

Notice that output of non-compliant DCs is a function of variable fines whereas output of compliant DCs is independent of fines (Lemmas 3.1 and 3.2).

Each DC maximizes its payoff by choosing whether to comply with DPR. Thus, a DC's choice of whether to comply and its consequent payoff is the maximum of (3.10) and (3.11):

$$\begin{aligned}&\max\{R(\theta,x^c_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))+\zeta(\theta,\rho,x^c_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))-C(x^c_\theta(\cdot)),\\&\quad[1-f][R(\theta,x^{nc}_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))-\omega(\theta,\rho,x^{nc}_\theta(\cdot),\vec{x}_{\setminus\theta}(\cdot))]-C(x^{nc}_\theta(\cdot))-F\}.\end{aligned}$$

To find the DC that is indifferent between complying and not complying with DPR, denoted by $\tilde{\theta}$, we equate the payoffs to compliance and non-compliance so that

$$\begin{aligned}&R(\tilde{\theta},x^c_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))+\zeta(\tilde{\theta},\rho,x^c_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))-C(x^c_{\tilde{\theta}}(\cdot))=\\&[1-f][R(\tilde{\theta},x^{nc}_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))-\omega(\tilde{\theta},\rho,x^{nc}_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))]-C(x^{nc}_{\tilde{\theta}}(\cdot))-F.\end{aligned}$$

Re-arranging so that the difference between the payoffs from compliance and non-compliance equal zero and reintroducing the reduced form profit function from (3.1) we have

$$PR(\tilde{\theta}, x^{c}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - PR(\tilde{\theta}, x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) + \zeta(\tilde{\theta}, \rho, x^{c}_{\theta}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) + f\ R(\tilde{\theta}, x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))$$
$$+ [1-f]\ \omega(\tilde{\theta}, \rho, x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) + F = 0 = \psi_3(\tilde{\theta}, \rho, x^{c}_{\tilde{\theta}}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot), F, f), \tag{3.12}$$

where $\psi_3(\tilde{\theta}, \rho, x^{c}_{\tilde{\theta}}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot), F, f) = 0 = \psi_3(\cdot)$ implicitly defines the indifferent DC, $\tilde{\theta}(\rho, x^{c}_{\tilde{\theta}}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot), F, f)$, which we denote by $\tilde{\theta}(\cdot)$. Our first theorem defines which DCs comply with DPR.

Theorem 3.1 *DCs that are more capable than $\tilde{\theta}(\cdot)$ comply with DPR.*

Proof: Re-arranging the terms and differentiating (3.12) with respect to $\tilde{\theta}(\cdot)$ yields

$$\frac{\partial\psi_3(\cdot)}{\partial\tilde{\theta}} = \frac{\partial PR(\tilde{\theta}(\cdot), x^{c}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))}{\partial\tilde{\theta}} - \frac{\partial PR(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))}{\partial\tilde{\theta}}$$
$$+ \frac{\partial\zeta(\tilde{\theta}(\cdot), \rho, x^{c}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))}{\partial\tilde{\theta}} + [1-f]\frac{\partial\omega(\tilde{\theta}(\cdot), \rho, x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))}{\partial\tilde{\theta}} + f\frac{\partial R(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}, \vec{x}_{\setminus\tilde{\theta}})}{\partial\tilde{\theta}}. \tag{3.13}$$

Treating $x^{c}_{\tilde{\theta}}(\rho, \vec{x}) - x^{nc}_{\tilde{\theta}}(\rho, \vec{x}, f)$ as a small increase in x from (3.9), we can rewrite the first two terms above as $\partial^2 PR(\cdot)/\partial\tilde{\theta}\partial x_{\theta}$, which is positive by Assumption 3.1(c). From Assumption 3.1(a) the last term is positive as well. Together the third and fourth terms are positive for all non-negative values of f from Assumption 3.4(b). Consequently, $\partial\psi_3(\cdot)/\partial\tilde{\theta} > 0$, and only DCs with $\theta \geqslant \tilde{\theta}(\cdot)$ comply with DPR. □

Our Theorem 3.1 yields the following corollary which we make use of in our analyses.

Corollary 3.1 *The increases in payoffs from compliance from increases in capability is greater than that from non-compliance.*

Proof: (3.13) can be rewritten as $\partial\Pi^{c}(\cdot)/\partial\tilde{\theta} - \partial\Pi^{nc}(\cdot)/\partial\tilde{\theta}$ where arguments in the payoffs are as per (3.2) and (3.5) except with outputs at optimal value functions, and which from the proof of Theorem 3.1 is positive. Theorem 3.1 holds for all values of $\tilde{\theta}(\cdot)$, thus all values of θ. Hence,

$$\frac{\partial\Pi^{c}(\cdot)}{\partial\theta} > \frac{\partial\Pi^{nc}(\cdot)}{\partial\theta}. \quad \square \tag{3.14}$$

To discern the intuition behind Theorem 3.1 and Corollary 3.1, consider the case where variable fines are zero, so that only fixed fines are in effect. Figure 3.4 provides an illustrative example of DC payoffs from compliance and non-compliance, $\Pi^{c}(\cdot)$ and $\Pi^{nc}(\cdot)$ from (3.2) and (3.5), respectively, profits pre-DPR ($PR(\cdot)$), the change in revenues from compliance ($\zeta(\cdot)$), losses in revenue from non-compliance ($\omega(\cdot)$), all which include outputs as optimal value functions, and fixed fines (F). The intersection of $\Pi^{c}(\cdot)$ and $\Pi^{nc}(\cdot)$

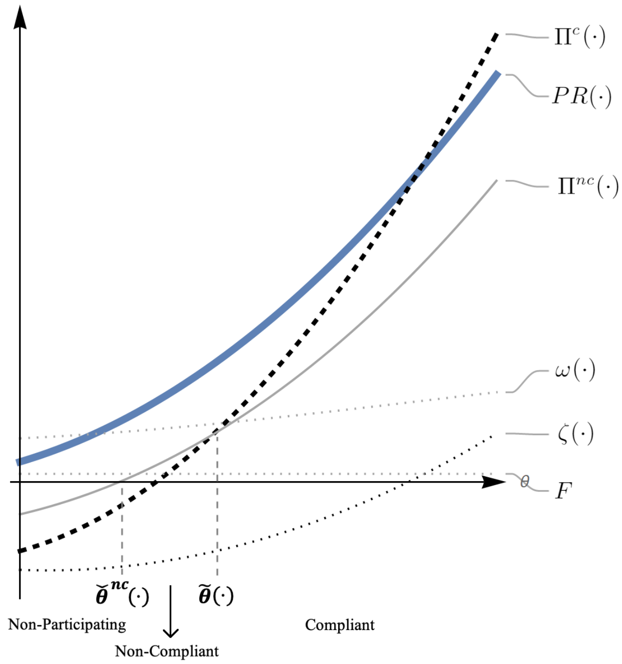


Figure 3.4: DC payoff from compliance and non-compliance: An illustrative example.

defines $\tilde{\theta}(\cdot)$. Because the compliant port function increases more than the non-compliant port function with an increase in capability (Assumption 3.4(b)), the payoff from compliance increases more with capability than does the payoff from non-compliance (Corollary 3.1). Therefore, the payoff from compliance for any DC that is more capable than $\tilde{\theta}(\cdot)$ is higher than its payoff from non-compliance, as shown in Figure 3.4.

With a positive variable fine, a proportion $f \in [0, 1]$ of net revenues from non-compliance is transferred to the policy-maker. In addition, the variable fine causes a decrease in output from non-compliance, and a corresponding decrease in payoffs. Together, this implies that the imposition of a variable fine decreases the payoff from non-compliance further from what is illustrated in Figure 3.4, thereby moving $\tilde{\theta}(\cdot)$ further to the left.

We can now define the proportion of DCs that comply as the set of DCs between $\tilde{\theta}(\cdot)$ and 1, so that with $\theta \sim U[0, 1]$ we can determine $\rho(\tilde{\theta}(\cdot)) = 1 - \tilde{\theta}(\cdot)$ and $d\rho(\cdot)/d\tilde{\theta} = -1 < 0$. With the proportion of DCs that comply defined, we can more fully explain our assumptions regarding the effects of $\rho(\cdot)$ on our port functions, Assumptions 3.2(c) and (e) and Assumptions 3.3(d) and (f). Consider our compliant port function, $\zeta(\cdot)$, for a given θ that complies, that is, $\theta > \tilde{\theta}(\cdot)$. As the proportion of DCs that comply increases, $\tilde{\theta}(\cdot)$ decreases. Thus, there is a potential gain in revenues for θ as a result of DSs porting their data from lower capability DCs that comply, both in total and at the margin. Next, consider the non-compliant port function, $\omega(\cdot)$. For a given θ that does not comply, as the proportion of DCs that comply increases there are more DCs with the portability feature that is attractive to DSs. Thus, there is a further loss of revenues in total and at the margin.

We now further explore the DC compliance decision by analyzing the impact of fixed and variable fines.

Lemma 3.3 *Both fixed and variable fines increase the proportion of compliant DCs.*

Proof: Differentiating (3.12) with respect to variable fines and simplifying using the envelope theorem yields

$$\frac{\partial \psi_3(\cdot)}{\partial f} = R(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - \omega(\tilde{\theta}(\cdot), \rho(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) > 0.$$

Then, using (3.13), and by the implicit function theorem

$$\frac{\partial \tilde{\theta}(\cdot)}{\partial f} = -\frac{\partial \psi_3(\cdot)/\partial f}{\partial \psi_3(\cdot)/\partial \tilde{\theta}} < 0. \tag{3.15}$$

Now, differentiating (3.12) with respect to fixed fines yields $\partial\psi_3(\cdot)/\partial F = 1$. By the implicit function theorem, it is straightforward that

$$\frac{\partial \tilde{\theta}(\cdot)}{\partial F} = -\frac{\partial \psi_3(\cdot)/\partial F}{\partial \psi_3(\cdot)/\partial \tilde{\theta}} < 0.$$

The result follows from $d\rho(\cdot)/d\tilde{\theta} < 0$. $\square$

Thus, for exposition we write $\partial\rho(\cdot)/\partial F, \partial\rho(\cdot)/\partial f > 0$. Using (3.15) and the above equation, we can quantify the relative effect of fixed and variable fines on compliance,

$$\frac{\partial \tilde{\theta}(\cdot)}{\partial f} = [R(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - \omega(\tilde{\theta}(\cdot), \rho(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))]\frac{\partial \tilde{\theta}(\cdot)}{\partial F}. \tag{3.16}$$

The impact of an increase in the fixed or variable fine is to decrease the payoffs for DCs that do not comply with DPR, whereas payoffs for DCs that comply are unaffected. Thus, the indifferent DC has lower capability, which results in a larger proportion of compliant DCs. The effect of the fixed fine can be inferred from Figure 3.4 - any increase in the fine decreases the payoffs from non-compliance, shifting the indifferent DC, $\tilde{\theta}(\cdot)$, to the left.

Participation

Here we analyze the mechanisms through which fines to enforce DPR impact DC participation, that is, the proportion of DCs that make positive payoffs. We denote the DC that generates zero payoffs from non-compliance by $\check{\theta}^{nc}$ - this DC is determined by setting (3.5) where outputs are stated as optimal value functions equal to zero so that

$$
\begin{aligned}
&\Pi^{nc}(\check{\theta}^{nc}, \rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot), F, f) = \\
&\quad [1-f][R(\check{\theta}^{nc}, x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot)) - \omega(\check{\theta}^{nc}, \rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot))] - C(x^{nc}_{\check{\theta}^{nc}}(\cdot)) - F = 0 \\
&\quad = \psi_4(\check{\theta}^{nc}, \rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot), F, f),
\end{aligned}
\tag{3.17}
$$

where $\psi_4(\check{\theta}^{nc}, \rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot), F, f)$ implicitly defines $\check{\theta}^{nc}(\rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot), F, f)$.

On the other hand, the compliant DC that is indifferent between participating and not participating also realizes zero net payoffs. This indifferent DC, $\check{\theta}^{c}$, is determined by setting (3.2) equal to zero but with outputs stated as optimal value functions so that

$$
\begin{aligned}
&\Pi^{c}(\check{\theta}^{c}, \rho(\cdot), x^{c}_{\check{\theta}^{c}}(\cdot), \vec{x}_{\setminus\check{\theta}^{c}}(\cdot)) = \\
&\quad R(\check{\theta}^{c}, x^{c}_{\check{\theta}^{c}}(\cdot), \vec{x}_{\setminus\check{\theta}^{c}}(\cdot)) + \zeta(\check{\theta}^{c}, \rho(\cdot), x^{c}_{\check{\theta}^{c}}(\cdot), \vec{x}_{\setminus\check{\theta}^{c}}(\cdot)) - C(x^{c}_{\check{\theta}^{c}}(\cdot)) = 0 = \psi_5(\check{\theta}^{c}, \rho(\cdot), x^{c}_{\check{\theta}^{c}}(\cdot), \vec{x}_{\setminus\check{\theta}^{c}}(\cdot)),
\end{aligned}
\tag{3.18}
$$

where $\psi_5(\check{\theta}^{c}, \rho(\cdot), x^{c}_{\check{\theta}^{c}}(\cdot), \vec{x}_{\setminus\check{\theta}^{c}}(\cdot)) = 0$ implicitly defines $\check{\theta}^{c}(\rho(\cdot), x^{c}_{\check{\theta}^{c}}(\cdot), \vec{x}_{\setminus\check{\theta}^{c}}(\cdot))$.

We begin by characterizing participation of DCs that do not comply using $\check{\theta}^{nc}(\cdot)$ in Lemmas 3.4 and 3.5.

Lemma 3.4 *DCs that are more capable than $\check{\theta}^{nc}(\cdot)$ generate positive payoffs from non-compliance.*

Proof: Partially differentiating (3.17) with respect to $\check{\theta}^{nc}(\cdot)$ yields

$$
\frac{\partial \psi_4(\cdot)}{\partial \check{\theta}^{nc}} = [1-f]\left[\frac{\partial R(\check{\theta}^{nc}(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot))}{\partial \check{\theta}^{nc}} - \frac{\partial \omega(\check{\theta}^{nc}(\cdot), \rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot)}{\partial \check{\theta}^{nc}}\right] > 0,
\tag{3.19}
$$

which is signed from Assumptions 3.1(a) and 3.3(a), and consists of higher capability resulting in increased revenue and decreased losses from the non-compliant port function. Thus, non-compliant DCs that are more capable than $\check{\theta}^{nc}(\cdot)$ generate positive payoffs. □

Lemma 3.5 *Fines decrease the proportion of DCs that generate positive payoffs from non-compliance.*

Proof: Partially differentiating (3.17) with respect to the variable fine f we have

$$
\frac{\partial \psi_4(\cdot)}{\partial f} = -[R(\check{\theta}^{nc}(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot)) - \omega(\check{\theta}^{nc}(\cdot), \rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot))] < 0,
$$

which is the direct effect of fines resulting in a transfer from non-compliant DCs to the policy-maker. The second term captures the indirect effect of fines increasing the proportion of compliant DCs. Then, using

(3.19) and the above equation, and by the implicit function theorem,

$$\frac{\partial \breve{\theta}^{nc}(\cdot)}{\partial f} = -\frac{\partial \psi_4(\cdot)/\partial f}{\partial \psi_4(\cdot)/\partial \breve{\theta}^{nc}} > 0. \tag{3.20}$$

Next, partially differentiating (3.17) with respect to the fixed fine F yields $\partial\psi_4(\cdot)/\partial F = -1 < 0$. Using (3.19) and the above equation, and by the implicit function theorem,

$$\frac{\partial \breve{\theta}^{nc}(\cdot)}{\partial F} = -\frac{\partial \psi_4(\cdot)/\partial F}{\partial \psi_4(\cdot)/\partial \breve{\theta}^{nc}} > 0. \tag{3.21}$$

From Lemma 3.4, DCs that are less capable than $\breve{\theta}^{nc}(\cdot)$ generate negative payoffs from non-compliance. Therefore, any increase in $\breve{\theta}^{nc}(\cdot)$ as a result of increased variable or fixed fines results in a decrease in the proportion of DCs that can generate positive payoffs from non-compliance. □

The effect of fixed fines on $\breve{\theta}^{nc}(\cdot)$ can be seen in Figure 3.4, where the payoff from non-compliance ($\Pi^{nc}(\cdot)$) shifts down with an increase in the fixed fine. This decreases the payoffs from non-compliance and shifts $\breve{\theta}^{nc}(\cdot)$ to the right. Therefore, the proportion of DCs generating positive payoff from non-compliance decreases in the fixed fine. The addition of a variable fine augments these effects and further decreases the net payoff from non-compliance. From the equations in Lemma 3.5 we can quantify the relative effect of fixed and variable fines on participation,

$$\frac{\partial \breve{\theta}^{nc}(\cdot)}{\partial f} = [R(\breve{\theta}^{nc}(\cdot), x^{nc}_{\breve{\theta}^{nc}}(\cdot), \vec{x}_{\setminus \breve{\theta}^{nc}}(\cdot)) - \omega(\breve{\theta}^{nc}(\cdot), \rho(\cdot), x^{nc}_{\breve{\theta}^{nc}}(\cdot), \vec{x}_{\setminus \breve{\theta}^{nc}}(\cdot))]\frac{\partial \breve{\theta}^{nc}(\cdot)}{\partial F}. \tag{3.22}$$

Broadly, both fixed and variable fines decrease payoffs from non-compliance. However, because less capable DCs pay the same fixed fine as more capable DCs, fixed fines disproportionately affect less capable DCs. In other words, the impact of variable fines on the extent to which participation decreases is partly tempered by $R(\cdot) + \omega(\cdot)$, the (low) net revenues from non-compliance of the least capable participating DC. The same is not the case with fixed fines.

Next we characterize participation of DCs that do comply using $\breve{\theta}^{c}(\cdot)$ in Lemma 3.6.

Lemma 3.6 *DCs that are more capable than $\breve{\theta}^{c}(\cdot)$ generate positive payoffs from compliance.*

Proof: Partially differentiating (3.18) with respect to $\breve{\theta}^{c}(\cdot)$ yields

$$\frac{\partial \psi_5(\cdot)}{\partial \breve{\theta}^{c}} = \left[\frac{\partial R(\breve{\theta}^{c}(\cdot), x^{c}_{\breve{\theta}^{c}}(\cdot), \vec{x}_{\setminus \breve{\theta}^{c}}(\cdot))}{\partial \breve{\theta}^{c}} + \frac{\partial \zeta(\breve{\theta}^{c}(\cdot), \rho(\cdot), x^{c}_{\breve{\theta}^{c}}(\cdot), \vec{x}_{\setminus \breve{\theta}^{c}}(\cdot))}{\partial \breve{\theta}^{c}}\right] > 0. \tag{3.23}$$

Each of the terms above are positive from Assumptions 3.1(a) and 3.2(a), and represent higher capability resulting in increased revenue and increased gains from the compliant port function. Thus, $\partial\psi_5(\cdot)/\partial\check{\theta}^c > 0$, and because $\check{\theta}^c(\cdot)$ generates zero payoffs from compliance, DCs that are more capable than $\check{\theta}^c(\cdot)$ generate positive payoffs from compliance. □

Unlike DCs that do not comply, DCs that comply do not pay fines and therefore there is no direct effect of fines on their participation.

In our next Lemma, we place $\check{\theta}^{nc}(\cdot), \check{\theta}^c(\cdot)$, and $\tilde{\theta}(\cdot)$ relative to each other on the capability line.

Lemma 3.7 *There are only two possible orderings for $\check{\theta}^{nc}(\cdot)$, $\check{\theta}^c(\cdot)$, and $\tilde{\theta}(\cdot)$.*

Case (a) : $\check{\theta}^{nc}(\cdot) < \check{\theta}^c(\cdot) < \tilde{\theta}(\cdot)$;

Case (b) : $\tilde{\theta}(\cdot) \leqslant \check{\theta}^c(\cdot) \leqslant \check{\theta}^{nc}(\cdot)$.

Proof: From Corollary 3.1, the DCs' payoffs from compliance increase faster with capability than do payoffs from non-compliance. Thus, the condition that defines $\tilde{\theta}(\cdot)$, equal payoffs in (3.12), defines $\tilde{\theta}(\cdot)$ uniquely. This is the single crossing condition.

Case (a): If (3.12) holds where payoffs are positive, then using Corollary 3.1, $\check{\theta}^{nc}(\cdot) < \check{\theta}^c(\cdot) < \tilde{\theta}(\cdot)$.

Case (b): If (3.12) holds where payoffs are (weakly) negative, then again using Corollary 3.1, $\tilde{\theta}(\cdot) \leqslant \check{\theta}^c(\cdot) \leqslant \check{\theta}^{nc}(\cdot)$. □

In what follows we refer to Case (a) as the interior solution and to Case (b) as the corner solution. Figure 3.5 illustrates the impact of increased fines on DC compliance and participation. We consider an initial regime with low-fines l and then a high-fines regime h. Using subscripts to denote the regime, the intersection of $\Pi_l^{nc}(\cdot)$ and the horizontal axis (where payoff is zero) gives $\check{\theta}_l^{nc}(\cdot)$ - the DC that generates zero payoffs from non-compliance, as seen in (3.17). By Lemma 3.4, DCs that are more capable than $\check{\theta}_l^{nc}(\cdot)$ generate positive payoffs from non-compliance. Similarly, the intersection of $\Pi^c(\cdot)$ with the horizontal axis gives $\check{\theta}^c(\cdot)$, which denotes the DC that generates zero payoffs from compliance as in (3.18). DCs that are more capable than $\check{\theta}^c(\cdot)$ generate positive payoffs if they comply with DPR by Lemma 3.6. Of central importance is the fact that payoffs from compliance increase faster with capability than do payoffs from non-compliance (Corollary 3.1). The intersection of $\Pi^c(\cdot)$ and $\Pi_l^{nc}(\cdot)$ gives the DC that is indifferent between complying and not complying under the low-fines regime, $\tilde{\theta}^l(\cdot)$.

An increase in fixed fines decreases payoffs from non-compliance (Lemma 3.5) from $\Pi_l^{nc}(\cdot)$ to $\Pi_h^{nc}(\cdot)$. As fines increase, $\tilde{\theta}(\cdot)$ moves to the left (Lemma 3.3), the payoffs to $\tilde{\theta}(\cdot)$ decrease, and $\check{\theta}^{nc}(\cdot)$ moves to the right (Lemma 3.5). With sufficient increases in fines, the corner solution is reached where payoffs to $\tilde{\theta}(\cdot)$ are zero ($\Pi_h^{nc}(\cdot)$ in Figure 3.5). By definition, at this point $\check{\theta}^{nc}(\cdot) = \check{\theta}^c(\cdot) = \tilde{\theta}(\cdot)$. Any further increases in fines

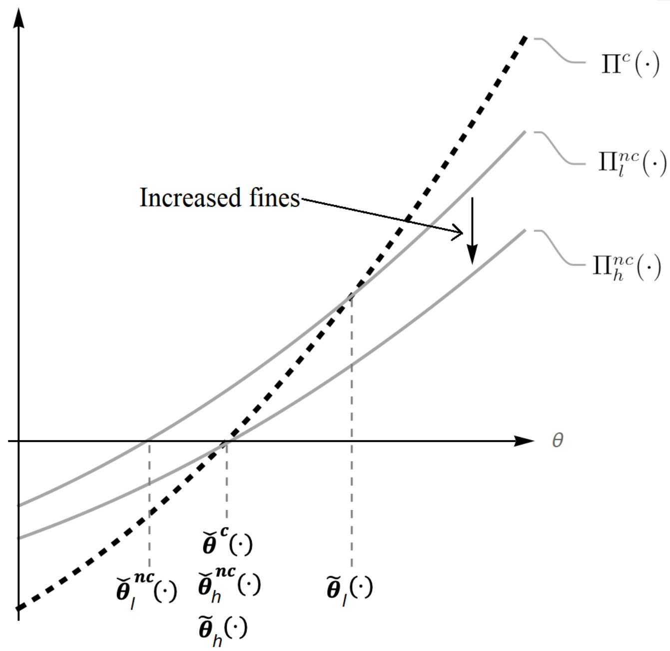


Figure 3.5: Impact of fines on compliance and participation: An illustrative example.

continue to have the same effects - $\tilde{\theta}(\cdot)$ moves further to the left and $\breve{\theta}^{nc}(\cdot)$ moves to the right. Then, we get the ordering $\tilde{\theta}(\cdot) < \breve{\theta}^c(\cdot) < \breve{\theta}^{nc}(\cdot)$, and obtain negative payoffs for the indifferent DC $\tilde{\theta}$, the corner solution.

Using these two cases, we define how the DC industry is segmented in our next Theorem.

Theorem 3.2 *In Case (a) DCs are segmented by capability into non-participating DCs, participating non-compliant DCs, and participating compliant DCs. In Case (b) DCs are segmented by capability into non-participating DCs, and participating compliant DCs. In Case (b) no non-compliant DCs participate.*

Proof: Consider first the interior solution, Case (a), where the payoff for the indifferent DC $\tilde{\theta}(\cdot)$ is positive. In this case, $\breve{\theta}^{nc}(\cdot) < \breve{\theta}^c(\cdot) < \tilde{\theta}(\cdot)$ by Lemma 3.7. Then, the least capable DCs, $\theta < \breve{\theta}^{nc}(\cdot)$, do not participate because of negative payoffs. The set of DCs defined by $\breve{\theta}^{nc}(\cdot) < \theta < \tilde{\theta}(\cdot)$ participate but do not comply because they make positive payoffs from non-compliance, and such payoffs are larger than their payoffs from compliance. Finally, DCs with $\tilde{\theta}(\cdot) < \theta$ participate and comply because their payoffs from compliance are positive, and larger than their payoffs from non-compliance.

Next, consider the corner solution, Case (b), where the payoffs for the indifferent DC $\tilde{\theta}(\cdot)$ are negative, and $\tilde{\theta}(\cdot) \leqslant \breve{\theta}^c(\cdot) \leqslant \breve{\theta}^{nc}(\cdot)$ by Lemma 3.7. Here, DCs with $\theta \leqslant \breve{\theta}^c(\cdot)$ do not participate because of negative payoffs, but DCs with $\breve{\theta}^c(\cdot) < \theta$ participate and comply because they have positive payoffs from compliance, and such payoffs are larger than those that can be achieved through non-compliance. $\square$

This segmentation for the interior solution is illustrated in Figure 3.6, where, by Theorem 3.1, compliant

DCs lie on the right side of $\tilde{\theta}(\cdot)$, and non-compliant DCs lie on the left side of $\tilde{\theta}(\cdot)$. If payoffs to non-compliance are positive, then the interior solution (Theorem 3.2 Case (a)) is obtained. Here, DCs between 0 and $\check{\theta}^{nc}(\cdot)$ do not participate because of negative payoffs, DCs between $\check{\theta}^{nc}(\cdot)$ and $\tilde{\theta}(\cdot)$ participate but do not comply with DPR, and DCs between $\tilde{\theta}(\cdot)$ and 1 participate and comply with DPR. Lemma 3.3 shows that fines shift $\tilde{\theta}(\cdot)$ to the left because fines increase the attractiveness of compliance as compared to non-compliance. Further, Lemma 3.5 shows that increasing fines shifts $\check{\theta}^{nc}(\cdot)$ to the right. This is because an increase in fines decreases payoffs to non-compliant DCs.

This segmentation for the corner solution is shown in Figure 3.7. Corner solutions occur when payoffs to $\tilde{\theta}(\cdot)$ are weakly negative, and are characterized by several properties, as described in Theorem 3.2 Case (b). Firstly, industry segmentation pivots around $\check{\theta}^{c}(\cdot)$, so that DCs that are more capable than $\check{\theta}^{c}(\cdot)$ participate and comply, while DCs that are less capable than $\check{\theta}^{c}(\cdot)$ do not participate. Thus, corner solutions are characterized by the absence of non-compliant participating DCs. As well, once fees are increased sufficiently for the corner solution to be reached, fines cease to have any effect on compliance or participation. With any further increase in fines, $\check{\theta}^{nc}(\cdot)$ moves further to the right and $\tilde{\theta}(\cdot)$ moves further to the left; but because $\check{\theta}^{c}(\cdot)$ remains unchanged, the industry segmentation is unaltered. So it is with output - fines cease to affect output of participating and compliant DCs or participation choices of non-participating DCs at the corner.

Directly from Theorem 3.2, the analyses of the corner solution, our Case (b) is straightforward because fines cease having any effect on the equilibrium. A special case of the corner solution is one where all DCs participate and comply, as we describe in our next theorem.

Theorem 3.3 *A necessary and sufficient condition for full compliance and full participation is a (weakly) positive change in revenues from compliance for the least capable DC.*

Proof: First consider the case where the change in revenues from compliance for the least capable DC is negative, so that $\zeta(\theta = 0, \rho(\cdot), x^c_{\theta=0}(\cdot), \vec{x}_{\backslash\theta=0}(\cdot)) < 0$. Then, payoffs from compliance to the least capable DC captured in (3.2) are negative because $PR(\theta = 0, \vec{x}(\cdot)) = 0$ as described prior to Assumption 3.1, and so the least capable DC cannot comply profitably. Thus, $\zeta(\theta = 0, \rho(\cdot), x^c_{\theta=0}(\cdot), \vec{x}_{\backslash\theta=0}(\cdot)) \geqslant 0$ is necessary for full compliance.

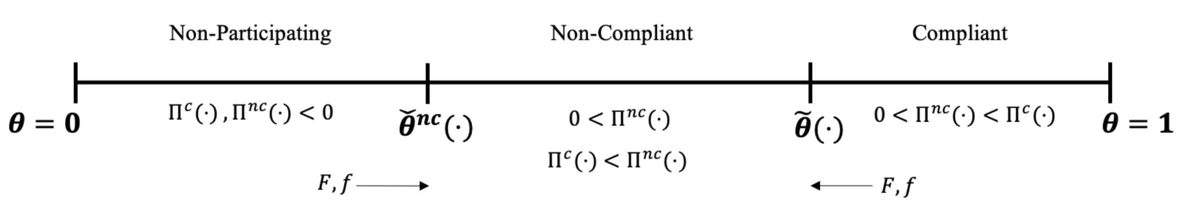


Figure 3.6: DC decision: An illustrative example of the interior solution.

We now show that $\zeta(\theta = 0, \rho(\cdot), x^c_{\theta=0}(\cdot), \vec{x}_{\backslash\theta=0}(\cdot)) \geqslant 0$ is sufficient. The payoffs from non-compliance for the least capable DC are negative because $PR(\theta = 0, \vec{x}(\cdot)) = 0$, and with the loss in revenues from non-compliance, $\omega(\theta, \rho(\cdot), \vec{x}(\cdot)) \in \mathbb{R}_{>0}$, (3.5) is negative. Because $\partial\Pi^c(\cdot)/\partial\theta > \partial\Pi^{nc}(\cdot)/\partial\theta$ from Corollary 3.1, the payoff from compliance increases more with capability than does the payoff from non-compliance. If the least capable DC generates negative payoffs from non-compliance and positive payoffs from compliance, then by Corollary 3.1, the payoffs from compliance are larger than the payoffs from non-compliance $\forall\theta > 0$, and all DCs comply. Further, with $\zeta(\theta = 0, \rho(\cdot), x^c_{\theta=0}(\cdot), \vec{x}_{\backslash\theta=0}(\cdot)) > 0$ and increasing in θ by Assumption 3.2, (3.2) is positive $\forall\theta > 0$, and there is full participation. $\square$

Full compliance and full participation is a special case of the corner solution described in Figure 3.7, but with $\check{\theta}^c(\cdot) = 0$. This requires that the change in revenues from compliance for the least capable DC is positive, and because this change in revenue is increasing in θ, and increasing faster than the non-compliant port function, there is full compliance and participation. The change in revenue from compliance can be high for example when DSs port their data from other industries to the focal industry captured by our capability line. If such incoming DSs have a strong preference for DCs that enable portability, then this can further increase the change in DC revenue from compliance.

For the remainder of our analysis we focus on an interior solution, Case (a) of Theorem 3.2. We begin by formalizing the parallel effect of fines on compliance and participation. Our next theorem shows that although fines increase compliance, they also collaterally decrease participation.

Theorem 3.4 *If $\check{\theta}^{nc}(\cdot) < \tilde{\theta}(\cdot)$ so that an interior solution (Case (a) of Theorem 3.2) holds, then any increase in compliance as a result of fines is accompanied by a decrease in participation.*

Proof: From Theorem 3.2 Case (a), the interior solution is characterized by $0 < \check{\theta}^{nc}(\cdot) < \tilde{\theta}(\cdot) < 1$. By Lemma 3.3, $\partial\tilde{\theta}(\cdot)/\partial f < 0$ and $\partial\tilde{\theta}(\cdot)/\partial F < 0$, so any increase in fines increases DC compliance. From Lemma 3.5, $\partial\check{\theta}^{nc}(\cdot)/\partial f > 0$ and $\partial\check{\theta}^{nc}(\cdot)/\partial F > 0$, and any increase in fines decreases participation. $\square$

3.4.3 Stage 1: Fine Imposition

In our setup, DS Surplus (DSS) is a measure of consumer surplus, and is increasing in aggregate output, $X(F, f) \in \mathbb{R}_{>0}$ that we define below, so that $\partial DSS(X(\cdot))/\partial X > 0$. DC surplus (DCS) is a measure of

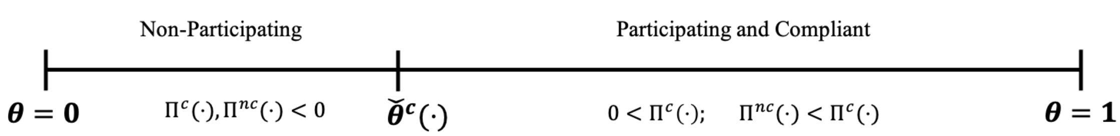


Figure 3.7: DC decision: An illustrative example of the corner solution.

producer surplus and is the aggregate payoffs to DCs after transfers to governments (such as fines). Fines are a transfer and do not affect social welfare directly. Therefore, for the purposes of calculating social welfare, we develop a measure of DCS without fines, DCS_{-f}. In this section, we first analyze the effect of fines on each of DSS and DCS, and then analyze the effect of fines on social welfare. We begin with DS surplus through output.

Aggregate Output

Aggregate output consists of the aggregate output of DCs that comply and the aggregate output of DCs that do not comply but that participate.

$$X(F,f) = \int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} x_{\theta}^{nc}(\cdot)d\theta + \int_{\tilde{\theta}(\cdot)}^{1} x_{\theta}^{c}(\cdot)d\theta. \tag{3.24}$$

The DCs that do not participate, $\theta \in [0, \check{\theta}^{nc}(\cdot)]$, do not produce any output. From the first term in (3.24), DCs between $\check{\theta}^{nc}(\cdot)$ and $\tilde{\theta}(\cdot)$ participate but do not comply with DPR so individual DC output is $x_{\theta}^{nc}(\cdot)$. From the second term in (3.24), the DCs between $\tilde{\theta}(\cdot)$ and 1 participate and comply with DPR so individual DC output is $x_{\theta}^{c}(\cdot)$.

We now totally differentiate aggregate output with respect to fixed fines by applying Leibnitz's rule, and drop the terms that are zero, to get

$$\frac{dX(\cdot)}{dF} = [x_{\tilde{\theta}}^{nc}(\cdot) - x_{\tilde{\theta}}^{c}(\cdot)]\frac{\partial\tilde{\theta}(\cdot)}{\partial F} - x_{\check{\theta}^{nc}}^{nc}(\cdot)\frac{\partial\check{\theta}^{nc}(\cdot)}{\partial F} + \int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \frac{\partial x_{\theta}^{nc}(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F}d\theta + \int_{\tilde{\theta}(\cdot)}^{1} \frac{\partial x_{\theta}^{c}(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F}d\theta. \tag{3.25}$$

From Lemma 3.3, $\partial\tilde{\theta}(\cdot)/\partial F < 0$, because an increase in fixed fines causes some DCs to switch from non-compliance to compliance. Such DCs have higher output from (3.9). A fine induced switch to compliance brings with it an increase in output, so the first term is positive. The second term is negative because fines decrease participation, $\partial\check{\theta}^{nc}(\cdot)/\partial F > 0$ by Lemma 3.5. This term captures lost output because the least capable DCs cease to participate as a result of increased fines. Next, in the third and fourth terms, $\partial\rho(\cdot)/\partial F > 0$ because $\partial\tilde{\theta}(\cdot)/\partial F < 0$ by Lemma 3.3 and recalling that $\rho(\cdot) = 1 - \tilde{\theta}(\cdot)$. However, this increase in compliance decreases non-compliant DCs' output through the third term (Lemma 3.2) and increases compliant DCs' output through the fourth term (Lemma 3.1).

Next, we totally differentiate (3.24) with respect to variable fines, to get

$$
\begin{aligned}
\frac{dX(\cdot)}{df} =& [x^{nc}_{\tilde{\theta}}(\cdot) - x^{c}_{\tilde{\theta}}(\cdot)]\frac{\partial\tilde{\theta}(\cdot)}{\partial f} - x^{nc}_{\check{\theta}^{nc}}(\cdot)\frac{\partial\check{\theta}^{nc}(\cdot)}{\partial f} \\
&+ \int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \left[\frac{\partial x^{nc}_{\theta}(\cdot)}{\partial \rho}\frac{\partial \rho(\cdot)}{\partial f} + \frac{\partial x^{nc}_{\theta}(\cdot)}{\partial f}\right] d\theta + \int_{\tilde{\theta}(\cdot)}^{1} \frac{\partial x^{c}_{\theta}(\cdot)}{\partial \rho}\frac{\partial \rho(\cdot)}{\partial f} d\theta. \qquad (3.26)
\end{aligned}
$$

Compared to fixed fines, variable fines have an additional effect on aggregate output. This is an infra-marginal effect captured by the second term within the second set of square brackets, $\partial x^{nc}_{\theta}(\cdot)/\partial f$. This term is negative from Lemma 3.2 - variable fines decreases output for each non-compliant DC that continues to participate. Thus, the term under the first integration sign is negative.

In summary, there are four mechanisms through which fixed fines affect DSS. First, fines decrease DC participation, which in turn leads to lost output from DCs that cease to participate. Second, an increase in fines converts some non-compliant DCs into compliant DCs. This eliminates the fine externality for these DCs and leads to a higher output. Third, fines increase compliance, which leads to a decrease in output from non-compliant DCs and an increase in output from compliant DCs. Fourth, variable fines have the additional effect of decreasing the output of participating non-compliant DCs, leading to decreased output from all participating non-compliant DCs. Thus, fines decrease aggregate output from non-compliant DCs, $\theta \leqslant \tilde{\theta}(\cdot)$, and increase the aggregate output from compliant DCs, $\tilde{\theta}(\cdot) < \theta$.

We now compare fixed and variable fines as tools to achieve compliance.

Theorem 3.5 *If $\check{\theta}^{nc}(\cdot) < \tilde{\theta}(\cdot)$ so that an interior solution (Case (a) of Theorem 3.2) holds, then relative to variable fines, fixed fines lead to lower participation.*

Proof: Multiplying (3.25) by $[R(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - \omega(\tilde{\theta}(\cdot), \rho(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))]$, and substituting the right-hand side of each of (3.16) and (3.22) with its respective left-hand side, we get

$$
\begin{aligned}
&[R(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - \omega(\tilde{\theta}(\cdot), \rho(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))]\frac{\partial X(\cdot)}{\partial F} = [x^{nc}_{\tilde{\theta}}(\cdot) - x^{c}_{\tilde{\theta}}(\cdot)]\frac{\partial\tilde{\theta}(\cdot)}{\partial f} \\
&- x^{nc}_{\check{\theta}^{nc}}(\cdot)\left[\frac{R(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - \omega(\tilde{\theta}(\cdot), \rho(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))}{R(\check{\theta}^{nc}(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot)) - \omega(\check{\theta}^{nc}(\cdot), \rho(\cdot), x^{nc}_{\check{\theta}^{nc}}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot))}\right]\frac{\partial\check{\theta}^{nc}(\cdot)}{\partial f} \\
&+ [R(\tilde{\theta}(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - \omega(\tilde{\theta}(\cdot), \rho(\cdot), x^{nc}_{\tilde{\theta}}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot))]\left[\int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \frac{\partial x^{nc}_{\theta}(\cdot)}{\partial \rho}\frac{\partial \rho(\cdot)}{\partial F} d\theta + \int_{\tilde{\theta}(\cdot)}^{1} \frac{\partial x^{c}_{\theta}(\cdot)}{\partial \rho}\frac{\partial \rho(\cdot)}{\partial F} d\theta\right].
\end{aligned}
$$

Now comparing the above equation with (3.26), the first terms in the right-hand side in both equations are identical, indicating the same level of compliance. The second terms in each of these equations indicates the collateral loss of participation suffered in order to achieve the level of compliance observed in the first term. Because the level of compliance achieved in each equation is the same, we can compare the second terms to compare the collateral effect on participation through variable and fixed fines.

The term in the first set of square brackets above is greater than unity because the net revenues of $\tilde{\theta}(\cdot)$ in the numerator are larger than the net revenues of $\check{\theta}^{nc}(\cdot)$ in the denominator. Therefore the second term in the above equation is larger than the second term in (3.26), and so the use of fixed fines instead of variable fines to achieve a given level of compliance leads to lower participation. $\square$

Separately, the direct effect of variable fines on non-compliant DC output in the second term in the second set of square brackets in (3.26) is negative. As can be seen in the above equation, fixed fines do not have this effect.

DC surplus

DCS is the aggregate payoff to compliant and non-compliant DCs, so that

$$DCS(F,f) = \int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \Pi^{nc}(\theta, x_{\theta}^{nc}(\cdot), \vec{x}_{\backslash\theta}(\cdot), F, f)d\theta + \int_{\tilde{\theta}(\cdot)}^{1} \Pi^{c}(\theta, x_{\theta}^{c}(\cdot), \vec{x}_{\backslash\theta}(\cdot))d\theta. \tag{3.27}$$

The payoffs from non-compliance consist of profits net of the non-compliant port function and fines as defined in (3.5), while the payoffs from compliance consist of profits plus the compliant port function as in (3.2). We now evaluate the effect of fixed fines and variable fines on DCS. Totally differentiating DCS with respect to fixed fines we have

$$\frac{dDCS(\cdot)}{dF} = -\int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \Big[[1-f]\frac{\partial\omega(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F} + 1\Big]d\theta + \int_{\tilde{\theta}(\cdot)}^{1} \frac{\partial\zeta(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F}d\theta. \tag{3.28}$$

Now totally differentiating DCS with respect to variable fines we have

$$\frac{dDCS(\cdot)}{df} = -\int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \Big[[1-f]\frac{\partial\omega(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial f} + [R(\cdot)-\omega(\cdot)]\Big]d\theta + \int_{\tilde{\theta}(\cdot)}^{1} \frac{\partial\zeta(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial f}d\theta. \tag{3.29}$$

Several terms cancel out or drop to zero leading to the above equations. First, payoffs from compliance and non-compliance for $\tilde{\theta}(\cdot)$ are the same by definition. Second, the decrease in participation with an increase in fines ($\partial\check{\theta}^{nc}(\cdot)/\partial f$), does not affect DCS because these DCs generate zero payoffs - thus, their non-participation has no effect on DCS. Third, the payoff function $\Pi^{nc}(\theta, x_{\theta}^{nc}(\cdot), \vec{x}_{\backslash\theta}(\cdot), F, f)$ contains $x_{\theta}^{nc}(\theta, \vec{x}_{\backslash\theta}(\cdot), f)$, which is an indirect maximal value function in variable fines, f. In other words, $\Pi^{nc}(\theta, x_{\theta}^{nc}(\cdot), \vec{x}_{\backslash\theta}(\cdot), F, f)$ is an envelope for $x_{\theta}^{nc}(\cdot)$, such that $\partial\Pi^{nc}(\cdot)/\partial x_{\theta}^{nc} = 0$ in order for $\Pi^{nc}(\cdot)$ to be maximal. Therefore, (3.29) quantifies the instantaneous rate of change in payoffs with respect to the variable fine, and is independent of the indirect effect of fines on payoffs through output. However, the decrease in output is a feature associated with the decrease in payoffs in a longer-term analysis (pp 190-197, Silberberg 1990).

In each of (3.28) and (3.29), the term under the second integral captures the increased compliance due to fines, which leads to increased changes in revenue from compliance. These gains apply to compliant (and more capable) DCs, $\tilde{\theta}(\cdot) < \theta \leqslant 1$. The first term under the first integral captures the increased compliance due to fines, which leads to increased losses in revenue from non-compliance. The second term under the first integration sign is the increased fines transferred to the policy-maker. In (3.29), this is $-[R(\cdot) + \omega(\cdot)]$ instead of -1 in (3.28), because the effect of an increase in variable fines depends on the net revenues from non-compliance. These losses apply to non-compliant (and less capable) DCs, $\check{\theta}^{nc}(\cdot) < \theta \leqslant \tilde{\theta}(\cdot)$.

Social Welfare

We defined DCS as the aggregate payoffs to DCs after subtracting fines. As such fines are a transfer, they do not impact social welfare. For this reason, we create a measure of DCS without fines,

$$\begin{aligned} DCS_{-f}(F,f) = &\int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} [R(\theta, x_\theta^{nc}(\cdot), \vec{x}_{\setminus\theta}, (\cdot)) - \omega(\theta, \rho(\cdot), x_\theta^{nc}(\cdot), \vec{x}_{\setminus\theta}(\cdot)) - C(x_\theta^{nc}(\cdot))]d\theta \\ &+ \int_{\tilde{\theta}(\cdot)}^{1} [R(\theta, x_\theta^{c}(\cdot), \vec{x}_{\setminus\theta}(\cdot)) + \zeta(\theta, \rho(\cdot), x_\theta^{c}(\cdot), \vec{x}_{\setminus\theta}(\cdot)) - C(x_\theta^{c}(\cdot))]d\theta, \end{aligned} \tag{3.30}$$

so that social welfare (SW) consists of DCS without fines, and DSS, $SW(F,f) = DCS_{-f}(F,f) + DSS(X(F,f))$. Thus, the effect of fixed fines on social welfare is $dSW(F,f)/dF = dDCS_{-f}(F,f)/dF + DSS'(X(F,f))dX(F,f)/dF$, which we calculate in the following equation by differentiating (3.30) with respect to the fixed fine, and substituting for $dX(F,f)/dF$ from (3.26),

$$\begin{aligned} \frac{dSW(F,f)}{dF} = &-\frac{\partial\check{\theta}^{nc}(\cdot)}{\partial F}[R(\check{\theta}^{nc}(\cdot), x_{\check{\theta}^{nc}}^{nc}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot)) - \omega(\check{\theta}^{nc}(\cdot), \rho(\cdot), x_{\check{\theta}^{nc}}^{nc}(\cdot), \vec{x}_{\setminus\check{\theta}^{nc}}(\cdot)) - C(x_{\check{\theta}^{nc}}^{nc}(\cdot))] \\ &+ \frac{\partial\tilde{\theta}(\cdot)}{\partial F}\Big[[R(\tilde{\theta}(\cdot), x_{\tilde{\theta}}^{nc}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - \omega(\tilde{\theta}(\cdot), \rho(\cdot), x_{\tilde{\theta}}^{nc}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - C(x_{\tilde{\theta}}^{nc}(\cdot))] \\ &\qquad - [R(\tilde{\theta}(\cdot), x_{\tilde{\theta}}^{c}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) + \zeta(\tilde{\theta}(\cdot), \rho(\cdot), x_{\tilde{\theta}}^{c}(\cdot), \vec{x}_{\setminus\tilde{\theta}}(\cdot)) - C(x_{\tilde{\theta}}^{c}(\cdot))]\Big] \\ &- \int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \frac{\partial\omega(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F}d\theta + \int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \frac{\partial\zeta(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F}d\theta \\ &+ DSS'(X(F,f))\Big[[x_{\tilde{\theta}}^{nc}(\cdot) - x_{\tilde{\theta}}^{c}(\cdot)]\frac{\partial\tilde{\theta}(\cdot)}{\partial F} - x_{\check{\theta}^{nc}}^{nc}(\cdot)\frac{\partial\check{\theta}^{nc}(\cdot)}{\partial F} \\ &\qquad + \int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)} \frac{\partial x_\theta^{nc}(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F}d\theta + \int_{\tilde{\theta}(\cdot)}^{1} \frac{\partial x_\theta^{c}(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial F}d\theta\Big]. \end{aligned} \tag{3.31}$$

The effect of variable fines on $SW(F,f)$ is

$$\begin{aligned}\frac{dSW(F,f)}{df} =& -\frac{\partial\check{\theta}^{nc}(\cdot)}{\partial f}[R(\check{\theta}^{nc}(\cdot),x^{nc}_{\check{\theta}^{nc}}(\cdot),\vec{x}_{\setminus\check{\theta}^{nc}}(\cdot))-\omega(\check{\theta}^{nc}(\cdot),\rho(\cdot),x^{nc}_{\check{\theta}^{nc}}(\cdot),\vec{x}_{\setminus\check{\theta}^{nc}}(\cdot))-C(x^{nc}_{\check{\theta}^{nc}}(\cdot))]\\ &+\frac{\partial\tilde{\theta}(\cdot)}{\partial f}\Big[[R(\tilde{\theta}(\cdot),x^{nc}_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))-\omega(\tilde{\theta}(\cdot),\rho(\cdot),x^{nc}_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))-C(x^{nc}_{\tilde{\theta}}(\cdot))]\\ &\qquad-[R(\tilde{\theta}(\cdot),x^{c}_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))+\zeta(\tilde{\theta}(\cdot),\rho(\cdot),x^{c}_{\tilde{\theta}}(\cdot),\vec{x}_{\setminus\tilde{\theta}}(\cdot))-C(x^{c}_{\tilde{\theta}}(\cdot))]\Big]\\ &-\int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)}\frac{\partial\omega(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial f}d\theta+\int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)}\frac{\partial\zeta(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial f}d\theta\\ &+DSS'(X(F,f))\Bigg[[x^{nc}_{\tilde{\theta}}(\cdot)-x^{c}_{\tilde{\theta}}(\cdot)]\frac{\partial\tilde{\theta}(\cdot)}{\partial f}-x^{nc}_{\check{\theta}^{nc}}(\cdot)\frac{\partial\check{\theta}^{nc}(\cdot)}{\partial f}\\ &\qquad+\int_{\check{\theta}^{nc}(\cdot)}^{\tilde{\theta}(\cdot)}[\frac{\partial x^{nc}_{\theta}(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial f}+\frac{\partial x^{nc}_{\theta}(\cdot)}{\partial f}]d\theta+\int_{\tilde{\theta}(\cdot)}^{1}\frac{\partial x^{c}_{\theta}(\cdot)}{\partial\rho}\frac{\partial\rho(\cdot)}{\partial f}d\theta\Bigg].\end{aligned}\tag{3.32}$$

The structure of (3.31) and (3.32) are similar. The first four lines of each equation contain the effect of fines on DCS_{-f}, and the last two lines are the effect of fines on DSS. The first terms in each of (3.31) and (3.32) capture decreases in participation and the resulting decrease in welfare. The second terms capture the marginal effect of an increase in compliance on $DCS_{-f}(F,f)$. The second line contains the payoffs to $\tilde{\theta}(\cdot)$ from non-compliance, but without subtracting fines, while the third line contains the payoffs to $\tilde{\theta}(\cdot)$ from compliance. From (3.12), after subtracting fines the payoffs to $\tilde{\theta}(\cdot)$ from non-compliance are equal to its payoffs from compliance. Thus, the second term is negative, and captures the decrease in welfare from adding a constraint of fines for non-compliance. The third and fourth terms capture the externalities caused by increased compliance due to fines. The third term captures the increased losses from non-compliance that are caused by increased compliance, while the fourth term captures the increased gains from compliance that are caused by increased compliance. The fifth term (on the fifth and sixth lines) is the change in DSS from a change in output caused by fines. This change in output occurs because of DCs' increased output when converting from non-compliance to compliance, lost output from DCs that cease to participate, decreased output from non-compliant DCs due to increased compliance and the fine externality, and increased output from compliant DCs due to increased compliance. We can now ascertain the condition under which it is welfare maximizing to eliminate non-compliance through the use of fines.

Theorem 3.6 *With increased fines, if the gains to social welfare from more capable DCs outweigh losses from less capable DCs, then eliminating non-compliance through fines is welfare maximizing.*

Proof: From (3.31) and (3.32), all the gains to SW from increased fines accrue from more capable DCs, $\tilde{\theta}(\cdot) < \theta$, while the losses to SW from increased fines are from less capable DCs, $\theta \leqslant \tilde{\theta}(\cdot)$. If the gains are larger than the losses, then (3.31) and (3.32) can be signed positive. By Theorem 3.4, fines increase compliance and decrease participation simultaneously, leading to the corner solution described in Theorem 3.2 Case (b), where all participating DCs are compliant. □

3.5 Discussion and Conclusion

We model the decisions of whether to comply with DPR by DCs that differ in their capability when a policy-maker has the ability to impose fixed and variable fines to enforce DPR. Whereas non-compliant DCs face a fixed and variable fine, they also face a loss of revenue from non-compliance due to DSs moving to DCs that comply with DPR, as portability is viewed by some DSs as an additional feature that is more valuable than the costs of switching DCs. On the other hand, DCs that comply can gain from DSs moving from non-compliant DCs but can also face a loss of DSs to more capable DCs.

Our model is formulated using general functions governed by curvature conditions and describes the choices by the policy-maker and incentives of DCs through a two-stage game. In the first stage the policy-maker sets a fixed and a variable fine for DCs that do not comply with DPR. In the second stage DCs choose whether to participate based on whether they can make positive payoffs, and participating DCs choose whether to comply with DPR. We find that DCs which are more capable comply with DPR, and that DCs which are less capable either do not comply or do not participate and drop out from the market altogether. Thus, for the policy-maker, increasing fines incentivizes non-compliant DCs to comply with DPR, but also decreases participation by forcing the least capable DCs to exit. Therefore, fines can be used to squeeze the set of non-compliant DCs from two sides - more capable non-compliant DCs are motivated to comply, whereas the least capable non-compliant DCs cease to participate. Fines cause a decrease in DC participation, thereby decreasing options for DSs to choose from.

If fines are sufficiently high, then a corner solution is reached where non-compliance is eliminated so that all DCs either participate and comply, or do not participate. Once this corner solution is reached, increases in fines cease to have any effect. An interesting corner solution is where all DCs participate and comply - this can occur when the changes to revenue from complying with portability are positive for all DCs. This can happen, for instance, with inter-industry porting where DSs from other industries port their data to the focal industry and when DSs highly value the feature of portability.

We also find that relative to variable fines, fixed fines disproportionately affect less capable DCs and lead to lower participation by DCs. Although less capable DCs with lower revenues face a proportionately small variable fine, the fixed fine can have a large effect on such DCs. A separate consequence of variable fines is a decrease in non-compliant DC output because variable fines act as an externality for such DCs, decreasing the marginal return of additional output – output that might serve additional DSs. Therefore, fines have the effect of decreasing the output and DC surplus from non-compliant (and thus, less capable) DCs, and they also have the effect of increasing output and DC surplus from compliant (and thus, more capable) DCs. If the gains to more capable DCs outweigh the losses from less capable DCs throughout the range of the fine, then increasing fines so as to eliminate non-compliance maximizes surplus. This outcome is possible in our setting if the value generated by more capable DCs is significantly greater than that generated by less capable DCs.

Although fines can increase welfare, this comes with collateral effects. First, fines force the least capable DCs to exit the market. The least capable DCs may be smaller and are often viewed by policy-makers as small businesses that are the source of innovation. Second, among the DCs that continue to participate, fines decrease output and DC surplus from less capable DCs and increase output and DC surplus from more capable DCs. Thus, for the policy-maker, the potential increase in welfare comes with a form of market concentration towards higher capability.

On the whole, the implication of our work is that the use of fines to enforce DPR could have unintended consequences. Of key relevance is the fact that in market based environments, individual DSs utilize their local knowledge to make decisions that are beneficial to themselves (Hayek 1945). Policy-makers can enforce data portability, but DSs may not have an incentive to port their data to less capable DCs - instead, they may port their data to DCs that can give them greater value in return for their personal data.

When assimilating the implications of our work, it is important to take note that we model competing DCs that differ in capability, which is the ability to generate more revenue from each unit of output. We interpret capability as the representation of a firm's production technology. Thus, capability is exogenously determined through sunk costs, and results in the ability to provide attractive features such as ease of use, reliability, quality of recommendations, and speed. However, regardless of how the dimension along which DCs differ is named and interpreted, it is the characterization of capability in the assumptions that leads to the results, namely increasing revenues in capability. Should the dimension along which DCs differ be named or interpreted differently, if it follows the assumptions the same results will obtain.

We do not study taxes that are focused on particular jurisdictions, or welfare that differs by jurisdictions such that welfare is considered in one jurisdiction and not another. In addition, even though the effects through porting are modeled using general functions and can encompass a variety of specific functional forms, our results do not hold if the actual effects of porting are different from our assumptions. Further, even though our general-form results point the policy-maker in the direction of the effects, the magnitudes depend on the specific situation, properties of the functions, and the parameter estimates. Finally, any model is an abstraction to study the interactions between a limited set of effects considered most important, and therefore, we have eliminated several effects that could have a different impact on the choices by the policy-maker, DCs, and DSs. Horizontally differentiated DCs and probability of detection of DPR violation are two examples of such effects.

Chapter 4

Energy Productivity Implications of IT Investments

Using a varying coefficient model, we build a structural econometric specification using the Cobb-Douglas to evaluate the effect of IT on energy productivity. In addition, given the large completed and planned investments in the Smart Electricity Grid, we posit that IT has a large impact on the output elasticity of electric energy. We use U.S. industry level data collected from the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Energy Information Administration. We find that IT increases energy productivity, and that IT's effect on electricity productivity is an order of magnitude larger than its effect on non-electric energy. In terms of magnitude, we find that a 1% increase in IT capital increases the output elasticity of energy by 0.13%.

4.1 Introduction

The slogan "Advancing American Energy" is at the heart of the U.S. energy policy, with strategic investments being made in order to advance clean energy, reduce American dependence on foreign oil, and crucially, increase energy productivity. In its 2014-2018 strategic plan, the U.S. Department of Energy identifies "Improving energy productivity by increasing efficiency" as its first strategic objective (U.S. Department of Energy 2014). Separately, sustainability and climate change have become central issues globally. At the United Nations COP21 meeting in Paris in 2015, 196 parties agreed to work to limit global temperature rise to below 2 degrees celsius (United Nations 2016). Rolling back the rate of growth of global warming requires a significant reduction in the emissions of greenhouse gases (GHG) by all signatories to the Paris accord (Brundtland and Khalid 1987, Davis et al. 2010). The Department of Energy for instance, aims to reduce U.S. carbon pollution by 3 billion metric tons by 2030.

There has been a resurgence of U.S. and global focus on energy productivity, and although several studies show that IT can augment the productivity of other inputs used in the production process, the impact of IT on energy productivity has not been studied extensively. Energy productivity is defined as the gross output per energy resource input (Steinberger and Krausmann 2011). Producing more output per unit energy is a strategy that aligns well with both the U.S. energy policy mentioned above and the global climate change initiatives. Several opportunities to increase energy productivity exist (Farrell et al. 2007). This chapter focuses on the opportunity that IT brings in terms of increasing energy productivity.

Firms have been viewed traditionally as a "technology" that converts inputs into output - this technology can be mathematically represented as a production function. In this context, energy productivity can be increased in two different ways. First is the "pure form" increase in energy productivity, which is characterized by an increase in gross output from an additional unit of energy, holding all other inputs constant. This happens, for instance, when innovation leads to improvements in production technology resulting in an improved direct effect of energy on gross output. Second is input complementarity, where there is an increase in gross output from an additional unit of energy resulting from an increase in another input. This complementarity between two inputs can result in an increase in productivity. In effect, our research question investigates the complementarity between IT and energy.

The complementarity between IT and energy happens in two broad ways. Firstly, there is a long history of the use of IT for operational efficiency - this includes process management, inter-organizational coordination, forecasting, and scheduling. For instance, manufacturing automation enabled by IT encodes manufacturing processes into highly efficient processes, thereby reducing the use of other inputs like energy, materials, and labor. IT systems like Enterprise Resource Planning and Supply Chain Management enable operation wide visibility to decision makers, apart from improving business processes and increasing reliability. Given that operational efficiency focuses on the efficient use of time and resources, it is plausible that such IT investments have a positive impact on energy productivity. Secondly, several works cited in subsection 4.1.1 describe the targeted use of IT for measuring, reporting, and improving energy efficiency. Given the widely prevalent external inducements for energy efficiency from consumers, policy-makers, and standard setters, such IT investments can further drive energy productivity.

Separately, we explore the effect of IT on the productivity of electric energy and of non-electric energy. The electricity supply chain consists of generators, transmitters, distributors, and consumers. With the advent of the smart grid, IT plays a significant role within this supply chain - the utilities sector has approximately three times the IT intensity of the manufacturing industry.[1] Smart grid investments have been shown to increase the energy efficiency of electricity producing industries, and pilot studies have shown smart grid investments to decrease demand by over 25% (Corbett et al. 2018). We explain more about the complementarity between IT and electric energy, and more broadly between IT and energy in section 4.1.1.

4.1.1 Related Literature

The productivity paradox of the 1980's was the inability of researchers to find a significant relationship between IT investments and productivity despite the rapidly increasing levels of IT investments by firms. However, since Brynjolfsson and Hitt (1996), several studies have shown the significance of this relationship. Lee and Barua (1999) and Dedrick et al. (2003) find that past studies that could not find a significant relationship between IT investments and productivity were analyzing data of poor quality, deflating their inputs and outputs incorrectly, or using modelling techniques that needed further refinement. Further, Lee and Barua (1999) find that firms investing in IT increased their allocative efficiency as their IT intensity increased. They confirm positive and significant returns to IT investments, that firms were systematically underinvesting in IT, and that an increase in the input share of IT would lead to a disproportionate increase in output. Researchers today agree to a large degree about the value of IT to overall productivity, and have moved on to more specific questions, for instance about the varied effects of different types of IT investments, the effect of a focal firm's IT investments on other firms, and how IT combines with other inputs to generate value.

Firm inputs consist of non-IT capital, labor, IT capital, energy, materials, and services. Each of the firm's inputs has a direct effect on output, and while output can increase with each input, the extent of the increase is of consequence. In particular, firms prefer to expend their costs on the most efficient input in order to get the most output per dollar spent on input. Several studies cited in the subsection "Direct Effects" evaluate

[1] IT Intensity is defined as the proportion of IT capital to gross output.

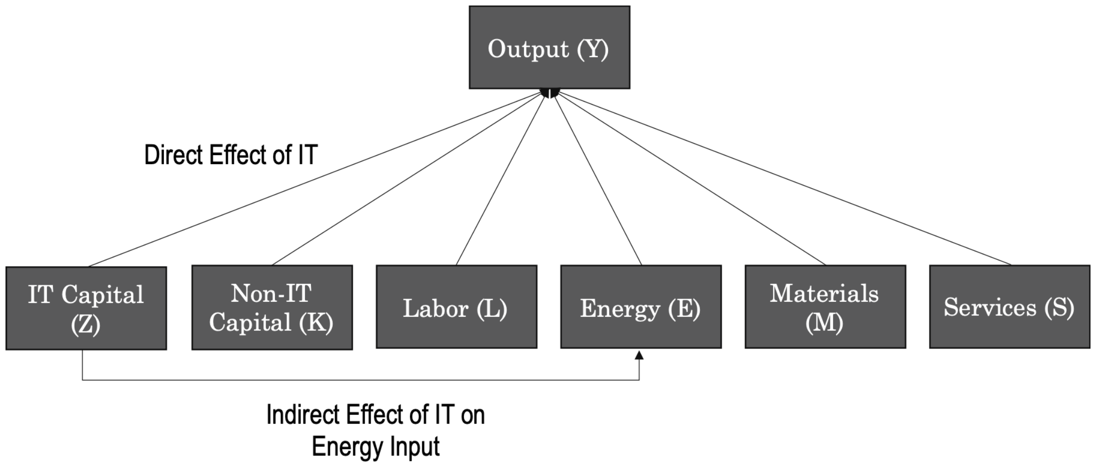


Figure 4.1: Direct and Indirect Effects of IT

the direct impact of IT on productivity, and compare this with the impact of non-IT capital and labor on productivity.

An input can also have an effect on other inputs (See Figure 4.1). These indirect effects from one input can change the output elasticity (OE) of another input, and result in a compounded effect on output. Several studies on indirect effects are cited in the below subsections. In our study, we explore the indirect effect of IT on the OE of energy.

Direct Effects

The use of IT capital in manufacturing has outpaced the use of other inputs by an annual growth rate of 3.46% in the period 1998-2018. Several research undertakings have helped explain the seemingly insatiable appetite for IT investments. Stiroh (2002) finds that there is a robust correlation between IT investments and productivity, that IT capital accumulation is important for business output, and that the sharp increase in U.S. productivity growth in the late 1990s can be explained by IT investments made in the early 1990s. He finds that virtually all the productivity gains seen by the U.S. during the time was due to industries that either produced IT or used IT intensively. Brynjolfsson and Hitt (1996) used firm level data to show that IT capital made substantial contributions to the output of firms, and that dollar for dollar, IT capital creates more value than non-IT capital. Their data consisted of five annual surveys of hundreds of large firms, supplemented with firm data from other sources. Brynjolfsson and Hitt (2003) delve further into this question, and find that while the contribution of computers to productivity is roughly equal to its cost in the short run, this contribution increases substantially over a longer period of time. In general, IT has been shown to be a significant and growing contributor to growth in GDP and in productivity (Dedrick et al. 2003).

Using an event study, Dehning et al. (2003) found that firms investing in IT for strategic reasons see an increase in firm value. Further, firms belonging to industries where IT plays a strategic role are rewarded by an increased firm valuation on the stock market when they make announcements to invest in IT. Firms that are industry leaders in implementing IT in a strategic role see positive and abnormal stock market returns when they make announcements to invest in IT. Anderson et al. (2006) find that firm valuation and earnings increase in its IT investments. Dewan and Kraemer (2000) analyze returns to IT investments at a country level and find that developed countries see positive and significant returns to IT investments, while returns on non-IT investments suggest that they have over-invested in non-IT capital. The opposite is true for developing countries.

Overall, the marginal product of IT has been shown to be both larger than the IT investment itself, and larger than the marginal product of any other input.

Indirect Effects of IT

In addition to the increase in productivity attributed directly to IT, IT augments capital and labor, thereby increasing the productivity of both capital and labor. This effect is even more pronounced in industries with high IT intensity (Mittal and Nault 2009). The high growth rates of average labor productivity in the 1990's can be attributed partially to IT capital (Oliner and Sichel 2000). As well, IT capital has traditionally been a substitute for non-IT capital. With an increasing number of units of IT substituting for labor with the passage of time, the relationship between non-IT capital and IT capital has evolved into a complementary one (Chwelos et al. 2010).

Zhang et al. (2015) show that for three or more inputs, the Morishima Elasticity of Substitution (MES) is a more informative and relevant measure of elasticity than the widely used Allan Elasticity of Substitution (AES). They find that declines in the prices of IT have led to an increase in the quality of IT even if the input share of IT was kept constant. Gurbaxani et al. (2000) evaluate the change in ratio of computing hardware to IT personnel as the input share of computing hardware increases, finding that this is a homothetic relationship. In other words, the hardware-personnel ratio is independent of the size of the IT operation within organizations.

In the first of two studies evaluating the effect of IT on Intermediate Inputs[2], Gong et al. (2016) find that IT has significantly reduced transaction costs, thereby enabling a move to the market in logistics. In a later study, Gong et al. (2021) find a more nuanced effect of IT on the make versus buy decision. They find that for the U.S. manufacturing sector, IT hardware tends to favor the internal provisioning of intermediate inputs, whereas software reduces the costs of external provisioning.

[2]Intermediate Inputs consist of energy, materials and services.

Broadly, while studies on both direct and indirect effects of IT are plentiful, the effect of IT on energy productivity has been relatively under-studied. Our next subsection surveys research that provides context in this area.

IT and Energy Productivity

One of the key applications of IT is to systematically codify and execute improvements in operational efficiency, a core objective of which has been time and waste minimization. Industry best practices such as total quality management, just in time, and efficient customer response are codified, bundled, and delivered to manufacturing firms by IT vendors. This has led to widespread adoption of operationally efficient process management techniques. Examples of IT capital in manufacturing includes systems for enterprise resource planning, supply chain management, materials and equipment management, real time analysis and decisions, and flexible manufacturing (Nota et al. 2020). For instance, enterprise resource planning systems enable organizations to plan inventories based on demand rather than on production, and automated supply chain systems are designed to move products with minimal wastage and minimal inventory. The resulting operational efficiencies include improved material handling (Yazdi et al. 2018), reduction in inventories (Rai et al. 2006), reduction in the length of time that products spend in the supply chain (Yao and Zhu 2012), improved forecasting and scheduling leading to lower input costs (Nagasawa et al. 2017), and improved inter-organizational collaboration and coordination (Bharadwaj et al. 2007). Further, operational efficiencies can make the production process more environmentally friendly (Kleindorfer et al. 2005) - this is exemplified by the use of energy efficiency as a measure of operational efficiency (Nota et al. 2020).

Separately, the induced innovation hypothesis that "the changes in relative factor prices should lead to innovations that reduce the need for the relatively expensive factor", has been studied extensively over nearly a century (p. 160 Popp 2002, Hicks 1932). Acemoglu et al. (2012) analyze this hypothesis and find that ".. taxes/subsidies redirect innovation toward clean inputs", "optimal policy involves carbon taxes and research subsidies", and that "the use of exhaustible resource in dirty input production helps the switch to clean innovation under laissez-faire" (p. 131). Popp (2002) shows empirically that government energy efficiency regulation and sudden upward volatility in energy prices had significant impact on energy efficiency. Over the recent decades, energy prices have been more volatile on a year-to-year basis than the other intermediate inputs, customer preferences have evolved towards environmentally friendly products, and governments have been actively regulating (and looking to further regulate) energy efficiency through punitive measures, taxes, and subsidies over the past several decades. This has created a set of inducements towards energy efficiency in the production process. In conjunction, IT has become an important tool for improving energy efficiency (Preuveneers and Ilie-Zudor 2017; Shrouf and Miragliotta 2015; Wang et al. 2018), and IT-enabled operational optimizations and scheduling can integrate energy input as part of the objective function (Bougain et al. 2015; Duflou et al. 2012). Today's IT can sense the current state of operation, analyze this information across a broad swath of operations, and make decisions in real-time thereby improving energy management at the device, machine, production line, facility, plant, and multi-factory levels (Nota et al. 2020). As well, hardware and software has been extensively used to measure and report on compliance with regulations and standards (Kleindorfer et al. 2005), thereby providing actionable measures which can drive improvements to energy productivity.

Energy input consists of electric energy and non-electric energy, and most of the above described phenomena are also applicable to IT's effect on electric energy. However, over the last decade, we have witnessed some very large IT investments in the electric grid by both governments and industry in order to increase energy productivity. This has led to a dramatic increase in the installed base of smart meters and advanced metering infrastructure, with 58.5 million smart meters installed in the U.S. by 2014 and penetration rates between 50% to 100% in OECD countries (See Appendix A.1).

Lack of transparency in electricity pricing has been criticized by economists because regulated fixed prices do not provide incentives to consumers to use electricity during off-peak hours. Ever since time-dependent electricity pricing has come into vogue in the U.S., electricity price volatility has become widespread. Kahn (1979) promises as early as the 1970's that prices would vary between 2.5 cents to 30 cents depending on demand and supply. One of the reasons for the efficiency of real-time pricing is that electric energy must be consumed as it is generated because large scale electricity storage solutions have not been viable over our study period. However, the cognitive load required to manually adjust consumption to off-peak hours is unreasonable (Ito 2014). Here, IT-enabled behavioral solutions can provide apt solutions by first providing a mechanism to deliver incentives to consumers, and secondly by automating the use of energy-consuming devices. IT investments in the Smart Grid have improved operational efficiency on the electricity supply side (generation, transmission, and distribution), but "the goal is to use smart meters to reduce electricity consumption at an overall level, as well as during critical peak load periods" (p. 253, Corbett et al. 2018). Initial estimates show that the smart grid has reduced peak demand by over 25%, and that the bulk of these reductions in energy consumption have come from an overall reduction in electricity consumption rather than from a shift of consumption to a different hour (Corbett et al. 2018).

In other words, given wide fluctuations in electricity prices and the requirement that produced electricity be consumed instantaneously, the real time nature of IT is particularly valuable. In this setting, IT can inform electricity consumers about electricity prices as well as programs run by utilities, and automate consumption patterns based on price. Historically, IT has had an important role in the electricity grid, and average IT intensity is over 40% higher among utilities than the average industry. While studies have shown that IT investments in the smart grid have improved efficiency among electricity producers, the effect of IT investments on electricity productivity among electricity consumers has not been assessed formally. An empirical measurement of the impact of IT on electricity consumption is valuable, especially given the large, planned, and completed IT investments in the electric grid.

Our study focuses on the manufacturing sector for three reasons. Firstly, there is renewed interest from policy-makers on revitalizing this sector. Secondly, the manufacturing sector is the largest consumer of energy, accounting for approximately a third of global first use of energy (Saygin et al. 2011). Thirdly, data about the first use of energy and its decomposition is available for the manufacturing sector from the U.S. Energy Information Administration. In sum, this chapter addresses the following research questions for the manufacturing sector - Firstly, what is the impact of IT investments on energy productivity? Secondly, what is the impact of IT on the productivity of electric energy and non-electric energy?

4.2 Model Specification

We use Cobb-Douglas (CD) as the foundational functional form from which to derive estimates of interest. In this section, we explain what the CD is, why our estimation is an effort to disentangle the effect of interest from Total Factor Productivity (TFP), and why the CD form is appropriate for industry-level analysis, before deriving our two estimation models.

4.2.1 The Cobb-Douglas for Productivity Analysis

Production at a firm level is $Y = f(X)$, where Y is the firm output and X is a vector of inputs. Inputs typically include non-IT capital (K), labor (L), IT capital (Z), and intermediate inputs (R). Intermediate inputs consist of energy, materials, and services. The marginal product MP_j measures the increase in firm output with an increase in an input j,

$$MP_j = \frac{\partial f(X)}{\partial X_j}.$$

The output elasticity (OE), σ_j, is the percentage change in output with a 1% increase in input j,

$$\sigma_j = MP_j \frac{X_j}{Y}.$$

Kundisch et al. (2014) show that the standard CD productivity function can be derived from the income accounting identity if input shares are held constant over time, and therefore is theoretically based. The income accounting identity states that gross output is the sum of the wage bill, operating surplus on IT capital, operating surplus on non-IT capital, and intermediate inputs. This can be represented in CD form,

$$Y = A_1 K^{\alpha_1} L^{\beta_1} Z^{\gamma_1} R^{\delta_1}, \tag{4.1}$$

where $\alpha_1, \beta_1, \gamma_1$, and δ_1 are the output elasticities of the respective input, and A_1 is TFP. TFP represents technical change and is described below.

The CD in (4.1) in log form is linearly additive, and is hence estimable using regressions,

$$y = a_1 + \alpha_1 k + \beta_1 l + \gamma_1 z + \delta_1 r + \epsilon_1. \tag{4.2}$$

Note that the lower-case letters represent the log of the relevant quantity and that the estimation is within the context of panel data so that each observation corresponds to an industry-year. As the CD form is derived from the Income Accounting Identity, the inputs and the output are by definition related in the CD, and it is common to see regression coefficients of determination that are upwards of 95%.

4.2.2 Total Factor Productivity

In (4.1), all the direct effects of inputs on output are captured in the elasticities of each input. In contrast, TFP represents effects that are not traced back to specific inputs given the functional form used for the estimation and can contain one of the following. Firstly, any variability in production that can systematically be attributed to a constant multiplier will enter the TFP term and be captured as an intercept in a regression estimation. Secondly, there can theoretically be factors outside K, L, Z, and R that drive firm productivity. For instance, Chang and Gurbaxani (2012) hypothesize that knowledge transfers from IT outsourcing providers affects firm productivity. Therefore, they enter this knowledge transfer as an input into the production function. Similarly, Cheng and Nault (2007, 2012) view supplier and customer IT investments as a missing input and enter these into the CD as a factor of production. They empirically verify their argument, finding that IT investments by customers and suppliers benefit the focal industry, an effect they describe as *IT spillovers*. Thirdly, TFP could be capturing a different way of combining L, K, Z, and R than the form assumed: If the inputs are interacting in a way that is not captured by (4.1), then some of these effects could enter the TFP term. For example, Mittal and Nault (2009) hypothesize that IT investments augment labor and capital, thereby making these inputs more productive. They enter the augmentation function into the CD to estimate the augmentation effect.

Several eminent works within the themes of IT productivity, substitution, and spillovers use the TFP to disentangle the effect of interest. The themes are thus deeply interconnected through the common base estimation form, with the purpose of understanding the inter-relationships between inputs and output. In our study, we hypothesize that IT and energy inputs combine in a way that their effect on output is not captured simply by their direct effects on output. Our work seeks to disentangle this effect from TFP.

4.2.3 Aggregation

The production function is a foundational element in the theory of the firm and it plays the role of transforming inputs into output at the firm level. However, the use of CD to study productivity at the industry or national level brings with it two key aggregation challenges. First, each firm input is an aggregation of a variety of sub-inputs. For instance, firm capital consists of equipment and structures, and in turn equipment consists of machinery and vehicles. Secondly, firms can have several lines of business so that its gross output and inputs are the sum of outputs and input associated with each of its business lines. In a similar vein to the second challenge above, inputs at the industry level consist of the aggregation of inputs across its constituent firms, and output at the industry level consists of the aggregation of firm outputs.

Aigner and Chu (1968) study these aggregation challenges as they pertain to the use of CD for industry level productivity analysis. They argue that the industry level production function represents the "average" firm in the industry so that such firms can use the estimates obtained from the industry level productivity analysis for their resource allocation decisions. In conjunction with their use of the income accounting identity, Kundisch et al. (2014) explain how the "Cobb-Douglas is valid for estimation at the firm, industry, sector, and economy levels, indicating that aggregation does not invalidate the use of the Cobb-Douglas form" (p. 450). This important result provides the theoretical underpinning for the use of CD in our industry-level analysis on energy productivity. Several papers have used CD for aggregated analysis. For instance, Stiroh (2002), Dewan and Kraemer (2000) and Gurbaxani et al. (2000) assess direct effects of IT on production at the industry level.

4.2.4 Main Model

Splitting intermediate inputs, R, into its constituent inputs energy (E), materials (M), and services (S), (4.1) becomes

$$Y = A_2 K^{\alpha_2} L^{\beta_2} Z^{\gamma_2} E^{\theta_2} M^{\delta_2} S^{\zeta_2}, \tag{4.3}$$

where $\alpha_2, \beta_2, \gamma_2$, θ_2, δ_2, and ζ_2 are the output elasticities of the respective input, and A_2 is the TFP. To arrive at the econometric specification to estimate direct effects of each input on output, we take logs so that

$$y = a_2 + \alpha_2 k + \beta_2 l + \gamma_2 z + \theta_2 e + \delta_2 m + \zeta_2 s + \epsilon_2. \tag{4.4}$$

Next, we use a varying coefficient model (Hastie and Tibshirani 1993) to measure the impact of IT on energy productivity. We consider the effect of IT by specifying the OE of energy as a linear function of IT, $\theta_2 = \theta_2(Z)$, which takes the form

$$\theta_2 = b_2 + \varphi_2 Z, \tag{4.5}$$

where θ_2 is the OE of energy and Z is the IT input. We choose a linear function for three reasons. Firstly, this is the simplest function that can capture the relationship between IT and the OE of energy. In other words, more complex and finely tuned non-linear functions are more likely to find a significant relationship as compared to a linear function. Secondly, the linear function allows for a test of our hypothesis - a zero estimate for φ_2 indicates the absence of a relationship, allowing for a structure for hypothesis testing. Finally, several prior works have used linear functions to estimate varying coefficient models including Hastie and Tibshirani (1993) and Gong et al. (2016, 2021), thereby providing a history of prior work that uses this approach.

Now taking the first partial derivative of (4.5) with respect to Z,

$$\varphi_2 = \frac{\partial \theta_2}{\partial Z} = \frac{\partial}{\partial Z}\left(\frac{\partial Y}{\partial E}\frac{E}{Y}\right) = \frac{E}{Y}\frac{\partial^2 Y}{\partial E \partial Z}.$$

The LHS in the above equation, φ_2, is IT Augmented Energy Productivity, measured as the change in the OE of energy with a unit increase in IT. In the second equality, we replace θ_2 with its definition as the OE of energy. Now, taking the partial derivative with respect to IT brings us to the RHS term above - φ_2 is the partial differential of output with respect to energy and IT, a measure of complimentarity between energy and IT, which when weighted by the input proportion of energy to output gives us a measure of *IT Augmented Energy Productivity.*

The sign of φ_2 is positive if Z and E are complements and negative if they are substitutes. In other words, the two inputs are complements if $\partial^2 Y/\partial E \partial Z > 0$, indicating that the marginal returns to energy increase in IT. Similarly, IT and energy are substitutes if $\partial^2 Y/\partial E \partial Z < 0$.

Substituting (4.5) in (4.3),

$$Y = A_2 K^{\alpha_2} L^{\beta_2} Z^{\gamma_2} E^{(b_2+\varphi_2 Z)} M^{\delta_2} S^{\zeta_2}.$$

To arrive at an econometric model specification for our base model, we take logs and change the coefficient subscript to allow for the estimates of the varying coefficient model to differ from (4.4), so

$$y = a_3 + \alpha_3 k + \beta_3 l + \gamma_3 z + b_3 e + \varphi_3 Z e + \delta_3 m + \zeta_3 s + \epsilon_3. \tag{4.6}$$

4.2.5 Split-Intermediates Model

Manufacturing industries often convert non-electric energy into electricity before consuming it. However, our measure of electricity does not include onsite generation of electricity from non-electric energy. This creates a mechanical relation between the two forms of energy: for a given level of required energy input, a higher level of non-electric energy leads to a lower level of electric energy input. In this economic context, using the base model to divide energy into electric and non-electric energy and studying them simultaneously presents econometric concerns and often leads to opposing results. For this reason, we use the split-intermediates model, which dis-aggregates a single variable of interest from intermediate inputs. This enables a study into the effects of IT on each of electric and non-electric energy, as well as validating results from our main model on IT's effects on net energy input.

We begin by re-assessing the effect of IT on the energy input using the split-intermediates model. To study our variable of interest, energy, we begin by isolating it from intermediate inputs, thereby splitting intermediate inputs into energy, E, and other intermediate inputs, $R_{\setminus E}$, leading to the CD form

$$Y = A_4 K^{\alpha_4} L^{\beta_4} Z^{\gamma_4} E^{\theta_4} {R_{\setminus E}}^{\delta_4}, \tag{4.7}$$

where θ_4 and δ_4 are the OE of energy and other intermediate inputs respectively. Similar to (4.5), the OE of energy is specified as a linear function of IT, $\theta_4 = \theta_4(Z) = b_4 + \varphi_4 Z$. Substituting in the above equation,

$$Y = A_4 K^{\alpha_4} L^{\beta_4} Z^{\gamma_4} E^{(b_4+\varphi_4 Z)} {R_{\setminus E}}^{\delta_4}.$$

To arrive at an econometric model specification, we take logs, so

$$y = a_4 + \alpha_4 k + \beta_4 l + \gamma_4 z + b_4 e + \varphi_4 Z e + \delta_4 r_{\setminus E} + \epsilon_4. \tag{4.8}$$

Given the high levels of IT intensity in electricity generation, transmission, distribution, and consumption, and the multi-billion dollar investments planned and completed in computerizing the electric grid, the effect of IT on the productivity of electric energy is an important question. In order to assess IT's effect on electricity productivity, we simply estimate the model described above to test this hypothesis, splitting intermediate inputs into electric energy ($\widetilde{E}$, where the tilde symbolizes the sine wave), and other intermediate inputs, $R_{\setminus \widetilde{E}}$. Then, we have the CD form

$$Y = A_5 K^{\alpha_5} L^{\beta_5} Z^{\gamma_5} \widetilde{E}^{\widetilde{\theta}_5} {R_{\setminus \widetilde{E}}}^{\delta_5}. \tag{4.9}$$

Similar to (4.5), the OE of electric and non-electric energy is specified as a linear function of IT, $\widetilde{\theta}_5 = \widetilde{\theta}_5(Z) =$

$b_5 + \varphi_5 Z$. Substituting for $\widetilde{\theta}$ in the above equation and taking logs, we get our estimation form

$$y = a_5 + \alpha_5 k + \beta_5 l + \gamma_5 z + b_5 \widetilde{e} + \varphi_5 Z \widetilde{e} + \delta_5 r_{\setminus \widetilde{E}} + \epsilon_5. \tag{4.10}$$

We now repeat the above process for non-electric energy, ($\check{E}$, where the inverted hat symbolizes other forms of energy like those extracted from the ground). This gives us the estimation form

$$y = a_6 + \alpha_6 k + \beta_6 l + \gamma_6 z + b_6 \check{e} + \varphi_6 Z \check{e} + \delta_6 r_{\setminus \check{E}} + \epsilon_6. \tag{4.11}$$

The interpretation of the varying coefficient remains unchanged, as described in Section 4.2.4. Next, we describe our data before proceeding to present our results.

4.3 Data

Our study uses 3-digit North American Industry Classification System (NAICS) level U.S. manufacturing industry data collected from the Bureau of Economic Analysis (BEA, an agency of the U.S. Department of Commerce), the U.S. Bureau of Labor Statistics (BLS, U.S. Department of Labor), and the U.S. Energy Information Administration (EIA, U.S. Department of Energy). For each industry, materials, and services in current dollars along with their chain-type price indexes were obtained from BEA's online database. We also obtained gross output in constant 2012 dollars from BEA. IT capital, non-IT capital, intellectual capital (all in constant 2012 dollars), and labor (hours of all persons) were sourced from the BLS online database. Energy data was obtained from the Manufacturing Energy Consumption Survey, available online from the EIA. This data accounts for the first use of energy (measured in trillion BTUs) for all purposes within each manufacturing industry, includes a breakdown into its electric and non-electric components, and is collected and provided once every four years by the EIA. Because BTU is a physical input similar to hours of all persons for labor, it does not suffer the additional measurement errors caused by deflating dollar values. Finally, capacity utilization data by industry was obtained from the Federal Reserve Economic Data, maintained by the St. Louis Fed.

Gross output consists of sales to final users in the economy, or sales to other industries, while inputs consist of IT capital, non-IT capital, labor, and intermediate inputs. IT capital consists of hardware capital and software capital. In turn, hardware is categorized as computers, peripheral equipment, communication equipment, and other IT, while software consists of pre-packaged, custom, and own-account. Non-IT capital consists of equipment and structures, less any IT Hardware embedded in the equipment account, while labor

Table 4.1: Mean Across Industries over Time

Year	Gross Output	Non-IT Capital	Labor	IT Capital	Materials	Services	Energy (EIA Measure, BTU)	Electricity (EIA Measure, BTU)	Energy (EIA Measure, BTU, Interpolated)	Electricity (EIA Measure, BTU, Interpolated)
1998	331,021.76	166,952.65	2,029.19	9,859.76	197,709.18	35,707.12	1,396.59	177.29	1,396.59	177.29
1999	337,736.59	169,950.47	2,007.86	10,743.71	198,116.57	35,819.05			1,378.64	173.09
2000	341,149.47	172,775.06	1,991.08	11,421.76	197,428.98	35,256.08			1,360.79	169.44
2001	328,536.18	175,350.65	1,863.83	11,912.47	183,827.09	36,447.08			1,343.15	166.89
2002	329,934.88	176,718.00	1,736.90	12,055.94	190,700.18	34,192.08	1,325.82	166.00	1,325.82	166.00
2003	327,541.88	176,826.06	1,655.20	11,834.82	183,178.86	34,467.67			1,308.58	166.96
2004	338,549.82	176,471.71	1,649.99	11,546.41	191,606.02	29,709.95			1,289.83	168.54
2005	351,998.59	176,366.29	1,633.55	11,538.94	198,972.12	31,862.26			1,267.65	169.16
2006	356,773.12	176,651.35	1,644.46	11,812.53	198,165.87	33,542.21	1,240.12	167.24	1,240.12	167.24
2007	367,058.41	177,711.71	1,619.50	12,382.29	203,116.98	34,647.12			1,206.37	161.80
2008	346,736.88	179,287.29	1,555.15	13,482.24	191,588.45	28,683.09			1,169.84	154.46
2009	306,864.71	179,493.71	1,352.43	14,308.41	163,892.48	28,523.67			1,135.00	147.43
2010	322,389.82	178,507.06	1,355.72	14,616.06	174,186.60	29,238.95	1,106.35	142.94	1,106.35	142.94
2011	331,519.00	178,197.12	1,382.98	15,138.82	183,728.04	28,003.74			1,087.33	142.56
2012	338,360.88	179,039.76	1,413.87	15,789.24	189,933.24	29,245.12			1,077.18	145.21
2013	348,081.47	180,489.35	1,427.09	16,402.06	195,506.66	29,597.62			1,074.10	149.17
2014	350,749.53	182,027.53	1,450.82	16,820.88	196,855.61	29,495.28	1,076.29	152.71	1,076.29	152.71
2015	352,446.00	183,516.65	1,464.17	17,156.47	198,944.49	28,205.04			1,082.12	154.48
2016	353,496.53	184,647.12	1,469.40	17,466.71	200,781.98	29,020.23			1,090.55	154.64
2017	354,671.65	185,502.47	1,482.84	17,638.00	193,538.05	32,621.02			1,100.73	153.73
2018	364,632.94	186,531.59	1,513.92	17,929.29	196,711.59	33,812.99	1,111.76	152.29	1,111.76	152.29

is measured as the hours of all persons. Intermediate inputs consist of energy (sourced from energy producing industries), materials (raw materials and semi-finished goods sourced from goods-producing industries) and services (sourced from service producing industries). Energy input is defined as the first use of energy, consisting of energy produced offsite and onsite, and excludes energy produced from other energy inputs, thereby avoiding intra-establishment double counting. Energy consists of electric energy and non-electric energy, where non-electric refers to residual fuel oil, distillate fuel oil, natural gas, HGL excluding gasoline, coal, coke, breeze (fine sized coke), and other sources. Electric energy consists of electricity purchases, transfers-in, and generation from non-combustible renewable resources less electricity sold or transferred out.

Because energy data is available once in four years and given that our industry level panel data is annual, we have two broad approaches to balance the data. The first is to reduce the frequency of analysis from yearly to four-yearly. This would mean that we ignore KLEMS data for the years in between benchmark years and lose some of the richness of the data in the process. The second approach, which we use, is to interpolate electricity consumption data for the non-benchmark years. There are two main types of interpolation techniques: linear and spline. Linear interpolation considers only two points at a time. This does not give us a smooth interpolation and it has been shown to provide less reliable estimates. Polynomial and spline interpolation techniques tend to provide more precise estimates. Given the six available data points from EIA, we use six degrees of freedom, and we use the R function *smooth.spline* in order to execute

Table 4.2: Outlier Statistics: Apparel, Leather, and Allied Products (Naics 315,316)

	Mean Across Industries		Apparel, Leather, and Allied Products (Naics 315,316)	
Variable	**Levels**	**Annual Growth Rate %**	**Levels**	**Annual Growth Rate %**
Gross Output	341,917	0.55	39818	-7.22
Non-IT Capital	178,239	0.56	25543	-2.27
Labor	1,605	-1.38	578	-7.03
IT Capital	13,898	3.08	968	0.27
Energy	1,209	-1.13	22	-10.28
Materials	191,833	0.10	19905	-8.54
Services	31,814	-0.04	7261	-6.22

so that the interpolated data equates to the original data whenever the original data is available (every four years). The mean values of the original and interpolated data across industries over time is presented in Table 4.1, and a visualization of this data is presented in Appendix B.1.

The interpolated data for the non-benchmark years induces a measurement error in the independent variables $\widetilde{e}$ and $\breve{e}$. This is a classical measurement error because there is no reason to believe that the measurement error in electricity input $(\widetilde{\epsilon_e})$ is systematically correlated to the true value $(\widetilde{e^\star})$, or that the measurement error in non-electricity input $(\breve{\epsilon_e})$ is systematically correlated to the true value $(\breve{e^\star})$. In other words, $COV(\widetilde{\epsilon_e}, \widetilde{e^\star}) = 0$ and $COV(\breve{\epsilon_e}, \breve{e^\star}) = 0$. This leads to an attenuation bias such that the estimated effect of interest $\hat{\beta} = \beta(\sigma^2_{\widetilde{e}^\star}/(\sigma^2_{\widetilde{e}} + \sigma^2_{\widetilde{e}^\star}))$. Thus our estimate for β will be lower than the true effect, and we can qualify our estimation as the lower bound on the effect of interest. As the extent of bias varies positively with the extent of measurement error, the basis of our choice of interpolation method is the minimization of these errors.

We limit our analysis to the period between 1998 and 2018 given that this is the period for which energy data is available from EIA. The mean values of key variables by year are reported in Table 4.1 and visualized in Appendix B.1. Of note is the fact that gross output grew at an average annual rate of 0.55% while energy contracted at -1.13%. Meanwhile, IT capital was the fastest growing input, growing at an average annual rate of 3.08%.

The smallest industry in our sample was also by far the most rapidly shrinking one (Naics 315,316 - Apparel, Leather, and Allied Products Manufacturing industry, see Table 4.2). The gross output from this industry shrank at over 7% a year over our sample period, so that in 2018, it represented 0.29% of the overall manufacturing sector's output (Each manufacturing industry generates on average 5.55% of the overall manufacturing sector's output, given the 18 constituent industries). The overall trend in apparel and leather manufacturing outsourcing has affected this U.S. industry significantly, to the point where this industry is structurally different, and not representative of the broad U.S. manufacturing sector that we

Table 4.3: List of industries being studied (all manufacturing industries), along with its North American Industry Classification System (NAICS) code.

NAICS	Description
321	Wood products
327	Nonmetallic mineral products
331	Primary metals
332	Fabricated metal products
333	Machinery
334	Computer and electronic products
335	Electrical equipment, appliances, and components
336	Transportation Equipment
337	Furniture and related products
339	Miscellaneous manufacturing
311,312	Food and beverage and tobacco products
313,314	Textile mills and textile product mills
322	Paper products
323	Printing and related support activities
324	Petroleum and coal products
325	Chemical products
326	Plastics and rubber products

are studying.[3]. For these reasons, we qualify this industry as an outlier and drop it from our analysis. Our sample thus contains seventeen of eighteen manufacturing industries, as listed in Table 4.3. Summary statistics for the key variables in our analysis across industries and years are shown in Table 4.4.

4.3.1 Controls and Econometric Adjustments

Our sample years contain several economic events that affected the overall U.S. economy. To control for such time-specific economy-wide events, we use time-fixed effects in our model. Further, as a result of production in an industry being fairly stable over time, we can expect that there is auto-correlation between yearly observations within an industry. We use the Wooldridge test for auto-correlation and find that we can reject the null hypothesis that autocorrelation is absent (F = 18.52). Further, as the autocorrelation can be expected to be specific to each industry, we use panel-specific AR1 (PSAR1). Next, we test for heteroskedasticity using the Likelihood Ratio (LR) test. We find that we can reject the null hypothesis of the absence of panel-level heteroskedasticity at all reasonable levels of significance ($\chi^2 = 387.74$), and control

[3]This industry's inputs and outputs are visualized in Appendix B.2

Table 4.4: Summary statistics for the Key Variables in our Analysis.

Variable	Obs.	Mean	Std. dev.	Median	Min.	Max.
Gross Output (Millions of 2012 Dollars)	357	341,916.67	290,155.72	227,845.00	47,619.00	1,037,226.00
Non-IT Capital (Millions of 2012 Dollars)	357	178,238.74	117,662.55	160,081.00	24,478.00	440,606.00
Labor (Hours of all persons)	357	1,604.76	1,002.50	1,262.60	225.81	4,280.48
IT Capital (Millions of 2012 Dollars)	357	13,897.94	21,197.05	5,350.00	675.00	110,730.00
Materials (Millions of 2012 Dollars)	357	191,832.81	191,533.16	112,404.63	24,295.85	669,461.24
Services (Millions of 2012 Dollars)	357	31,814.16	24,341.58	21,438.57	4,706.00	95,097.25
Energy (BTU)	357	1,201.37	1,872.09	342.26	34.50	7,320.00
Electricity (BTU)	357	158.86	135.10	128.09	15.29	577.00

for heteroskedasticity using the Huber-White process in our model specification.

Capacity is the greatest output a plant can generate given its existing IT and non-IT capital. Capacity utilization then is the proportion of actual output to capacity, capturing fluctuations in economic activity due to business cycles. In other words, in periods of slower economic activity, the level of energy input for a given level of IT and non-IT capital may decrease and vice versa. Further, the effect of business cycles may be different for different industries. Following modern IT Productivity work (Gong et al. 2021), we use capacity utilization as a control for business cycles.

Our use of panel-specific autocorrelation controls produces an auto-correlation coefficient for each industry, thereby controlling for some industry specific characteristics. Further, the use of capacity utilization controls for industry-specific business cycles. As an additional control for structural differences between industries, we add sector fixed effects (SFE) to differentiate durable manufacturing industries from non-durable manufacturing industries, recognizing that industries within each sector share relatively similar production structures.

A potential concern in typical productivity analyses is omitted correlated variables, particularly given the wide variety of factors that can affect both the independent and dependent variables. In this chapter, an example of such a variable is state-level regulation for energy efficiency. Assuming certain industries are over-represented in the affected states, such regulation can be a possible confounder if states with strong energy legislation also happened to have had IT expenditure increases. There are several ways our work allays this concern. Firstly, if the affected industries invest in IT to increase energy productivity, this is part of our hypothesis. Secondly, given that the regulation affects industries differently, we test for such industry-specific effects using IFE in our robustness test. Thirdly, in our base model, we include several controls that are industry specific including PSAR1 and capacity utilization, as well as sector-specific controls (SFE) which capture some industry-specific or sector specific effects. Finally, if the regulations came in the midst of the time series, we use GMM to test for this, where we use lagged variables as instruments.

4.4 Results

We begin by validating our dataset against prior studies. We then show the results from our main and split-intermediates models. We perform a variety of robustness checks to validate our results, and finally we interpret our results and quantify the magnitude of the effects in dollar terms.

Table 4.5: Comparison with Elasticity Estimates from Prior Studies.

Description	*Non-IT Capital*	*Labor*	*IT*	*Intermediate Inputs*	*Returns to Scale*	*Notes*
Our Study	0.229***	0.0734***	0.0985***	0.605***	1.0059	N = 357
	(0.0285)	(0.0205)	(0.0121)	(0.0212)		Years = 1998 to 2018
Gong, Nault and Rahman 2016	0.196***	0.122***	0.062**	0.611***	0.991	N = 162
	(0.009)	(0.028)	(0.0273)	(0.024)		Years = 2000 to 2008

Notes: Results of Cobb-Douglas regression in the Gross Output form, compared to Gong et al. (2016). Standard Errors are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively.

Table 4.6: Main Model Results

Model	*Non-IT Capital*	*Labor*	*IT*	*Energy*	*Materials*	*Services*	*IT Augmented Energy Productivity*
(1)	0.0928**	0.141***	0.146***	0.0838***	0.483***	0.0713***	
	(0.0374)	(0.0268)	(0.0152)	(0.0135)	(0.0202)	(0.0114)	
(2)	0.125***	0.138***	0.0858***	0.0599***	0.481***	0.0682***	5.47e-07***
	(0.0371)	(0.0263)	(0.0226)	(0.0154)	(0.0203)	(0.0116)	(2.00e-07)

Notes: Results from the Main Model. Analysis includes 17 3-digit NAICS manufacturing industries over 1998-2018. Number of observations = 357. Standard Errors are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively. Dependent variable is log of Gross Output. Includes econometric controls for Year Fixed Effects, Sector Fixed Effects, Panel-level Auto-correlation, and Heteroskedasticity. Control variables and fixed effects estimates are suppressed for brevity.

4.4.1 Dataset Validity

To validate our dataset against prior studies, we compare the results of the standard CD in (4.2), with energy, materials, and services combined into a single measure of intermediate inputs, as executed by Gong et al. (2016). As can be seen in Table 4.5, our results are consistent with the previous study, particularly given that our dataset includes newer data from 2009 to 2018, a period where the proportion of IT investments increased. This leads to a higher estimate of IT elasticity and a lower estimate for labor, which is consistent with IT capital deepening. The estimated returns to scale (calculated as the sum of the output elasticities of each input) is 1.0059, a close approximation of the underlying CD assumption of constant returns to scale.

4.4.2 Main Model Results

Table 4.6 Model (1) contains the estimates of the direct effects of each input on gross output as specified in our base model in in (4.4). Non-IT capital, labor, IT capital, energy, materials and services are positive as expected. Model (2) shows the results from the varying coefficient model in (4.6) - IT Augmented energy productivity (φ_3) is positive and significant at the 1% level, and thus, IT augments the OE of energy.

Table 4.7: Split Intermediates Model - Results

VARIABLES	(1)	(2)	(3)
Non-IT Capital	0.0655** (0.0333)	0.0572* (0.0306)	0.248*** (0.0358)
Labor	0.117*** (0.0233)	0.125*** (0.0236)	0.0358** (0.0169)
IT	0.0874*** (0.0190)	0.0993*** (0.0176)	0.0309 (0.0216)
Energy	0.0662*** (0.0134)		
Non-Electric Energy		0.0578*** (0.0105)	
Electricity			-0.0594*** (0.0164)
IT Augmented Energy Productivity	4.82e-07*** (1.75e-07)		
IT Augmented Non-Electric Energy Productivity		4.56e-07** (1.78e-07)	
IT Augmented Electricity Productivity			8.63e-07*** (3.10e-07)
Other Intermediate Inputs	0.616*** (0.0195)	0.621*** (0.0193)	0.634*** (0.0208)

Notes: Results from the Split-Intermediates Model. Analysis includes 17 3-digit NAICS manufacturing industries over 1998-2018. Number of observations = 357. Standard Errors are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively. Dependent variable is log of Gross Output. Includes econometric controls for Year Fixed Effects, Sector Fixed Effects, Panel-level Auto-correlation, and Heteroskedasticity. Control variables and fixed effects estimates are suppressed for brevity.

4.4.3 Split Intermediates Model Results

We display our results from the split intermediates model in Table 4.7. Model (1) shows the estimates from (4.8), validating the results from our main model whereby IT has a positive and significant effect on the productivity of energy. Next, using (4.11) and (4.10), we evaluate the effect of IT on non-electricity and electricity productivity in Models (2) and (3). In Model (2), the results show that IT augmented non-electric energy productivity is positive and significant at the 1% level. In Model (3) again, IT augmented electricity productivity is positive and significant. However, the direct effect of IT becomes insignificant (p-value = 15.9%) revealing deep inter-connections between IT and electric energy. In other words, at the industry level the rise in the use of IT has coincided with the use in electricity to such a degree that the interaction effect is highly significant, but this also results in a decrease in the direct effect of IT on output. These deep interconnections include the use of electricity to run IT and the use of both IT and electricity to operate non-IT capital (such as equipment and structures). Thus, IT and electricity are highly complementary in nature and have their greatest impact on output when they are available together.

4.4.4 Robustness

We now proceed to further test the robustness of our results using different econometric adjustments as well as by using another broadly different approach that we term the Constant Output Model to further validate the effect of IT on energy productivity.

Base Model - Energy Productivity

We begin by validating the IT-augmented energy productivity results from our base model (Table 4.8) by first testing it with GMM. Because estimation with GMM is based on moment conditions and not on the entire distribution of the data, GMM is widely used in panel data settings with autocorrelation and heteroskedasticity and where lagged variables are used as instruments. We instrument with one-year lags of energy, IT and the interaction of energy and IT, as well as the second lag for the key variable of interest - energy input. This is in addition to the controls for year-fixed effects, sector fixed effects, and capacity utilization as described earlier. We test for the exogeneity of our instrumental variables using the Sargan test and find that we cannot reject the null hypothesis that the instruments are exogenous ($\chi^2 = 0.483$). We find that IT Augmented Energy Productivity continues to be positive when GMM is used. As with prior studies in IT productivity, we find that non-IT capital and labor become insignificant with GMM. However, we also find that this remains fairly consistent in the GMM runs of the base CD as well as in the energy augmentation specification. We also run our model with industry fixed effects and display those results in Models (3) and (4), where our earlier results continue to hold.

Constant Output Model

The models thus far evaluate the effect of IT on the OE of energy. Our results indicate that with greater levels of IT, energy becomes more productive: IT increases the marginal product of energy and more output is produced per unit of energy. An alternate perspective is analogous to cost minimization, that is, does IT decrease the amount of energy required to produce a given level of output? To evaluate this question, we solve for e in (4.4), and substitute $\phi_2 = 1/\theta_2$,

$$e = -\phi_2 a_2 + \phi_2 y - \phi_2 \alpha_2 k - \phi_2 \beta_2 l - \phi_2 \gamma_2 z - \phi_2 \delta_2 m - \phi_2 \zeta_2 s - \epsilon_2. \tag{4.12}$$

To evaluate the relationship between z and e, we can run the estimation form

$$e = \psi_{10} + \psi_{11} y + \psi_{12} k + \psi_{13} l + \psi_{14} z + \psi_{15} m + \psi_{16} s + \epsilon_7. \tag{4.13}$$

In this simple estimation form we ask about the impact of IT on energy input given a level of output, while controlling for the other inputs. The effect of IT on energy is estimated as ψ_{14} and we expect ψ_{14} to be negative as IT decreases the amount of energy required to produce a certain level of output. The results of

Table 4.8: Robustness - Disaggregated Intermediate Inputs

	GMM		IFE	
DV: Gross Output	***(1)***	***(2)***	***(3)***	***(4)***
Non-IT Capital	0.0334	0.0413	0.241***	0.215***
	(0.0268)	(0.0444)	(0.0522)	(0.0568)
Labor	-0.00185	0.0128	0.0489**	0.123***
	(0.0111)	(0.0263)	(0.0227)	(0.0268)
IT	0.143***	0.145***	0.0769***	0.0410*
	(0.0126)	(0.0248)	(0.0151)	(0.0216)
Energy		0.0258*		-0.0246
		(0.0147)		(0.0211)
Materials		0.647***		0.440***
		(0.0196)		(0.0210)
Services		0.0700*		0.0433***
		(0.0408)		(0.00975)
Intermediate Inputs	0.788***		0.537***	
	(0.0267)		(0.0226)	
IT Augmented Energy Productivity		2.83e-07*		6.64e-07**
		(1.65e-07)		(2.85e-07)
Econometric Controls	YFE, SFE	YFE, SFE	PSAR1, He, IFE	PSAR1, He, IFE

Notes: Main Model Robustness Results. Analysis includes 17 3-digit NAICS manufacturing industries over 1998-2018. Number of observations = 357. Standard Errors are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively. Control variables and fixed effects estimates are suppressed for brevity.

Table 4.9: Robustness - Constant Output Model - Main Model

	Non-IT Capital	*Labor*	*IT*	*Materials*	*Services*	*Output*
DV: Energy	1.912***	-0.742***	-0.682***	-0.138	-0.0596**	0.522***
	(0.0850)	(0.0657)	(0.0492)	(0.0873)	(0.0253)	(0.118)

Notes: Robustness Results: Main Model Constant Output. Analysis includes 17 3-digit NAICS manufacturing industries over 1998-2018. Number of observations = 357. Standard Errors are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively. Dependent variable is log of Energy Input. Includes econometric controls for Year Fixed Effects, Sector Fixed Effects, Panel-level Auto-correlation, and Heteroskedasticity. Control variables and fixed effects estimates are suppressed for brevity.

this constant output model derived from our main model are displayed in Table 4.9. As we expect, IT has a negative effect on energy given a level of output and other inputs.

Next, we follow a similar process for the split-intermediates model by taking logs of (4.7), solving for e and substituting $\phi_4 = 1/\theta_4$ to get

$$e = -\phi_4 a_4 + \phi_4 y - \phi_4 \alpha_4 k - \phi_4 \beta_4 l - \phi_4 \gamma_4 z - \phi_4 \delta_4 r_{\setminus E}. \tag{4.14}$$

We run the estimation form for the split-intermediates constant output model

$$e = \psi_{20} + \psi_{21} y + \psi_{22} k + \psi_{23} l + \psi_{24} z + \psi_{25} r_{\setminus E} + \epsilon_8. \tag{4.15}$$

We now repeat the above process to derive the split intermediates constant output model for electricity input and non-electric energy input. Taking logs and solving for $\widetilde{e}$ in (4.9), we get the constant output specification for the effect of IT on electric energy input,

$$\widetilde{e} = \psi_{30} + \psi_{31} y + \psi_{32} k + \psi_{33} l + \psi_{34} z + \psi_{35} r_{\setminus \widetilde{E}} + \epsilon. \tag{4.16}$$

Similar to the above equation, we can find the specification to evaluate the effect of IT on non-electric energy where from (4.16), $\widehat{e}$ is substituted for $\widetilde{e}$ to get

$$\breve{e} = \psi_{40} + \psi_{41} y + \psi_{42} k + \psi_{43} l + \psi_{44} z + \psi_{45} r_{\setminus \breve{E}} + \epsilon. \tag{4.17}$$

The results of the robustness check from the specification in (4.15) are shown in the first column in Table 4.10. Here, IT capital has a significant negative effect on the consumption of energy given a level of output. The second and third columns show results from (4.16) and (4.17). Similar to the first column, we see a negative effect of IT on energy given a certain level of output and controlling for the other inputs.

Table 4.10: Constant Output - Split Intermediates Model: Results

	DV: Energy	*DV: Electricity*	*DV: Non-Electric Energy*
Non-IT Capital	1.953***	1.264***	2.096***
	(0.0854)	(0.0743)	(0.0915)
Labor	-0.631***	0.129***	-0.686***
	(0.0860)	(0.0430)	(0.0898)
IT	-0.743***	-0.383***	-0.949***
	(0.0473)	(0.0363)	(0.0525)
Other Intermediate Inputs:			
Materials and Services	-0.315***		
	(0.0915)		
Materials, Services and Non-Electricity		-0.00277	
		(0.0481)	
Materials, Services and Electricity			-0.487***
			(0.104)
Gross Output	0.689***	-0.0606	1.050***
	(0.119)	(0.0726)	(0.133)

Notes: Robustness Results: Split-Intermediates Constant Output Model. Analysis includes 17 3-digit NAICS manufacturing industries. Standard Errors are reported in parenthesis. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively. Includes econometric controls for Year Fixed Effects, Sector Fixed effects,, Panel-Level Auto-correlation, and Heteroskedasticity. Control variables and fixed effects estimates are suppressed for brevity.

4.4.5 Interpretation of the results

To calculate the magnitude of the effects for the average industry, we note that the mean levels of gross output, IT, and energy are approximately \$341,916, \$13,897, and 1,201 respectively (all dollar values are in millions of 2012 U.S. dollars). We now focus on Model (2) in Table 4.6, where the OE of IT is 0.0858, implying that the marginal product of a 1% increase IT is \$293.36. This is calculated by multiplying the OE of IT (0.0858/100) with gross output (\$341,916).[4] In other words, a 1% increase in IT (amounting to \$138.79 for the mean industry) delivers an increase of \$293.36 in gross output for the mean industry, a large increase in output by any measure, confirming prior findings about the large magnitude of IT's marginal product.

The OE of energy is 0.0599 (Model (2) in Table 4.6), amounting to a marginal product of 204.81 from a 1% increase in energy. We have estimated IT Augmented Energy Productivity at $5.47 * 10^{-7}$ from Model (2) in Table 4.6. To calculate this main effect of interest, a 1% increase in IT leads to an augmentation to the OE of energy equal to $138.79 * 5.47 * 10^{-7}$, which equals 0.00007602, and is an increase in the marginal product of energy of $0.00007602 * 341916.67/100 =$ \$0.26. This means that a 1% increase in IT is associated with a 0.13% increase in the OE of energy due to IT augmentation. Note that these augmentation effects are in addition to the direct effects of IT that were quantified earlier.

[4] MP from 1 unit $= \gamma Y/X$. MP from 1% increase in IT $= \gamma Y/100$

On the other hand, we estimate IT Augmented Electricity Productivity at $8.63 * 10^{-7}$ (Model (3), Table 4.7). Here, augmentation to the OE of electric energy from a 1% increase in IT capital is calculated to be 0.0001327. Assuming that the initial OE of electricity is proportionate to the OE of the energy input, the initial OE of electricity can be calculated as $(158.86/1201.36 * 0.0599 = 0.0079)$. In other words, a 1% increase in IT is related to a 1.52% increase in the OE of electricity from IT augmentation. Using the same method, the initial OE of non-electric energy is 0.0520 and is augmented by a 1% increase of IT to 0.0521, a 0.12% increase in its OE.

4.5 Conclusion

We developed a varying coefficient model based on the CD to estimate the effect of IT on the productivity of energy, electric energy, and non-electric energy. We began by developing the base model to test the presence of IT augmented energy productivity and then the split-intermediates model to evaluate whether IT augments electricity and non-electric energy productivity. Our results indicate that IT has a positive and significant effect on the OE of energy as well as on electric and non-electric energy. We also find that the effect of IT on electric energy is an order of magnitude larger than the effect of IT on the OE of non-electric energy - a 1% increase in IT increases the OE of non-electric energy by 0.12%, but increases the OE of electric energy by 1.52%. Considering energy more broadly, a 1% increase in IT leads to a 0.13% increase in the OE of energy input. Thus, our findings show the effects of IT on energy productivity are small but positive and they provide an important perspective for policy-makers that continue to invest heavily in IT to increase energy and electricity productivity.

One of the key limitations of our study is that we use industry-level data, not fine-grained firm-level data. This also leads to a statistically small sample size. However, our results are strong results despite the small sample size. Secondly, while most firm-level analyses focus on a small subset of firms, our data captures all U.S. manufacturing firms including small, large, public, and private, which provides greater generalizeability. A second limitation of our data is that it pertains to U.S. manufacturing firms only, meaning results may not port to other countries or other industries that have different production structures or IT intensities.

Chapter 5

Conclusion

IT today plays a central and growing role both in our daily lives and in industry. This has implications for the structure of the overall economy as well as for market power, innovation, and competition within the IT industry, which has contributed to policy-maker interest in regulating the IT industry. While at the broadest level it is straightforward to make a case for regulating IT, the details of how to regulate are complex because regulation could have undesired or unforeseen consequences. Each of our papers exposes such consequences of specific planned or implemented regulations, which we describe below.

In our first study, we look at intermediated platforms, where firms operating on a monopolistic platform sell their platform-compatible wares to end-customers. We find that a monopolistic platform's choice of transfers maximizes customer adoption by coordinating prices and investments on each side of the market. In the case of the SG, the objectives of the platform and policy-maker are aligned as far as transfers are concerned. When a platform increases fees towards profit maximizing levels, firm profits and participation decrease but effects on firm profits, participation, and end-customer adoption are more nuanced and depend on the shape of demand facing each firm. If demand is convex, less responsive to own price or more responsive to cross price, then prices can increase faster than fees, a phenomenon we term P-Prices. Further, in the presence of P-prices, it is inevitable that investments increase in fees; we term this effect P-investments. In this case, because fees increase both prices and investments, the effect on customer adoption is apriori indeterminate. In the absence of P-Prices and P-Investments, however, prices increase while investments and customer adoption decrease with an increase in fees. Thus, the decision to regulate such a monopolist platform with Ramsey-Boiteux pricing (or mechanisms) depends on the absence of P-Prices and P-Investments.

Despite the widespread prevalence of intermediated platforms, such settings have not been studied widely.

They differ from standard two-sided platforms in that another set of decisions about prices and investments are made by firms operating on the platform. We model a setting where the platform first sets the fees and transfers, and then firms choose their prices and investments in the second stage after observing the platform's actions. Our novel contribution here is the study of how multi-product firms operating on a platform set prices and investments, and how the platform can influence firm prices, investments, profits, participation, and customer adoption. Further, we theoretically articulate a new consideration as the policy-maker considers platform regulation - the presence or absence of P-Prices and P-Investments. Thus, we provide a different, or intermediated, perspective through which platforms and platform players can be evaluated, and provide insights to platforms or regulators looking to influence firm participation, investments, profits, prices, or customer adoption.

In our second study, we study the effect of the use of fines to enforce Data Portability Regulations (DPR) on the structure of the DC industry. DPR requires that DCs such as Amazon and Expedia allow DSs to download their personal data in a widely accepted format, so that DSs can port their data to other competing DCs. The intent behind this law is to improve DS rights, improve consumer surplus, and increase competition in the DC industry. We find that fines used to enforce DPR divides DCs into three groups: High capability DCs participate and comply with the DPR, mid-capability DCs participate but do not comply, and the least capable DCs cease to participate. Thus, with an increase in fines, compliance increases in two ways - some participating non-compliant DCs begin to comply, while other, less capable DCs cease participating. As well, fines increase the surplus from more capable DCs while potentially decreasing surplus from less capable participating DCs. If the increase in surplus from high capability DCs is larger than the lost surplus from less capable DCs, then total surplus increases. Thus, some of the objectives of DPR are met: improved DS rights, greater compliance and possibly greater surplus. However, this is in conjunction with some negative collateral effects: low capability DCs cease participating, the output from mid-capability DCs decreases, and the high capability DCs benefit. These results run counter to the DPR objective of reducing concentration.

We model DCs as differing in capability, which is the ability to generate more revenue per unit of output, so that the profit functions of each DC depends on its capability. Our innovation here are port functions, which capture effects from the imposition of DPR. These effects include fines and loss of revenues from non-compliance, and compliance costs as well as gains/losses in revenue from compliance. In the first stage, the policy-maker sets the fine for non-compliance with DPR, and in the second stage DCs choose whether to participate, and if so, whether to comply with the DPR. Thus, DCs decision to comply considers the effect of the compliant and non-compliant port functions as part of the payoff function.

In our third study about the effects of IT on energy productivity, we find first that over the period 1998-2018, the direct effect of IT on productivity remains strong - a 1% increase in IT capital investment amounting to \$138.79M is associated with a marginal product of \$293.36M. In addition, IT has the effect of improving energy productivity - a 1% increase in IT investment increases the OE of energy by 0.13%. As well, there have been large SG investments to modernize and computerize the electric grid, and a history of tightly integrated IT investments in the electricity supply chain. We find that IT has a large and significant effect on the OE of electricity : a 1% increase in IT increases the OE of electricity by 1.52%. On the other hand, the same increase in IT increases the OE of non-electric energy by 0.12%.

We use a structural econometric model based on the CD to derive our econometric specifications. In our main model, we use a varying coefficient on the standard CD, where inputs are Non-IT capital, IT capital, labor, energy, materials, and services, and where output is measured as gross output. Our novel contribution here is in the split-intermediates model, where we split the variable under study (energy, electricity, or non-electric energy) from intermediate inputs. Our IT-augmented energy productivity results from this model are consistent with the main model, and next, we use this model to estimate IT-augmented electricity and non-electric energy productivity. We perform a variety of robustness checks, including the constant output model which measures the effect of IT on energy input given a level of output, and find that the results are consistent across the specifications.

Thus, regulation could have undesired, but expected consequences. For example, we find across our studies that regulation can decrease IT investments, and profits of participating firms. The decrease in IT investments has consequences for overall productivity and for energy productivity as we show in our third paper. Secondly regulation can have unforeseen consequences. For example, regulators may perceive DPR to be of benefit to innovation, competition, and local DCs. However, when the DC industry is modeled as distributed by capability, one can see how DPR can instead have the effect of decreasing competition and increasing concentration. Similarly, it is possible that regulating the fees charged by the SG platform can have a negative effect on investments and potentially customer adoption. In summary, policy-makers should weigh the benefits of IT regulation against its undesired and unexpected consequences. Through our studies, we have exposed some of these consequences.

Appendix A

Appendix: Platform Intermediation and Firm Investments

A.1 The Smart Grid

The legacy electric grid is an inter-connected supply chain for electricity, consisting of generators, transmitters, distributors and customers. Generators, transmitters and distributors are generally classified as utilities, and the ownership of these can take a variety of forms including ownership by public stockholders, various levels of government, or private parties. Customers, on the other hand, include the vast number of residences and businesses. The last wave of significant North American investments in electricity infrastructure took place in the 1970's (Joskow 2012). Since then, electricity consumption has increased, consumption patterns have changed, and production technology has improved.

The U.S. Energy Independence and Security Act of 2007 states "*It is the policy of the United States to support the modernization of the nation's electricity transmission and distribution system to maintain a reliable and secure electricity infrastructure that can meet future demand growth*" (U.S. Energy Independence and Security Act 2007). To this end, the U.S. government invested over $7B through Smart Grid (SG) stimulus investments (Science Applications International Corporation 2011). The Canadian government is investing $21.6B CAD on green infrastructure over 2017-2027, a significant portion of which is slated to be used to modernize the electric grid under a set of technologies called the SG (Government of Canada, Department of Finance 2017). The European Commission expects 80% of European electricity meters to be SG compatible by 2020 (European Commission 2019).

The SG is characterized by the use of digital information and control technology to make the electric grid reliable, efficient and secure. The technology consists of sensors collecting ambient information, data storage solutions, intelligent systems to analyze the data and make decisions in real time, and control systems to effect changes in the SG. Gellings (2011) states that *"The term 'Smart Grid' refers to a modernization of the electricity delivery system so that it monitors, protects, and automatically optimizes the operation of its interconnected elements – from the central and distributed generator through the high-voltage transmission network and the distribution system, to industrial users and building automation systems, to energy storage installations, and to end-use customers, and their thermostats, electric vehicles, appliances, and other household devices"* (p. 1-1).

There are two primary reasons for the current widespread investments in the SG. Firstly, several large economies are investing in energy productivity (Vijairaghavan 2017). For instance, the U.S. Department of Energy's 2014-2018 strategic plan identifies "Improving energy productivity" as its first strategic objective (U.S. Department of Energy, 2014). The SG is able to deliver energy productivity improvements by bringing about efficiencies in electricity generation, transmission, distribution and use. Secondly, several countries are making investments in order to reduce greenhouse gas emissions. This is particularly true of signatories to the Paris Accord like the European Union, Canada, China and India (United Nations 2015). A fully implemented SG is projected to reduce greenhouse gas emissions by integrating renewable energy sources into the grid, as well as by increasing efficiency in the production and transmission of electricity. The economic benefit of high levels of adoption of the SG is estimated at between 2.8 to 6.0 times its cost (Gellings 2011), in addition to a 12% reduction in electricity sector emissions (Pratt et al. 2010).

The engineering aspects of the SG have been well-researched, and several standards relevant to the SG are available (IEEE 2017). However, the economic structure required to incentivize investments by key players in the SG such as utilities, technology companies, and customers remains an open question. The challenge lies in the nature of SG technologies and that of the electric grid - individual investments in the SG by firms in the electricity supply chain (generators, transmitters, distributors, and customers) are not as effective as coordinated investments across the supply chain. In other words, an investment by one utility, technology company, or customer in a part of the SG may not deliver the expected benefits if complementary investments are not made.

Customer demand for electricity varies with both season and time of day. Although base electricity load is supplied by thermal and nuclear generators, peaker plants are utilized when demand peaks. Even though peaker plants account for between 10 and 15 percent of generating capacity, they are utilized only for an average of 22 minutes a day (Joskow 2012). This represents a large investment in under-utilized capacity. We focus on a specific SG initiative termed Demand Response (DR), the purpose of which is to shift electricity consumption from a peak (high demand) hour to an off-peak hour through real-time pricing. The most

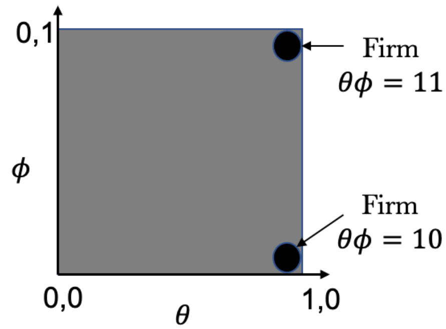


Figure A.1: Firm Distribution: $\theta\phi = 10$ and $\theta\phi = 11$

critical advantage to DR is reduced infrastructure costs and increased capacity utilization - generation, transmission, and distribution infrastructure can then be planned for a lower level of peak demand. These form compelling reasons for the policymaker's interest in ensuring wide adoption of the SG in general and DR in particular. Their intent is evidenced by the large investments by governments worldwide.

Customers were historically charged a flat electricity rate because legacy electricity meters were capable of only recording cumulative consumption over the billing period. Thus, electricity supply and demand did not reach equilibrium dynamically - the opaqueness of electricity pricing did not give customers the necessary incentives to shift consumption to an off-peak hour with lower rates. Smart meters, however, record the amount of electricity consumed, as well as when it is consumed. In addition, smart meters have a two-way communication system that can exchange usage, price, and operational control signals. These meters are typically owned, installed and operated by utilities. A combination of government mandates and subsidies has led to a penetration rate of between 50% and 100% in most OECD countries. The U.S. penetration rate of smart meters was 40.6% in 2014, and had been growing at an average of 4.9 percentage points a year between 2014 and 2016 (Federal Energy Regulatory Commission 2016). Meanwhile, China had contracted for the production of 425 million smart meters through 2015 (International Energy Agency 2018). However, the mere installation of smart meters does not enable customers to react to real-time prices. EMSs and DR-apps are also required, as described in Section 2.1.2.

A.2 Demands in a 2-firm case

In a one-sided market, a downward sloping demand implies that demand decreases in own prices. When customers purchase a solution across two sides of a market, demand depends on own-prices and cross-side prices. For instance, the demand for Microsoft Windows compatible hardware such as laptops and PCs depends on the own price set by the hardware manufacturer, but also the cross-side software prices set by the Windows compatible software makers. We model such a 2-sided market in EMS and DR-apps. Additionally, we model investments, so that demand depends on own price, own investment, cross-side price and cross-side investment. Thus, for a firm $\theta\phi$, the EMS demand generated is $d^E(\theta, \phi, P^E_{\theta\phi}, \omega^E_{\theta\phi}, P^A, \omega^A)$ and the DR-app demand generated is $d^A(\theta, \phi, P^A_{\theta\phi}, \omega^A_{\theta\phi}, P^E, \omega^E)$.

Consider a 2-firm case where the first firm has $\theta\phi = 10$ (high EMS capability = 1 and low DR-app capability = 0) and the second firm has $\theta\phi = 11$ (high EMS capability =1 and high DR-app capability = 1). This example firm distribution is depicted visually in Figure A.1. EMS Demand for the firm $\theta\phi = 10$ is then $d_E(1, 0, P^E_{10}, \omega^E_{10}, P^A_{10}, \omega^A_{10}, P^A_{11}, \omega^A_{11})$. In other words, EMS demand depends on own EMS capability, own EMS price, own EMS Investment, and cross side prices and investments in DR-apps. In a 2-firm case, there are two cross-side prices and investments, one each for $\theta\phi = 10$ and $\theta\phi = 11$. The EMS demand for the firm $\theta\phi = 10$ is shown in Figure A.2.

The DR-app demand for firm $\theta\phi = 10$ is $d_A(1, 0, P^A_{10}, \omega^A_{10}, P^E_{10}, \omega^E_{10}, P^E_{11}, \omega^E_{11})$. For the firm $\theta\phi = 11$, EMS and DR-app demands are $d_E(1, 1, P^E_{11}, \omega^E_{11}, P^A_{11}, \omega^A_{11}, P^A_{10}, \omega^A_{10})$ and $d_A(1, 1, P^A_{11}, \omega^A_{11}, P^E_{11}, \omega^E_{11}, P^E_{10}, \omega^E_{10})$ respectively.

A.3 Marginal Firm Curve

$\Upsilon(\cdot)$ is the proportion of firms that participate in the platform. The shape of the marginal firm curve can be categorized into four possible cases, and we can calculate the proportion of participating firms for each case.

- *Case 1:* If the curve intersects the lines $\theta = 1$ and $\phi = 1$ as shown in Figure 2.6, then $\hat{\phi}(1, f, t)$ represents the DR-app capability of the marginal firm with an EMS capability value of unity and given a f and t. Similarly, $\hat{\theta}(1, f, t)$ represents the EMS capability of the marginal firm with a DR-app

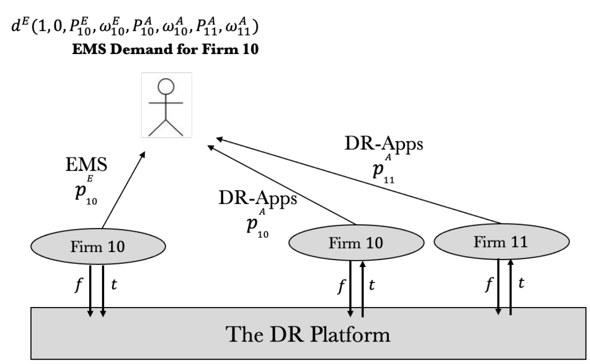


Figure A.2: EMS demand for firm $\theta\phi = 10$

capability value of unity and given f and t. The proportion of participating firms then is calculated as

$$\Upsilon_1(\cdot) = \int_{\hat{\phi}(1,f,t)}^{1} \int_{\hat{\theta}(\phi,f,t)}^{1} d\theta d\phi.$$

- *Case 2:* If the curve intersects the lines $\theta = 1$ and $\theta = 0$, then the proportion of participating firms is calculated as

$$\Upsilon_2(\cdot) = \int_{0}^{1} \int_{\hat{\theta}(\phi,f,t)}^{1} d\theta d\phi.$$

- *Case 3:* If the curve intersects the lines $\theta = 0$ and $\phi = 0$, then the proportion of participating firms is calculated as

$$\Upsilon_3(\cdot) = \int_{0}^{\hat{\phi}(0,f,t)} \int_{\hat{\theta}(\phi,f,t)}^{1} d\theta d\phi + (1 - \hat{\phi}(0,f,t)).$$

- *Case 4:* If the curve intersects the lines $\theta = 1$ and $\theta = 0$, then the proportion of participating firms is calculated as

$$\Upsilon_4(\cdot) = \int_{\hat{\phi}(1,f,t)}^{\hat{\phi}(0,f,t)} \int_{\hat{\theta}(\phi,f,t)}^{1} d\theta d\phi + (1 - \hat{\phi}(0,f,t)).$$

Note that the proportion of participating firms is decreasing in the EMS capability of the marginal firm $\partial\Upsilon(\cdot)/\partial\hat{\theta} < 0$. Similarly, $\partial\Upsilon(\cdot)/\partial\hat{\phi} < 0$. Because $\partial\Upsilon(\cdot)/\partial\hat{\theta} < 0$ and $\partial\Upsilon(\cdot)/\partial\hat{\phi} < 0$ for all four cases, our results hold generally across all the cases. We illustrate the proofs for *Case 1* where the curve intersects the lines $\theta = 1$ and $\phi = 1$ as shown in Figure 2.6.

A.4 Generalizability

Our model is generalizable to several intermediated platform settings (where firms intermediate the interaction between a platform and end customers). An example of such a platform is Microsoft Windows. Here, firms set prices for and investments in Windows compatible hardware or software (or both). Customers value maximize by making purchase decisions of the Windows compatible hardware and software. Microsoft makes decisions about the extent of cross-subsidization (transfers) between hardware providers and software providers, as well as the level of fees (See Figure A.3). Table A.1 summarizes the similarities between the DR platform and Microsoft Windows.

Our model can be used to explain the impact of Microsoft's fees and transfers on firm prices, firm investment and customer adoption. In other words, our model provides a theoretical approach to estimate the appropriate setting of fees and transfers if Microsoft's objective was to maximize investments, maximize customer adoption, minimize firm prices, or any combination thereof.

Table A.1: Windows and DR: Similarities between Intermediated Platforms

Characteristic	DR platform	Microsoft Windows platform
2-Sided Market: Customers need both in order to get value from the platform.	EMS and DR-applications are the two sides of the platform.	Microsoft Windows compatible hardware and software form the two sides of the platform
Intermediated Platform: Platform intermediates the interaction between firms, and between firms and end customers.	Firms operating on the platform choose their prices and level of investment on EMS and DR-Applications.	Firms choose their prices and level of investment on Windows compatible hardware and software
Player compensation: Players keep a portion of the price paid by the end customer, and this is determined through fees and transfers	Consumers pay firms for the purchase of EMS or DR-apps. The platform provider keeps fee, and adjusts the transfer of each side in order to cross-subsidize.	Consumers pay firms for the purchase of hardware or software. Microsoft keeps fee, and adjusts the transfer to each side in order to cross-subsidize.
Decisions: Profit/Utility Maximization by each of the players	Firms choose prices for and investment in EMS and DR-applications, the platform owner chooses fees and transfer, customers choose whether to participate.	Firms choose prices for and investment in PC hardware and Windows applications, Microsoft chooses fees and transfer, customers choose whether to participate.

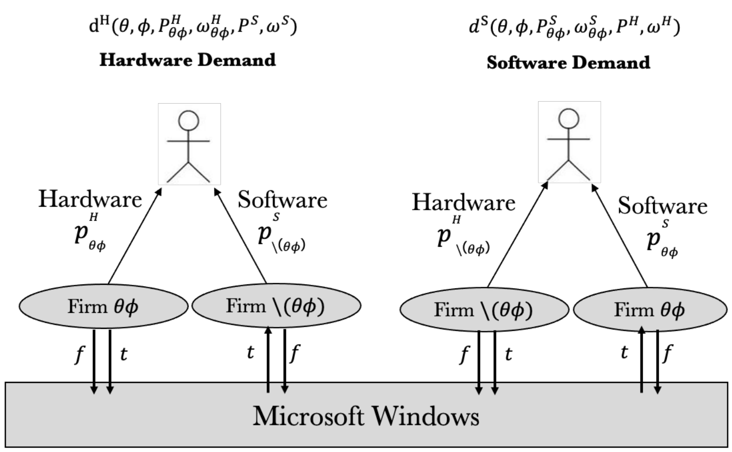


Figure A.3: Generalizability of our model to other platform intermediated settings: An example

Appendix B

Appendix: Energy Productivity Implications of IT Investments

B.1 Visual Representation of the Mean of the Key Variables over Time

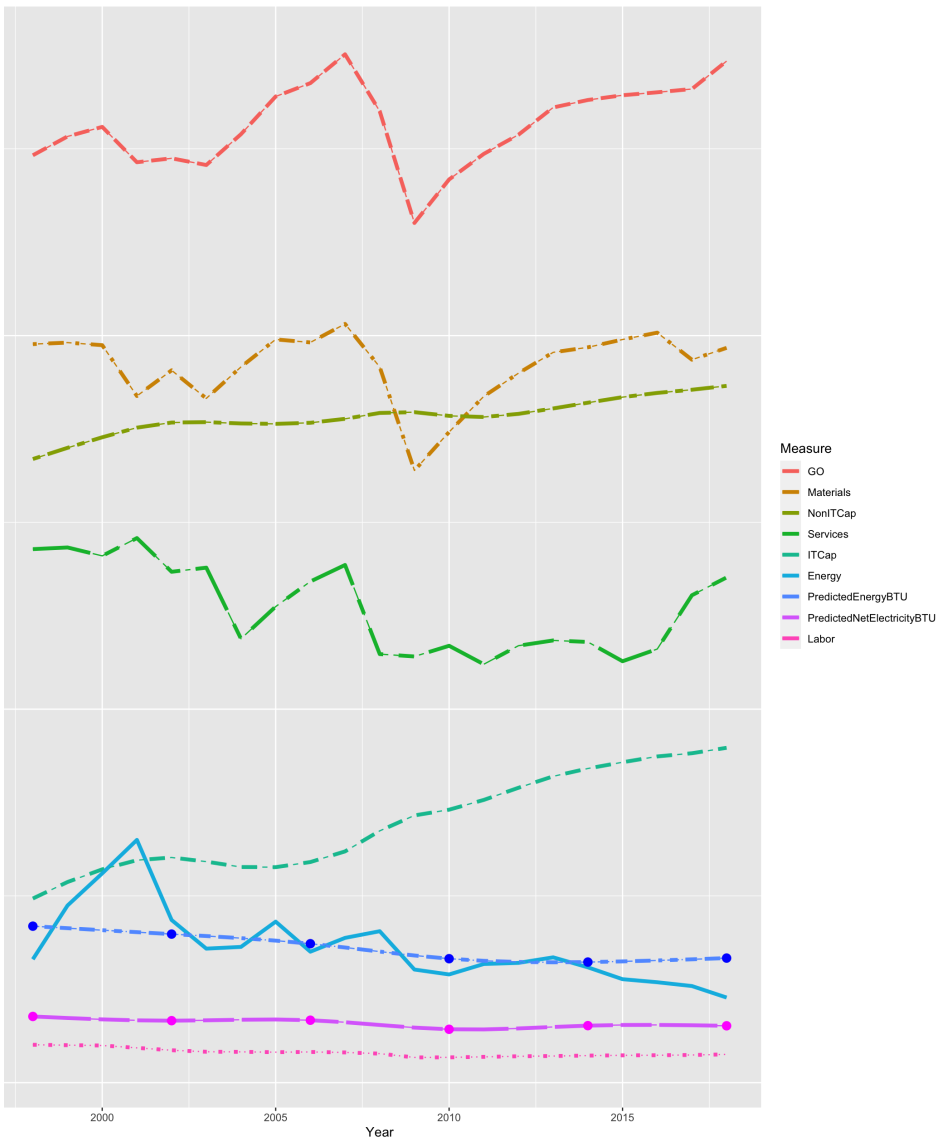


Figure B.1: KLEMS : Mean Across Industries over Time.

Notes: The six points on the lines representing *PredictedEnergyBTU* and *PredictedNetElectricityBTU* correspond to the original data from the EIA. The rest of the line consists of interpolated data. Each variable is scaled differently for visualization purposes. Thus, this chart visualizes growth in the mean value of each variable over time, but cannot be used to compare one variable against another.

B.2 Apparel, Leather and Allied Products (Naics 315,316) - Inputs and Outputs Over Time

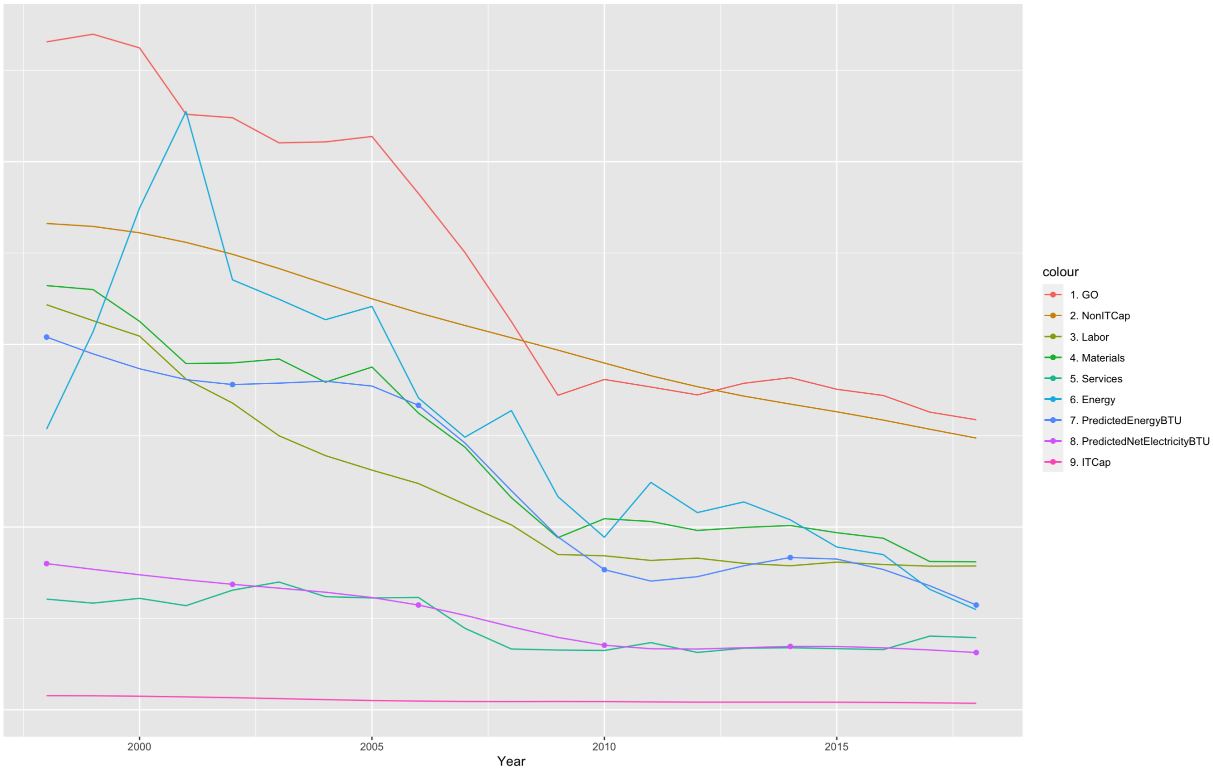


Figure B.2: Naics 315,316 - Inputs and Outputs Over Time.

Notes: Each variable is scaled differently for visualization purposes. Thus, this chart visualizes growth in the mean value of each variable over time, but cannot be used to compare one variable against another.

Appendix C

Appendix: Copyright Forms

Copyright Authorization

I, Barrie Nault, as a co-author of the paper "Platform Intermediation and Firm Investments", give permission to Vaarun Vijairaghavan to include the work in a thesis titled "How Should Information Technology be Regulated?". I am aware that Vaarun Vijairaghavan will sign a Thesis Non-Exclusive License that authorizes Library and Archives Canada to reproduce, communicate to the public on the Internet, loan, distribute or sell copies of this thesis.

B.R. NAULT
Name

B.R. Nault
Signature

June 2, 2021
Date

Copyright Authorization

I, Barrie R. Nault, as a co-author of the paper "Winner Take All Through Data Portability", give permission to Vaarun Vijairaghavan to include the work in a thesis titled "How Should Information Technology be Regulated?". I am aware that Vaarun Vijairaghavan will sign a Thesis Non-Exclusive License that authorizes Library and Archives Canada to reproduce, communicate to the public on the Internet, loan, distribute or sell copies of this thesis.

B.R. NAULT

Name

B.R. Nault

Signature

June 2, 2021

Date

Copyright Authorization

I, Hooman Hidaji, as a co-author of the paper “Winner Take All Through Data Portability", give permission to Vaarun Vijairaghavan to include the work in a thesis titled “How Should Information Technology be Regulated?". I am aware that Vaarun Vijairaghavan will sign a Thesis Non-Exclusive License that authorizes Library and Archives Canada to reproduce, communicate to the public on the Internet, loan, distribute or sell copies of this thesis.

Hooman Hidaji

Name

Signature

2021/5/27

Date

Bibliography

Daron Acemoglu, Philippe Aghion, Leonardo Bursztyn, and David Hemous. The environment and directed technical change. *American economic review*, 102(1):131–66, 2012.

Inaki Aguirre, Simon Cowan, and John Vickers. Monopoly price discrimination and demand curvature. *American Economic Review*, 100(4):1601–15, 2010.

Dennis J Aigner and Shih-Fan Chu. On estimating the industry production function. *The American Economic Review*, 58(4):826–839, 1968.

Ebtesam Alomari, Ahmed Barnawi, and Sherif Sakr. Cdport: A framework of data portability in cloud platforms. In *Proceedings of the 16th International Conference on information integration and web-based applications & services*, volume 04-06, pages 126–133. ACM, 2014. ISBN 9781450330015.

Edward Anderson, Geoffrey G Parker, and Burcu Tan. Platform performance investment in the presence of network externalities. *Information Systems Research*, 25(1):152–172, 2014.

Mark C Anderson, Rajiv D Banker, and Sury Ravindran. Value implications of investments in information technology. *Management Science*, 52(9):1359–1376, 2006.

Baris Ata, A Serasu Duran, and Ozge Islegen. An analysis of time-based pricing in retail electricity markets. *Available at SSRN 2826055*, 2018.

Sulin Ba and Barrie R Nault. Emergent themes in the interface between economics of information systems and management of technology. *Production and Operations Management*, 26(4):652–666, 2017.

J Yannis Bakos and Barrie R Nault. Ownership and investment in electronic networks. *Information Systems Research*, 8(4):321–341, 1997.

Sundar Bharadwaj, Anandhi Bharadwaj, and Elliot Bendoly. The performance effects of complementarities between information systems, marketing, manufacturing, and supply chain processes. *Information systems research*, 18(4):437–453, 2007.

Uldis Bojārs, Alexandre Passant, JG Breslin, and Stefan Decker. Social network and data portability using semantic web technologies. In *Proceedings of the BIS 2008 Workshop on Social Aspects of the Web, Innsbruck, Austria (May 2008)*, 2008.

S. Bougain, D. Gerhard, C. Nigischer, and S. Uĝurlu. Towards energy management in production planning software based on energy consumption as a planning resource. *Procedia CIRP*, 26:139–144, 2015. ISSN 2212-8271. doi: https://doi.org/10.1016/j.procir.2014.07.093. URL `https://www.sciencedirect.com/science/article/pii/S2212827114009068`. 12th Global Conference on Sustainable Manufacturing – Emerging Potentials.

Francesco Clavorà Braulin and Tommaso Valletti. Selling customer information to competing firms. *Economics Letters*, 149:10–14, 2016.

GH Brundtland and M Khalid. Our common future: Brundtland commission world commission on environment and development (wced), 1987.

Erik Brynjolfsson. Information assets, technology and organization. *Management Science*, 40(12):1645–1662, 1994.

Erik Brynjolfsson and Lorin Hitt. Paradox lost? firm-level evidence on the returns to information systems spending. *Management science*, 42(4):541–558, 1996.

Erik Brynjolfsson and Lorin Hitt. Computing productivity: Firm-level evidence. *The Review of Economics and Statistics*, 85(4):793–808, 2003.

California State Legislature. California consumer privacy act of 2018, 2018. `http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5`.

Young Bong Chang and Vijay Gurbaxani. Information technology outsourcing, knowledge transfer, and firm productivity: An empirical analysis. *MIS quarterly*, 36(4), 2012.

Zhuo Cheng and Barrie R Nault. Industry level supplier-driven it spillovers. *Management Science*, 53(8): 1199–1216, 2007.

Zhuo Cheng and Barrie R Nault. Relative industry concentration and customer-driven it spillovers. *Information Systems Research*, 23(2):340–355, 2012.

Laurits Christensen, Andrea Colciago, Federico Etro, and Greg Rafert. The impact of the data protection regulation in the EU. *Intertic Policy Paper*, 2013.

Paul Chwelos, Ronald Ramirez, Kenneth L Kraemer, and Nigel P Melville. Research note?does technological progress alter the nature of information technology as a production input? new evidence and new results. *Information Systems Research*, 21(2):392–408, 2010.

Cloud Standards Customer Council. Interoperability and Portability for Cloud Computing: A Guide. Technical report, Cloud Standards Customer Council, 12 2017.

Jacqueline Corbett, Katherine Wardle, and Chialin Chen. Toward a sustainable modern electricity grid: The effects of smart metering and program investments on demand-side management performance in the us electricity sector 2009-2012. *IEEE Transactions on Engineering Management*, 65(2):252–263, 2018.

Council of European Union. Regulation 2016/679 of the European Parliament and of the council, 2016. `https://publications.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en`.

Bennet Cyphers and Danny O'Brien. Facing Facebook: Data portability and interoperability are anti-monopoly medicine. *EFF*, 2018. URL `https://www.eff.org/deeplinks/2018/07/facing-facebook-data-portability-and-interoperability-are-anti-monopoly-medicine`.

Steven J Davis, Ken Caldeira, and H Damon Matthews. Future co2 emissions and climate change from existing energy infrastructure. *Science*, 329(5997):1330–1333, 2010.

Alexandre de Corniere and Greg Taylor. Data and competition: A general framework with applications to mergers, market structure, and privacy policy. *CEPR Discussion Paper No. DP14446*, 2020.

Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, and Ignacio Sanchez. The right to data portability in the gdpr: Towards user-centric interoperability of digital services. *Computer Law & Security Review*, 34(2):193–203, 2018.

Jason Dedrick, Vijay Gurbaxani, and Kenneth L. Kraemer. Information technology and economic performance: A critical review of the empirical evidence. *ACM Comput. Surv.*, 35(1):1–28, March 2003. ISSN 0360-0300. doi: 10.1145/641865.641866. URL `http://doi.acm.org/10.1145/641865.641866`.

Bruce Dehning, Vernon J Richardson, and Robert W Zmud. The value relevance of announcements of transformational information technology investments. *Mis Quarterly*, pages 637–656, 2003.

Sanjeev Dewan and Kenneth L Kraemer. Information technology and productivity: evidence from country-level data. *Management Science*, 46(4):548–562, 2000.

Avinash Dixit and Victor Norman. Advertising and welfare. *Bell Journal of Economics*, 9(1):1–17, 1978.

Joost R. Duflou, John W. Sutherland, David Dornfeld, Christoph Herrmann, Jack Jeswiet, Sami Kara, Michael Hauschild, and Karel Kellens. Towards energy and resource efficient manufacturing: A processes and systems approach. *CIRP Annals*, 61(2):587–609, 2012. ISSN 0007-8506. doi: https://doi.org/10.1016/j.cirp.2012.05.002. URL `https://www.sciencedirect.com/science/article/pii/S0007850612002016`.

R Eisenmann, T. Platform mediated networks: Definitions and core concepts. *Harvard Business School*, 9-807-049, 2007.

Electric Choice. Map of deregulated energy states and markets, 2018. URL `https://www.electricchoice.com/map-deregulated-energy-markets/`. last accessed on 2020-01-31.

Eric Emch and T Scott Thompson. Market definition and market power in payment card networks. *Review of Network Economics*, 5(1), 2006.

Barbara Engels. Data portability among online platforms. *Internet Policy Review*, 5(2), 2016.

European Commission. Smart grids and meters, Jul 2019. URL `https://ec.europa.eu/energy/en/topics/markets-and-consumers/smart-grids-and-meters`. last accessed on 2019-06-01.

Diana Farrell, Scott S Nyquist, and Matthew C Rogers. Making the most of the world's energy resources. *McKinsey Quarterly*, 1:20, 2007.

Joseph Farrell and Carl Shapiro. Dynamic competition with switching costs. *The RAND Journal of Economics*, pages 123–137, 1988.

Federal Energy Regulatory Commission. Assessment of demand response and advanced metering. Technical report, 2016. URL `https://www.ferc.gov/legal/staff-reports/2016/DR-AM-Report2016.pdf`. last accessed on 2020-01-31.

Esther Gal-Or, Ronen Gal-Or, and Nabita Penmetsa. The role of user privacy concerns in shaping competition among platforms. *Information Systems Research*, 29(3):698–722, 2018.

C Gellings. Estimating the costs and benefits of the smart grid: A preliminary estimate of the investment requirements and resultant benefits of a fully functioning smart grid. Technical Report 1022519, Electric Power Research Institute, Palo Alto, CA., 2011.

Xianjun Geng, Yinliang Tan, and Lai Wei. How add-on pricing interacts with distribution contracts. *Production and Operations Management*, 27(4):605–623, 2018.

Anindya Ghose and Sang Pil Han. Estimating demand for mobile applications in the new economy. *Management Science*, 60(6):1470–1488, 2014.

Fengmei Gong, Barrie R Nault, and Mohammad S Rahman. Research note - an internet-enabled move to the market in logistics. *Information Systems Research*, 27(2):440–452, 2016.

Fengmei Gong, Zhuo Cheng, and Barrie R Nault. The different effects of hardware and software on production interdependence in manufacturing. *Decision Support Systems*, page 113521, 2021.

Government of Canada, Department of Finance. Budget in brief, Mar 2017. URL `http://www.budget.gc.ca/2017/docs/bb/brief-bref-en.html`. last accessed on 2020-01-31.

Inge Graef, Jeroen Verschakelen, and Peggy Valcke. Putting the right to data portability into a competition law perspective. *Law: The Journal of the Higher School of Economics, Annual Review*, pages 53–63, 2013.

Timothy R Graeff and Susan Harmon. Collecting and using personal data: consumers' awareness and concerns. *Journal of consumer marketing*, 19(4):302–318, 2002.

Alok Gupta. Editorial thoughts: What and how isr publishes. *Information Systems Research*, 28(1):1–4, 2017.

Vijay Gurbaxani, Nigel Melville, and Kenneth Kraemer. The production of information services: a firm-level analysis of information systems budgets. *Information systems research*, 11(2):159–176, 2000.

Oliver Hart and John Moore. Property rights and the nature of the firm. *Journal of political economy*, 98 (6):1119–1158, 1990.

Trevor Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779, 1993.

Friedrich August Hayek. The use of knowledge in society. *The American economic review*, 35(4):519–530, 1945.

John Hicks. *The theory of wages*. Macmillan, London, 1932.

Bengt Hölmstrom. Moral hazard and observability. *The Bell journal of economics*, pages 74–91, 1979.

IEEE. Smart grid standards. Online, 2017. URL `https://smartgrid.ieee.org/resources/standards`. last accessed on 2020-01-31.

International Energy Agency. Global EV Outlook 2018. Technical report, 2018. URL `https://www.iea.org/reports/global-ev-outlook-2018`. last accessed on 2020-01-31.

ISO Central Secretary. Information technology — cloud computing — interoperability and portability. Standard ISO/IEC 19941:2017, International Organization for Standardization, Geneva, CH, 2017. URL `https://www.iso.org/standard/66639.html`.

Koichiro Ito. Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *American Economic Review*, 104(2):537–63, 2014.

Geoffrey A Jehle and Philip J Reny. *Advanced microeconomic theory*. Financial Times/Prentice Hall, Harlow, England, 3rd edition, 2011. ISBN 9780273731917.

Justin P Johnson and David P Myatt. On the simple economics of advertising, marketing, and product design. *American Economic Review*, 96(3):756–784, 2006.

Paul L Joskow. Creating a smarter U.S. electric grid. *The Journal of Economic Perspectives*, 26(1):29–47, 2012.

Alfred E Kahn. Applications of economics to an imperfect world. *The American Economic Review*, 69(2): 1–13, 1979.

Michael L Katz and Carl Shapiro. Network externalities, competition, and compatibility. *American economic review*, 75(3):424–440, 1985.

Brent Kendall and John D. McKinnon. Congress, enforcement agencies target tech. *The Wall Street Journal*, 2019. URL `https://www.wsj.com/articles/ftc-to-examine-how-facebook-s-practices-affect-digital-competition-11559576731?mod=searchresults&page=1&pos=1`.

Jin-Hyuk Kim, Liad Wagman, and Abraham L Wickelgren. The impact of access to consumer data on the competitive effects of horizontal mergers and exclusive dealing. *Journal of Economics & Management Strategy*, 28(3):373–391, 2019.

Paul R Kleindorfer, Kalyan Singhal, and Luk N Van Wassenhove. Sustainable operations management. *Production and operations management*, 14(4):482–492, 2005.

Paul Klemperer. Markets with consumer switching costs. *The quarterly journal of economics*, 102(2):375–394, 1987.

Dennis O Kundisch, Neeraj Mittal, and Barrie R Nault. Research commentary - using income accounting as the theoretical basis for measuring it productivity. *Information Systems Research*, 25(3):449–467, 2014.

Wing Man Wynne Lam and Xingyi Liu. Does data portability facilitate entry? *International Journal of Industrial Organization*, 69:102564, 2020.

Richard A Lambert. Contracting theory and accounting. *Journal of accounting and economics*, 32(1-3): 3–87, 2001.

Byungtae Lee and Anitesh Barua. An integrated assessment of productivity and efficiency impacts of information technology investments: Old data, new analysis and evidence. *Journal of Productivity Analysis*, 12(1):21–43, 1999.

Maurice D Levi and Barrie R Nault. Converting technology to mitigate environmental damage. *Management Science*, 50(8):1015–1030, 2004.

Litos Communication. The smart grid: An introduction. Technical report, International Trade Administration, Department of Commerce, Washington, D.C., 2010.

James Mackintosch. Behemoths have dominated the market before, but tech is different. *The Wall Street Journal*, 2018. URL `https://www.wsj.com/articles/behemoths-have-dominated-the-market-before-but-tech-is-different-1528997537`.

Neeraj Mittal and Barrie R Nault. Research note: investments in information technology: Indirect effects and information technology intensity. *Information Systems Research*, 20(1):140–154, 2009.

T Nagasawa, C Pillay, G Beier, K Fritzsche, F Pougel, T Takama, and I Bobashev. Accelerating clean energy through industry 4.0: Manufacturing the next revolution. *United Nations Industrial Development Organization*, 2017.

Barrie R Nault. Equivalence of taxes and subsidies in the control of production externalities. *Management Science*, 42(3):307–320, 1996.

Barrie R Nault and Albert S Dexter. Agent-intermediated electronic markets in international freight transportation. *Decision Support Systems*, 41(4):787–802, 2006.

Barrie R Nault and Mark B Vandenbosch. Disruptive technologies—explaining entry in next generation information technology markets. *Information Systems Research*, 11(3):304–319, 2000.

Barrie R Nault and Steffen Zimmermann. Balancing openness and prioritization in a two-tier internet. *Forthcoming in Information Systems Research*, 2019.

Giancarlo Nota, Francesco David Nota, Domenico Peluso, and Alonso Toro Lazo. Energy efficiency in industry 4.0: The case of batch production processes. *Sustainability*, 12(16):6631, 2020.

Stephen D Oliner and Daniel E Sichel. The resurgence of growth in the late 1990s: Is information technology the story? *The Journal of Economic Perspectives*, 14(4):3, 2000.

Geoffrey G Parker and Marshall W Van Alstyne. Two-sided network effects: A theory of information product design. *Management science*, 51(10):1494–1504, 2005.

A Mitchell Polinsky and Steven Shavell. The economic theory of public enforcement of law. *Journal of economic literature*, 38(1):45–76, 2000.

David Popp. Induced innovation and energy prices. *American Economic Review*, 92(1):160–180, March 2002.

Robert G Pratt, Patrick J Balducci, Clint Gerkensmeyer, Srinivas Katipamula, Michael CW Kintner-Meyer, Thomas F Sanquist, Kevin P Schneider, and Thomas J Secrest. The smart grid: an estimation of the energy and co_2 benefits. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2010.

Davy Preuveneers and Elisabeth Ilie-Zudor. The intelligent industry of the future: A survey on emerging trends, research challenges and opportunities in industry 4.0. 9:287–298, 2017.

Arun Rai, Ravi Patnayakuni, and Nainika Seth. Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS quarterly*, pages 225–246, 2006.

Amrit Rajkumar. Alphabet becomes fourth u.s. company to reach $1 trillion market value. *The Wall Street Journal*, 2020. URL `https://www.wsj.com/articles/alphabet-becomes-fourth-u-s-company-to-ever-reach-1-trillion-market-value-11579208802`.

Liana Razmerita, Martynas Jusevičius, and Rokas Firantas. New generation of social networks based on semantic web technologies: The importance of social data portability. In *Proceedings of the international workshop on adaptation and personalization for web*, volume 2, pages 79–87, 2009.

Jean-Charles Rochet and Jean Tirole. Two-sided markets: a progress report. *The RAND journal of economics*, 37(3):645–667, 2006.

D Saygin, E Worrell, MK Patel, and DJ Gielen. Benchmarking the energy use of energy-intensive industries in industrialized and in developing countries. *Energy*, 36(11):6661–6673, 2011.

Science Applications International Corporation. Smart Grid Around the World: Selected Country Overviews. Technical report, Energy Information Administration, Washington, DC., 2011.

Robert Seamans and Washington Bytes. Data portability and competition between technology platforms. *Forbes*, 2018. URL `https://www.forbes.com/sites/washingtonbytes/2018/03/06/data-portability-and-competition-between-technology-platforms`.

Steven Shavell. The optimal use of nonmonetary sanctions as a deterrent. *The American Economic Review*, pages 584–592, 1987.

Mahdi Negahi Shirazi, Ho Chin Kuan, and Hossein Dolatabadi. Design patterns to enable data portability between clouds' databases. In *2012 12th International Conference on Computational Science and Its Applications*, pages 117–120. IEEE, 2012.

F Shrouf and G Miragliotta. Energy management based on internet of things: practices and framework for adoption in production management. *Journal of Cleaner Production*, 100:235–246, 2015. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2015.03.055. URL `https://www.sciencedirect.com/science/article/pii/S0959652615002760`.

Eugene Silberberg. *The structure of economics: A mathematical analysis.* McGraw-Hill/Irwin, second edition, 1990.

Julia K Steinberger and Fridolin Krausmann. Material and energy productivity. *Environmental Science and Technology*, pages 11690–1176, 2011.

Kevin J Stiroh. Information technology and the us productivity revival: what do the industry data say? *The American economic review*, 92(5):1559–1576, 2002.

Peter Swire and Yianni Lagos. Why the right to data portability likely reduces consumer welfare: antitrust and privacy critique. *Md. L. Rev.*, 72:335, 2012.

Y. R. Tan and J. Carrillo. The agency model for digital goods: Strategic analysis of dual channels in electronic publishing industry. In *Proceedings of PICMET '14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration*, pages 646–657, 2014.

Jean Tirole. *The theory of industrial organization.* MIT press, 1988.

United Nations. Adoption of the paris agreement. Number FCCC/CP/2015/L.9/Rev.1, Paris, 2015. United Nations.

United Nations. The Paris Agreement, 2016. URL `https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement`. Last accessed on 2021/05/03.

United States House of Representatives. Majority staff report and recommendations on investigation of competition in digital markets, 2020.
`https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf`.

U.S. Department of Energy. Strategic plan 2014-2018. *U.S. Department of Energy*, Mar 2014. URL `https://www.energy.gov/sites/default/files/2014/04/f14/2014_dept_energy_strategic_plan.pdf`.

U.S. Energy Independence and Security Act. 42 USCS Chapter 152 Subchapter 9 Section 17381, 2007.

Vaarun Vijairaghavan. Energy productivity implications of IT investments. *Working Paper.*, 2017. Haskayne School of Business, University of Calgary.

Wenbo Wang, Haidong Yang, Yingfeng Zhang, and Jianxue Xu. Iot-enabled real-time energy efficiency optimisation method for energy-intensive manufacturing enterprises. *International Journal of Computer Integrated Manufacturing*, 31(4-5):362–379, 2018. doi: 10.1080/0951192X.2017.1337929.

Wikipedia. Ramsey problem, 2020. URL `https://en.wikipedia.org/wiki/Ramsey_problem`. last accessed on 2021-06-01.

Michael Wohlfarth. Data portability on the Internet. *Business & Information Systems Engineering*, 61(5): 551–574, 2019.

Yuliang Yao and Kevin Xiaoguo Zhu. Research note—do electronic linkages reduce the bullwhip effect? an empirical analysis of the us manufacturing supply chains. *Information Systems Research*, 23(3-part-2): 1042–1055, 2012.

Ghafoorpoor Poorya Yazdi, Aydin Azizi, and Majid Hashemipour. An empirical investigation of the relationship between overall equipment efficiency (oee) and manufacturing sustainability in industry 4.0 with time study approach. *Sustainability*, 10(9):3031, 2018.

Dawei Zhang, Zhuo Cheng, Hasan A Qurban H Mohammad, and Barrie R Nault. Research commentary - information technology substitution revisited. *Information Systems Research*, 26(3):480–495, 2015.