Sources and Consequences of Market Power: Insights from Antitrust and Tax Policy

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

2024

vorgelegt von
Robin Mamrak

Referentin:	Prof. Dr. Dr. h.c. Monika Schnitzer
Korreferent:	Prof. Dr. Martin Watzinger
Promotionsabschlussberatung:	10. Juli 2024

Tag der mündlichen Prüfung: 26. Juni 2024

Namen der Berichterstatter: Monika Schnitzer, Martin Watzinger, Fabian Waldinger

Acknowledgements

First and foremost, I would like to thank my supervisor, Monika Schnitzer, for her continuous support over the past years. Her excellent questions, insightful feedback, and helpful advice crucially shaped this dissertation. I learnt a lot from her and I am particularly thankful for having had the chance to be part of her research group. I am also very grateful to my second advisor, Martin Watzinger, who taught me how to do research at the beginning of my doctoral studies. Conversations with him were an important source of inspiration for my projects. Finally, I would like to thank Fabian Waldinger for completing my dissertation committee. His critical questions and comments during seminars greatly improved the chapters of this dissertation.

My co-authors Felix Montag, Alina Sagimuldina, Monika Schnitzer, Martin Watzinger, and Kathrin Wernsdorf deserve special credit for this dissertation. I thank all of them for the productive collaboration as well as the numerous insights I gained from our joint work and discussions, even though not all projects were successful. I also wish to thank Alessandra Allocca and Markus Nagler, who were excellent mentors. They provided very helpful advice, guidance, and feedback, including on early-stage ideas or drafts.

My time at the Munich Graduate School of Economics would not have been the same without my fellow doctoral students, some of whom became good friends. I am extremely grateful to Lena Greska, Sebastian Hager, and Bernhard Kassner, whose company made the last few years so much more fun. Their encouragement helped me overcome even the most frustrating phases of the doctorate. I also extend my gratitude to my (former) colleagues at the Seminar for Comparative Economics, Marie-Louise Arlt, Philipp Brewing, Anna Gumpert, Vanja Milanovic, Fabrice Naumann, Maximilian Schader, Ines Steinbach, and Xuan Teng for the always friendly environment. In addition, I thank the student assistants for their excellent support.

During my doctoral studies, I was lucky to spend two periods abroad. Working as a trainee in the Chief Economist Team at DG COMP at the European Commission was an extremely insightful experience and gave me additional motivation to write this dissertation. I also had the chance to spend four months as a visiting researcher at the Laboratory for Innovation Science at Harvard, which allowed me to meet many fascinating people and discover the vibrant atmosphere there. I thank both institutions for their hospitality and for making my stays abroad so enjoyable. Moreover, I am indebted to Monika Schnitzer for making both these experiences possible.

I gratefully acknowledge financial support by the German Research Foundation through CRC TRR 190, by the Elite Network of Bavaria through the international doctoral programme 'Evidence-Based Economics', and by Fulbright Germany through a doctoral scholarship for my research stay at Harvard. This dissertation also benefited from numerous comments and discussions at conferences and seminars.

Most of all, I thank my family, who has been instrumental in shaping my journey. I am incredibly grateful to my parents for their unwavering support, patience, and inspiration, and for the countless opportunities they gave me. My deepest gratitude goes to my wife, Lisa, for always being there for me. Without her endless love, encouragement, and support, this dissertation would not have been possible.

Robin Mamrak
March 2024

Contents

List of Figures

List of Tables

Preface

For several decades, economists have acknowledged that market power is present in many markets (e.g., Bresnahan, 1989). Microeconomic textbooks define market power as the ability of a firm to raise its price above marginal cost, hence allowing the firm to charge a mark-up on its output (e.g., Goolsbee et al., 2016). Market power not only reduces consumer welfare, but it also has important implications for different areas of policymaking. Some policies, such as antitrust enforcement, directly aim to curtail excessive market power by promoting competition. Many other policies, such as taxation, have objectives that are unrelated to market power. Yet, their effectiveness may still be influenced by the presence of market power.

A number of highly influential recent studies document an increase in aggregate market power over the last decades (De Loecker and Eeckhout, 2018; De Loecker et al., 2020). This development has been made responsible for secular trends such as the decline in business dynamism (Eggertsson et al., 2021; Akcigit and Ates, 2023) or the fall of the labour share of output (Autor et al., 2020; Barkai, 2020). These findings have sparked an intense debate in the economics literature (e.g., Basu, 2019; Berry et al., 2019; Syverson, 2019), highlighting the increasingly important role of market power throughout the economy.

This dissertation provides further insights into the sources and consequences of market power from a microeconomic perspective. The chapters in this dissertation each focus on an individual industry – copiers, consumer electronics, and retail fuel, respectively – and address policy questions in settings where firms' market power, and partly the source thereof, is important.

The first two chapters deal with (quasi-)monopolists whose market power stems from patents, whereas the third chapter studies imperfect consumer information about prices as a source of market power for firms in an oligopolistic market. Chapter 1 analyses the effect of antitrust enforcement on innovation and shows that extensive market power can lead to a lack of innovation. Chapter 2 investigates the impact of technology transfer on innovation in an emerging market, leveraging a case where technology transfer was the result of an antitrust intervention. Finally, chapter 3 shows that imperfect price information reduces commodity tax pass-through by softening competition between firms.

The first chapter investigates how antitrust enforcement against patent-based monopolies affects innovation. Patents and other types of intellectual property (IP) are an important source of market power. For example, firms may strategically use their patents to block entry by refusing to license their technology to potential competitors. In fact, the use of IP for exclusionary purposes has been suggested as one driving force behind the rise in aggregate market power (Berry et al., 2019; Akcigit and Ates, 2023).

Specifically, I study the antitrust case against Xerox Corporation in the early 1970s. Xerox was the de facto monopolist in the US market for plain-paper copiers, primarily because of its large patent portfolio. Xerox refused to grant licenses to other copier producers and used patent infringement suits to block entry of potential competitors. In 1972, the Federal Trade Commission charged Xerox with monopolization of the copier market. The case was settled by a consent decree in 1975. As the key remedy, Xerox had to license all its domestic and foreign copier-related patents to any third parties at reasonable rates.

To investigate the impact of the Xerox case on subsequent innovation, I look at the number of patents filed in the US by firms other than Xerox or its subsidiaries. I employ a difference-in-differences strategy that compares patenting across technology classes with a differential exposure to compulsory licensing of Xerox's patents.

I find that antitrust enforcement against Xerox had an overall positive effect on subsequent innovation by other firms in the copier industry. This effect is primarily driven by Japanese applicants. Among them, the positive innovation effect was most pronounced for firms with extensive prior patenting experience in copier technologies. Therefore, the main beneficiaries of access to Xerox's technology were potential competitors from Japan.

These findings suggest that society lacked some innovations during Xerox's patent-protected monopoly. In that way, antitrust intervention against Xerox also brought enormous benefits to American consumers. They experienced lower prices and more innovative copiers, irrespective of whether the innovators were Japanese.

The chapter contributes to a growing empirical literature on antitrust and innovation (e.g., Watzinger et al., 2020; Kang, 2021; Poege, 2022; Watzinger and Schnitzer, 2022) by analysing the differential impact of antitrust enforcement on innovation by domestic and foreign firms. This relates to recent policy concerns in the US that stricter antitrust enforcement against domestic incumbents could undermine the American dominance of the high-technology sector, as it may particularly help foreign competitors to catch up. The evidence from the first chapter shows that, in the case of Xerox, US antitrust intervention primarily helped Japanese competitors to build on Xerox's technology. This is because Japanese entrants employed a different business model that allowed them to expand the market for plain-paper copiers to the lower-volume segment, which had been ignored by Xerox. In line with this narrative evidence, I show that Japanese innovation became more novel and diverse following the antitrust case.

The chapter also adds to the literature on compulsory licensing of patents (e.g., Acemoglu and Akcigit, 2012; Moser and Voena, 2012; Watzinger et al., 2020). It shows that compulsory licensing as an antitrust remedy can promote innovation by other firms *within* the target industry if patents are the main entry barrier. Most closely related, Watzinger et al. (2020) study compulsory licensing in the antitrust case against Bell in the 1950s. They find no effect in the target sector, because Bell could continue to foreclose its rivals even after the loss its patents. Conversely, in the case of Xerox, compulsory licensing removed the main barrier to entry. Thus, the chapter shows how the source of market power impacts the effectiveness of compulsory licensing as an antitrust remedy.

The second chapter studies the consequences of another important US antitrust case against a monopolist who derived its market power from patents. The Radio Corporation of America (RCA) was one of the leading innovators in the field of radio and television technology since the 1920s. Unlike Xerox, RCA did not use its large patent portfolio to block market entry. Instead, RCA generated substantial revenues through licensing. As RCA licensed its patents exclusively in packages, the company was repeatedly accused of monopolizing the market for radio and television technology. In 1958, RCA signed a consent decree with the US Department of Justice that settled an antitrust lawsuit and obliged RCA to make the majority of its patents available to domestic firms free of charge. As this eliminated the possibility for RCA to earn royalties in the US, RCA increasingly licensed its patents to firms in Japan.

Leveraging this antitrust case as a natural experiment, the second chapter investigates the impact of voluntary technology transfer from the US on Japanese innovation. Japan experienced an unprecedented rise in the field of consumer electronics in the 1960s and 1970s. Narrative evidence suggests that the transfer of US technology may have contributed to the rapid growth of the Japanese electronics industry (e.g., Shih and Dieterich, 2014).

The 1958 antitrust settlement induced a voluntary technology transfer from the US to Japan predominantly in those fields where RCA was an active innovator prior to 1958. In my empirical approach, I exploit this differential likelihood of licensing by RCA by comparing Japanese patenting across technology classes over time.

I find that there was a disproportionate increase in Japanese patenting in RCA's main fields. This effect is driven by Japanese inventions that built on RCA's technology. There was no corresponding increase in patenting by other non-US countries, indicating that the estimates do not reflect general patenting trends across technologies. A complementary approach on the firm level corroborates these results. Using data on individual licensing agreements, I show that receiving a license from RCA was associated with a substantial surge in follow-on innovation by the licensees.

The chapter extends prior research on technology transfer (e.g., Giorcelli, 2019; Giorcelli and Li, 2021) by providing empirical evidence on patent licensing as a channel for technology diffusion. The findings establish technology transfer from the US as one important factor that contributed to the rise of the Japanese consumer electronics industry from the early 1960s onwards. More broadly, the chapter shows that licensing state-of-the-art technology can promote innovation and technological progress in an emerging market.

The chapter also contributes to the literature on the effect of (compulsory) licensing on follow-on innovation (e.g., Moser and Voena, 2012; Watzinger et al., 2020; Nagler et al., 2022; Mamrak, 2023). In particular, it sheds light on how *voluntary* licensing agreements at market-level royalty rates affect subsequent innovation. The key difference to the setting in chapter 1 is that compulsory licensing of RCA's patents applied only to domestic firms. Therefore, RCA could continue to exert market power over foreign firms. This allowed RCA to earn substantial royalties from voluntary patent licensing in Japan throughout the 1960s.

The third chapter, which is based on joint work with Felix Montag, Alina Sagimuldina, and Monika Schnitzer, turns to a different source of market power. We study how market power resulting from imperfect consumer information about prices affects commodity tax pass-through in oligopolistic markets. Understanding how sellers pass through taxes is fundamental for the design of optimal tax policies. Most literature on tax pass-through assumes that consumers possess complete price information (e.g., Weyl and Fabinger, 2013), but this assumption is often unrealistic in practice. Imperfect information is a common feature in many markets. It affects consumers' sensitivity to price differences across sellers, hence softening competition.

In our empirical application, we study retail fuel products, which have a high degree of price transparency and homogeneity compared to other products. Nevertheless, they exhibit significant price dispersion that is random and unpredictable to consumers, consistent with imperfect consumer information. We build on the theoretical consumer search model by Stahl (1989) featuring some consumers that know all prices and others that have to search for prices sequentially. We adapt this model to derive how tax pass-through depends on the share of well-informed consumers and the number of sellers.

The theoretical model yields three testable predictions. First, the pass-through rate increases with the share of well-informed consumers. Second, pass-through to the price paid by well-informed consumers is higher than pass-through to the price paid by uninformed consumers in the same market. Third, when market power is derived from imperfect information, the relationship between the number of competitors and pass-through to the average price is non-monotonic.

We test these predictions empirically using multiple tax changes in the German retail fuel market since 2020. To estimate pass-through, we employ a difference-in-differences strategy with French fuel stations as a control group, using the universe of station-level prices in Germany and France. A key feature of the setting is that the fuel type that consumers purchase is highly correlated with their incentive to become informed about prices. Therefore, we can test how pass-through varies between fuel types with different shares of well-informed consumers. Similarly, we can test how pass-through varies across stations with different numbers of competitors. Our estimates support all three model predictions.

The chapter contributes to both the theoretical (e.g., Weyl and Fabinger, 2013) and the empirical literature (e.g., Miravete et al., 2018; Genakos and Pagliero, 2022; Hollenbeck and Uetake, 2021) on tax pass-through by accounting for imperfect price information. Only a few existing studies depart from the full information assumption, but most of them focus on the salience of taxes (e.g., Chetty et al., 2009; Kroft et al., 2024). We also discuss alternative models but conclude that none of them can explain the empirical findings jointly as well as a consumer search model with imperfect information.

Our findings are widely applicable beyond the retail fuel market and have important implications for tax policy. They highlight that the extent to which taxes are passed on to consumers depends on the degree of consumer information. Imperfect information reduces the effectiveness of tax policy to change consumers' behaviour. Moreover, taxation affects different consumer groups differently. This can have distributional implications and limit the possibility to stimulate the economy through (temporary) tax reductions. Overall, these implications may change the attractiveness of taxes relative to other policy instruments such as regulation. Finally, our results show that the number of sellers is not necessarily a good predictor for the intensity of competition, unlike in models with full information. Pass-through can be small even when there are many sellers in a market. Therefore, understanding tax pass-through requires a stronger emphasis on consumer information, rather than the number of competitors.

Overall, this dissertation offers new insights into the sources and consequences of market power. The individual chapters illustrate how market power may affect innovation, technology transfer, and commodity tax pass-through, respectively. Thus, the dissertation contributes to the extensive literature in industrial organisation that aims to better understand the origins of market power and its implications for policy. The findings are intended to inform the design of economic policies that account for the increasing importance of market power in the economy.

A key insight from this dissertation is that it is not only relevant for policymaking *whether* firms have market power, but the specific source of a firm's market power also matters. For example, the results from chapter 1 indicate that compulsory licensing can be an effective antitrust remedy to promote innovation in the target industry if patents are the main source of the incumbent's market power. Chapter 3 shows that market power resulting from imperfect consumer information about prices may limit the effectiveness of tax policy. These findings have different policy implications from prior literature, where firms derive market power from other sources. Therefore, one important takeaway from this dissertation is that the source of market power matters for the impact of policies in fields as diverse as antitrust and taxation.

Chapter 1

Antitrust and (Foreign) Innovation: Evidence from the Xerox Case

1.1 Introduction

Competition authorities in many countries have tightened their antitrust policy in recent years. In the US, this has raised concerns that stricter antitrust enforcement against domestic incumbents could undermine the American dominance of the high-technology sector, as it may particularly help foreign competitors to catch up.[1] This is an important policy concern, since promoting innovation is increasingly becoming a key objective of antitrust policy, especially in high-technology industries (Gilbert, 2022).

This chapter investigates how antitrust enforcement against patent-based monopolies affects innovation by domestic and foreign firms. Patents – and intellectual property (IP) more broadly – are an important source of market power. On the one hand, this is the intended effect of patents. They incentivize innovation by granting patentees the right to exclude others from using the patented invention. On the other hand, this market power can be abused if patentees engage in exclusionary practices. For example, dominant firms may strategically use their patents to block entry by refusing to license their technology to potential competitors.[2] This can give rise to a conflict between patent and antitrust laws, which may warrant intervention by antitrust au-

This chapter is based on single-authored work (Mamrak, 2023).

[1]See, for example, 'Antitrust Can Hurt U.S. Competitiveness' in *The Wall Street Journal* (`https://www.wsj.com/articles/antitrust-can-hurt-u-s-competitiveness-11625520340`, last accessed: 11 February 2024) or 'Big Tech: Breaking Us Up Will Only Help China' in *Wired* (`https://www.wired.com/story/big-tech-breaking-will-only-help-china`, last accessed: 11 February 2024).

[2]For example, *The Economist* notes that 'patents should spur bursts of innovation; instead, they are used to lock in incumbents' advantages' (see `https://www.economist.com/leaders/2015/08/08/time-to-fix-patents`, last accessed: 11 February 2024).

thorities (e.g., Carrier, 2002). However, there is still little evidence about the impact of antitrust enforcement on innovation when patents are the main barrier to entry.

To fill this gap, I study the antitrust case against Xerox Corporation in the early 1970s. Xerox was the monopolist in the market for plain-paper copiers in the US throughout the 1960s. Xerox had developed and commercialised a novel copier technology that is still widely used today. It had protected its technology through more than 2,000 patents and strictly refused to grant licenses to potential competitors. Xerox also used patent infringement suits to block entry by competitors who developed their own patented technologies. In 1972, the Federal Trade Commission (FTC) charged Xerox with monopolization of the copier market through strategic abuse of the patent system. The case was settled by a consent decree in 1975 and Xerox was ordered to license all its domestic and foreign copier-related patents to any third parties at reasonable rates (FTC, 1975).

The case against Xerox is particularly well suited for addressing my research question. It ‘defined what may have been a peak in antitrust prosecution directed toward patent-based monopolies’ (Scherer, 2005, p. 300). Therefore, it was one of the most important American antitrust cases in the 20th century. The FTC’s intervention was widely perceived as success and triggered a transition to competition in the market for plain-paper copiers (Bresnahan, 1985a; Tom, 2001). As many of the entrants were foreign firms, the case allows me to study the impact of antitrust enforcement on both domestic and foreign innovation.

To empirically estimate the effect of the antitrust case on innovation, I use data on patent applications and employ a difference-in-differences strategy across technology classes (Moser and Voena, 2012; Moser et al., 2014). My main approach uses a continuous treatment variable that exploits variation in the share of patents in a six-digit technology class (based on the Cooperative Patent Classification) that were subject to compulsory licensing. Specifically, I compare the annual number of US patent applications by applicants other than Xerox across differentially affected six-digit classes within the same four-digit class, controlling for a range of fixed effects.

I find that antitrust enforcement against Xerox had an overall positive effect on subsequent innovation by other firms in the copier industry. There was a disproportionate increase in patenting in technologies with a greater exposure to compulsory licensing of Xerox’s patents. My estimates indicate that the antitrust case led to an additional 160 patent applications per year. Event-study analyses illustrate that these estimates are not driven by any differences in pre-trends across groups. The result is also robust to a wide range of alternative model specifications and treatment definitions.

Moreover, I show that the number of forward citations received by compulsorily licensed Xerox patents increased disproportionately after 1975. This complementary analysis follows Watzinger et al. (2020) and employs a matching strategy to find a control patent for every compulsorily licensed Xerox patent. Overall, the findings support the conclusion that the antitrust case against Xerox spurred technological progress, as other firms used the newly available technology for follow-on innovation.

Interestingly, the main beneficiaries of increased access to Xerox's technology were competitors from Japan. In my main approach, when splitting the number of patent applications by applicant country, the positive effect of compulsory licensing is almost entirely driven by Japanese firms. In contrast, the estimated effect on patenting by US applicants is quantitatively small and statistically indistinguishable from zero. I further show that there was great heterogeneity in the effect of the antitrust case even among Japanese applicants. Only established firms increased their patenting, whereas small firms and start-ups did not benefit from access to Xerox's technology. Moreover, the positive innovation effect is driven by Japanese firms with extensive prior patenting experience in copier technologies – that is, (potential) competitors to Xerox in the copier market.

The finding that Japanese rivals increased their innovation following the antitrust case is in line with historical narratives about the development of the copier industry. Scherer (2005) notes that several Japanese copier producers (e.g., Canon, Konica, Ricoh) successfully entered the American market after 1975 and became important competitors to Xerox. Japanese entrants strategically focused on the lower end of the copier market. That is, they produced machines that were cheaper, smaller, and designed for lower volumes than existing plain-paper copiers. This business model was different from that of most American copier producers and is considered one of the key reasons for the Japanese success (e.g., Jacobson and Hillkirk, 1986).

Consistent with this narrative, I show that patents filed by Japanese competitors were more likely to contain words in their title or abstract that can be associated with smaller copiers. Moreover, I find that innovation became more novel and diverse following the antitrust case. These changes in the direction of innovation are again driven by Japanese patent applicants. Their innovation activity expanded to new technology fields, while there was no reduction in the quality of inventions. Therefore, the results are in line with the historical evidence suggesting that Japanese entrants focused on smaller desktop copiers, which were more differentiated from existing products.

Finally, I also investigate how Xerox's own patenting activities reacted to the removal of most of its IP. Relative to a synthetic control group (Abadie et al., 2010, 2015), Xerox and its subsidiaries reduced their patenting by no more than 30 patents per year after 1975. This effect is much smaller than the increase in patenting by other firms, which I find in my main approach. Therefore, my estimates indicate that the overall impact of the antitrust case on subsequent innovation was positive.

The first important takeaway from this chapter is that compulsory licensing can promote innovation by other firms *within* the target industry if patents are the main entry barrier. This finding is different from prior results by Watzinger et al. (2020), who study compulsory licensing in the antitrust case against Bell in the 1950s. They find no effect of compulsory licensing in the target industry. This is because Bell was a vertically integrated monopolist that could continue to foreclose its rivals even after the loss of most of its IP. In contrast, Xerox's monopoly was primarily based on the strategic use of its patent portfolio to block entry. Consequently, compulsory licensing removed the main barrier to entry in the case of Xerox, while it did not in the case of Bell. As my results show, such differences in the market structure of the target industry can be crucial. I find the largest increase in patenting among firms whose prior experience overlaps with Xerox's technology. Therefore, this chapter goes beyond prior literature by highlighting how the source of market power impacts the effectiveness of compulsory licensing as an antitrust remedy.

The second key takeaway from this chapter is that the antitrust case against Xerox particularly benefited Japanese competitors. This result speaks to current debates about antitrust policy. For example, a comment in *The Wall Street Journal* warned that '[a]ntitrust action against leading U.S. tech companies would shrink American dominance of the world's fastest-growing industry'.[3] Based on the historical evidence from the Xerox case, concerns that antitrust could benefit foreign competitors may be worth investigating. However, drawing appropriate policy conclusions from this finding requires a nuanced view. Although US antitrust intervention against Xerox primarily helped foreign competitors to build on the incumbent's technology, it is unlikely that this result is due to some of the competitors being Japanese. Instead, Japanese entrants employed a different business model that allowed them to expand the market for plain-paper copiers to the lower-volume segment, which had been ignored by Xerox. This suggests that society lacked some innovations during Xerox's patent-protected monopoly. In that way, the antitrust case against Xerox also brought enormous benefits to American consumers. They experienced lower prices and more innovative copiers, irrespective of whether the innovators were Japanese.

This chapter contributes to the literature on the effect of antitrust on innovation by estimating the differential impact of antitrust intervention on innovation by domestic

[3] See 'The Misguided Antitrust Attack on Big Tech' in *The Wall Street Journal* (`https://www.wsj.com/articles/the-misguided-antitrust-attack-on-big-tech-11600125182`, last accessed: 11 February 2024).

and foreign firms. To the best of my knowledge, this aspect has not been studied in the economic literature so far, despite its relevance for current policy debates. While most of the literature on antitrust and innovation is theoretical (Segal and Whinston, 2007; Cabral, 2018; Federico et al., 2020), there has been an increasing number of empirical contributions in recent years (Watzinger et al., 2020; Cunningham et al., 2021; Kang, 2021; Poege, 2022; Watzinger and Schnitzer, 2022). I further complement these studies by providing empirical evidence on the impact of one of the most important US antitrust cases in the 20th century. This is relevant for policymaking due to the limited number of high-profile cases in the last century.

The chapter also adds to prior studies on compulsory licensing and the protection of IP rights (Chien, 2003; Acemoglu and Akcigit, 2012; Moser and Voena, 2012; Galasso and Schankerman, 2015; Nagler et al., 2022). For instance, Moser and Voena (2012) analyse compulsory licensing in the chemical industry following the 'Trading with the Enemy Act' after World War I. Compulsory licensing is also a frequently used remedy in competition cases (Delrahim, 2004). I contribute to the literature by studying the effectiveness of compulsory licensing in the context of an antitrust case where the targeted monopolist used its patents as an entry barrier. My estimates show that Xerox's patents exerted a blocking effect on follow-on innovation by other firms during the monopoly period. This effect is consistent with a rent dissipation theory (Arora and Fosfuri, 2003; Gaessler et al., 2019). Xerox likely refused to grant licenses to its competitors, because it feared that revenues from licensing would be lower than the loss in profits due to increased product market competition.

Finally, the chapter extends previous research on the case against Xerox (Bresnahan, 1985a,b; Tom, 2001). Most prominently, Bresnahan (1985a) describes the transition to competition in the copier market and discusses potential innovation effects. However, to the best of my knowledge, I am the first to provide empirical evidence on the impact of the antitrust case on subsequent innovation.

The remainder of the chapter is structured as follows. Section 1.2 explains the historical background on Xerox and the antitrust case. In section 1.3, I introduce the data and my empirical strategy. The main results are presented in section 1.4. In sections 1.5 and 1.6, I investigate which firms benefited from the antitrust case and study the underlying mechanism. Finally, section 1.7 analyses the effect on Xerox and section 1.8 concludes.

1.2 Historical Background

Xerox Corporation was the de facto monopolist in the market for plain-paper copiers in the US until the early 1970s. At the start of the antitrust case in 1972, its share in the rapidly growing market was close to 95%. Xerox was the 17th most profitable American company based on return on stockholders' equity (FTC, 1975). This section provides a brief historical overview of Xerox's rise and the antitrust case.

1.2.1 Xerox's Monopoly in the Plain-Paper Copier Market

The foundations for Xerox's success were laid in 1938, when the American physicist Chester Carlson invented the process of electrophotography. This technology, which was later called 'xerography' (Greek for 'dry writing'), allowed to print images using an electrostatic process and a dry powder (i.e., toner).

Although xerography forms the basis of the technology used in copiers and laser printers still today, it took two decades to transform Carlson's invention into a marketable product. In 1946, the Haloid Photographic Company, a small manufacturer of photographic equipment from Rochester, NY, agreed to commercialise xerography.[4] Haloid introduced its first xerographic office copier in 1949. Despite the machine's limited initial success, the company continued investing in its xerographic technology and was later renamed Xerox Corporation.

Xerox achieved its major breakthrough in 1959 when it launched the Xerox 914 office copier. The fully automated machine could produce a xerographic copy within seconds, was easy to operate, and could be used with ordinary (plain) paper. This made Xerox's technology superior to that of competitors, as the 914 did not require using special (coated) paper. The Xerox 914 became an immediate commercial success and Xerox's annual sales increased by a factor of 25 from 1959 to 1968, making Xerox the fastest company to reach $1 billion in sales (Jacobson and Hillkirk, 1986; Gomes-Casseres and McQuade, 1991).

The success of the 914 rewarded Xerox for its large and risky investment into commercialising xerography. Despite widespread initial scepticism regarding the technology's potential, Xerox spent more than its total earnings throughout the 1950s on the development of the 914 (Jacobson and Hillkirk, 1986). The copier market also grew

[4] As Carlson lacked financial resources to develop xerography himself, he approached more than a dozen major US firms to commercialise his invention, but none of them was interested. In 1944, the Battelle Memorial Institute, a non-profit research organisation, agreed to invest in the technology's development and continued searching for a corporate partner. Battelle then entered into an agreement with Haloid. More details on the development of xerography and the origins of Xerox can be found in Kearns and Nadler (1992) and Owen (2005).

rapidly. The annual number of copies made in the US increased from approximately 20 million in the mid-1950s to around 10 billion a decade later (Jacobson and Hillkirk, 1986). In the 1960s, Xerox introduced several new copiers that could operate at higher speed or contained additional features.

Xerox had protected its copier technology through more than 2,000 patents and strictly refused to grant licenses to any other manufacturers of plain-paper copiers. Therefore, throughout the 1960s, no other firms could sell plain-paper copiers. Xerox also sold its plain-paper copiers and protected its IP abroad. To this end, the company had established two foreign subsidiaries, Rank Xerox in Europe and Fuji Xerox in Japan, which acted as Xerox's international sales organisations.[5]

In 1970, Xerox's monopoly in the market for plain-paper copiers was challenged for the first time when International Business Machines (IBM) introduced its first xerographic copier. Despite the inferior quality of IBM's copier, Xerox immediately sued for patent infringement, initiating a legal battle that lasted for several years (Jacobson and Hillkirk, 1986). Similarly, in 1972, Litton Industries launched a plain-paper copier and was sued by Xerox.

1.2.2 FTC Complaint and 1975 Consent Decree

In late 1972, the Federal Trade Commission (FTC) filed an antitrust complaint against Xerox that alleged monopolization of the plain-paper copier market.[6] Xerox was accused of violating Section 5(a) of the FTC Act by hindering effective competition in plain-paper copiers through its strategic use of the patent system and anti-competitive pricing policies (FTC, 1975). As outlined by Bresnahan (1985b), the theory of harm regarding Xerox's patent practices was twofold. On the one hand, entry into the market may have been inhibited by the size, complexity, and obscurity of Xerox's patent thicket in combination with the threat of infringement litigation. On the other hand, Xerox was accused of strategically building, maintaining, and using parts of its patent portfolio for the sole purpose of denying access to relevant technologies to its competitors. The second part of allegedly anti-competitive practices related to Xerox's pricing policies. In particular, Xerox pursued a lease-only policy with tied maintenance for its

[5] Rank Xerox was established in 1956 as joint venture between Xerox and the Rank Organization from the UK; Fuji Xerox was established in 1962 as joint venture between Rank Xerox and the Japanese company Fuji Photo Film (Gomes-Casseres and McQuade, 1991).

[6] The FTC defined the relevant market as the sale and lease of office copiers and supplies in the US (FTC, 1975). Plain-paper copiers represented the most important submarket. At the time of the complaint, there were 25 firms active in the American copier market, but only three firms (i.e., Xerox, IBM, and Litton) distributed plain-paper copiers. Amongst them, Xerox accounted for 95% of revenues.

copiers and it used various ways of price discrimination between customers, based on the number of copies or the number of Xerox machines installed.

In 1975, Xerox settled with the FTC by signing a consent decree whose main remedy was compulsory licensing of all of Xerox's copier-technology patents.[7] Compulsory licensing also applied to Xerox's foreign patents, including the ones held by Rank Xerox and Fuji Xerox. Xerox had to grant the first three licenses to each firm royalty-free and could then ask for reasonable royalties that were not to exceed 1.5% of the licensee's revenues. Future Xerox patents issued until 1981 were also covered by the licensing requirement (FTC, 1975). Moreover, Xerox was ordered to cease all patent infringement suits as well as one pricing policy that provided discounts on individual rental rates to very high-volume customers.

The features of the consent decree indicate that the FTC viewed Xerox's use of its patent portfolio as the main barrier to entry. The initial FTC complaint had also proposed to ban a long list of Xerox's pricing policies, out of which only one was prohibited by the final consent decree. In contrast, the FTC made no concessions on its claim that Xerox licenses its copier-related patents. Accordingly, Bresnahan (1985b, p. v) argues that 'the FTC's emphasis on patents over pricing practices was wise. [...] [T]he [pricing] practices were price discrimination devices – i.e., the fruits not the causes of monopoly power.'

The market for plain-paper copiers in the US became competitive in the 1970s. More than a dozen other firms – such as Eastman Kodak, Savin, or Smith Corona (SCM) from the US as well as Canon, Konishiroku (Konica), or Ricoh from Japan – entered the market. From 1972 to 1977, prices for plain-paper copiers declined by more than 30% and Xerox's market share in net new placements fell from close to 100% to less that 20% (Bresnahan, 1985b).

1.3 Data and Empirical Strategy

I use data on patent applications to empirically measure innovation. Patents are well-suited for my empirical approach for several reasons. First, patent data are consistently available throughout the relevant period. Second, as patents are assigned to hierarchical technology classes, I can compare patenting across different technologies within the same field. Third, patent citations allow me to measure follow-on innovation to Xerox's

[7] Xerox had initially denied all allegations but eventually decided to settle. In parallel to the FTC case, Xerox had also been involved in several other antitrust lawsuits brought forward by competitors. Therefore, the company incurred high costs – both in terms of legal expenses and the time of its executives. As later explained by Peter McColough, then Xerox's chief executive officer, the reason for settling with the FTC was not to admit any wrongful acts but rather to find a way forward allowing Xerox to focus on its business again (Jacobson and Hillkirk, 1986).

patents (Jaffe and Trajtenberg, 1996), since firms had to cite any prior art irrespective of the licensing order.[8]

My main data source is the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. In addition, I use data from the 'HistPat' database (Petralia et al., 2016) as well as the 'patentCity' project by Bergeaud and Verluise (2022) to identify the applicants' country of origin, which is not yet consistently reported in PATSTAT throughout the 1970s.[9]

To identify which Xerox patents were subject to compulsory licensing, I use a list published in the Official Gazette of the United States Patent and Trademark Office (USPTO, 1975). This list, which was published in compliance with the consent decree, reports the publication number and title of more than 2,600 patents owned by Xerox and Fuji Xerox as of 1975. One caveat is that the list does not allow to exactly determine which patents were licensable. Instead, according to its description, 'the list [...] is believed to include all of the patents available for licensing', but 'there are several patents included in the list to which the consent order is not applicable' (USPTO, 1975, p. 1665). To approximate the set of licensable Xerox patents, therefore, I consider all patents on the 1975 list as if they were subject to compulsory licensing.

1.3.1 Class-Level Analysis of Cumulative Innovation

For my main empirical approach, I construct a panel dataset from 1970 until 1985 that counts the annual number of patent applications in the US on the level of six-digit technology classes based on the Cooperative Patent Classification (CPC). Patent applications by Xerox and its subsidiaries are excluded from the sample. I exploit the classes' differential exposure to compulsory licensing, depending on the share of patents that were licensable in each class. Since the classification system is hierarchical, I can compare patenting across differentially affected six-digit technology (sub)classes within the same four-digit class.[10]

[8]Despite these advantages, using patent data also has some drawbacks. On the one hand, patents may be an imperfect measure of innovation, because not all innovations are patentable and inventors may opt for secrecy as an alternative means of protection (e.g., Moser, 2012). However, this concern is mitigated by the fact that patent protection played an important role in the copier industry, as indicated by the historical background. On the other hand, patent citations may not accurately measure follow-on innovation, as citations may have been added by the examiner even in the absence of any knowledge flow (Alcácer and Gittelman, 2006).

[9]I thank Cyril Verluise for sharing their data.

[10]This class-level approach follows Moser and Voena (2012) and Moser et al. (2014). To highlight the hierarchical structure of the technology classes, I henceforth refer to a four-digit CPC class (e.g., G03G) as 'class' and to a six-digit CPC class (e.g., G03G 15) as 'subclass', although this slightly deviates from the official terminology. For an overview of the classification system, see `https://worldwide.espacenet.com/classification` (last accessed: 11 February 2024).

I use the following difference-in-differences (DiD) regression model to estimate the effect of compulsory licensing on cumulative innovation:

$$\text{Patents}_{c,s,t} = \beta \cdot \text{Share}_s \cdot \text{Post}_t + \alpha_s + \lambda_{c,t} + \epsilon_{c,s,t}, \tag{1.1}$$

where $\text{Patents}_{c,s,t}$ is the number of patent applications in year t assigned to six-digit subclass s within four-digit class c. Share_s is a continuous treatment variable that measures the exposure of a subclass to the antitrust intervention. It is defined as the share of unexpired patents per subclass (as of 1975) that were subject to compulsory licensing. That is, for each subclass, the treatment variable captures the number of Xerox patents on the USPTO list relative to the overall size of the technology class. The variable Post_t is a dummy that equals one in years after 1975. The regression also includes subclass fixed effects (α_s) as well as year $\times$ class fixed effects ($\lambda_{c,t}$). This controls for time-invariant differences across subclasses and allows the classes to experience idiosyncratic shocks over time. As a consequence, the DiD estimate $\hat{\beta}$ is identified only from variation over time across subclasses *within* the same class. Standard errors are clustered accordingly at the four-digit class level.

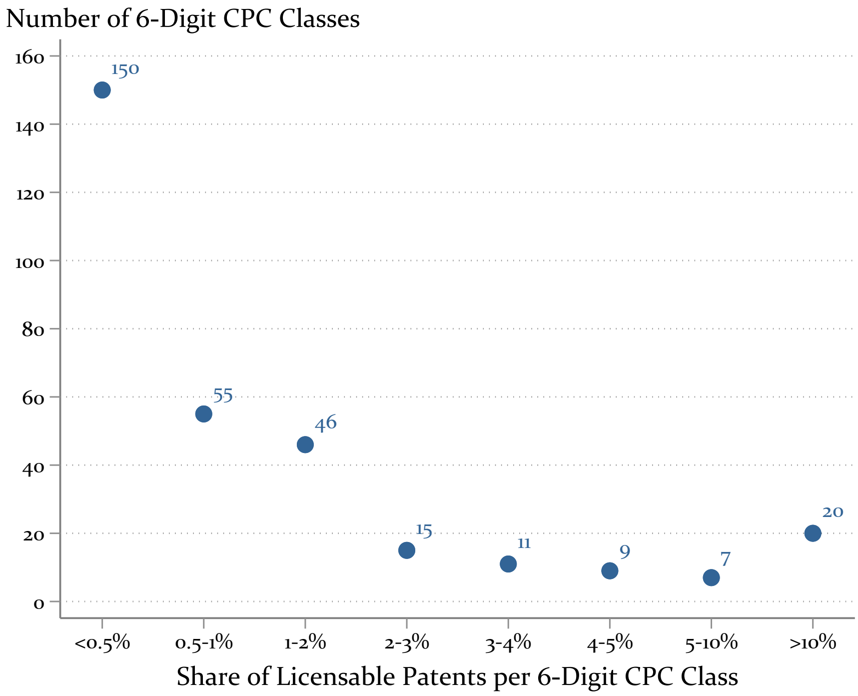
I estimate the DiD regression on a sample of 2,210 six-digit subclasses that belong to 141 four-digit classes. On average, there were 15 patent applications per year per subclass. 313 subclasses contain at least one Xerox patent that was subject to compulsory licensing – such that these subclasses are 'treated' (with $\text{Share}_s > 0$). The remaining 1,897 subclasses are 'untreated', as they had no exposure to compulsory licensing in the absence of spillovers.[11] I use the weights by Iacus et al. (2012) to adjust for the different number of untreated subclasses per treated subclass. Figure 1.1 shows the distribution of the treatment variable across subclasses. Among the 313 treated subclasses, there are 150 subclasses with a share of licensable patents below 0.5%, while there are 20 subclasses where more than 10% of all patents were subject to compulsory licensing. Overall, the classes in the sample contain a total of 2,479 compulsorily licensed Xerox patents that were unexpired as of 1975.[12] Appendix A.1.1 presents additional summary statistics of the sample and for the treatment variable as well as more information on Xerox's patent portfolio.

There are several advantages to my approach of using the share of compulsorily licensed patents per subclass as treatment variable. First, the approach builds on the simple economic intuition that any effect of compulsory licensing on innovation

[11] This sample is obtained after applying two restrictions. First, a six-digit subclass must have at least one patent application in the pre-treatment period. Second, every four-digit class must have at least one treated subclass.

[12] This represents 95% of the patents on the USPTO list, because the list contains some patents that had expired by 1975, which I do not consider for the treatment definition.

Figure 1.1. Class-Level Analysis: Distribution of Treatment Variable (Share$_s$)



Notes: The figure depicts the distribution of the treatment variable (Share$_s$, i.e., the share of patents per six-digit subclass that were subject to compulsory licensing) across treated subclasses. It shows the number of subclasses in the sample (on the vertical axis) that have a given share of compulsorily licensed patents (on the horizontal axis). Example: there are 150 subclasses with a share of licensable patents below 0.5%.

should be more pronounced in technologies with a greater exposure to the antitrust measure. In contrast, using a binary treatment indicator would handle all treated subclasses equally. Second, using shares instead of the absolute number of licensable patents also takes into account the size of each subclass.[13] An additional nice feature of the approach is that the treatment variable can also be interpreted as capturing the propensity of a given subclass for containing xerography-related patents.

My empirical strategy identifies the causal effect of the antitrust measure under the assumption that, within the same four-digit technology class, patenting in subclasses with little or no exposure to compulsory licensing provides a valid counterfactual for patenting in subclasses with a greater share of licensable patents. The counterfactual represents a situation in which other firms could not license Xerox's patents and Xerox would have continued to block entry through patent infringement suits. Therefore, in other words, the number of patent applications in subclasses with different values

[13]Nevertheless, I show in appendix A.1.4 that my results are robust to using a binary treatment variable or alternative continuous measures, where the treatment variable captures the number (as opposed to the share) of licensable Xerox patents per subclass.

of the treatment variable must have followed a common trend in the absence of the antitrust intervention.

To assess the common trend assumption, I also estimate the following event-study variation of equation (1.1):

$$\begin{aligned} \text{Patents}_{c,s,t} = &\sum_{\tau=1970}^{1974} \delta_\tau \cdot \text{Share}_s \cdot \mathbb{1}[\text{Year}_t = \tau] \\ &+ \sum_{\tau=1976}^{1985} \beta_\tau \cdot \text{Share}_s \cdot \mathbb{1}[\text{Year}_t = \tau] + \alpha_s + \lambda_{c,t} + \epsilon_{c,s,t}, \end{aligned} \tag{1.2}$$

where the coefficients of interest are the lags (β_τ) that estimate the post-1975 treatment effects. In contrast, the leads (δ_τ) represent anticipatory effects and should not be statistically different from zero.

There are two potential concerns with my identification strategy. First, subclasses that were more exposed to compulsory licensing may have been different from less exposed subclasses in terms of unobserved characteristics, which may cause different patenting trends over time. For instance, Xerox may have chosen to patent in technology classes that had a higher likelihood of future innovation activity. This would violate the common trend assumption. Second, firms may have strategically substituted their patenting from untreated to treated technology classes in response to the antitrust case. This would violate the stable unit treatment value assumption (SUTVA).

I believe that both of these potential concerns are unlikely to substantially bias my estimates. First, the event-study analysis allows to assess whether patenting in treated technologies already developed differently prior to the antitrust case. Second, a violation of SUTVA could only be caused by strategic changes to firms' actual innovation investment but not their patent filing behaviour, because technology classes are assigned to applications by the patent examiner. This raises the question whether firms had any incentive to substitute their innovation activity towards technologies more exposed to compulsory licensing. Yet, as the market for copiers became more contestable following the antitrust intervention, if anything, expected profits should be lower (and not higher) in treated technology fields.

1.3.2 Patent-Level Analysis of Follow-on Innovation

In a complementary approach, I use a different empirical set-up on the patent level to study direct follow-on innovation building on Xerox's technology. In particular, I investigate whether the number of forward citations to compulsorily licensed Xerox patents increased after 1975. Following Watzinger et al. (2020), I use exact matching to construct a control group for every licensable Xerox patent. I match on grant year, four-digit CPC class, and the aggregate number of citations until 1972 (i.e., the start of the antitrust case) or for at least two years.[14] Conditioning on the first two variables controls for differences in citations patterns over the patent term and across technologies. Conditioning on the number of previous citations additionally controls for how much a patent is used by other firms, although it does not necessarily imply that matched patents are of the same underlying quality.

My final sample consists of 1,311 compulsorily licensed Xerox patents matched to 25,899 control patents in 445 distinct strata. The number of matched Xerox patents is lower than the total number of compulsorily licensed patents for several reasons. Most importantly, I only match compulsorily licensed Xerox patents that were granted until 1972 (instead of 1975). This restriction circumvents concerns that Xerox may have strategically changed its patenting behaviour after publication of the FTC complaint. Similarly, matching only on citations until 1972 ensures that the matching procedure is unlikely to be confounded by endogenous changes in citation patterns. Moreover, I drop self-citations and only consider patents that received at least one citation during their term of validity. Among the 1,346 resulting Xerox patents that had not expired by 1975, 97% can be matched to at least one control patent. For the sample of matched patents, I then construct a panel that counts the number of forward citations received by every patent in every year from 1970 until 1985 or until patent expiry.

I estimate the following DiD regression model to estimate the effect of compulsory licensing on subsequent citations:

$$\text{Citations}_{i,t} = \beta \cdot \text{Xerox}_i \cdot \text{Post}_t + \alpha_i + \lambda_t + \epsilon_{i,t}, \tag{1.3}$$

where $\text{Citations}_{i,t}$ is the number of forward citations received by patent i in year t. Xerox_i is a dummy variable that equals one for compulsorily licensed Xerox patents and zero for matched control patents. The variable Post_t again is a dummy that equals one in years after 1975. The regression also includes patent fixed effects (α_i) and year fixed effects (λ_t). This controls for time-invariant differences in the number of citations across patents and allows for year-specific shocks that affect all patents equally. I again use the weights by Iacus et al. (2012) to adjust for the different number of matched control patents per treated Xerox patent. Standard errors are clustered at the four-

[14]That is, for Xerox patents granted in 1971 and 1972, I match on the aggregate number of citations until 1973 and 1974, respectively. This is necessary because matching on citations becomes meaningless when using less than two years of data. However, my results are robust to only using Xerox patents granted until 1970, which I match on the aggregate number of citations until 1972.

digit technology class level to account for potential serial correlation in citations within such a class.

The identifying assumption for this patent-level analysis is that Xerox patents would have received the same number of citations as their matched control patents in the absence of compulsory licensing. As pointed out by Watzinger et al. (2020), one concern regarding this common trend assumption is that the authorities may have chosen to license patents with a high potential for follow-on innovation. However, official FTC documents on the Xerox case do not support this concern. The FTC did not attempt to promote innovation; instead, its main objective was to stop Xerox from using its patents to block entry into the product market for plain-paper copiers (FTC, 1975). To further address the concern, I also estimate an event-study variation of equation (1.3) with leads and lags.

1.4 Effect of the Antitrust Case on Innovation

This section introduces the main results of my empirical analysis. I first present estimates from the main approach on the class level, followed by estimates from the complementary approach on the patent level. The following section then investigates which firms benefited from access to Xerox's technology.

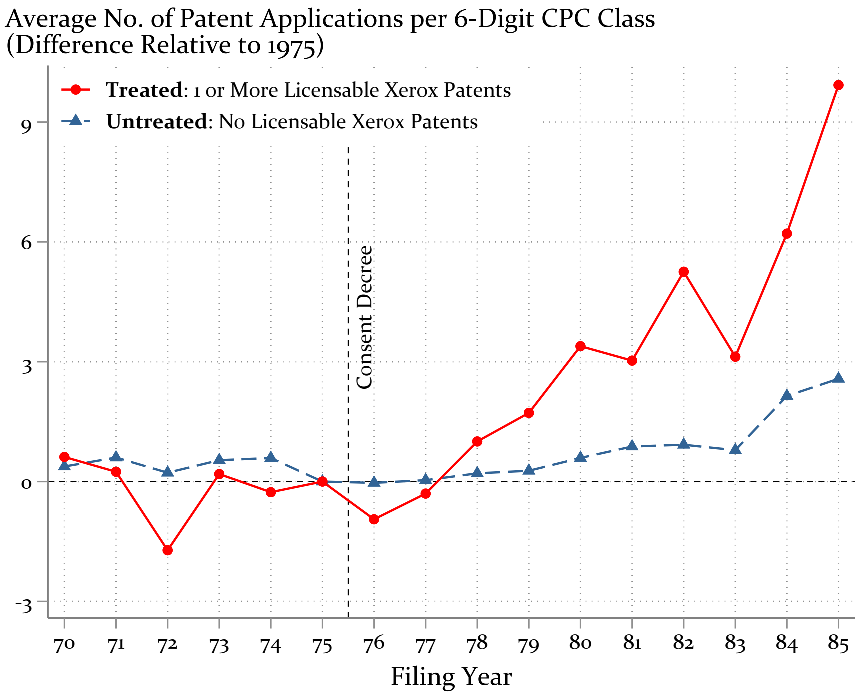
1.4.1 Cumulative Innovation

Panel (A) of Figure 1.2 depicts the average number of patent applications per six-digit technology subclass separately for (treated) subclasses with at least one compulsorily licensed Xerox patent and the remaining (untreated) subclasses in the sample. Although this figure solely presents averages, it shows that treated subclasses experienced a relative increase in the number of patent applications after 1975. In contrast, in the years preceding the consent decree, patenting in both groups followed a relatively common trend – consistent with the identifying assumption underlying my empirical strategy.

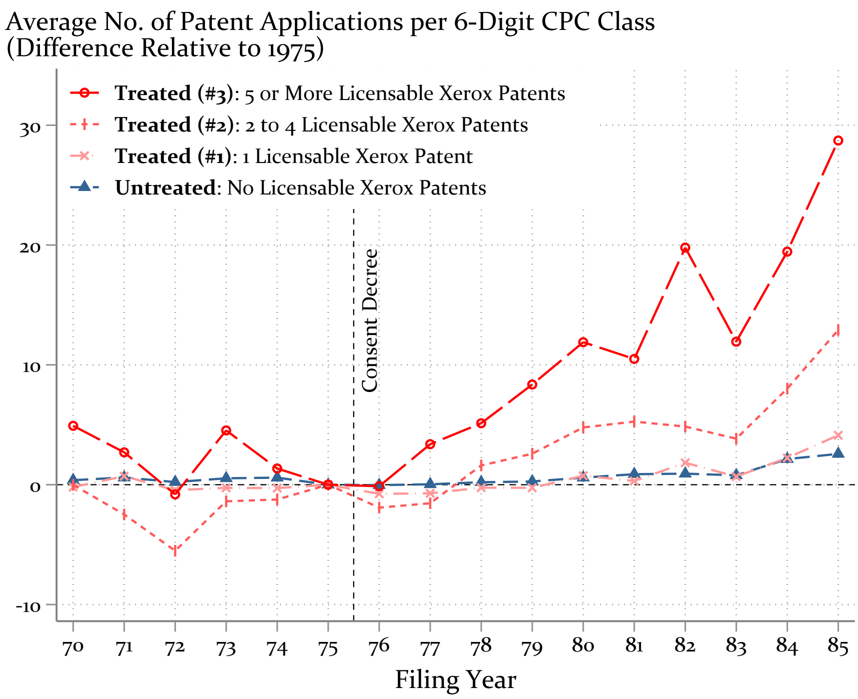
Next, I assess which treated subclasses were responsible for the relative increase in patenting following the antitrust intervention. To this end, in panel (B) of Figure 1.2, I split the treated subclasses into three subgroups, respectively containing one, two to four, and five or more licensable Xerox patents. As is evident from the figure, the relative increase in patenting is most pronounced in subclasses in which five or more Xerox patents were compulsorily licensed. It is smaller in subclasses with two to four licensable Xerox patents, whereas there is virtually no change in patenting around 1975 in subclasses with only one licensable Xerox patent. Finally, it is reassuring that pre-

Figure 1.2. Class-Level Analysis: Patenting Trends Across Technology Classes

(A) Treated vs. Untreated Subclasses



(B) By Number of Licensable Patents



Notes: The figure depicts the average number of patent applications by firms other than Xerox per six-digit subclass relative to 1975. In panel (A), averages are computed separately for treated and untreated subclasses, where a subclass is defined as treated if it contained at least one compulsorily licensed Xerox patent. In panel (B), treated subclasses are further divided into three subgroups, containing (#1) one, (#2) two to four, and (#3) five or more licensed Xerox patents, respectively. In both panels, the subclasses are aggregated using the weights by Iacus et al. (2012). Note that the scale differs between panels (A) and (B).

Table 1.1. Class-Level Analysis: Baseline Estimates

	(1)	(2)	(3)
$\text{Share}_s \cdot \text{Post}_t$	0.210***	0.189**	
	(0.045)	(0.094)	
$\text{Share}_s \cdot \text{Post}_t \cdot \mathbb{1}[\text{Lic}_s = 1]$			-0.035
			(0.080)
$\text{Share}_s \cdot \text{Post}_t \cdot \mathbb{1}[2 \leq \text{Lic}_s \leq 4]$			0.085
			(0.097)
$\text{Share}_s \cdot \text{Post}_t \cdot \mathbb{1}[\text{Lic}_s \geq 5]$			0.377***
			(0.141)
Additional Patents per Year	180	163	160
Relative Increase	1.6%	1.5%	1.4%
Subclass FE	✓	✓	✓
Year FE	✓		
Year × Class FE		✓	✓
Mean of Outcome	15.13	15.13	15.13
No. of 6-Digit CPC Classes	2,210	2,210	2,210
No. of 4-Digit CPC Classes	141	141	141
Observations	35,360	35,360	35,360

Notes: The table shows the results from difference-in-differences regressions following variations of equation (1.1). The outcome variable in all regressions is the number of patent applications by firms other than Xerox in a given six-digit CPC subclass and year. In column (3), the treatment variable is interacted with indicators for subclasses with one, two to four, and five or more compulsorily licensed Xerox patents, as indicated by the variable Lic_s. Standard errors clustered at the four-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

trends remain mostly parallel even across the three treated subgroups with differential exposure to the antitrust measure.

I now turn to a regression framework to investigate the impact of compulsory licensing on innovation more systematically. As outlined above, my main approach compares patenting across six-digit subclasses with differential exposure to the antitrust measure within the same four-digit class. I use a continuous treatment specification, inspired by the pattern shown in panel (B) of Figure 1.2.

Table 1.1 presents the results from estimating different variations of the DiD regression model in equation (1.1). Column (1) shows the estimate when only controlling for subclass and year fixed effects. The point estimate is positive and highly statistically significant. My baseline specification is given by column (2). Adding year × class fixed effects slightly reduces the magnitude of the DiD estimate. The point estimate in column (2) indicates that, on average, a one percentage-point higher share of compulsorily licensed Xerox patents in a subclass is associated with 0.19 additional patent applications per year in that subclass after 1975.[15] This baseline estimate is statistically significant at the 5% level. In column (3), I again split the treated subclasses into three subgroups. Consistent with the pattern presented in panel (B) of Figure 1.2, the estimates show that the increase in patent applications is driven by subclasses in which five or more Xerox patents were compulsorily licensed. This result is reassuring, as it implies that the share of compulsorily licensed patents in a given subclass only affected subsequent patenting if the absolute number of licensable patents in that subclass was sufficiently large.

The magnitude of my baseline estimate in column (2) of Table 1.1 corresponds to around 163 additional patents per year following the antitrust case.[16] This represents an economically meaningful effect, given that the average monetary value of a single patent in treated technologies was more than \$20 million in today's dollars (Kogan et al., 2017). Relative to the overall number of patent applications in treated subclasses, my estimates imply that patenting increased by around 1.5%. This increase may seem relatively small at first glance. Yet, it should be taken into account that patents in treated subclasses, which represent the denominator of the percentage estimate, also cover inventions for a variety of products other than copiers. The percentage increase would likely be much larger if it were possible to isolate copier-specific patents in the denominator.

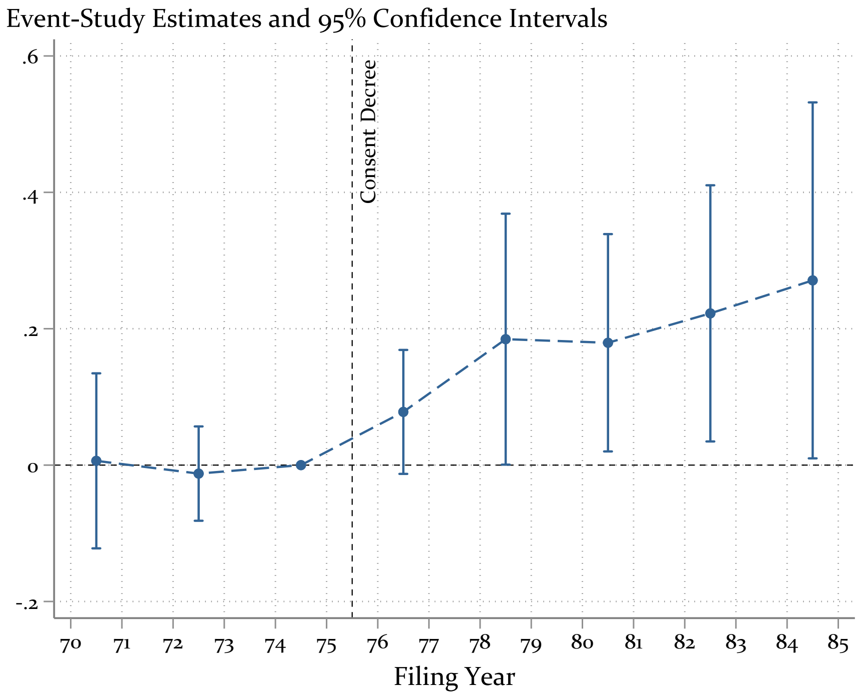
Figure 1.3 graphically depicts the point estimates and 95% confidence intervals from the event-study analysis in equation (1.2), corresponding to the simple DiD estimate in column (2) of Table 1.1. The figure shows that there was a disproportionate increase in patenting in treated technologies following the 1975 consent decree. That is, on average, subclasses with a greater exposure to compulsory licensing experienced a greater relative increase in the number of patent applications. In contrast, before 1975, the number of patent applications across differentially treated subclasses followed a relatively common trend. As can be seen in Figure A.1 in the appendix, pre-trends also remain parallel when extending the panel back to 1960. This supports the identifying assumption underlying my empirical strategy.

In the next step, I restrict the outcome variable to patent applications that built on compulsorily licensed Xerox patents through citations. This serves to investigate whether the estimated increase in patenting is, in fact, related to compulsory licensing of Xerox's patents. The corresponding DiD estimates are shown graphically in Fig-

[15] I define the variable Share_s in percentage terms (i.e., $\times$ 100). Therefore, the estimate $\hat{\beta}$ in equation (1.1) can be interpreted as the average annual post-1975 increase in patenting per subclass that corresponds to a one percentage-point increase in the share of licensable patents.

[16] This estimate is obtained by multiplying the point estimate with the share of licensable patents (i.e., the treatment variable) in every subclass and then aggregating.

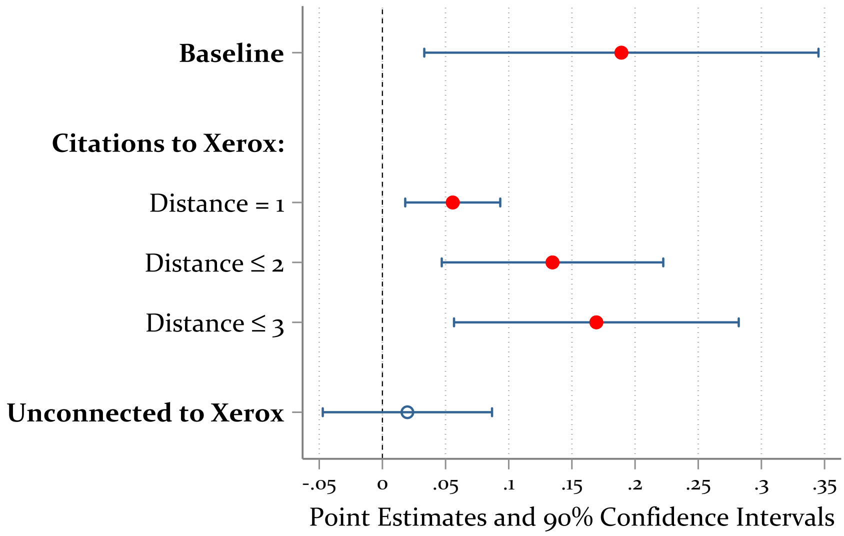
Figure 1.3. Class-Level Analysis: Event-Study Estimates



Notes: The figure depicts point estimates and 95% confidence intervals from the event-study analysis in equation (1.2). Patent applications are binned in two-year groups to reduce noise in the estimates. The regression uses the weights by Iacus et al. (2012). Standard errors are clustered at the four-digit CPC technology class level.

ure 1.4. The first row repeats the baseline estimates from column (2) of Table 1.1. In the second row, the outcome variable is restricted to patent applications that directly cite a compulsorily licensed Xerox patent. Following the framework by Ahmadpoor and Jones (2017), these patents are defined to have a distance $= 1$ to Xerox. Although the point estimate remains statistically significant, its magnitude is only around 30% of that of the baseline. This highlights that only a fraction of the additional patents filed in treated subclasses after 1975 directly built on Xerox's technology. Therefore, in the following two rows, I include patent applications that are related to compulsorily licensed Xerox through higher-degree citations. Patents with distance $= 2$ ($= 3$) do not directly cite Xerox but cite a patent with distance $= 1$ ($= 2$). The estimated coefficients increase when higher-degree citations are included. When using patents with distance ≤ 3 as outcome variable, the magnitude of the point estimates is close to that of the baseline. This result suggests that only looking at direct citations to Xerox's patents may not capture the entire impact of the antitrust measure, highlighting the cumulative nature of innovation. Finally, the last row of Figure 1.4 shows the point estimate when restricting the outcome variable to patent applications that are

Figure 1.4. Class-Level Analysis: (Indirect) Citations to Licensable Xerox Patents



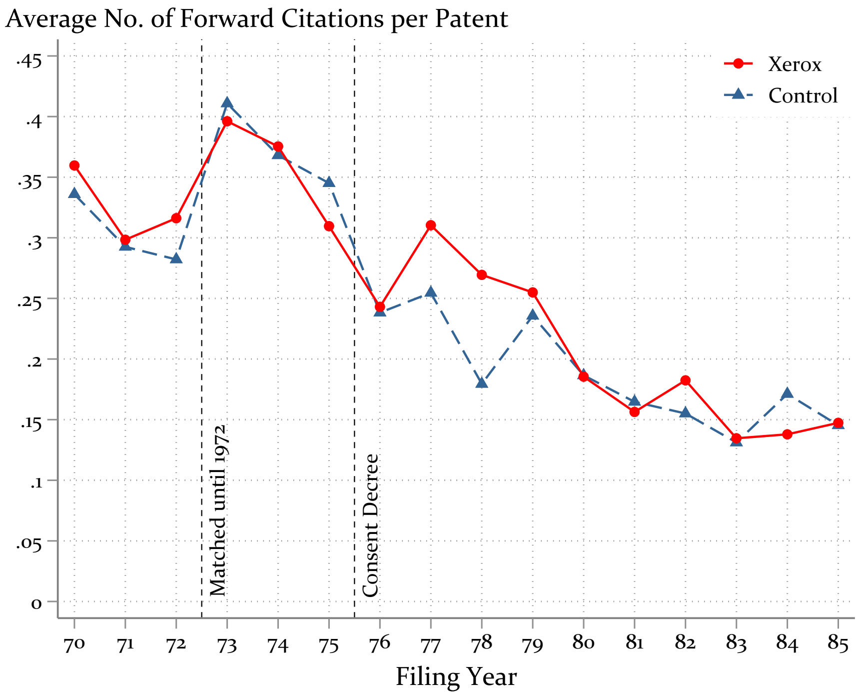
Notes: The figure depicts point estimates and 90% confidence intervals from estimating the regression model in equation (1.1). In the baseline, the outcome variable is the overall number of patent applications by firms other than Xerox per subclass and year. In the remainder of the figure, the outcome variable only considers patent applications that built on compulsorily licensed Xerox patents through citations of different degrees, following the distance framework by Ahmadpoor and Jones (2017). I refer to patents as 'unconnected to Xerox' if they have a distance ≥ 4 to Xerox. Red dots (blue circles) indicate statistically significant (insignificant) point estimates. Standard errors are clustered at the four-digit CPC technology class level.

unconnected to Xerox (i.e., with distance ≥ 4). This serves as a placebo check and, reassuringly, the estimate is close to zero and statistically insignificant.

I run a number of additional robustness checks to ensure that my main results are not driven by a specific model, treatment, or sample specification. The various results are reported in appendix A.1. For example, I find that my estimates are robust to estimating a Poisson pseudo-likelihood regression (instead of an ordinary least squares model) and to excluding the subclasses with the highest share of compulsorily licensed patents from the sample. In addition, I employ both a binary treatment measure and another continuous treatment specification that is based on the number (as opposed to the share) of compulsorily licensed Xerox patents per subclass. Using these alternative treatment definitions does not affect the main results. I also show that my findings are robust to aggregating treated and untreated six-digit subclasses within a four-digit class as in Watzinger et al. (2020) and Watzinger and Schnitzer (2022).

Overall, the empirical evidence suggests that compulsory licensing of Xerox's patents had a positive and statistically significant effect on subsequent innovation. There was a disproportionate increase in patenting in technologies where a larger share of Xerox patents became available for licensing. I now turn to the results of my comple-

Figure 1.5. Patent-Level Analysis: Citations to Xerox vs. Control Patents



Notes: The figure depicts the average number of forward citations per patent. Self-citations are not taken into account. Averages are computed separately for compulsorily licensed Xerox patents and matched control patents. The patents are aggregated using the weights by Iacus et al. (2012).

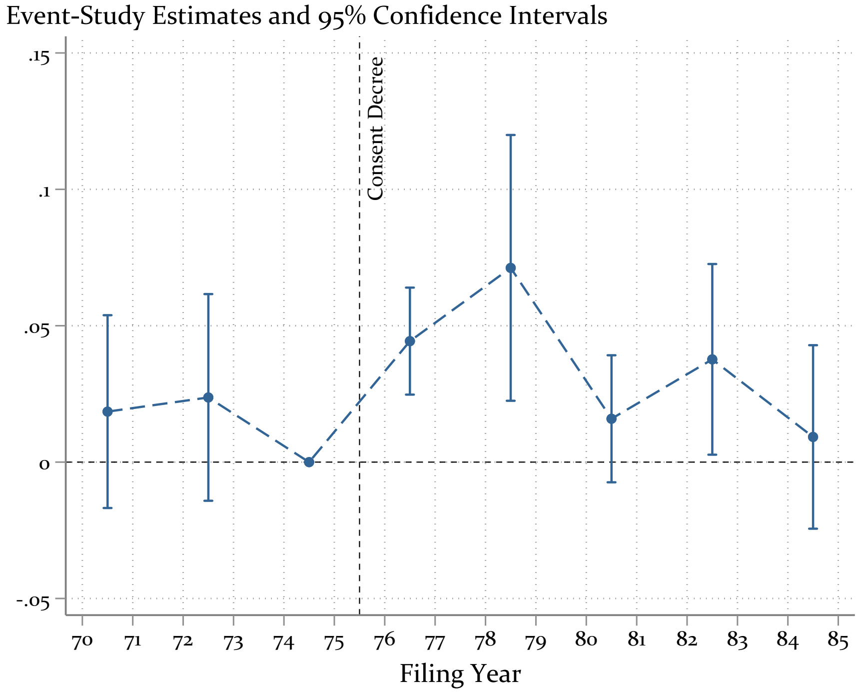
mentary analysis on the patent level, which investigates direct follow-on innovation to compulsorily licensed Xerox patents.

1.4.2 Direct Follow-on Innovation

Figure 1.5 depicts the average number of forward citations per patent separately for compulsorily licensed Xerox patents and matched control patents. The figure suggests that Xerox patents experienced a relative increase in citations in the years 1977 to 1979. For instance, the difference in 1978 represents an increase in citations by roughly 50% and implies that, on average, there was one additional citation for every ten Xerox patents in that year. In most other years, the average number of citations across both groups closely tracks each other. The common trend until 1972 is mostly by construction due to the matching technique. Yet, the development until 1975 speaks in favour of the identifying assumption that the number of citations would have followed a common trend even in the absence of compulsory licensing.

This first visual impression of an increase in citations to Xerox patents is also confirmed by the event-study estimates shown in Figure 1.6. While the pre-treatment estimates are not statistically distinguishable from zero, the estimates in the years fol-

Figure 1.6. Patent-Level Analysis: Event-Study Estimates



Notes: The figure depicts point estimates and 95% confidence intervals from an event-study variation of the model in equation (1.3). Citations are binned in two-year groups to reduce noise in the estimates. The regression uses the weights by Iacus et al. (2012). Standard errors are clustered at the four-digit CPC technology class level.

lowing the consent decree show a positive effect of compulsory licensing on the number of forward citations received by Xerox patents. The effect then fades out again, indicating that the impact of compulsory licensing on direct follow-on innovation may be rather short-lived. This pattern is consistent with the findings by Watzinger et al. (2020). It also goes in line with my previous result from Figure 1.4, showing that direct citations to compulsorily licensed Xerox patents only explain a fraction of the overall effect on cumulative innovation.

The baseline DiD estimate from the patent-level analysis is shown in column (1) of Table A.7 in the appendix. The point estimate indicates that, on average, every compulsorily licensed Xerox patent received an additional 0.02 citations per year after 1975 relative to matched control patents. The estimate is statistically significant at the 1% level.

Overall, the results from this complementary analysis on the patent level are in line with the findings from my main approach on the class level. After the 1975 consent decree, there was not only a disproportionate increase in patenting in fields where Xerox's technology became available for licensing to competitors; Xerox patents also experienced a relative increase in citations. I conclude that compulsory licensing of

Xerox's patents had a positive effect on innovation, as other firms used the newly available technology for follow-on research and either directly or indirectly built on Xerox's patents.

1.5 Which Firms Benefited?

I now investigate which firms benefited from access to Xerox's technology. In the first step, I split the number of patent applications by the applicant's country of origin, hence analysing where follow-on innovators were located. In the second step, I look at heterogeneity by firms' previous patenting behaviour. This allows me to asses whether the effect is driven by firms with prior experience in copier technologies – that is, firms that could become direct competitors to Xerox.

1.5.1 Effects by Applicant Country

Table 1.2 presents estimates of the effect of compulsory licensing on patenting by applicants from different countries. The table shows DiD estimates from my main class-level approach where the outcome variable (i.e., the number of patent applications) is split by applicant country. The results are striking: column (2) indicates that compulsory licensing of Xerox's patents had a quantitatively small and statistically insignificant effect on patenting by US applicants. In contrast, the positive and significant baseline estimate from column (1) is entirely driven by patenting by non-US firms, as shown in column (3). Amongst them, applicants from Japan were the driving force behind the increase in patenting. The estimate in column (4) indicates that, among the foreign applicants, Japanese firms accounted for more than 85% of the additional patent applications after 1975. This effect is statistically significant at the 5% level. The magnitude of the effect implies that compulsory licensing of Xerox's patents increased Japanese patenting in treated technologies by around 5.7% per year. In absolute terms, this corresponds to 123 additional Japanese patents per year after 1975. Column (5) shows that other non-US countries, if anything, experienced a quantitatively small increase in patenting.[17] The results are similar when repeating this exercise with my complementary approach on the patent level, as shown in Table A.7 in the appendix.

These estimates suggest that, in terms of subsequent innovation, Japanese firms were the main beneficiaries of antitrust action against Xerox. Yet, one concern about

[17] The important role of Japanese applicants becomes even more apparent when comparing the DiD estimates to the average number of patent applications per country, which is even smaller for Japan than for other non-US countries. That is, in percentage terms, the effect on Japanese patenting is even more pronounced.

Table 1.2. Class-Level Analysis: Heterogeneity by Applicant Country

	Baseline	Applicant Country			
		USA	Non-USA	Among Non-USA	
				Japan	Others
	(1)	(2)	(3)	(4)	(5)
$\text{Share}_s \cdot \text{Post}_t$	0.189**	0.029	0.162**	0.143**	0.020
	(0.094)	(0.038)	(0.073)	(0.064)	(0.013)
Additional Patents per Year	163	25	140	123	17
Relative Increase	1.5%	0.4%	3.2%	5.7%	0.8%
Mean of Outcome	15.1	8.9	5.7	2.3	3.5
No. of 6-Digit CPC Classes	2,210	2,210	2,210	2,210	2,210
No. of 4-Digit CPC Classes	141	141	141	141	141
Observations	35,360	35,360	35,360	35,360	35,360

Notes: The table shows the results from difference-in-differences regressions following equation (1.1). Column (1) repeats the baseline estimates from Table 1.1. In columns (2) to (5), the outcome variable is restricted to patent applications filed by assignees from selected countries. All regressions include subclass and year × class fixed effects. Standard errors clustered at the four-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

these estimates could be that increased patenting by Japanese firms in the US may not necessarily represent novel innovation. It may also reflect that Japanese firms started to seek protection for their existing technologies in a foreign country where they previously did not enter the product market. If this were the case, I should not find any treatment effect when using data on domestic patent applications in Japan. To address this concern, therefore, I repeat my analysis with data on patent applications at the Japanese Patent Office (instead of at the USPTO). The data were obtained from the Japanese Institute of Intellectual Property (Goto and Motohashi, 2007). As shown in Table A.6 in the appendix, the disproportionate increase in patenting in technologies where Xerox patents became available for licensing is also present in domestic patent applications in Japan. In addition, this effect is again driven by increased patenting by Japanese firms, whereas the coefficient on patenting by US firms is small in magnitude and statistically insignificant. All in all, this robustness check confirms the interpretation of my results as increased innovation by Japanese firms.

Finally, it should be noted that interventions by the Japanese government – such as the Ministry of International Trade and Industry – played a very limited role in the development of the Japanese copier industry (Jacobson and Hillkirk, 1986). Therefore, Japanese copier producers equally had to rely on patents to protect their IP both domestically and abroad, even prior to the antitrust case against Xerox.

1.5.2 The Role of Prior Experience in Copier Technologies

Next, I investigate whether the firms that benefited from compulsory licensing were (potential) competitors to Xerox in the copier market. This need not necessarily be the case, as Xerox's patents covered some basic technologies that could also be used outside of the copier industry. For example, one application entirely unrelated to copiers is xeroradiography, a specific X-ray technique. Moreover, Watzinger et al. (2020) show that compulsory licensing in the case of Bell led to an increase in innovation only outside of telecommunications, which was Bell's core industry. As Bell was a vertically integrated company, it could still foreclose rivals in the telecommunications industry. However, the market structure in the case of Xerox was fundamentally different. Xerox's patents were the main entry barrier in the plain-paper copier market. Therefore, one would hypothesize that the removal of patent protection should allow Xerox's competitors to use the newly available technology for follow-on innovation.

To identify potential entrants into the copier market, I compute a measure of firms' closeness to Xerox based on their prior patenting experience. I define the variable

$$\text{Closeness}_i = \sum_s \text{w}_{i,s} \cdot \text{Share}_s, \tag{1.4}$$

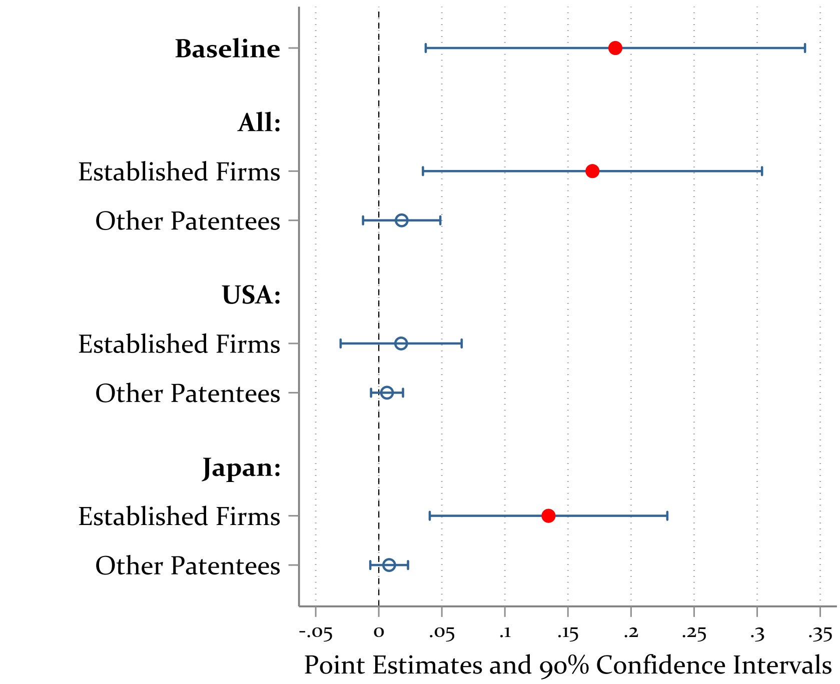
where $\text{w}_{i,s}$ is the share of firm i's unexpired patents (as of 1975) that are in subclass s. Share_s is the treatment variable from equation (1.1) and represents the share of patents in a given subclass that were subject to compulsory licensing. Therefore, the firm-level variable Closeness_i represents the degree to which a firm's existing patent portfolio overlaps with the set of compulsorily licensed Xerox patents. Summary statistics for the closeness measure are presented in Table A.3 in the appendix. I make two sample restrictions to use this closeness measure. First, I only consider patents by firms, hence excluding patent applications filed by individuals, universities, or government bodies. Second, I only consider firms that filed at least ten patent applications from 1970 until 1975 – which I define as 'established' firms.[18] The resulting firm sample consists of 1,635 firms.

Panel (A) of Figure 1.7, which depicts DiD estimates from my main class-level approach, shows that the sample of established firms accounts for almost the entire post-1975 increase in patenting. This is evident from the point estimate in the second row, where the outcome variable is restricted to patent applications by established

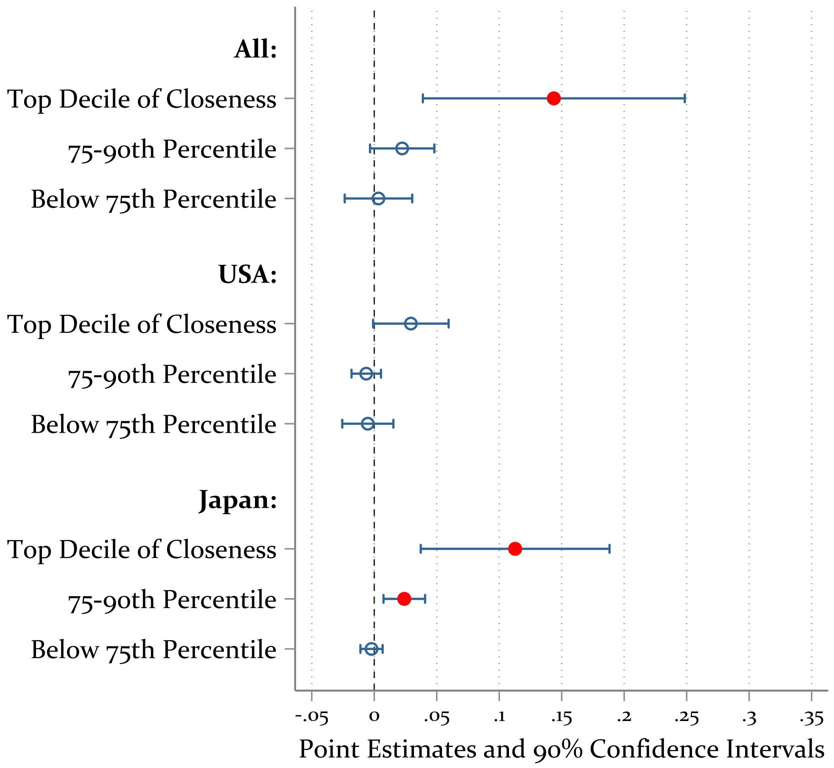
[18] For firms with very few patent applications in the pre-treatment period, the weighted sum in equation (1.4) becomes meaningless. However, my results do not hinge on this specific definition of the closeness measure or the sample. Moreover, in appendix A.1.7, I present an alternative approach to estimating heterogeneity by prior patenting experience, which leads to similar results as Figure 1.7.

Figure 1.7. Class-Level Analysis: Heterogeneity by Firms' Patenting Experience

(A) Established Firms vs. Other Patentees



(B) Closeness to Xerox



Notes: The figure depicts point estimates and 90% confidence intervals from estimating the regression model in equation (1.1). In panel (A), the outcome variable (i.e., the number of patent applications) is split by applicant country as well as by whether the assignee is part of the restricted sample of 'established firms' with at least ten patent applications from 1970 until 1975. All remaining patent applications are labelled as coming from 'other patentees'. Panel (B) employs the closeness measure from equation (1.4). The outcome variable is split by applicant country and by the applicant's percentile in terms of the distribution of the closeness measure across all firms in the restricted sample. Red dots (blue circles) indicate statistically significant (insignificant) point estimates. Standard errors are clustered at the four-digit CPC technology class level.

firms. This result is not obvious, since established firms only account for around 57% of all patent applications in my baseline sample. The estimate in the third row of panel (A) further shows that all remaining patentees did not experience a significant increase in patenting in technologies exposed to compulsory licensing. This indicates that antitrust action against Xerox did not benefit start-ups or other small firms. The remaining estimates in panel (A) of Figure 1.7 highlight that the increase in patenting among established firms is almost exclusively driven by Japanese firms.

Panel (B) of Figure 1.7 now employs the closeness measure to further investigate which firms benefited from compulsory licensing. For that purpose, I split firms into three groups according to their percentile in the distribution of the closeness measure (across all countries). I then repeat my class-level DiD approach with the outcome variable restricted to patent applications by firms from each group. The estimates in the top three rows of panel (B) indicate that the observed increase in the number of patent applications after 1975 is due to firms with a high degree of prior experience in technologies related to Xerox. In other words, compulsory licensing of Xerox's patents promoted innovation primarily within the target industry.[19] The remainder of panel (B) of Figure 1.7 repeats the analysis separately for applicants from the US and Japan. Among the American firms, the point estimate is positive only for firms with the highest technology overlap with Xerox, but this effect is quantitatively small and not statistically significant. For Japan, the estimates highlight that there is great heterogeneity in the effect of the antitrust case even among Japanese patent applicants. The positive innovation effect is particularly driven by firms in the top decile of the distribution of the closeness measure. In contrast, firms below the top quartile of the closeness measure did not experience any change in their patenting in either country. Overall, Figure 1.7 highlights that the main beneficiaries of the antitrust case were those Japanese firms that had extensive prior knowledge in copier technologies. Accordingly, they could use the technology available for licensing from 1975 onwards for follow-on innovation.

The increase in innovation among Japanese competitors is in line with the historical events in the copier industry. Starting in the mid-1970s, several Japanese companies – including Canon, Konica, Minolta, Ricoh, Sharp, or Toshiba – entered the US copier

[19] Strictly speaking, the results from Figure 1.7 are not indicative about whether the firms that benefited from compulsory licensing were, in fact, active in the same product market as Xerox. However, there are two reasons why I define closeness to Xerox based on prior technology experience rather than product market activity. First, I believe that firms' prior patenting better captures their potential to compete in the copier market. Second, and relatedly, empirically identifying Xerox's (potential) competitors based on industry assignment (e.g., Compustat segments data) is complicated. This is because entrants into the copier market originally operated in very different industries, ranging from photographic equipment (e.g., Canon, Kodak, Konica) to computing (e.g., IBM) to consumer electronics (e.g., Sharp, Toshiba).

market with great success and became important competitors to Xerox (e.g., Jacobson and Hillkirk, 1986; Gomes-Casseres and McQuade, 1991; Scherer, 2005). Reassuringly, these Japanese firms are all located in the top decile of the distribution of the closeness measure, hence supporting my interpretation of this measure as identifying potential competitors in the copier market.[20]

In summary, the heterogeneity analyses reveal two important results. First, compulsory licensing in the case of Xerox was effective at increasing innovation in the copier industry. That is, the antitrust intervention allowed established firms with prior experience in copier technologies to increase their patenting. However, second, this is true primarily for Japanese competitors, which were the main beneficiaries of the antitrust measure in terms of their subsequent innovation performance. To further corroborate this result, I present additional firm-level evidence in the next step.

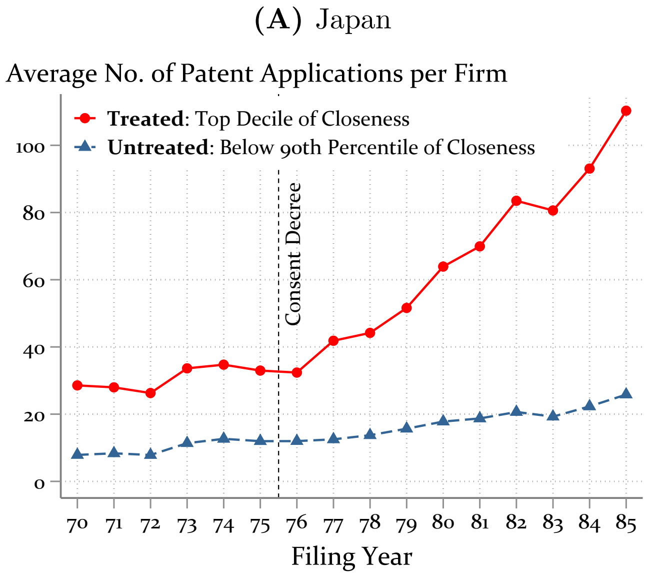
1.5.3 Firm-Level Results on the Effect on Japanese Competitors

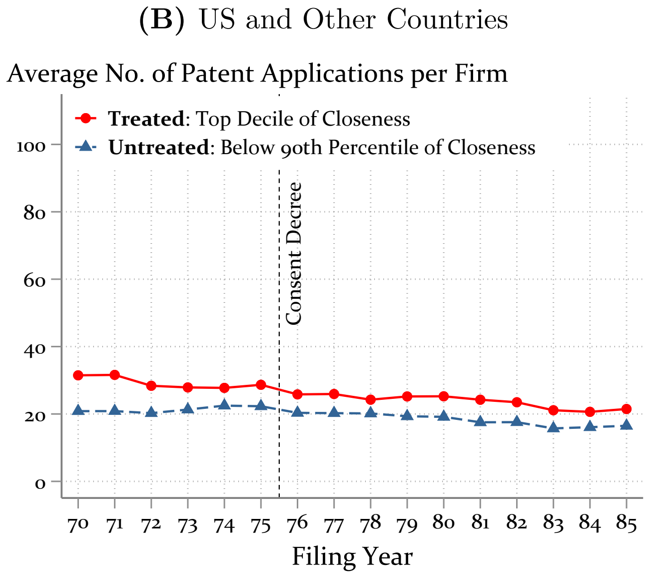
I turn to the firm level to further investigate how Japanese competitors changed their patenting behaviour following the consent decree. Figure 1.8 depicts the average number of patent applications per firm separately for Japan and all other countries (including the US) in panels (A) and (B), respectively. Within each panel, I split firms into two groups and define a firm as treated if it is in the top decile of the distribution of the closeness measure. The figure reveals that Japanese firms were on a different patenting trend from the rest of the sample throughout the relevant period. While the average number of patent applications per firm in the US and other countries was mostly constant or even slightly decreased over time, patenting among Japanese firms steadily increased from 1970 until 1985. However, Japanese firms in the top decile of the distribution of the closeness measure experienced a much stronger increase in patenting after 1975 than the remaining Japanese firms in the sample. In other words, Japanese firms with a large amount of prior experience in copier technologies disproportionately increased their innovation activities after Xerox's patents became available for licensing. In other countries, in contrast, patenting by firms with a greater technology overlap with Xerox did not evolve differently over time.

These descriptive results demonstrate that the effect on Japanese competitors in my class-level approach does not simply reflect an aggregate increase in Japanese patenting. This is reassuring and corroborates my key empirical findings. In appendix A.3.2,

[20] Among the 163 firms (i.e., 10%) that are in the top decile of the distribution of the closeness measure, 26 firms are from Japan. A list of these Japanese firms is provided in Table A.8 in the appendix.

Figure 1.8. Firm-Level Analysis: Patenting Trends Across Firms





Notes: The figure depicts the average number of patent applications per firm. Panel (A) includes firms from Japan, whereas panel (B) includes all remaining firms in the sample of established firms. Averages are computed separately for treated and untreated firms, where a firm is defined as treated if it is located in the top decile of the distribution of the closeness measure defined in equation (1.4).

I further investigate whether differences in observable firm characteristics may explain the heterogeneous effect across countries, but I find no evidence supporting this hypothesis. Therefore, my additional analyses on the firm level indicate that the positive effect on Japanese innovation represents, in fact, a phenomenon that is idiosyncratic to a specific group of Japanese firms that had extensive prior knowledge in copier technologies.

1.6 Mechanism

Why were Japanese copier producers more successful in building on Xerox's technology than their American counterparts? I now address this question by investigating the mechanisms underlying my results. First, I discuss historical narratives suggesting that Japanese entrants focused on producing smaller desktop copiers. I then study how compulsory licensing of Xerox's patents affected the quality and diversity of innovation.

1.6.1 The Japanese Focus on Smaller Desktop Copiers

Historical narratives indicate that American and Japanese firms entered the copier market with different strategies. On the one hand, American entrants such as IBM or Kodak started competing with Xerox in the same (high-volume) market segment where

Xerox was dominant. For IBM, in particular, entering the high-volume segment may be explained by the potential to exploit economies of scale from the company's existing distribution network for mainframe computers. On the other hand, Japanese entrants strategically focused on the low end of the copier market by producing smaller and lower-volume plain-paper copiers than Xerox and most American competitors (e.g., Jacobson and Hillkirk, 1986; Porter, 1988; Gomes-Casseres and McQuade, 1991).[21] According to Jacobson and Hillkirk (1986, p. 105), key determinants of the Japanese success included entering the right market segment at the right time, standardising and externalising the production of inputs, and using automation to exploit economies of scale – all with the objective of building a 'value-added product that's simpler and cheaper to build and use'.

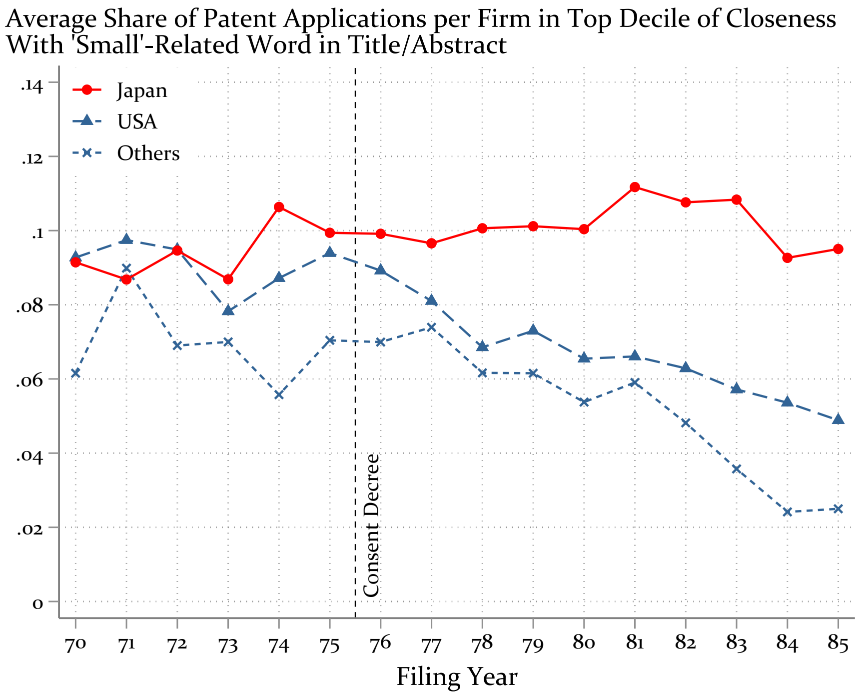
This indicates that the heterogeneous effect across countries is unlikely to be driven by some of the competitors being Japanese. Instead, in line with the arguments by Chesbrough and Rosenbloom (2002), Japanese entrants employed a different business model that expanded the market for plain-paper copiers to the lower-volume segment. This distinct competitive strategy is also well established in the management literature, where the Japanese entry into the American copier market is frequently named as an example of a successful attack on a dominant market leader (e.g., Porter, 1985; Paley, 2006).

Consistent with this possible mechanism, I show in Figure 1.9 that patents filed by Japanese competitors more frequently contained words associated with smaller desktop copiers. I focus on firms in the top decile of the distribution of the closeness measure and search the titles and abstracts of their patents for (variations of) one of the following words: compact, desktop, efficient, energy-saving, miniature, minuscule, portable, scale, small, simple, size, and tiny. As shown in Figure 1.9, the share of patent applications per firm with such 'small'-related words was roughly equal across countries in the early 1970s. Then, however, there was a divergence. The average share of Japanese patent filings that contained any 'small'-related word slightly increased up to around 12%, whereas the share steadily declined for competitors from the US and other countries.[22] Although this result is purely descriptive, it supports the narrative evidence about the role of the Japanese entrants' distinct business model – suggesting that a

[21] See also, 'Small Is Better, Xerox-san' in *The Washington Post* from 1981 (`https://www.washingtonpost.com/archive/business/1981/02/15/small-is-better-xerox-san/bb081ecb-1061-4492-8bd2-29aa5f91212d/`, last accessed: 11 February 2024).

[22] Even if the Japanese's focus on lower-volume copies fully explained the heterogeneous effect across countries, it is unclear how this should affect the pattern in Figure 1.9. After all, building smaller copiers required inventions far beyond miniaturisation of existing technologies. Therefore, I consider Figure 1.9 as indicative that the Japanese entrant's distinct business model likely played a role; but it is inconclusive regarding the importance of this channel.

Figure 1.9. Firm-Level Analysis: Use of 'Small'-Related Words in Patents



Notes: The figure depicts the average share of patent applications per firm that contain a 'small'-related word in the patent title or abstract. The following words (and variations of them) are considered: compact, desktop, efficient, energy-saving, miniature, minuscule, portable, scale, small, simple, size, and tiny. The sample is restricted to firms in the top decile of the distribution of the closeness to Xerox. Averages are computed separately by country. A three-year moving average is applied to these averages to reduce noise.

greater degree of product differentiation from existing copiers may (at least partly) explain the higher rate of innovation among Japanese competitors.

1.6.2 Effect on the Quality and Diversity of Innovation

Thus far, the chapter focused on the effect of the antitrust case on the intensity of innovation. Yet, simple patenting numbers may not necessarily be informative about the quality and content of the underlying inventions. Therefore, I now analyse how compulsory licensing of Xerox's patents affected the quality and diversity of innovation.

To study patent quality, Table 1.3 makes use of two different patent measures: the number of forward citations that a patent has received and the ten-year quality measure by Kelly et al. (2021, KPST). The latter is constructed using textual analysis and based on the idea that high-quality patents should be novel relative to prior art but must have a high impact on future inventions. Citations, in contrast, only capture how much a given patent is used by subsequent patents. In columns (2) and (3) of Table 1.3, I restrict the outcome variable to patents in the top 10% of the distribution of these two patent measures. The DiD estimates indicate that the newly filed patents

Table 1.3. Class-Level Analysis: Patent Quality

		Patents in Top 10%		Mean	
	Baseline	Forward Citations	Quality (KPST)	Forward Citations	Quality (KPST)
	(1)	(2)	(3)	(4)	(5)
$\text{Share}_s \cdot \text{Post}_t$	0.189**	0.004	0.142**	0.106	0.002*
	(0.094)	(0.009)	(0.057)	(0.126)	(0.001)
Additional Patents per Year	163	4	122	n/a	n/a
Relative Increase	1.5%	0.3%	5.8%	n/a	n/a
Mean of Outcome	15.1	1.6	2.0	14.1	-0.0
No. of 6-Digit CPC Classes	2,210	2,210	2,210	2,210	2,210
No. of 4-Digit CPC Classes	141	141	141	141	141
Observations	35,360	35,360	35,360	35,360	35,360

Notes: The table shows the results from difference-in-differences regressions following equation (1.1). Column (1) repeats the baseline estimates from Table 1.1. In columns (2) and (3), the outcome is the number of patent applications in the top 10% of the distribution of forward citations and the quality measure by Kelly et al. (2021, KPST), respectively. In columns (4) and (5), the outcome variable is the average number of forward citations per patent and the average KPST measure per patent, respectively. All regressions include subclass and year × class fixed effects. Standard errors clustered at the four-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

after 1975 were of higher quality in terms of the KPST measure but not in terms of forward citations. In columns (4) and (5), the outcome variable is the average number of forward citations per patent and the average KPST quality measure per patent, respectively. This set-up leads to a similar result, finding a small positive effect only on the average KPST measure.

The estimates suggest that the additional patents filed after 1975 were more dissimilar from prior art but did not have a greater impact on subsequent patents, relative to the average patent. Both citations and the KPST measure capture how much a patent is used in subsequent patent filings. While citations rely on explicit references, the KPST measure identifies a forward similarity through textual analysis (Kelly et al., 2021). However, KPST then divide this impact measure by a patent's backward similarity (i.e., an indicator for novelty). A high KPST quality measure can, therefore, result from a low backward similarity or a high forward similarity (or both). As I do not find any effect on citations, it is plausible to assume that the higher KPST measure among the additional post-1975 patents stems from their low backward similarity, suggesting that these patents were more novel.

Next, I turn to the diversity of innovation and analyse whether overall patenting activity expanded to new technologies. Watzinger and Schnitzer (2022) propose to measure the diversity of innovation by looking at the number of technology subgroups with at least some patenting activity. In the hierarchical patent classification system, subgroups are the level below six-digit technology classes, which I use in my main

Table 1.4. Class-Level Analysis: Active Subgroups

	All	Applicant Country			
		USA	Non-USA	Among Non-USA	
				Japan	Others
	(1)	(2)	(3)	(4)	(5)
$\text{Share}_s \cdot \text{Post}_t$	0.037**	0.006	0.044**	0.043**	0.006
	(0.015)	(0.007)	(0.020)	(0.021)	(0.006)
Additional Subgroups per Year	32	5	37	37	5
Relative Increase	1.1%	0.2%	2.3%	4.2%	0.4%
Mean of Outcome	4.8	3.6	2.6	1.1	2.0
No. of 6-Digit CPC Classes	2,210	2,210	2,210	2,210	2,210
No. of 4-Digit CPC Classes	141	141	141	141	141
Observations	35,360	35,360	35,360	35,360	35,360

Notes: The table shows the results from difference-in-differences regressions following a variation of equation (1.1). Unlike in the main approach, the outcome variable now is the number of 'active' subgroups (aggregated to two dots) within a six-digit technology class. Column (1) reports the baseline estimates. In columns (2) to (5), only patent applications by assignees from selected countries are counted to determine whether a subgroup was 'active'. All regressions include subclass and year × class fixed effects. Standard errors clustered at the four-digit IPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

approach. Therefore, they represent the most disaggregate technology classification of a patent.[23] Following Watzinger and Schnitzer (2022), I refer to a subgroup as 'active' if it contained at least one patent application in a given year. To estimate the effect of the antitrust case on the diversity of innovation, I then use the same DiD approach as in my baseline analysis. Now, however, the outcome variables counts the annual number of active subgroups within each six-digit technology class.

Table 1.4 presents the resulting DiD estimates. Column (1) shows that compulsory licensing of Xerox's patents had an overall positive effect on the number of active subgroups. This increase in the diversity of innovation is again driven by patent applications from Japan. The estimate in column (4) indicates that, on average, a one percentage-point higher share of compulsorily licensed Xerox patents in a subclass is associated with 0.04 additional subgroups with at least one Japanese patent per year in that subclass after 1975. This effect is statistically significant at the 5% level. It corresponds to around 37 additional active subgroups among Japanese applicants per year, which represents an increase by around 4.2%. The corresponding event-study estimates are presented in Figure A.8 in the appendix. Reassuringly, there were no significant pre-trends in the number of active subgroups with at least one Japanese patent application.

[23] In fact, subgroups have their own hierarchical order that is referred to as 'dot'-hierarchy. I follow Watzinger and Schnitzer (2022) and aggregate subgroups to the two-dot level, which is the second highest level. On average, there are around 4.8 subgroups at the two-dot level per six-digit technology class.

These results demonstrate that (Japanese) innovation became more diverse after 1975, as patenting activity expanded to a larger number of distinct technologies. In appendix A.1.8, I further investigate whether firms also changed the direction of their innovation, following the approach by Kang (2021). I find that firms patented relatively more outside of their primary technology fields. That is, firms shifted the focus of their innovation activities towards previously peripheral technologies.

Overall, my analyses indicate that antitrust enforcement against Xerox made innovation more novel and diverse. Patenting activity expanded to new technology fields, while there was no reduction in the quality of inventions. These changes in the direction of innovation are again driven by Japanese patent applicants. Therefore, the results are in line with narrative evidence suggesting that Japanese entrants focused on smaller and lower-volume desktop copiers, which were more differentiated from existing products.

Finally, I show in appendix A.4 that there was also a disproportionate increase in Japanese copier exports to the US after 1975, relative to exports in other industries and by other countries. This indicates that Japanese competitors benefited from the antitrust case not only in terms of increased innovation; they were able to generate higher revenues in the product market as well.

1.7 Effect on Xerox

While the chapter so far focused on innovation by firms other than Xerox and its subsidiaries, I now study how Xerox's own patenting activities reacted to the removal of most of its IP. Estimating the effect on Xerox is challenging, because it requires to find a good counterfactual, indicating how much Xerox would have innovated in the absence of compulsory licensing.

I address this challenge by using the synthetic control method by Abadie et al. (2010, 2015) to find a control group for patenting by Xerox and its subsidiaries. I match on the yearly number of patent applications from 1960 until 1972 (i.e., until the start of the antitrust case). As donor pool, I use the sample of established firms. I exclude firms in the top decile of the distribution of the closeness measure, as these potential competitors may have been affected by the antitrust case themselves. Therefore, they do not represent a suitable counterfactual. The resulting synthetic control group consists of 66.6% Siemens, 17.4% Bell, and 16.1% Westinghouse.[24] That is, it

[24]Bell also faced an antitrust lawsuit in the late 1970s and was ultimately broken up in 1984 (Watzinger and Schnitzer, 2022). In appendix A.5, I show that excluding Bell from the donor pool leads to quantitatively similar albeit slightly more negative effects on Xerox.

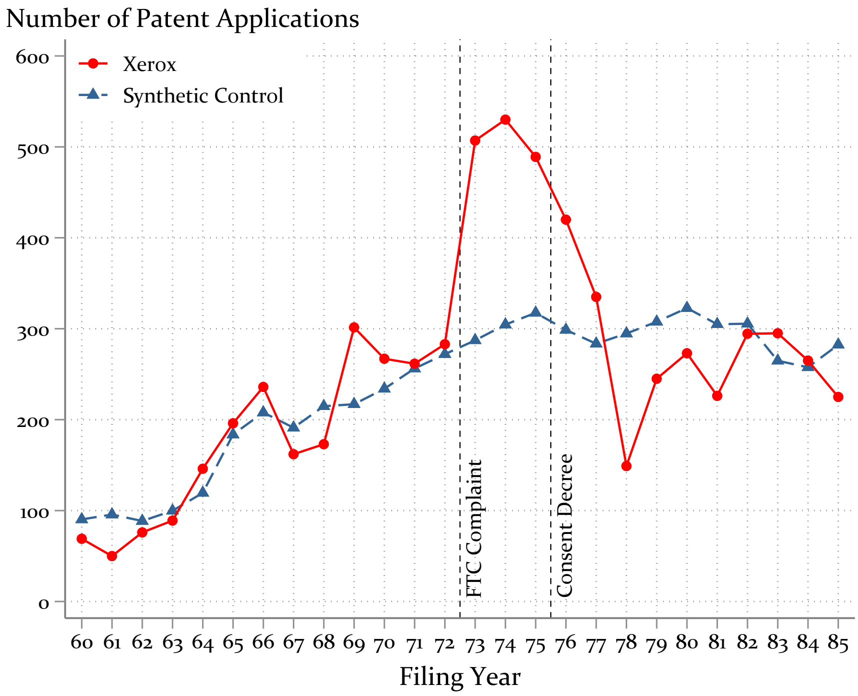
represents a weighted average of other companies active in high-technology sectors but not in the copier industry.

Figure 1.10 depicts the annual number of patent applications by Xerox and the synthetic control group. Mechanically, patenting by Xerox and its synthetic control closely tracks each other until 1972, when the FTC published its antitrust complaint. Then, in 1973, Xerox increased its patenting by around two thirds relative to the previous year, whereas there was no such trend break among the synthetic control. It is unclear what explains this sudden rise in Xerox's patenting, although the timing suggests that it may be linked to the antitrust case. One possibility is that Xerox may have reacted to the FTC complaint by filing patents for disclosure purposes – that is, to prevent others from obtaining patent protection on similar inventions. Finally, following the consent decree, Xerox's patenting declined until 1978 and then slowly increased again. In contrast, patenting by the synthetic control remained relatively constant after 1975.

A simple (hand-computed) DiD estimate indicates that patenting by Xerox and its subsidiaries decreased by around 16 patents per year on average after 1975, relative to the synthetic control and relative to the pre-1972 period. Excluding Bell from the donor pool leads to a slightly larger reduction in Xerox's patenting by around 30 patents, as shown in appendix A.5. Although these simple estimates should be taken with caution, they suggest that the increase in innovation among (Japanese) competitors may have been partly attenuated by a decrease in Xerox's own patenting. An alternative interpretation of the pattern shown in Figure 1.10 is that, rather than showing a negative effect, patenting by Xerox and its subsidiaries may have reversed back to the level of the synthetic control after 1975. Regardless of which interpretation one prefers, the synthetic control approach shows that any potential reduction in Xerox's patenting was much smaller in magnitude than the increase in innovation by other firms of around 160 patents per year, which I find in my main approach. Therefore, the overall effect of the antitrust case on innovation remains largely positive.

There is one important caveat regarding these results on the effect on Xerox. After 1975, a decline in Xerox's patenting may not necessarily represent less innovation. This is because the consent decree also imposed compulsory licensing on future Xerox patents issued until 1981. Therefore, unlike for other firms, the incentives for Xerox to file for patent protection were significantly reduced after 1975. As a consequence, the estimated decrease in Xerox's patenting likely represents an upper bound of the actual reduction in Xerox's innovation activities.

Figure 1.10. Effect on Xerox: Patenting by Xerox vs. Synthetic Control



Notes: The figure depicts the number of patent applications per year for Xerox and its subsidiaries (in red) and a synthetic control group (in blue). The synthetic control group is computed using the algorithm by Abadie et al. (2010, 2015) and consists of 66.6% Siemens, 17.4% Bell, 16.1% Westinghouse.

1.8 Conclusion

In this chapter, I study how antitrust enforcement against patent-based monopolies affects subsequent innovation by domestic and foreign firms. To answer this question, I analyse the impact of compulsory patent licensing in the context of the antitrust case against Xerox in the 1970s. I find a positive effect of compulsory licensing on innovation by other firms, measured by a disproportionate increase in patenting in technologies where Xerox patents became available for licensing. Moreover, Xerox patents were cited more frequently after 1975 relative to a set of matched control patents. Therefore, antitrust action against Xerox was successful not only by promoting competition on the product market (Bresnahan, 1985a). My results indicate that it also significantly spurred innovation and promoted technological progress in copier technologies. Yet, this positive effect is mainly driven by increased innovation by Japanese competitors. They started developing smaller desktop copiers and their innovation became more diverse.

The positive innovation effect raises the question why Xerox did not license its technologies to other firms in return for royalties prior to the consent decree. In theory, efficient bargaining between the owner of an upstream technology and downstream innovators leads to ex-ante licensing such that any surplus-enhancing follow-on innovation should be developed (Green and Scotchmer, 1995). However, patents may exert a blocking effect on follow-on innovation if upstream and downstream parties fail to reach a licensing agreement. In the case of Xerox, the most plausible mechanism explaining the absence of ex-ante licensing is a rent dissipation theory (Arora and Fosfuri, 2003; Gaessler et al., 2019). That is, licensing likely would have been unprofitable for Xerox, as its loss in profits due to increased product market competition would have exceeded possible licensing revenues. Appendix A.6 presents a brief conceptual framework that outlines this point in greater detail. In particular, I point out how several specificities of the Xerox case make the rent dissipation theory more likely than alternative explanations for the absence of ex-ante licensing such as asymmetric information (Bessen and Maskin, 2009) or coordination failure (Galasso and Schankerman, 2015).

While this chapter identifies the blocking effect of Xerox's patents on follow-on innovation, it remains an open question whether compulsory licensing affected firms' incentives to innovate. The Xerox case was one out of more than 100 compulsory licensing cases since the 1940s (Scherer and Watal, 2014). Therefore, it is unlikely that there were large disincentive effects, as the case probably did not alter expectations about the probability of future compulsory licensing orders. Moreover, I show that Xerox itself only modestly reduced its patenting in response to the antitrust case.

There are two important insights from the chapter that can inform policymakers far beyond the Xerox case. First, I show that compulsory licensing can be a suitable antitrust measure to increase innovation *within* the target industry if it removes the main barrier to entry. This result complements prior evidence from the antitrust case against Bell (Watzinger et al., 2020). Since Bell was a vertically integrated monopolist that could continue to foreclose its competitors after the antitrust case, there was no increase in innovation in the target industry. In contrast, compulsory licensing of Xerox's patents removed the main barrier to entry and allowed competitors to build on its technology. This shows that a similar policy can have very different impacts, depending on the source of market power of the incumbent. Thus, the chapter extends prior literature and provides a more complete picture of the effectiveness of compulsory licensing as an antitrust remedy.

Second, the finding that Japanese competitors particularly benefited from antitrust action against Xerox is relevant in light of recent concerns that stricter antitrust enforcement by US authorities may weaken American competitiveness in high-technology sectors. The Xerox case provides an example that illustrates why such concerns may be warranted. However, this does not imply that, in general, antitrust intervention benefits foreign competitors. Instead, the chapter shows how a specific group of entrants, in this case from Japan, most successfully built on the incumbent's technology. They invented a more differentiated product and employed a business model that allowed them to serve a market segment previously ignored by the monopolist. In that way, consumers experienced lower prices, greater product variety, and higher quality. I leave the quantifications of the overall welfare effects of antitrust enforcement as a promising topic for future research.

Chapter 2

Technology Transfer and the Rise of Japan

2.1 Introduction

In the 1960s and 1970s, Japan's consumer electronics sector experienced a significant expansion, particularly in the television industry. Major Japanese television manufacturers – such as Hitachi, Panasonic, or Toshiba – achieved significant import penetration into the US market and competed fiercely with domestic producers. The number of American television manufacturers dropped from approximately 30 in 1960 to only one by 1990. While several factors likely contributed to this shift, there is little empirical evidence on the determinants of the rise of Japan in televisions and consumer electronics more broadly. Historical narratives suggest that voluntary transfer of US technology to Japan may have played a role (e.g., Shih and Dieterich, 2014).[1] Yet, it remains unclear to what extent Japanese firms not only adopted but may also have enhanced American technology.

This chapter aims to fill this gap by empirically studying the effect of technology transfer from the US on Japanese innovation in the 1960s. In doing so, the chapter not only provides insights into the causes of Japan's surge in consumer electronics; it also addresses the broader question of whether technology transfer to emerging markets can promote innovation in the target country. Since innovation is a key driver of long-term growth, this is an important policy issue from a development economics

This chapter is based on single-authored work. I thank Martin Watzinger for sharing the initial idea for this chapter.

[1]See also 'Japan and the Big Squeeze' in *The Washington Post* from 1990 (`https://www.washingtonpost.com/archive/opinions/1990/09/30/japan-and-the-big-squeeze/0fb1617e-8756-4390-a776-f1619d59869a/`, last accessed: 11 February 2024) or 'How Japan Picks America's Brains' in *Fortune Magazine* from 1987 (`https://money.cnn.com/magazines/fortune/fortune_archive/1987/12/21/69996/index.htm`, last accessed: 11 February 2024).

perspective. The question is also relevant for policymakers in developed countries, as more innovative products are an important component of consumer welfare in high-technology markets.

Estimating the causal effect of technology transfer on innovation is challenging for two reasons. First, technology transfer typically results from endogenous decisions by firms. Unobservable factors may simultaneously affect the profitability of both technology transfer and innovation across different fields. This makes it difficult to isolate the effect of technology transfer. Second, data on licensing of technology are often unavailable.

I address these challenges by leveraging the 1958 settlement of a US antitrust case against the Radio Corporation of America (RCA) as a natural experiment. The settlement eliminated the possibility for RCA to earn royalties in the US and induced a large wave of voluntary patent licensing agreements between RCA and Japanese firms. Thus, it led to a voluntary technology transfer from the US to Japan predominantly in those fields where RCA was an active innovator prior to 1958. In my main empirical approach, I exploit this differential likelihood of licensing by RCA by comparing Japanese innovation – measured by the number of US patents filed by applicants from Japan – across technologies over time. In addition, I employ a complementary approach on the firm level to study how patenting by RCA's licensees developed over time, relative to that of other Japanese firms that did not obtain a patent license. This firm-level analysis is possible because there are data on which Japanese firms actually licensed RCA's technology in the 1960s.

The results from both empirical approaches consistently indicate that patent licensing by RCA significantly promoted Japanese innovation after 1958. Consequently, the chapter establishes technology transfer from the US as one important factor that contributed to the rise of the Japanese consumer electronics industry from the early 1960s onwards. More broadly, the chapter shows that licensing state-of-the-art technology can spur technological progress in an emerging market.

RCA was a large American electronics company and one of the leading innovators in the field of radio and television technology, until it was acquired by General Electric in 1986. Among RCA's major inventions were electronic colour television in the 1950s and the liquid crystal display (LCD) in 1968. Since the company's founding in 1919, patent licensing was one of RCA's major revenue streams. As RCA licensed its patents exclusively in packages, the company was repeatedly accused of monopolizing the market for radio and television technology. In 1958, RCA signed a consent decree with the US Department of Justice that settled an antitrust lawsuit and ended the package licensing practice. Importantly, the settlement obliged RCA to make the majority of its patents available to domestic firms free of charge. In an attempt to replace the source of income eliminated by the antitrust settlement, RCA increasingly licensed its technology to Japanese firms from 1958 onwards. This allowed RCA to continue earning substantial royalties over many years, and it induced a large wave of technology transfer to Japan.

To estimate the aggregate effect of RCA's technology transfer on Japanese innovation, I employ a difference-in-differences (DiD) approach on the technology class level (Moser and Voena, 2012; Moser et al., 2014). Specifically, I compare the annual number of US patent applications by Japanese applicants across four-digit technology classes (based on the Cooperative Patent Classification) with a differential likelihood of licensing by RCA. Importantly, I can make this comparison *within* the same three-digit class, exploiting the hierarchical nature of the patent classification system. I use both a binary and a continuous treatment specification to capture how the likelihood of licensing by RCA varies across technologies.

I find that there was a disproportionate increase in Japanese patenting in RCA's main fields. The point estimate when using the more conservative specification with a continuous treatment variable implies that Japanese patenting increased by around 16% in response to RCA's technology transfer. In absolute terms, Japanese applicants filed around 110 additional patents in the US per year after 1958. Event-study analyses show that these estimates are not driven by any differences in pre-trends across technologies. Moreover, the increase in the intensity of Japanese innovation does not reflect inventions of lower quality. I show that the number of high-quality Japanese patents also increased and Japanese innovation became more diverse.

To ensure that these results do not reflect technology-specific trends that may be unrelated to patent licensing by RCA, I extend the empirical set-up to a triple DiD model, using Germany as an additional comparison country. Germany also recovered from the destruction caused by World War II, but there is no evidence that RCA licensed its patents to any German firms. The estimates of the triple DiD analysis reveal that Japanese patenting increased disproportionately in RCA's technologies also relative to patenting by German applicants, while German patenting did not evolve differently in RCA's main fields after 1958. This shows that the *entire* surge in Japanese patenting in technologies with a higher likelihood of licensing by RCA represents an effect that is specific to Japan. It cannot be explained by more general patenting trends across technologies. This corroborates the interpretation of my findings as the impact of RCA's technology transfer.

In the next step, I establish that the increase in Japanese patenting represents direct follow-on innovation building on RCA's technology. To this end, I disaggregate the additional Japanese patents filed after 1958 by their citation distance to RCA, using the framework by Ahmadpoor and Jones (2017). I find that the majority of the additional Japanese patents either directly or indirectly built on RCA's technology. Conversely, patents unconnected to RCA only account for a negligible fraction of the overall effect.

An additional advantage of the setting is that there are data on which Japanese firms entered into licensing agreements with RCA, available from US Congress hearings (US Congress, 1970). I can incorporate these data into the main empirical approach to disaggregate the additional Japanese patents filed after 1958 by whether the applicant received a license from RCA. The estimates indicate that around two thirds of the overall increase in Japanese patenting is driven by RCA's licensees, while the remaining increase may be due to spillover effects to other firms.

Finally, I employ a complementary empirical approach on the firm level to estimate the direct effect of receiving a patent license from RCA on follow-on innovation by the licensees. If the increase in Japanese patenting is indeed the result of technology transfer by RCA, then one should observe that RCA's licensees disproportionately increased their patenting, compared to other Japanese firms that did not obtain a patent license. Therefore, I construct a control group for every Japanese licensee, using the synthetic control method by Abadie et al. (2010, 2015). Then, I estimate a simple DiD model to investigate how the licensees' patenting changed in response to receiving a license from RCA, relative to patenting by the control firms. The results show that the licensees substantially increased their patenting upon receiving a license from RCA. This is true both in absolute terms and relative to the synthetic control firms. Again, the increase is driven by patents that either directly or indirectly built on RCA's technology through citations. Moreover, the licensees filed more high-quality patents and their innovation activity expanded to new technologies. Overall, these firm-level results further support the conclusion that RCA's patent licensing promoted follow-on innovation among its Japanese licensees.

This chapter contributes to several strands of the economics literature. First, it adds to prior research on the impact of technology transfer on firms in the receiving country by providing empirical evidence on the effectiveness of patent licensing as a channel for technology diffusion. There is a large literature on the diffusion of technology across countries that, however, mostly focusses on other diffusion channels such as trade in intermediate inputs (e.g., Goldberg et al., 2009) or foreign direct investment (e.g., Keller and Yeaple, 2009; Bai et al., 2020).[2] More closely related, Giorcelli (2019) studies the effect of management and technology transfer on firm performance in the

[2]See Keller (2004) for an early review of the literature on international technology diffusion.

context of a US programme that supported Italian firms in the 1950s. Giorcelli and Li (2021) analyse the Sino-Soviet Alliance around the same time, which transferred machinery and know-how to Chinese plants. The chapter adds to these prior contributions by investigating the effect of technology transfer through patent licensing (as opposed to the transfer of machines) on follow-on innovation (as opposed to other firm-level outcomes). In addition, by studying the rise of Japan in consumer electronics after World War II, the results from this chapter indicate that licensing of patents can promote innovation and technological progress in an emerging market. Thus, the chapter also complements other studies that highlight the importance of foreign knowledge for the growth of an emerging country's industry (e.g., Mostafa and Klepper, 2018).

Second, the chapter relates to the literature on the effect of (compulsory) licensing on follow-on innovation by the recipients of the licensed technology (Acemoglu and Akcigit, 2012; Moser and Voena, 2012; Watzinger et al., 2020; Nagler et al., 2022; Mamrak, 2023). While most prior work studies compulsory licensing that was partly royalty-free, I focus on how *voluntary* licensing agreements at market-level royalty rates affect subsequent innovation.[3] This is similar to Nagler et al. (2022), who investigate the licensing of the transistor by Bell in the 1950s. I add to their research by studying the impact of non-standardized patent licensing that transferred US technology to a foreign country. Moreover, I can exploit data on licensing agreements to study the innovation response by individual licensees on the firm level.

Third, the chapter extends previous work on the role of technology transfer for the rise of Japan in consumer electronics, which mostly presents narrative evidence by economic historians (e.g., Curtis, 1994; Johnstone, 1999; Chandler, 2005; Choi, 2008). To the best of my knowledge, I am the first to provide thorough empirical evidence on the role of US technology transfer for Japanese innovation in the 1960s.

The findings from this chapter also inform industrial policy. The Japanese government – through its Ministry of International Trade and Industry (MITI) – protected the nascent Japanese consumer electronics industry in several ways. For example, it blocked access to the domestic market to foreign entrants, restricted foreign direct investment in Japan, and promoted the acquisition of foreign technology.[4] As a consequence, foreign manufacturers such as RCA could only make business in Japan by licensing their technology. Although this chapter does not aim to directly evaluate the

[3]Note that I use compulsory licensing of RCA's patents, which applied exclusively to domestic firms in the US, only as a shock that shifted RCA's voluntary licensing activity to Japan.

[4]The overall importance of Japanese industrial policy has been subject to an intense debate. Some scholars argue that MITI played a crucial role for Japan's rapid economic growth since the mid-1950s (e.g., Johnson, 1982). Conversely, others highlight the contribution of individual entrepreneurs and contend that Japanese firms became successful not because of but *despite* government interventions (e.g., Johnstone, 1999).

effect of the Japanese industrial policy, my estimates suggest that Japan may have benefited from protecting its nascent industry, as this allowed Japanese firms to build on state-of-the-art American technology while being shielded from foreign competition. Thus, the chapter also relates to a small but growing literature on infant industry protection (Juhász, 2018; Liu, 2020; Lane, 2021) and industrial policy more broadly (see, e.g., the comprehensive review by Juhász et al., 2023). In particular, my results are consistent with Liu (2020) who finds that industrial development requires foreign technology as a complement to trade protection.

The remainder of the chapter is structured as follows. Section 2.2 explains the historical background on RCA, its licensing policy, and the antitrust case. Section 2.3 introduces the data. In section 2.4, I estimate the aggregate effect of RCA's technology transfer on Japanese innovation and show the corresponding results. Section 2.5 then presents a complementary analysis that studies the direct effect of receiving a patent license from RCA on follow-on innovation by the licensees. Finally, section 2.6 concludes.

2.2 Historical Background

This section introduces the historical background and explains how US antitrust enforcement in the 1950s induced a large wave of technology transfer to Japan.

2.2.1 RCA, Patent Licensing, and the Antitrust Case

The Radio Corporation of America (RCA) was founded in 1919 as a subsidiary of General Electric (GE). Its primary objective was to conduct research and development in radio communication technology, which had primarily been used for military purposes until then. RCA initially acted as a patent trust that pooled the radio-related patents of GE and other major American companies in exchange for a share in equity. In 1932, RCA turned into an independent company.[5]

RCA quickly became the leading innovator in the field of radio and television technology. In the 1920s, RCA was at the forefront of the rapid advancement of radio broadcasting and receiving. It set up the first nationwide American radio network, the National Broadcasting Company (NBC). Later, RCA also played a key role in the development of colour television by creating a technology that allowed colour broadcasts to be received on existing black-and-white televisions. In 1953, RCA's colour technology was adopted as the US industry standard.

[5]More details on the origins and history of RCA can be found in Graham (1986), Bilby (1986), and Shih and Dieterich (2014).

Ever since RCA's founding, patent licensing was one of the company's major revenue streams. For example, between 1952 and 1956, RCA received a total of around $1 billion (in 2023 dollars) in royalties, representing more than three quarters of all industry royalty payments (Levy, 1981). Although RCA also generated revenues from selling radio and television receivers and components, the licensing income was crucial to fund RCA's research and development activity at its inhouse laboratories (Graham, 1986). Licensing was such a lucrative source of income for RCA for two reasons. First, RCA had filed more than 10,000 patents on radio and television technology. Second, RCA only licensed its patents in packages, hence making any licensee pay for an entire bundle of RCA patents. The royalties were computed as a percentage of the final product's selling price, regardless of how many patents the licensee actually needed and for what part of the product they were used (US v. RCA, 1954).

Because of this package licensing practice, the US Department of Justice (DoJ) filed a civil antitrust complaint against RCA in 1954. The DoJ charged RCA with monopolizing the market for radio and television technology. It argued that RCA's package licensing practice restrained research investments and production by other radio and television manufacturers (US v. RCA, 1954). In early 1958, the DoJ additionally filed a criminal indictment against RCA for violating Sections 1 and 2 of the Sherman Act (Levy, 1981).

Both antitrust cases were settled by a consent decree in 1958. As the key remedy, RCA was required to make all its radio-purpose (including television) patents available to domestic firms free of charge.[6] Consequently, the settlement ended RCA's package licensing practice (US v. RCA, 1958). In addition, RCA pleaded 'no defence' and accepted a fine to settle the criminal charges.

2.2.2 Technology Transfer to Japan

The 1958 antitrust settlement fundamentally changed RCA's licensing policy and ultimately induced a large wave of technology transfer to Japan. The consent decree eliminated the majority of RCA's licensing income. However, it did not limit the possibility to earn royalties from foreign firms. Therefore, RCA increasingly licensed its technology to firms in Japan (e.g., Johnstone, 1999; Shih and Dieterich, 2014). From 1960 to 1968, RCA entered into licensing agreements with a total of 169 Japanese firms (US Congress, 1970). According to Curtis (1994, p. 107), these agreements 'more than replaced the threatened licensing fees extracted from American manufacturers.'

[6] The consent decree established an exception for 100 colour television patents, which had to be placed in a patent pool. RCA had to license these patents royalty-free to the other members of the pool, whereas it could charge reasonable royalties to non-members (US v. RCA, 1958).

Japan presented a lucrative opportunity for RCA's patent licensing business. Despite the economic challenges faced by Japan in the aftermath of World War II, the country had a substantial industrial foundation with unused potential. Consequently, Japanese demand for American technology was high and the government promoted the acquisition of foreign technology. Moreover, RCA did not compete with its licensees in the Japanese product market and seemed unconcerned about potential future import competition in the US (Johnstone, 1999). There is some narrative evidence that RCA also licensed its patents to firms in other foreign countries – such as Philips in the Netherlands (Chandler, 2005). Yet, historians agree that Japan became the most important source of foreign licensing income for RCA and estimate that RCA's licensing revenues in Japan amounted to several hundred million dollars (in 2023 dollars) per year (e.g., Bilby, 1986; Curtis, 1994; Johnstone, 1999; Chandler, 2005).[7] Jim Clingham, a former executive at RCA, said that licensing technology to Japanese firms became 'one of the best businesses RCA had' (Johnstone, 1999, p. 13).

It should be noted that RCA licensed its patents to a very limited number of Japanese firms already prior to the US antitrust settlement. Choi (2008) names four Japanese firms (Hitachi, Kobe Kogyo, Sony, and Toshiba) that had agreements with RCA before 1958. However, he also states that '[t]hen, after 1958, there was a flood of Japanese companies establishing relationships with RCA' (Choi, 2008, p. 108).

It is an open question why RCA did not license its patents to Japanese firms to the same extent prior to 1958. Strictly speaking, the consent decree did not increase RCA's potential benefits from entering into licensing agreements with Japanese firms. However, the ban on receiving royalties from domestic licensees may have increased the *relative* attractiveness of moving abroad. This may have played a role, for example, if RCA was capacity constrained in its licensing business, although this seems unlikely. Alternatively, it could be that RCA, being the quasi-monopolist in the radio and television market, initially did not exhaust all available opportunities to generate revenues.

In addition to licensing its patents, RCA also entered into technical aid agreements with several Japanese licensees. These agreements entailed on-site training and thus led to a transfer of know-how that went beyond the use of patents (Choi, 2008; Shih and Dieterich, 2014). RCA licensing executives frequently travelled to Japan and the company even stationed a small number of licensing people in Tokyo on a permanent basis. Japanese licensees also visited RCA's research facilities in the US (Johnstone, 1999). The Japanese gratefully acknowledged RCA's technology transfer and support

[7]Exact numbers on the size of RCA's licensing revenues are unavailable due to RCA's restricted reporting policy. According to former RCA researcher Richard Williams, the foreign licensing income 'was always insulated [...] so that nobody really knew' (Johnstone, 1999, p. 14).

for its emerging consumer electronics industry. When RCA's long-time CEO, David Sarnoff, travelled to Japan in 1960, he was honoured by the Japanese Emperor with the 'Order of the Rising Sun', the highest distinction ever given to a foreign businessman (e.g., Bilby, 1986; Johnstone, 1999).

RCA was not the only American company that transferred its technology to Japan. For example, General Electric and Western Electric, a subsidiary of Bell, also licensed their patents. Yet, there are two reasons why this chapter focusses on technology transfer by RCA. First, RCA was the most important licensor in terms of the number of agreements (US Congress, 1970). This is because RCA was the American company whose business, and especially its research activities, most crucially depended on royalty income. This made RCA seek foreign licensees very aggressively after the antitrust settlement. Second, the 1958 consent decree provides a unique shock that I can exploit to empirically estimate the impact of technology transfer on Japanese innovation.

2.3 Data

To empirically measure Japanese innovation, I use data on the annual number of patent applications in the US filed by applicants from Japan. I focus on Japanese patent applications at the United States Patent and Trademark Office (USPTO) for two reasons. First, in contrast to patents filed at the Japanese Patent Office, data on patent applications at the USPTO are consistently available throughout the relevant period. Second, patents filed by Japanese firms in the US likely cover the subset of Japanese inventions that had the highest commercial value and the potential for worldwide application.[8] An additional advantage of using patents as a measure of innovation is that they are assigned to hierarchical technology classes. This is useful for my main empirical approach, because it allows me to compare patenting across different technologies within the same field. Moreover, I can measure follow-on innovation to RCA's patents through patent citations (Jaffe and Trajtenberg, 1996), since patent applicants have to cite any prior art.[9]

[8]This argument is akin to that for using 'triadic' patents (i.e., patents filed at the USPTO, the European Patent Office, and the Japanese Patent Office) with more recent data (e.g., Aghion et al., 2016).

[9]Nevertheless, patents represent an imperfect and potentially noisy measure of innovation, since not all innovations are patentable and inventors may opt for secrecy as an alternative means of protection (e.g., Moser, 2012). Moreover, patent citations may not accurately measure follow-on innovation, as citations may have been added by the examiner even in the absence of any knowledge flow (Alcácer and Gittelman, 2006).

My main data source is the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. In addition, I use data from the 'HistPat' database (Petralia et al., 2016) as well as the 'patentCity' project by Bergeaud and Verluise (2022) to identify the applicants' country of origin, which is not yet consistently reported in PATSTAT before the late 1970s.[10]

In addition, I exploit a list of RCA's licensing agreements in Japan between 1960 and 1968 from US Congress hearings (US Congress, 1970). The list contains the name of the licensee, the year of the agreement, and a vague description of the licensed know-how. These data serve as the basis for the firm-level analysis in section 2.5, where I study whether Japanese firms changed their innovation activity in response to receiving a license from RCA. I match the data to PATSTAT based on the licensee's name. Most licensees never patented in the US. Therefore, among RCA's 169 Japanese licensees, I restrict attention to the 42 firms that filed at least five patents in the US between 1950 and 1980. While this is a selected subset of the licensees, the sampling restriction allows me to track these firms' patenting over time. Whenever RCA entered into licensing agreements with the same licensee in several years, I focus on the year of the first license.

2.4 Aggregate Effect on Japanese Innovation

In this section, I estimate the aggregate effect of RCA's technology transfer on Japanese innovation. Before moving on to the empirical analysis, it is useful to briefly discuss why, conceptually, giving patent licenses to Japanese firms may affect their subsequent innovation activity.

If the owner of a patented technology (i.e., RCA) grants another firm a patent license, this may create an incentive for the licensee to invest in research and development. The patent license entitles the licensee to sell a product that utilizes the patented technology. Consequently, it gives the licensee 'freedom to operate' and ensures that the licensee can reap the benefits of any potential follow-on invention (e.g., Gaessler et al., 2019). In contrast, in the absence of a licensing agreement, a product involving a follow-on invention can only be sold in the market if it does not infringe upon the original patent. Whether that will be the case is unclear ex ante, since innovation is a highly incremental and cumulative process. Therefore, patents have been shown to exert a blocking effect on follow-on innovation in the absence of licensing agreements (e.g., Williams, 2013; Galasso and Schankerman, 2015; Gaessler et al., 2019).

[10]These data sources are identical to those in chapter 1. To make each chapter self-contained, I report them here again.

Licensing agreements may also promote follow-on innovation by enhancing knowledge diffusion to licensees. In principle, any patent should disclose all relevant information related to the patented invention. However, several factors may have limited knowledge diffusion from RCA's patents to Japanese firms. First, access to US patent documents was far from universal even within the US (Furman et al., 2021), and it was likely even more restricted in Japan. Second, language barriers may have hindered knowledge flows based solely on patent texts (Hegde et al., 2023; Higham and Nagaoka, 2023). Licensing may have helped to overcome both these barriers to knowledge diffusion. What is more, as described above, RCA's licensing agreements partly also included technical aid, which may have transferred additional (tacit) knowledge to licensees (Johnstone, 1999).

2.4.1 Empirical Approach

I employ an empirical approach on the technology class level, following Moser and Voena (2012) and Moser et al. (2014).[11] Intuitively, the empirical strategy compares Japanese patenting across technologies with a differential likelihood of licensing by RCA. I construct a panel dataset that counts the annual number of Japanese patent applications in the US on the level of four-digit technology classes based on the Cooperative Patent Classification (CPC). The hierarchical classification system allows me to compare Japanese patenting across four-digit technology (sub)classes within the same three-digit class.

I use the following difference-in-differences (DiD) regression model to estimate the effect of RCA's technology transfer on cumulative innovation in Japan:

$$\text{asinh}(\text{Patents}_{c,s,t}) = \beta \cdot \text{Treat}_s \cdot \text{Post}_t + \pi_s + \lambda_{c,t} + \epsilon_{c,s,t}, \qquad (2.1)$$

where $\text{Patents}_{c,s,t}$ is the number of Japanese patent applications in the US in year t assigned to four-digit subclass s within three-digit class c. I employ the inverse hyperbolic sine (IHS) transformation to this outcome variable, which is denoted by $\text{asinh}(\bullet)$. The IHS transformation allows me to focus on percentage (rather than absolute) changes in Japanese patenting. This is useful because the number of Japanese patent applications in the US continuously increased since the late 1950s, irrespective of RCA's patent licensing. The IHS transformation is also preferable over other non-linear transformations (e.g., logs), since it allows for zeros in the outcome variable (Bellemare and Wichman, 2020). The main treatment variable is denoted by Treat_s.

[11] This section uses the same empirical strategy as part of chapter 1. Therefore, the description of the empirical approach and the exposition of the results closely follow that in chapter 1.

In the simplest specification, Treat_s takes on the value of one for subclasses with a least one unexpired RCA patent. As further discussed below, I also employ a continuous treatment measure to account for the different likelihood of licensing by RCA across subclasses, reflected by a different number of RCA patents. The variable Post_t is a dummy that equals one in years after 1958. The regression also includes subclass fixed effects (π_s) as well as year × class fixed effects ($\lambda_{c,t}$). This controls for time-invariant differences across subclasses and allows the classes to experience idiosyncratic shocks over time. As a consequence, the DiD estimate $\hat{\beta}$ is identified only from variation over time across subclasses *within* the same class. Standard errors are clustered accordingly at the three-digit class level.

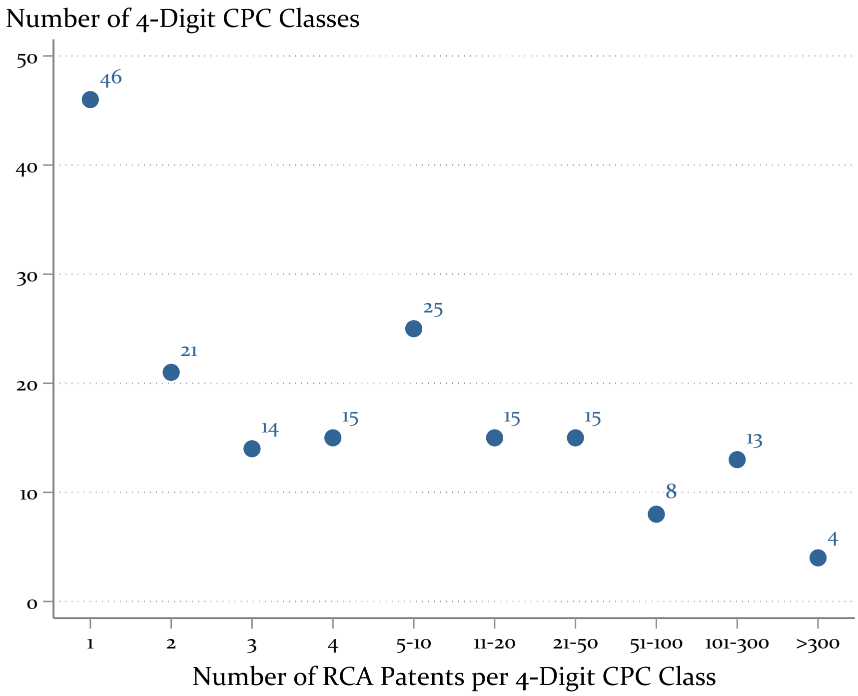
I estimate the DiD regression on a sample of 350 four-digit subclasses that belong to 63 three-digit classes. The dataset covers 16 years from 1953 until 1968. On average, there were 1.9 Japanese patent applications per year per subclass. 176 subclasses contain at least one RCA patent. Therefore, these subclasses are 'treated' to some degree. The remaining 174 subclasses are untreated, since RCA did not hold any patents in these technologies, which it could have licensed to Japanese firms. I use the weights by Iacus et al. (2012) to adjust for the potentially different number of untreated subclasses per treated subclass. Figure 2.1 shows how the number of RCA patents per subclass is distributed across subclasses. Among the 176 treated subclasses, 46 subclasses contain only one RCA patent, while four subclasses contain more than 300 RCA patents each. Overall, the classes in the sample contain a total of 5,776 RCA patents that were unexpired as of 1958.[12] Appendix B.1.1 presents additional summary statistics of the sample and more details on RCA's patent portfolio.

I employ two different treatment specifications to account for the heterogeneous number of RCA patents across subclasses, as shown in Figure 2.1. The simplest approach is to use a binary treatment variable with $\text{Treat}_s = \mathbb{1}[\text{RCA}_s > 0]$, where RCA_s is the number of unexpired RCA patents in subclass s. This specification (labelled as *binary* treatment) considers all subclasses with at least one RCA patent as treated, because these are the subclasses where RCA could potentially license its technology. An alternative approach is to assume that the likelihood of licensing increases in the number of patents that RCA holds in a given subclass. Therefore, in the second specification (labelled as *intensity* treatment), I use the number of unexpired RCA patents per subclass as a continuous treatment variable (i.e., $\text{Treat}_s = \text{RCA}_s$).[13]

[12] This sample is obtained after applying two restrictions. First, a four-digit subclass must have at least one Japanese patent application in the sample period. Second, every three-digit class must have at least one treated subclass.

[13] I discuss the advantages and drawbacks of each of these specifications below. In Appendix B.1.3, I also present robustness checks using alternative treatment definitions.

Figure 2.1. Aggregate Effect: Distribution of the Number of RCA Patents (RCA_s)



Notes: The figure depicts the distribution of the number of RCA patents per subclass (RCA_s) across treated subclasses. It shows the number of subclasses in the sample (on the vertical axis) that have a given number of RCA patents (on the horizontal axis). Example: there are 46 subclasses with one RCA patent.

My empirical strategy identifies the causal effect of RCA's technology transfer under the assumption that, within the same three-digit technology class, Japanese patenting in subclasses with no (or fewer) RCA patents provides a valid counterfactual for Japanese patenting in subclasses where RCA held (many) patents. In other words, the number of Japanese patent applications in subclasses with different numbers of RCA patents must have followed a common trend in the absence of patent licensing by RCA. The main concern with this identification strategy is that subclasses with (many) RCA patents may have been different from subclasses with no (or fewer) RCA patents in terms of unobserved characteristics. This may cause different trends in Japanese patenting over time. For instance, RCA may have chosen to patent in technology classes that had a higher likelihood of future innovation activity. To address this concern and assess the common trend assumption, I also estimate the following event-study variation of equation (2.1):

$$
\begin{aligned}
\operatorname{asinh}(\text{Patents}_{c,s,t}) = & \sum_{\tau=1953}^{1957} \delta_\tau \cdot \text{Treat}_s \cdot \mathbb{1}[\text{Year}_t = \tau] \\
& + \sum_{\tau=1959}^{1968} \beta_\tau \cdot \text{Treat}_s \cdot \mathbb{1}[\text{Year}_t = \tau] + \pi_s + \lambda_{c,t} + \epsilon_{c,s,t},
\end{aligned} \tag{2.2}
$$

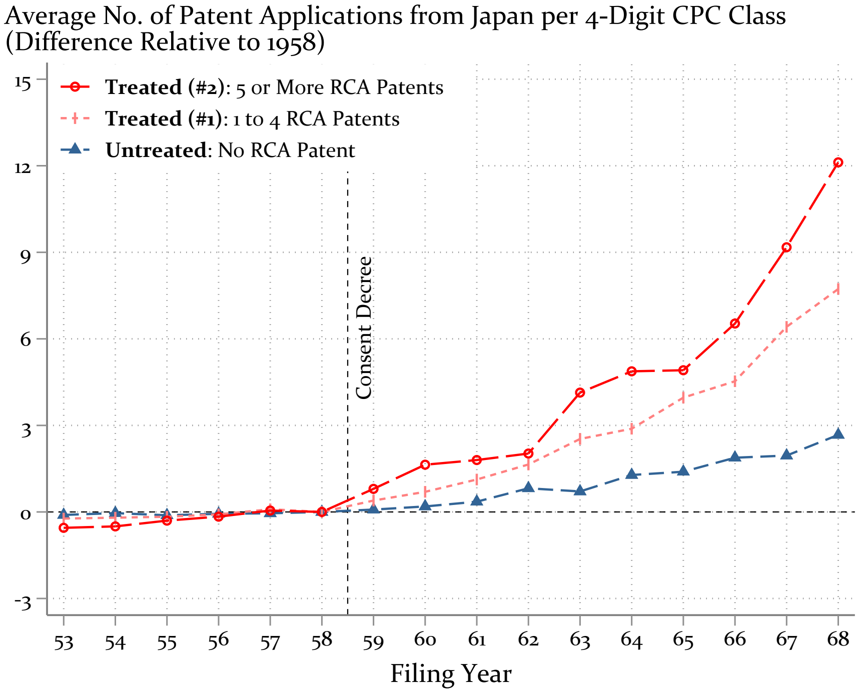
where the coefficients of interest are the lags (β_τ) that estimate the annual treatment effects in the ten years after 1958. In contrast, the leads (δ_τ) represent anticipatory effects in the five years prior to 1958 and should not be statistically different from zero.

2.4.2 Baseline Results

In Figure 2.2, I begin by descriptively showing the average yearly number of Japanese patent applications per four-digit technology subclass. The figure depicts patenting separately for three different groups: (treated) subclasses with five of more RCA patents, (treated) subclasses with one to four RCA patents, and the remaining (untreated) subclasses with no RCA patents. As is evident from the figure, treated subclasses experienced a relative increase in the number of Japanese patent applications after 1958. The increase is more pronounced in subclasses containing a larger number of RCA patents, which motivates using the intensity treatment in the regression analysis. In contrast, prior to 1958, patenting was relatively constant across all three groups. This is reassuring, although it does not directly speak to the identifying assumption underlying my empirical strategy. This is because the DiD model focuses on percentage (rather than absolute) changes in Japanese patenting.

I now investigate the effect of RCA's technology transfer on Japanese innovation more systematically by estimating the regression model in equation (2.1). Column (1) of Table 2.1 presents the baseline DiD estimate when using the binary treatment specification. Since the outcome variable is subject to an inverse hyperbolic sine (IHS) transformation, the point estimate cannot be interpreted quantitatively. Therefore, I follow Bellemare and Wichman (2020) to compute (semi-)elasticities, which allow me to derive the absolute and relative change in the overall number of Japanese patent applications. More details are provided in appendix B.1.2. As the table shows, Japanese applicants filed a total of 261 additional patents per year in subclasses with at least one RCA patent. This represents a relative increase in Japanese patenting by around 90%. The estimate is not only statistically significant at 1% level, but its magnitude is also economically large. Column (2) of Table 2.1 shows the corresponding DiD es-

Figure 2.2. Aggregate Effect: Patenting Trends Across Technology Classes



Notes: The figure depicts the average number of Japanese patent applications in the US per four-digit subclass relative to 1958. Averages are computed separately for treated and untreated subclasses, where a subclass is defined as treated if it contained at least one RCA patent. Treated subclasses are further divided into two subgroups, containing (#1) one to four and (#2) five or more RCA patents, respectively. The subclasses are aggregated using the weights by Iacus et al. (2012).

timate when using the intensity treatment. The implied semi-elasticity is 0.3%. This indicates that, on average, every additional RCA patent in a subclass is associated with an annual increase in Japanese patenting in that subclass by 0.3% after 1958. Taking into account the number of RCA patents in each subclass and aggregating this effect across subclasses, the semi-elasticity translates to 111 additional Japanese patents per year. This represents an aggregate relative increase in Japanese patenting by around 16%. Again, the estimate is highly statistically significant.

The estimates in Table 2.1 reveal that the number of Japanese patent applications increased disproportionately in RCA's technologies after 1958, regardless of the treatment definition. Yet, the magnitude of the estimated increase in Japanese patenting differs across the two specifications. It is much larger when using the binary treatment than when using the intensity treatment. Intuitively, the binary treatment considers all subclasses with at least one RCA patent as equally treated. Thus, it may capture changes in Japanese patenting in subclasses with very few patents, which are unlikely to result from RCA's patent licensing. In contrast, the intensity treatment uses more of the available information, as it also exploits variation in the number of RCA patents

Table 2.1. Aggregate Effect: Regression Estimates

	Binary	Intensity
	(1)	(2)
$\text{Treat}_s \cdot \text{Post}_t$	0.500***	0.003***
	(0.073)	(0.001)
Implied Semi-Elasticity	n/a	0.3%
Additional Patents per Year	261	111
Relative Increase	89.5%	16.3%
Mean of Outcome (w/o IHS)	1.7	1.9
No. of 4-Digit CPC Classes	307	350
No. of 3-Digit CPC Classes	54	63
Observations	4,912	5,600

Notes: The table shows the results from difference-in-differences regressions following equation (2.1). Column (1) uses the binary treatment with $\text{Treat}_s = \mathbb{1}[\text{RCA}_s > 0]$, where RCA_s is the number of unexpired RCA patents in subclass s. Column (2) uses the intensity treatment with $\text{Treat}_s = \text{RCA}_s$. All regressions include subclass and year × class fixed effects. Standard errors clustered at the three-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

across the treated subclasses. This specification gives a greater weight to changes in Japanese patenting in subclasses where RCA was a more active innovator and, hence, where the likelihood of licensing by RCA was higher.[14] In that sense, the intensity treatment is more conservative, which is reflected by a smaller change in the number of Japanese patent applications. However, the estimated increase in Japanese patenting by around 16% still represents an economically large and meaningful effect.

An additional drawback of the binary treatment specification is that classes where RCA held patents in all subclasses do not contribute to the identification of the treatment effect. Therefore, when using the binary treatment, I restrict the sample to contain only technology classes with both treated and untreated subclasses. This explains the different sample sizes across the two specifications in Table 2.1. As a consequence, the sample in column (1) does not include some of the three-digit CPC classes (e.g., H04) that contain part of RCA's key technology.[15] Conversely, identification of the

[14] The intensity treatment is also somewhat restrictive by assuming a linear relationship between the number of RCA patents in a subclass and the proportional change in Japanese patenting. In Table B.3 in the appendix, I run a robustness check that additionally includes the square of the number of RCA patents, hence allowing for a non-linear relationship. The resulting estimate is statistically significant and its magnitude lies in between those reported in Table 2.1.

[15] One potential solution to this problem is to define the binary treatment differently. For example, one could consider subclasses that contain a minimum of two RCA patents (as opposed to one) as treated. In Table B.3 in the appendix, I present estimates of a robustness check that uses several such alternative cut-offs to define the treatment. The overall result of a large and statistically significant increase in Japanese patenting in RCA's technologies remains unchanged.

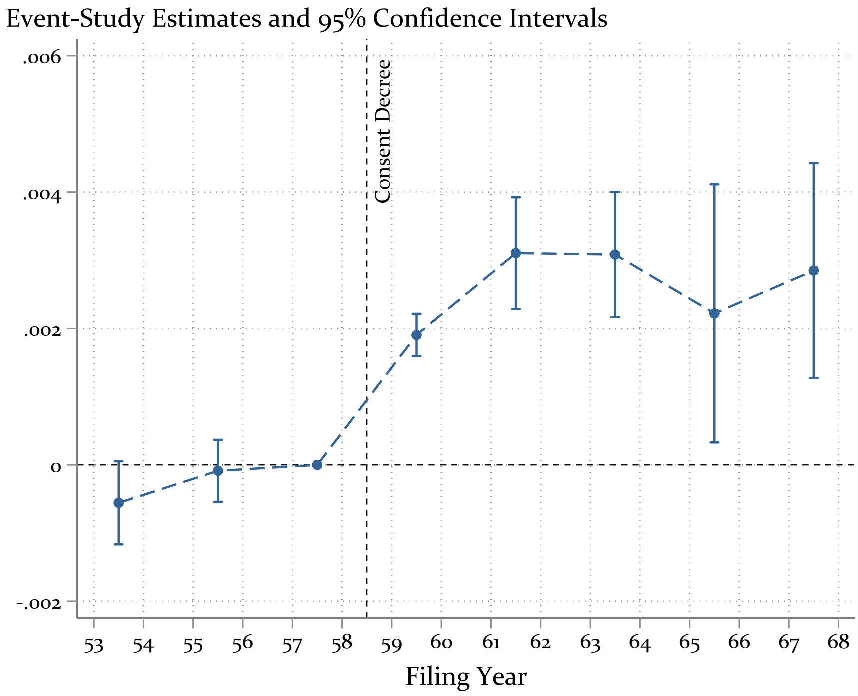
treatment effect in the intensity specification is possible whenever there is some variation in the number of RCA patents across the subclasses within a class.

Given these considerations, I henceforth focus on using the intensity treatment. Results from robustness checks with the binary treatment are reported in appendix B.1. They are qualitatively similar to the ones using the intensity treatment. In appendix B.1.3, I also employ an alternative continuous treatment variable that captures the share (as opposed to the number) of RCA patents per subclass, following chapter 1 (Mamrak, 2023). However, I argue that the absolute number of RCA patents better approximates the likelihood of licensing, because this likelihood should not directly be affected by the size of a technology class. In contrast, in chapter 1, the treatment variable does not aim to capture the likelihood of a formal licensing agreement. Instead, it approximates the propensity of a given technology class for containing copier-related patents, for which the overall number of patent applications in that technology class also matters. Finally, appendix B.1.4 presents additional robustness checks with alternative model and sample specifications. For example, I show that the magnitude of the baseline estimate is robust to rescaling the outcome variable, hence addressing recent concerns about IHS and other 'log-like' transformations (Chen and Roth, forthcoming). Overall, the various estimates indicate that the main finding of a disproportionate increase in Japanese patenting in RCA's main technologies is not driven by a particular specification of the model, treatment, or sample.

Figure 2.3 depicts the point estimates and 95% confidence intervals from the event-study analysis in equation (2.2). The figure shows graphically that there was a disproportionate increase in Japanese patenting in RCA's technologies after 1958. On average, subclasses with a higher number of RCA patents experienced a greater relative increase in the number of Japanese patent applications. In contrast, before 1958, the number of Japanese patent applications did not significantly differ across differentially treated subclasses. This supports the identifying assumption underlying my empirical strategy. Yet, the estimate for the years 1953/54 suggests that there may have been slightly diverging pre-trends across subclasses. Although these are not statistically significant at the 5% level, the common trend assumption may be violated. I will get back to discussing these potential pre-trends below.

In the next step, I study whether the additional Japanese patents filed after 1958 also were of high quality. To measure patent quality, I consider the number of forward citations that a patent has received. This captures how much a given patent is used by subsequent patents. In addition, I use the ten-year quality measure by Kelly et al. (2021, KPST), which is constructed using textual analysis of the patent documents. The KPST measure is based on the idea that high-quality patents should be novel

Figure 2.3. Aggregate Effect: Event-Study Estimates



Notes: The figure depicts point estimates and 95% confidence intervals from the event-study analysis in equation (2.2). Japanese patent applications are binned in two-year groups to reduce noise in the estimates. The figure uses the intensity treatment with $\text{Treat}_s = \text{RCA}_s$, where RCA_s is the number of unexpired RCA patents in subclass s. The regression uses the weights by Iacus et al. (2012). Standard errors are clustered at the three-digit CPC technology class level.

relative to prior art but must have a high impact on future inventions. Columns (2) and (3) of Table 2.2 presents DiD estimates from my main approach across technology classes, where the outcome variable is restricted to Japanese patent applications in the top 10% of the distribution of each of the quality measures. The estimates indicate that there was a relative increase in the number of high-quality Japanese patent applications in RCA's main technologies. Therefore, the increase in Japanese patenting found in the baseline is not driven by inventions of lower quality.

Finally, I turn to estimating the effect of RCA's technology transfer on the diversity of Japanese innovation. Following the approach by Watzinger and Schnitzer (2022), I measure the diversity of innovation by counting the number of 'active' technology subgroups within each subclass.[16] That is, the outcome variable in column (4) of Table 2.2 counts for each four-digit subclass how many technologies at the most disaggregate classification level contain at least one Japanese patent application in a given year. The

[16]In the hierarchical patent classification system, subgroups are the most disaggregate technology classification of a patent. Yet, subgroups have their own hierarchical order that is referred to as 'dot'-hierarchy. I follow Watzinger and Schnitzer (2022) and aggregate subgroups to the two-dot level.

Table 2.2. Aggregate Effect: Quality and Diversity of Innovation

		Patents in Top 10%		Diversity
	Baseline	Forward Citations	Quality (KPST)	Active Subgroups
	(1)	(2)	(3)	(4)
$\text{Treat}_s \cdot \text{Post}_t$	0.003***	0.001***	0.002**	0.003***
	(0.001)	(0.000)	(0.001)	(0.000)
Implied Semi-Elasticity	0.3%	0.3%	0.3%	0.3%
Additional Patents per Year	111	5	24	75
Relative Increase	16.3%	10.8%	22.3%	14.1%
Mean of Outcome (w/o IHS)	1.9	0.1	0.3	1.5
No. of 4-Digit CPC Classes	350	350	350	350
No. of 3-Digit CPC Classes	63	63	63	63
Observations	5,600	5,600	5,600	5,600

Notes: The table shows the results from difference-in-differences regressions following variations of equation (2.1). Column (1) repeats the baseline estimate from column (2) of Table 2.1. In columns (2) and (3), the outcome variable is restricted to Japanese patent applications in the top 10% of the distribution of forward citations and the quality measure by Kelly et al. (2021, KPST), respectively. In column (6), the outcome variable counts the number of 'active' technology subgroups within each four-digit subclass. Therefore, in that column, the number in the row 'Additional Patents per Year' refers to the number of additional subgroups (as opposed to patents). The table uses the intensity treatment with $\text{Treat}_s = \text{RCA}_s$, where RCA_s is the number of unexpired RCA patents in subclass s. All regressions include subclass and year × class fixed effects. Standard errors are clustered at the three-digit CPC technology class level. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

positive and statistically significant estimate indicates that Japanese innovation also became more diverse, as patenting activity expanded to new technologies.

Overall, the empirical evidence shows that there was a disproportionate increase in Japanese patenting in technologies with a higher likelihood of licensing by RCA. This increase in the intensity of Japanese innovation does not reflect inventions of lower quality. Instead, the number of high-quality Japanese patents also increased in RCA's main fields and Japanese innovation became more diverse.

2.4.3 Japan vs. Germany

Based on the results presented so far, one potential concern is that the increase in Japanese patenting may reflect technology-specific trends that could be unrelated to patent licensing by RCA. For example, RCA may have patented in technologies where innovation activity would have expanded even in the absence of any technology transfer. In that case, one should observe a similar increase in the number of patent applications in RCA's main fields by applicants from other countries.

Therefore, I investigate whether Japanese patenting increased disproportionately in RCA's technologies also relative to patenting by German applicants. Germany represents a suitable comparison country for several reasons. Like Japan, Germany was recovering from the destruction caused by World War II and had a solid industrial base. Some German firms (e.g., Telefunken) were manufacturing radio and television receivers, too. Yet, to the best of my knowledge, there is no evidence that RCA licensed its patents to any German firms. Consequently, comparing patenting by Japanese and German applicants should isolate the part of the increase in Japanese innovation that is due to RCA's technology transfer.[17]

I extend my empirical set-up in equation (2.1) to the following triple difference-in-differences model:

$$\begin{aligned} \text{asinh}(\text{Patents}_{i,c,s,t}) = {} & \alpha \cdot \text{Treat}_s \cdot \text{Post}_t \\ & + \beta \cdot \text{Treat}_s \cdot \text{Post}_t \cdot \text{Japan}_i + \pi_{i,s} + \lambda_{i,c,t} + \epsilon_{i,c,s,t}, \end{aligned} \tag{2.3}$$

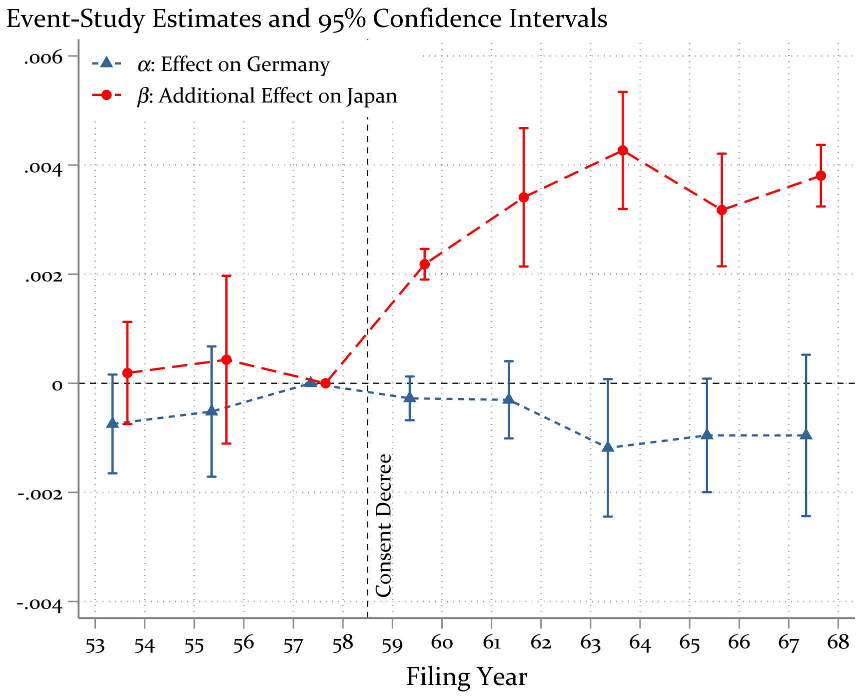
where the additional i subscript indexes the applicant country. The outcome variable $\text{Patents}_{i,c,s,t}$ counts the number of patent application in the US in subclass s, class c, and year t that is filed by applicants from either Japan or Germany. The variable Japan_i is a dummy that equals one when the applicant country is Japan. Consequently, the estimate $\hat{\alpha}$ captures how German patenting changed in RCA's main technologies after 1958. This estimate for Germany (i.e., when the triple interaction is zero) is completely analogous to that for Japan in the baseline DiD model in equation (2.1). In contrast, the triple DiD estimate $\hat{\beta}$ identifies the *additional* change in the number of patent applications in RCA's fields by Japanese applicants, relative to German applicants. The regression now includes subclass × country ($\pi_{i,s}$) as well as year × class × country ($\lambda_{i,c,t}$) fixed effects. The first type of fixed effects allows every subclass to have a different patenting level in each country, as long as this level does not change over time. The second type of fixed effects allows the four-digit classes to experience distinct idiosyncratic shocks in each country over time, as long as these shocks affect all subclasses within a class equally.

As for my baseline approach, I also estimate an event-study variation of the triple DiD model in equation (2.3). The event-study estimates are depicted in Figure 2.4. The figure shows that German patenting did not change differentially in RCA's main technologies after 1958. If anything, German applicants filed fewer patents in these fields. In contrast, Japanese patenting increased disproportionately after 1958 in technologies where RCA held more patents, even relative to German patenting.

Figure 2.4 also highlights that, relative to Germany, Japanese patenting did not follow any different trend before 1958 across technologies with different numbers of RCA

[17] I cannot exclude with certainty that RCA entered into licensing agreements with German firms. However, if German patenting also increased in response to licensing by RCA, my estimates would represent a lower bound of the true impact of RCA's technology transfer.

Figure 2.4. Aggregate Effect: Triple DiD



Notes: The figure depicts point estimates and 95% confidence intervals from an event-study variation of the triple DiD model in equation (2.3). Patent applications are binned in two-year groups to reduce noise in the estimates. The figure uses the intensity treatment with $\text{Treat}_s = \text{RCA}_s$, where RCA_s is the number of unexpired RCA patents in subclass s. The regression uses the weights by Iacus et al. (2012). Standard errors are clustered at the three-digit CPC technology class level.

patents. This can be seen by looking at the pre-1958 estimates for the additional effect on Japan, which are close to zero and statistically insignificant. Therefore, the slightly diverging pre-trends in the baseline DiD approach in Figure 2.3 likely reflect a development that also affected countries other than Japan. This supports the identifying assumption underlying my empirical strategy.

The point estimates for the triple DiD analysis are shown in column (2) of Table B.5 in the appendix. Consistent with Figure 2.4, the estimate of the effect on Germany ($\hat{\alpha}$) is very small in magnitude and statistically insignificant, whereas the estimate of the additional effect on Japan ($\hat{\beta}$) is positive and statistically significant at the 1% level. The magnitude of the triple DiD estimate implies that Japanese applicants filed 118 additional patents per year after 1958, which corresponds to a relative increase by 18%. Note that these numbers are very close in magnitude to my baseline DiD estimate that does not include Germany as a comparison country.

These results show that the increase in Japanese patenting cannot be explained by more general patenting trends across technologies. As I show in appendix B.1.5, this conclusion remains unchanged when I use all other non-US countries as a comparison category instead of Germany. Therefore, the observed increase in the number of Japanese patent applications in RCA's technologies represents an effect that is specific to Japan. This corroborates the interpretation that my estimates indeed capture the impact of RCA's technology transfer on Japanese innovation.

2.4.4 Discussion of Potential Concerns and Mechanisms

Could it be that Japanese innovation may have increased in RCA's fields even in the absence of any technology transfer? Of course, I cannot completely rule out the possibility that Japan experienced a positive innovation shock in specific technologies just after 1958. However, two aspects make it highly unlikely that my estimates do *not* reflect the impact of technology transfer. First, it seems extremely improbable that a shock specific to Japan would be confined only to those technologies where RCA held many patents. This is especially true in light of the results of my triple DiD analysis, which show that firms in other countries did not increase their patenting in RCA's main fields. Second, if the rise in Japanese innovation were unrelated to RCA, this raises the question as to why Japanese firms would have paid substantial licensing fees to RCA over many years.

It is also worth discussing the role of Japanese industrial policy in generating demand for foreign technology. The Japanese government actively supported a variety of industries, including consumer electronics. Government interventions involved granting generous subsidies and low-interest loans, as well as blocking access to the domestic market to foreign firms. Importing foreign technology also played an important role. For instance, the Japanese Ministry of International Trade and Industry (MITI) controlled all technology flows into the country and even promoted the licensing of foreign technology (Choi, 2008).

This raises the question as to whether Japanese demand for American technology could be a threat for the validity of my main empirical analysis. Note that, by default, a patent licensing agreement is an equilibrium outcome based on supply of and demand for technology. In the main empirical approach, I exploit a sudden change in the supply of technology, induced by the 1958 antitrust settlement with RCA. Even if Japanese demand for technology changed simultaneously and led to additional licensing agreements with RCA, this would only represent a concern for the empirical strategy if it had led to a disproportionate increase in Japanese innovation in RCA's main fields also *absent* any technology transfer. That is, there must have been a positive innovation shock in the counterfactual scenario that was confined to RCA's technologies, only affected Japan, and kicked in just after 1958. As discussed above, this is very implausible but cannot be ruled out. In any case, Japanese industrial policy is unlikely to have fostered such a shock, since MITI supported many different industries beyond consumer electronics, ranging from semiconductors to automobiles to steel.

In the remainder of this chapter, I present several additional pieces of evidence to illustrate that the most likely mechanism underlying the increase in Japanese patenting is technology transfer by RCA. First, in section 2.4.5, I disaggregate the overall increase in Japanese patenting, as estimated in my main empirical approach, by the additional patents' distance to RCA and by the applicant's licensee status. If technology transfer is the key mechanism, then one would expect the additional Japanese patents filed after 1958 to build on RCA's patents. One would also expect that those Japanese firms that entered into licensing agreements with RCA account for a large fraction of the increase in Japanese patenting. Second, in section 2.5, I turn to a complementary empirical approach on the firm level to analyse the direct effect of receiving a patent license from RCA on follow-on innovation by the licensees. If technology transfer explains the aggregate increase in Japanese patenting, then one should also observe a surge in the licensees' patenting, relative to other Japanese firms that did not receive a patent license.

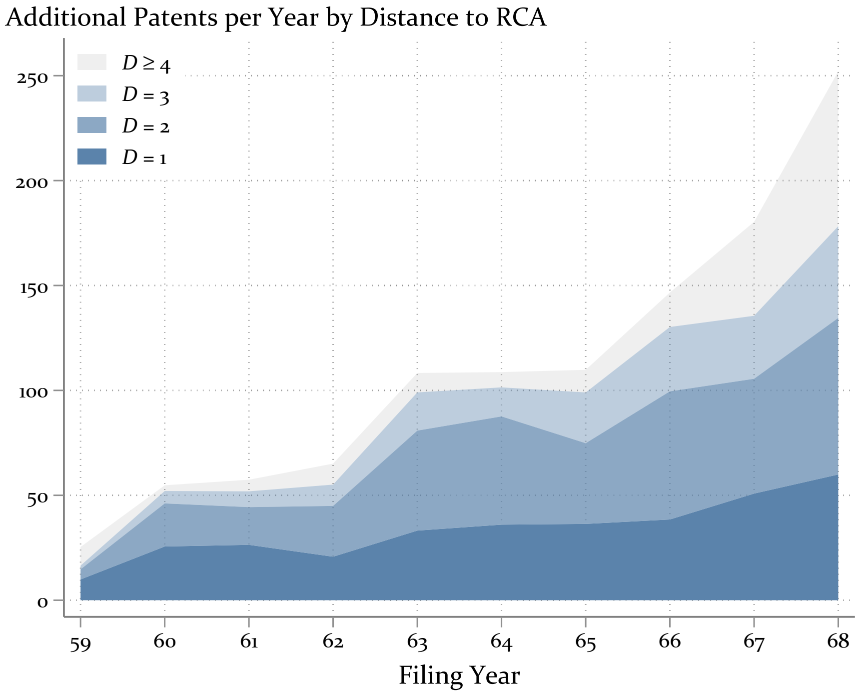
2.4.5 Disaggregation of Additional Japanese Patents Filed After 1958

I now disaggregate the additional Japanese patent filed in each year after 1958 by their distance to RCA and by the applicant's licensee status. Specifically, I estimate variations of the event-study model in equation (2.2), where the outcome variable is restricted to different subsets of Japanese patent applications. For each outcome, I then compute the number of additional patents per year implied by the regression estimates. The corresponding DiD regression estimates are presented in appendix B.1.6.

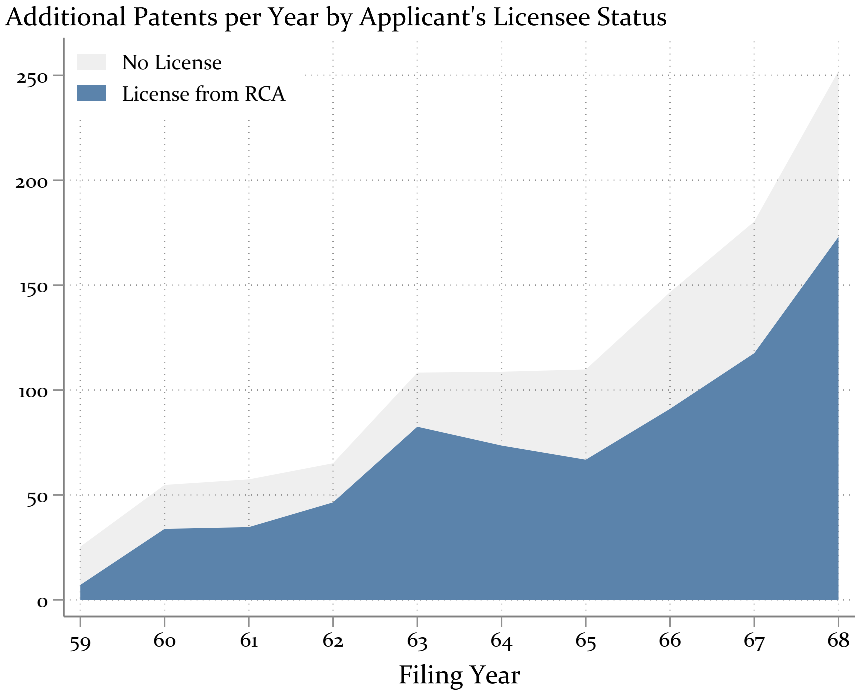
Panel (A) of Figure 2.5 shows that the majority of the additional Japanese patents filed after 1958 either directly or indirectly built on RCA's technology. To this end, I distinguish patents by their citation distance to RCA, following the framework by Ahmadpoor and Jones (2017). Patents that directly cite an RCA patent are defined to have a distance = 1 to RCA, while patents with distance = 2 (= 3) do not directly cite RCA but cite a patent with distance = 1 (= 2). As can be seen in the figure, Japanese patents that directly cited RCA (i.e., with distance = 1) account for around half of the increase in Japanese patenting in the first few years after the consent decree. Then, over time, the share of additional patents that is connected to RCA through indirect citations (i.e., with distance = 2 or distance = 3) increases. This pattern reflects the cumulative nature of innovation. Finally, and reassuringly, patents unconnected to

Figure 2.5. Aggregate Effect: Disaggregation of Additional Japanese Patents

(A) Distance to RCA



(B) Licensees vs. Other Firms



Notes: The figure depicts a disaggregation of the annual number of additional Japanese patents, as implied by the estimates from running variations of the event-study regression in equation (2.2). In panel (A), Japanese patents are disaggregated by their citation distance to RCA, following the framework by Ahmadpoor and Jones (2017). In panel (B), Japanese patents are disaggregated by whether the applicant received a patent license from RCA. Both panels use the intensity treatment with $\text{Treat}_s = \text{RCA}_s$, where RCA_s is the number of unexpired RCA patents in subclass s. Further details on how I compute the additional number of patents per year are provided in appendix B.1.2.

RCA (i.e., with distance ≥ 4) only represent a negligible fraction of the increase in Japanese patenting.

In panel (B) of Figure 2.5, I split the additional Japanese patents by whether the applicant received a license from RCA, using the information from US Congress (1970). The figure shows that around two thirds of the increase in Japanese patenting is driven by RCA's licensees. The remainder of the additional patents comes from Japanese firms that did not license any RCA patents. This may reflect spillover effects to other Japanese firms. For example, suppliers of licensees may have benefited from RCA's patent licensing even if they did not receive a license themselves.

Overall, the disaggregation exercise further supports the conclusion that RCA's technology transfer promoted Japanese innovation after 1958. The results indicate that the increase in Japanese patenting in RCA's main fields is driven by patents that built on RCA through forward citations. Moreover, patent applications by RCA's licensees account for the majority of the additional Japanese patents. These results are consistent with technology transfer as the main mechanism explaining the increase in Japanese patenting.

2.5 Direct Effect on Follow-on Innovation by Licensees

In this section, I estimate the direct effect of receiving a patent license from RCA on follow-on innovation by the licensees. This approach on the firm level complements the main empirical strategy on the technology class level. It serves to further explore the mechanism underlying the aggregate increase in Japanese patenting, as documented in the previous section.

2.5.1 Empirical Approach

Estimating how the Japanese licensees changed their innovation activity in response to receiving a license from RCA requires a suitable control group. Naturally, the set of Japanese firms that entered into licensing agreements with RCA was not randomly selected. This makes it challenging to find a good counterfactual, indicating how the licensees' patenting would have evolved without receiving a license from RCA.

I use the synthetic control method to obtain a control group for every Japanese licensee. This statistical method, proposed by Abadie et al. (2010, 2015), constructs a 'synthetic' control group for a treated unit by using a weighted combination of untreated units. The control group is built through a data-driven procedure based on the units' pre-treatment characteristics. The researcher only needs to specify the characteristics to be matched. Here, I match on the yearly number of patent applications in the six years until receiving the (first) license from RCA. The donor pool includes all other Japanese firms that did not receive any license.

The resulting sample consists of the 42 Japanese firms that received a license from RCA at some point between 1960 and 1968, along with their 42 synthetic control firms.[18] I then construct a balanced panel on the firm level that counts the annual number of patent applications in the US for each firm, covering six years before and after the licensing agreement.

I run the following simple DiD regression to estimate the impact of receiving a license from RCA on the licensee's patenting:

$$\text{Patents}_{i,t} = \beta \cdot \text{Treat}_i \cdot \text{Post}_t + \pi_i + \text{Year FE}_t + \epsilon_{i,t}, \tag{2.4}$$

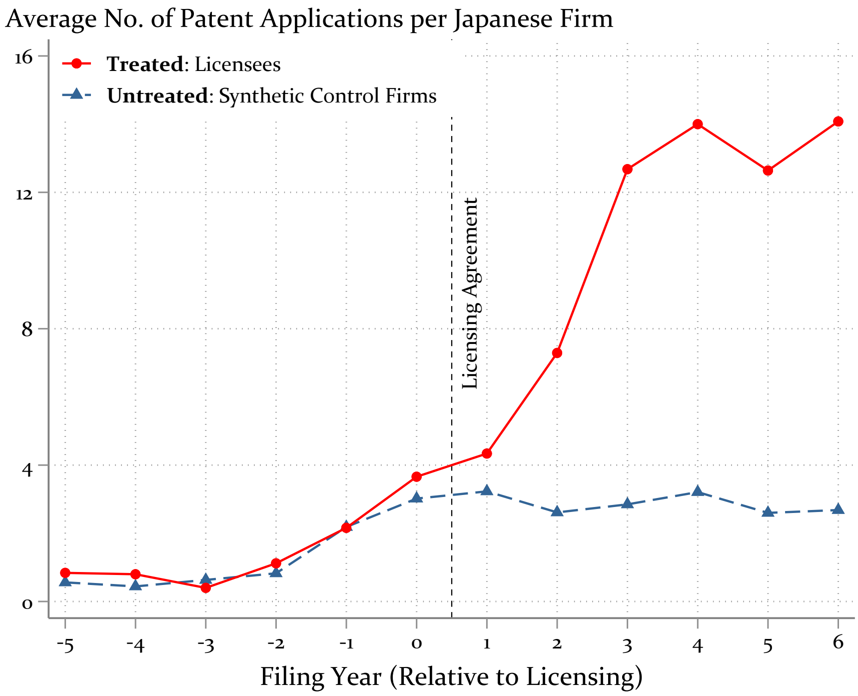
where i indexes a firm and t indexes the year *relative* to the licensing agreement.[19] The variable $\text{Patents}_{i,t}$ represents the number of patent applications in the US filed by firm i in relative year t. Treat_i is a dummy variable that equals one for firms that received a license from RCA and zero for the synthetic control firms. The variable Post_t is a dummy that equals one in years after the licensing agreement (i.e., if $t > 0$). The regression also includes firm fixed effects (π_i) as well as absolute year fixed effects (Year FE_t). This controls for time-invariant differences in patenting across firms and allows for year-specific shocks that affect all firms equally. Standard errors are clustered at the firm level to allow a firm's patenting to be correlated over time.

The identifying assumption underlying this empirical strategy is that Japanese firms that received a license from RCA would have filed the same number of patents as their synthetic control firms in the absence of any licensing agreement. The synthetic control method ensures that the licensees and their control firms had similar patenting trends prior to licensing. Yet, one potential concern is that the two groups may have been different in terms of unobserved characteristics. Consequently, the licensees may have patented more (or less) in subsequent years even without receiving a license from RCA. Given that the treatment is based on the firms' endogenous decision whether or not to license RCA's technology, it is unlikely that the synthetic control method fully controls for the selection into treatment. Therefore, I consider the present analysis primarily as descriptive and the magnitude of the estimates should be interpreted with caution.

[18] Both treated and control firms are restricted to the set of Japanese firms that filed at least five patents in the US between 1950 and 1980, as discussed in section 2.3.

[19] That is, $t = 0$ denotes the year of the agreement. For the synthetic control firms, t is relative to the year in which the corresponding treated firm received a license from RCA.

Figure 2.6. Direct Effect: Patenting by Licensees vs. Control Firms



Notes: The figure depicts the average number of patent applications in the US separately for RCA's Japanese licensees (in red) and their synthetic control firms (in blue) in the six years before and after the licensing agreement. The synthetic control groups for each licensee are computed using the algorithm by Abadie et al. (2010, 2015).

2.5.2 Results

Figure 2.6 depicts the average number of patent applications in the US filed by RCA's Japanese licensees and their synthetic control firms in the six years before and after the licensing agreement. By virtue of the algorithm underlying the synthetic control method, the average patenting of the control firms closely tracks that of the licensees prior to licensing. After receiving a license from RCA, however, the licensees disproportionately increased the number of patents that they filed. This is true both in absolute terms and relative to the synthetic control firms. The magnitude of the increase is substantial: three years after receiving an RCA licensee, the average Japanese licensee filed three times more patents than in the year of the licensing agreement. In contrast, patenting by the synthetic control firms stayed relatively constant over time.

The baseline DiD estimate from running the regression model in equation (2.4) is shown in column (1) of Table 2.3. On average, receiving a license from RCA was associated with an increase in the licensee's patenting by 4.3 patents per year. This estimate is statistically significant at the 5% level. Overall, it corresponds to around 179 additional patents per year. This magnitude implies that the licensees increased

Table 2.3. Direct Effect: Regression Estimates

	Baseline	Distance to RCA		Patents in Top 10%		Diversity
		$D \leq 3$	$D \geq 4$	Forward Citations	Quality (KPST)	Active Subgroups
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Treat}_i \cdot \text{Post}_t$	4.266**	3.388**	0.878	0.155**	0.953**	3.583**
	(1.844)	(1.324)	(0.608)	(0.077)	(0.433)	(1.543)
Additional Patents per Year	179.2	142.3	36.9	6.5	40.0	150.5
Relative Increase	174.1%	305.1%	65.6%	99.3%	194.9%	166.0%
Mean of Outcome	2.6	1.3	1.2	0.1	0.5	2.2
No. of Treated Firms	42	42	42	42	42	42
Clusters	84	84	84	84	84	84
Observations	1,008	1,008	1,008	1,008	1,008	1,008

Notes: The table shows the results from difference-in-differences regressions following variations of equation (2.4). In column (1), the outcome variable is the number of patent applications in the US per firm. In columns (2) and (3), the outcome variable is restricted to patent applications that are connected (with distance ≤ 3) and unconnected (with distance ≥ 4) to RCA through citations, respectively, following the distance framework by Ahmadpoor and Jones (2017). In columns (4) and (5), the outcome variable is restricted to patent applications in the top 10% of the distribution of forward citations and the quality measure by Kelly et al. (2021, KPST), respectively. In column (6), the outcome variable counts the number of 'active' technology subgroups (based on the Cooperative Patent Classification) per firm. Therefore, in that column, the number in the row 'Additional Patents per Year' refers to the number of additional subgroups (as opposed to patents). All regressions include firm and absolute year fixed effects. Standard errors clustered at the firm level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

their patenting by 174%, relative to the synthetic control firms. This number is in line with the large absolute rise in the licensees' patenting shown in Figure 2.6.

In columns (2) and (3) of Table 2.3, I investigate whether the increase in patenting among RCA's licensees truly reflects follow-on innovation that builds on RCA's technology. As in the class-level analysis, I distinguish patents by their citation distance to RCA, following the framework by Ahmadpoor and Jones (2017). In column (2), the outcome variable is restricted to patents that are either directly or indirectly connected to RCA (i.e., with distance ≤ 3). Compared to the baseline, the estimate shows that around 80% of the increase in the licensees' patenting comes from patents that are connected to RCA. In contrast, the estimate in column (3) suggests that there was no statistically significant change in the licensees' patenting when only taking into account patents unconnected to RCA (i.e., with distance ≥ 4). This is reassuring and supports the interpretation that the overall increase in the licensees' patenting primarily represents follow-on innovation that builds on RCA's technology.

The remainder of Table 2.3 moves beyond patenting intensity and instead focuses on the quality and diversity of innovation. As in the class-level analysis above, I measure patent quality through the number of forward citations that a patent has received and the ten-year quality measure by Kelly et al. (2021, KPST). In columns (4) and (5) of Table 2.3, the outcome variable is restricted to patents in the top 10% of the distribution of the two patent quality measures. The DiD estimates indicate that there was a statistically significant increase in the number of high-quality patents filed by Japanese firms after receiving a license from RCA. Interestingly, this effect is larger in magnitude for the KPST measure than for citations. Note that, in the firm sample, around 18% of the Japanese patents are among the top 10% of the (overall) distribution of the KPST measure, whereas only 6% are in the top 10% in terms of citations. This indicates that Japanese patents are over-represented among high-quality patents based on KPST's text-based measure, but they are under-represented in terms of citations. The latter could result from cultural barriers that may reduce citations to Japanese patents from non-Japanese applicants. This may also explain the different magnitude of the estimates in columns (4) and (5). In any case, the results suggest that the positive baseline effect is not driven by inventions of lower quality.

Finally, I analyse whether the licensees' patenting activity also expanded to new technologies. Again, I follow Watzinger and Schnitzer (2022) and measure the diversity of innovation by counting the yearly number of technology subgroups in which a firm filed at least one patent. I refer to this as the number of 'active' subgroups and use it as an outcome variable in column (6). The point estimate shows that, after receiving a license from RCA, the licensees' patents covered an additional 3.6 technology subgroups per year, compared to the synthetic control firms. That is, the Japanese licensees did not only increase their overall patenting, but their innovation activity also became more diverse.

As discussed above, these estimates can only be interpreted as the causal effect of RCA's licensing if the synthetic control firms represent a valid counterfactual. In appendix B.2.1, I present the results of a robustness check that uses an alternative control group. Specifically, I employ cell matching based on a firm's aggregate number of patent applications prior to the licensing agreement, the year of its first patent, and its primary technology field. The results are qualitatively similar, but the effects are smaller in magnitude. Relative to the alternative control group, the licensees filed 2.9 (instead of 4.3) additional patents per year after receiving a licensing from RCA. This translates to 120 (instead of 179) additional patents per year. One possible reason for the smaller effect size is that matching on a firm's primary technology field may control for different patenting trends across industries. Therefore, the robustness check again highlights that the magnitude of the estimates in this section should be interpreted with caution. Yet, the overall finding of a disproportionate increase in the licensees' patenting remains unchanged.

The increase in the licensees' patenting primarily occurred around two to three years after the licensing agreement, as can be seen in Figure 2.6. This represents a reasonable time frame for developing follow-on inventions that build on RCA's technology. Nevertheless, one concern with this sudden surge in the licensees' patenting is that it may not necessarily represent novel innovation. Instead, Japanese firms may have started to seek patent protection in the US for existing technologies, which they could not previously use in the product market without a license from RCA. If this were true, the newly filed patents should build on prior art that was older, hence exhibiting backward references with a higher average age. I show in Figure B.4 in the appendix that this was not the case. On the contrary, the mean and median age of backward citations was slightly lower among the licensees' patents in the years after licensing.

In summary, the firm-level analysis indicates that Japanese firms increased their patenting after receiving a license from RCA. This increase is driven by patents that either directly or indirectly cited RCA. The licensees also filed more high-quality patents and their innovation became more diverse. Although the exact magnitude of the effect is contingent upon the selection of an appropriate control group, the overall evidence suggests that RCA's patent licensing promoted follow-on innovation among its Japanese licensees. These findings are consistent with my analysis on the technology class level.

2.6 Conclusion

In this chapter, I estimate the impact of voluntary technology transfer from the US on subsequent innovation by Japanese firms. I leverage the 1958 settlement of a US antitrust case against RCA, which expanded RCA's patent licensing to Japan. I employ two complementary empirical strategies. In the main analysis in section 2.4, I estimate the aggregate effect of RCA's technology transfer on Japanese innovation, using a DiD approach across technology classes. The estimates indicate that there was a disproportionate increase in the number of Japanese patent applications in fields with a higher likelihood of licensing by RCA. I show that this effect is driven by Japanese inventions that built on RCA's technology and that there was no corresponding increase in patenting by other non-US countries. Finally, in section 2.5, I also use a firm-level approach to show that receiving a patent license from RCA was associated with a significant increase in follow-on innovation by the licensees.

A limitation of the chapter is that I cannot differentiate whether the increase in Japanese innovation was caused by RCA's patent licensing as opposed to additional technical aid. As outlined above, narrative evidence indicates that RCA did not only license its patents to Japanese firms; some of its agreements also entailed on-site training for the licensees (e.g., Shih and Dieterich, 2014). This may have transferred additional tacit know-how to Japanese firms. Unfortunately, there is no available information about the depth of each of RCA's licensing agreements, hence making it impossible to separately identify the effect of technical aid from that of licensing per se. In theory, both channels can be relevant. Nagler et al. (2022) document this empirically in the context of Bell's licensing of the transistor in the 1950s. They find that attendees of dedicated training symposia increased their follow-on innovation, but the overall effect was even larger for non-attendees.

Overall, the chapter shows that RCA's technology transfer promoted follow-on innovation by Japanese firms. This finding establishes technology transfer from the US as one important factor that contributed to the rise of Japan in consumer electronics in the 1960s and 1970s. Subsequent developments indicate that American firms – including RCA – ultimately faced intense import competition from Japan. This raises questions about the net benefits for the US from sharing its technology. Therefore, quantifying the impact of technology transfer to emerging markets on the originating country's industry remains an important task for future research.

Chapter 3

Imperfect Price Information, Market Power, and Tax Pass-Through

3.1 Introduction

Understanding how sellers pass through taxes is fundamental for the design of optimal tax policies. When firms have market power, pass-through can affect the impact of Pigouvian taxes, the effectiveness of unconventional fiscal policy, and the distributional consequences of commodity taxes. Competitive conduct is a key determinant of pass-through. Weyl and Fabinger (2013) present a unified theoretical framework to study pass-through under imperfect competition, where competition is captured by a conduct parameter. A key assumption of this framework is that consumers possess complete price information. No such framework exists when there is imperfect information.

In this chapter, we investigate how market power resulting from imperfect price information affects commodity tax pass-through. We show that imperfect consumer information about prices best explains the pass-through behaviour of firms observed in our empirical application. We study the retail fuel market, which represents an ideal setting to investigate the role of imperfect price information for two reasons. First, within a fuel type, products are homogeneous, which eliminates other sources of imperfect information. Second, retail fuel has a high degree of price transparency relative to that of other products. A finding of effects here would suggest that they are rampant across the economy.

This chapter is based on joint work with Felix Montag, Alina Sagimuldina, and Monika Schnitzer (Montag et al., 2023a).

Despite their high degree of price transparency, we show that retail fuel markets exhibit significant price dispersion that is random and unpredictable to consumers. We also show that some consumers use price comparison apps and are perfectly informed about prices whereas others do not and are not. Guided by these stylized facts, we modify the Stahl (1989) consumer search model to study tax pass-through. The model features some consumers who know all prices and others who have to search for prices sequentially. The degree of market power depends on the number of competitors and the share of well-informed consumers, as a larger share incentivizes firms to compete on prices.

Our theoretical analysis has key implications for the analysis of tax pass-through. First, we show that the share of well-informed consumers is positively related to consumers' average price sensitivity (i.e., the price elasticity of residual demand) and the pass-through rate. Second, since firms play mixed strategies, the pass-through to the price paid by well-informed consumers is higher than the pass-through to the price paid by uninformed consumers in the same market. Third, we find that, when market power is derived from imperfect information, the relationship between the number of competitors and the pass-through to the average price is non-monotonic. This suggests that the number of competitors can be a poor predictor of pass-through. Fourth, the full information conduct parameter approach (see Weyl and Fabinger, 2013) cannot nest models where market power derives from imperfect information. Its results on the determinants of pass-through therefore do not extend to settings with imperfect information.

To empirically test the model predictions, we analyse the pass-through of multiple recent tax changes in the German retail fuel market, using detailed price data at the station level. A unique aspect of the setting is that we can separately study fuel products that differ in how well their users are informed about prices. We find that tax pass-through is higher for fuel types with a larger share of well-informed consumers. We also show that pass-through to the minimum price is higher than pass-through to the average posted price for nearly all tax changes that we study. Finally, we find a non-monotonic relationship between the number of competitors in a local market and pass-through to the average price. Together, the different findings strongly suggest that the observed pass-through patterns are best explained by imperfect consumer information about prices.

These results are widely applicable beyond the retail fuel market, because markets with both well-informed and uninformed consumers are widespread across the economy. For instance, models of competition with imperfect information are used to explain price differences between online and brick-and-mortar stores (Baye et al., 2006).

Our chapter shows that, in such settings, understanding commodity tax pass-through requires a stronger emphasis on information, rather than the number of competitors.

Our findings are crucial to assess the impact of tax policy. The lower the share of well-informed consumers, the lower is the pass-through of commodity taxes. This can make Pigouvian taxes less effective, as there will be a smaller output response from consumers than in a setting with full information.[1] Since pass-through differs between well-informed and uninformed consumers, output reactions across consumer groups will be different, which has distributional implications. Therefore, Pigouvian taxes may induce stronger quantity reactions among well-informed than among uninformed consumers. Furthermore, if few consumers are well informed about prices, this lowers tax pass-through and limits the possibility of stimulating the economy by means of unconventional fiscal policy.

As in the canonical Stahl (1989) model, the equilibrium of our theoretical model is characterized by a distribution of prices, because firms set prices using mixed strategies. Well-informed consumers ('shoppers') always buy from the seller offering the lowest price. Uninformed consumers ('non-shoppers') do not search in equilibrium and instead pay the first price they draw. From an ex ante perspective, shoppers pay the expected minimum price, whereas non-shoppers pay the expected price. A key result that sets our findings apart from perfect information settings is that the relationship between the rate of pass-through to the expected price and the number of sellers is non-monotonic. This is because, above a certain threshold of competitors, it becomes increasingly unlikely for a particular firm to attract shoppers. Consequently, firms are more likely to charge a higher price and cater only to uninformed non-shoppers. Therefore, a larger number of sellers does not necessarily lead to lower average prices.

We focus on market power derived from imperfect price information to explain tax pass-through. Alternative models in the literature produce distinct predictions. We discuss these different predictions and contrast them with our empirical findings. Ultimately, we conclude that none of the alternative hypotheses can explain the empirical findings jointly as well as a consumer search model.

The assumption of perfectly competitive markets is widespread in the empirical literature on tax pass-through (e.g., Chetty, 2009; Chetty et al., 2009). In contrast, a growing theoretical literature considers how firms with market power pass through taxes (e.g., Sumner, 1981; Bulow and Pfleiderer, 1983; Stern, 1987; Hamilton, 1999), with Weyl and Fabinger (2013) providing a general model to capture the intensity of

[1] Although output can also be reduced with market power, Conlon and Rao (2023) demonstrate that limiting competition to address negative externalities results in much higher welfare costs than using taxes.

competition.[2] All of these models assume that consumers are fully informed about prices.

Some studies depart from the full information assumption. Many of these models assume that consumers are aware of posted net prices but that the tax component applied at checkout is less salient (e.g., Chetty et al., 2009; Kroft et al., 2024). Similarly, Busse et al. (2006) analyse differences in the pass-through by auto manufacturers of promotions that differ in how salient they are to consumers. Approaches that rely on differences in salience between the net price and taxes cannot explain the findings in our empirical application, where the gross price including taxes is posted.

We test the predictions empirically using multiple tax changes in the German retail fuel market. This industry is ideal because the fuel type that consumers purchase is highly correlated with their incentive to become informed about prices. In Germany, there is strong evidence suggesting that diesel drivers are better informed about prices than gasoline drivers.[3] Moreover, drivers fuelling with *E5* gasoline are less informed about prices than those fuelling with *E10*. We use search data from a price comparison smartphone app to confirm these hypotheses. Using differences between fuel types, we can test the predictions about the relationship between pass-through and the share of well-informed consumers. Since fuel stations sell all three fuel types, we can disentangle different components of market power. We can test how pass-through varies across consumer groups with different levels of information while we hold the station network fixed. Similarly, we can test how pass-through varies across stations with different numbers of competitors while we hold the consumer type fixed.

Our empirical analysis examines the impact of a temporary decrease in the value-added tax (VAT) and the introduction of a carbon emissions price in Germany in 2020/21. We estimate pass-through rates separately for each fuel type by comparing daily prices of stations in Germany with those in France, using a difference-in-differences (DiD) design. As a robustness check, we test a subset of these predictions exploiting additional tax changes in France.[4] We consider three French tax changes in 2022/23 to analyse the difference in pass-through to the minimum price and the average posted price in a market.

The first empirical finding is that tax pass-through is higher for fuel types with a higher share of well-informed consumers. The empirical literature on tax pass-through and market power (e.g., Miravete et al., 2018; Hollenbeck and Uetake, 2021; Nakamura

[2]Adachi and Fabinger (2022) generalize this to allow for richer governmental intervention, while Kroft et al. (2021) allow for free entry and love-of-variety preferences.

[3]Johnson (2002) makes a similar argument for why diesel drivers are more price sensitive.

[4]We cannot study differences in pass-through between fuel types in France, as all of the French tax changes occurred in 2022, where different fuel types were hit differently by global energy shocks.

and Zerom, 2010) has so far ignored imperfect information. Most closely related to our mechanism, Duso and Szücs (2017) find higher cost pass-through for electricity tariffs that consumers actively need to choose than for default tariffs. Similarly, Kosonen (2015) finds that Finnish hairdressers pass on VAT decreases more for advertised services. Our results also relate to Eizenberg et al. (2021), who find that spatial frictions and differences in the sensitivity to lower prices between different neighbourhoods lead to spatial differences in market power and price levels.

The second empirical finding is that pass-through to the minimum price, paid by well-informed consumers, is higher than pass-through to the average posted price, paid by uninformed consumers. This extends the literature on the distributional implications of tax pass-through. For example, Harju et al. (2022) find lower fuel tax pass-through in high-income areas. Conlon et al. (forthcoming) show that the sin tax burden is concentrated among few households that exhibit similar purchasing patterns. Our understanding of who searches is restricted to differences between consumers buying different fuel types. Byrne and Martin (2021) review the literature and document an inverse-U-shaped relationship between household income and consumer search.

The third empirical finding is a non-monotonic relationship between the number of sellers and the rate of pass-through to the expected price, or average price. The results in existing empirical literature are mixed, with evidence that tax pass-through increases (e.g., Genakos and Pagliero, 2022), decreases (e.g., Miller et al., 2017; Hindriks and Serse, 2019), or is uncorrelated (e.g., Kopczuk et al., 2016) with the number of competitors.

We find higher pass-through for tax increases than for tax decreases. However, as we never observe a symmetric tax increase and decrease of the same magnitude, this is not evidence of asymmetric tax pass-through beyond what can be explained by standard economic models. In contrast, Benzarti et al. (2020) find higher pass-through for tax increases than for decreases when tax changes are symmetric.[5] There are a number of studies estimating the pass-through of different tax changes, but none of them study its relation with market power.[6]

Finally, by emphasizing how imperfect consumer information affects tax pass-through in oligopolistic markets, we contribute to a literature on the supply-side determinants of tax pass-through in fuel markets (e.g., Marion and Muehlegger, 2011;

[5]Heim (2021) finds that, when search is high, pass-through is high for cost decreases and low for cost increases.

[6]These include single-industry studies (e.g., Fabra and Reguant, 2014; Li and Stock, 2019; Ganapati et al., 2020; Büttner and Madzharova, 2021; Dubois et al., 2020; Harding et al., 2012; Conlon and Rao, 2020) and cross-industry studies (e.g., Benedek et al., 2020).

Fischer et al., 2023).[7] An earlier literature studies fuel market cost pass-through using error correction models.[8] Relatedly, Borenstein et al. (1997) conclude that asymmetric pass-through may be explained by tacit collusion or imperfect information. Deltas and Polemis (2020) show that results from error correction models strongly depend on the research design and data features.

The remainder of the chapter is structured as follows. Section 3.2 describes the data and derives stylized facts about the fuel market. Section 3.3 outlines the theoretical model. Section 3.4 introduces the tax changes and provides descriptive evidence. Section 3.5 presents the empirical strategy. Section 3.6 discusses the estimation results. Section 3.7 contrasts the empirical results with alternative hypotheses and section 3.8 concludes.

3.2 Consumer Information in the Retail Fuel Market

We begin by describing the data and highlighting the key features of the retail fuel markets in Germany and France.

3.2.1 Data

Our comprehensive dataset includes real-time price changes for almost all fuel stations in Germany and France, along with various station characteristics. German stations are mandated to report price changes to the Market Transparency Unit at the Federal Cartel Office.[9] Similarly, in France, a government agency requires stations to report price changes, providing researchers access to these data.[10] We construct daily weighted average prices for each station, using the time of price changes. See Appendix C.1.1 for details on our dataset construction.

We analyse data from January 2019 to February 2023. We calculate summary statistics for 2019 to capture pre-intervention markets, as all tax changes occurred between 2020 and 2023. The top panel of Table 3.1 presents station-level summary statistics.

[7]Imperfect information is known to play an important role in pricing more generally in these markets (see, e.g., Chandra and Tappata, 2011; Byrne and de Roos, 2017; Luco, 2019; Pennerstorfer et al., 2020; Martin, forthcoming; Montag et al., 2023b).

[8]For a review of the literature, see Eckert (2013).

[9]Tankerkönig, a price comparison website, provides access to these data. See `https://dev.azure.com/tankerkoenig/_git/tankerkoenig-data` (last accessed: 19 February 2024).

[10]See `https://www.prix-carburants.gouv.fr/rubrique/opendata/` (last accessed: 19 February 2024). In France, fuel stations selling less than 500 m^3 of fuel per year are exempt from reporting price changes.

To analyse local price dispersion and competitive dynamics, we group fuel stations into non-overlapping local markets using a hierarchical clustering algorithm based on driving time, as in previous studies (e.g., Carranza et al., 2015; Lemus and Luco, 2021; Assad et al., 2020). The idea underlying this approach is to find clusters of stations that are naturally separated from each other. The details of our clustering method are explained in Appendix C.1.2. Table 3.1 shows that we assign the 14,648 German stations to 3,479 local markets with an average size of 4.2 stations. In France, there are 9,075 fuel stations assigned to 2,769 local markets. France has fewer markets and stations per market, a fact most likely related to its smaller population and lower population density.

An alternative way of defining local markets is to center markets on each station and include all competitors within a predefined radius around the centroid station. This approach has also been used in previous studies (e.g., Luco, 2019), but it has the drawback that it leads to overlapping markets, where some stations are assigned to multiple markets. We show that our results are robust to using this alternative market definition in Appendix C.4.6.

Ultimately, we are interested in the number of competitors in a local market. We measure the number of competitors by the number of competing price setters. That is, if there are two stations for which the same entity sets prices, we want to treat it as a single price setter. For Germany, we have two data sources that allow us to establish a common price setter across stations. First, the station dataset contains information on the brand of a fuel station. Prices at stations belonging to a brand of the vertically integrated fuel producers (e.g., Aral, Shell, etc.) are set centrally by the brand's headquarters, irrespective of whether the station is owned by the fuel producer or by a third-party owner-operator.[11] Moreover, some firms operate fuel stations under different brands or even without brands. 'Wer-zu-wem' is a database that contains ownership information for many such stations and allows us to group brands together with a common price setter (e.g., the brand Elan belonging to TotalEnergies). Ultimately, we compute the number of competing price setters in a local market. On average, there are 3.6 different price setters per market in Germany. Furthermore, 16% of markets contain only a single price setter and are thus monopoly markets. The median monopoly station is 10.6 minutes away from its closest competitor, whereas the me-

[11]We conducted several interviews with market participants. All our interviewees confirmed that prices are set at the headquarters level for both large integrated conglomerates and most small- and medium-sized station operators. See, e.g., for Shell: `https://support.shell.de/hc/de/articles/360010715077-Wer-bestimmt-die-Kraftstoffpreise-an-den-Shell-Stationen-` (last accessed: 19 February 2024).

Table 3.1. Summary Statistics

	Germany	France
Station Level		
(A) Station Characteristics		
No. of Stations	14,648	9,075
Median Driving Time to Closest Competitor	2.4min	n/a
Median Driving Distance to Closest Competitor	1.4km	n/a
(B) Prices, *E5*		
Mean Gross Price	1.41	1.53
Mean Price Net of Taxes and Duties	0.53	0.58
(C) Prices, *E10*		
Mean Gross Price	1.39	1.49
Mean Price Net of Taxes and Duties	0.51	0.57
(D) Prices, Diesel		
Mean Gross Price	1.25	1.45
Mean Price Net of Taxes and Duties	0.57	0.60
Market Level		
(E) Market Characteristics		
No. of Markets	3,479	2,769
Mean No. of Stations in Market	4.21	3.28
Mean No. of Competitors	3.60	n/a
(F) Local Monopolies		
Share of Monopoly Markets	16%	n/a
Median Driving Time to Closest Competitor	10.6min	n/a
Median Driving Distance to Closest Competitor	9.4km	n/a
(G) Prices, *E5*		
Mean Average Posted Price	1.42	1.53
Mean Minimum Price	1.38	1.51
(H) Prices, *E10*		
Mean Average Posted Price	1.40	1.48
Mean Minimum Price	1.35	1.46
(I) Prices, Diesel		
Mean Average Posted Price	1.25	1.45
Mean Minimum Price	1.21	1.43

Notes: The table shows summary statistics for 2019 (i.e., before all tax changes). The top panel presents data at the station level, whereas the bottom panel presents data at the market level. Non-overlapping markets are defined with a hierarchical clustering algorithm, as explained in Appendix C.1.2. A competitor is defined as a competing price setter (available only for Germany). All prices are in euro per liter.

dian driving time to the closest competitor across all stations is only 2.4 minutes. For France, our dataset does not contain information on fuel station brand or ownership.

Finally, we leverage data on search queries in 2015 from a major German smartphone app that enables users to compare fuel prices across stations. For each fuel price search, the dataset contains the searcher's location, a time stamp, a unique user ID, and the fuel type searched. This allows us to document intensive- and extensive-margin differences in search intensity between consumers who search for different fuel types.

3.2.2 Fuel Types

Diesel and gasoline are the two primary fuel types for passenger vehicles with combustion engines. In Germany, diesel accounts for 43% of the volume share and gasoline accounts for the remaining 57%.[12] Diesel and gasoline can be considered separate markets in the short term because of the high cost of substitution on both the demand and the supply side.

Gasoline can be classified according to its octane rating and ethanol share. Standard gasoline (called *Super*) has an octane rating of 95 and can be further distinguished by its ethanol share. Gasoline with a 5% share of ethanol is referred to as *E5*, while *E10* has a 10% ethanol share.[13] Although *E5* and *E10* are not taxed differently, *E10* is typically 4-6 eurocents cheaper in Germany because of a minimum biofuels quota. Since *E5* and *E10* have the same octane rating, there are no quality differences between the fuel types.

3.2.3 Price Dispersion

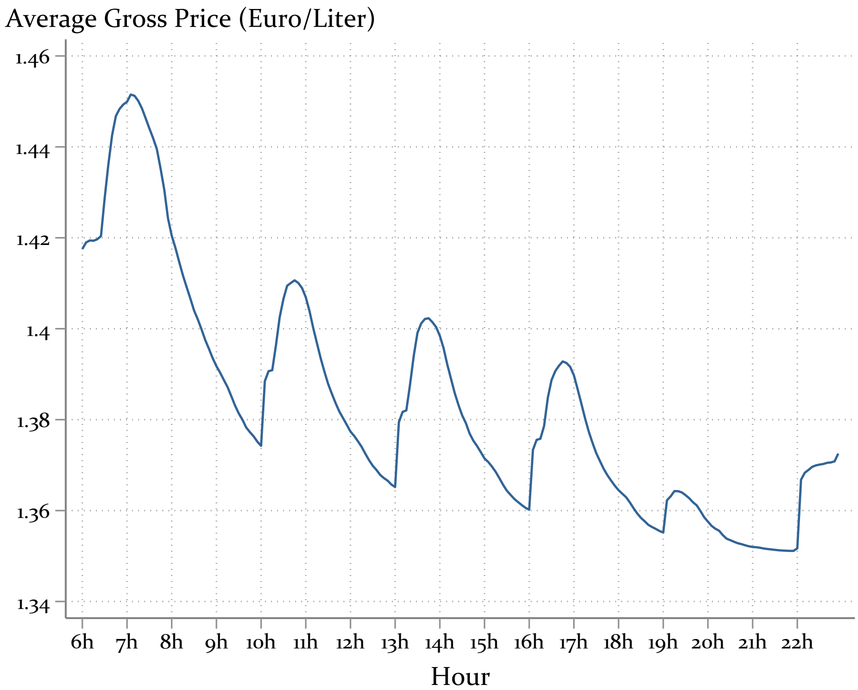
The summary statistics in Table 3.1 reveal substantial price dispersion within local markets on a particular day for fuel stations selling the same products. To understand the sources of this variation, we decompose it into components related to intertemporal differences in demand or product differentiation, as well as random and unpredictable price changes.

Figure 3.1 illustrates that the average price of *E10* in Germany at different times of the day varies significantly, with prices at approximately 7.30am being more than 10 eurocents higher than prices at approximately 10pm. On average, there are 14 daily price changes for *E10* at German fuel stations in 2019. As noted by Holt et al. (2023), these price cycles are different from those in other countries (e.g., Australia). They are not cost driven, as costs can be assumed not to vary within a day. Instead, the price cycles serve two purposes. First, high prices in the morning and lower prices in the evening are consistent with intertemporal price discrimination, where prices are high when drivers have little time to search for better prices. Second, frequent price changes

[12]Based on 2019 data from the German Ministry of Transportation's *Verkehr in Zahlen 2020/2021*. Truck diesel prices are not included, as they are not reported to the Market Transparency Unit.

[13]In addition, there are other types of *Super* with an octane rating of 98, but their market share in Germany is only approximately 5%.

Figure 3.1. Average Daily Price Cycles for *E10* in Germany, 2019



Notes: The figure shows average prices of *E10* in 2019 across fuel stations in Germany at different times of the day. Fuel prices are updated in five-minute intervals.

during the day make it difficult for drivers to learn which station is the cheapest at a particular point in time. This makes it more likely for stations to be able to sell at a price higher than the minimum price in the market.

Although the theoretical predictions in section 3.3 do not feature price cycles, these are not inconsistent with the model. Cason et al. (2020) show that, in experimental markets with features similar to those of the Burdett and Judd (1983) or Stahl (1989) search models, the observed pricing patterns feature price cycles. After we collapse the prices over time, the mixed strategy equilibrium distribution in the theoretical search models does well at explaining the observed price distribution.

To identify the random and unpredictable price dispersion for consumers, we narrow our focus to a particular time of day, 5pm, and calculate the absolute price deviation of fuel stations from the mean price in their local market for all non-monopolistic stations. We do this by regressing daily 5pm prices on market $\times$ date fixed effects and computing the resulting residuals. However, some stations may always deviate from the mean price in the same way because of product differentiation. They may, for example, be in a particularly attractive location or offer better amenities. To isolate the non-constant part of the deviation from the market mean, we further include station fixed

Table 3.2. Within-Market Price Residual at 5pm, 2019

	Stations	Markets		
	Mean Abs. Deviation	p25-p75	p10-p90	Range
(A) *E5*				
Market × Date FE	.0162	.0271	.0408	.0417
Market × Date FE and Station FE	.0104	.0177	.0272	.0279
(B) *E10*				
Market × Date FE	.0173	.0291	.0439	.0449
Market × Date FE and Station FE	.0105	.0181	.0275	.0281
(C) Diesel				
Market × Date FE	.0161	.0269	.0407	.0416
Market × Date FE and Station FE	.0103	.0175	.0272	.0278
No. of Stations/Markets	3,507,612	775,431	775,431	775,431
Observations	14,140	2,971	2,971	2,971

Notes: The table shows the distribution of the average absolute deviation of a fuel station's price from the average price in the same market on the same day at 5pm for each fuel type and for all stations that are not local monopolists. We use data for all weekdays in 2019. The table also shows the distribution of this absolute deviation after we include station fixed effects. The mean absolute deviation shows the average across all fuel stations. We compute the different range measures by calculating the range for each individual market on a particular day and then averaging across days and markets. All prices are in euro per liter.

effects. The remaining price variation is unpredictable even to the most sophisticated consumers.

Table 3.2 decomposes the observed price dispersion into its predictable and unpredictable components. On average, the absolute price deviation from the market mean is 1.6 eurocents for *E5* and diesel, and 1.7 eurocents for *E10*. The mean absolute deviation from the mean after we include station fixed effects, which is the unpredictable component, remains above 1.0 eurocent for all fuel types. In Appendix C.1.4, we present these within-market price residuals graphically.[14] Furthermore, the average difference between the cheapest and the most expensive fuel station in a local market in terms of the unpredictable component is approximately 2.8 eurocents for all fuel types, which is substantial.

Stylized Fact 1. *A substantial share of price dispersion is random and unpredictable to consumers.*

3.2.4 Consumer Information

Fuel stations in Germany and France are required to immediately report price changes, hence allowing real-time price information to be made available to consumers via smart-

[14] We also show price cycles at a more disaggregated level by zooming in on one local market and individual days.

phone apps. These apps provide perfect information on prices to users, whereas non-users can discover prices only by driving from station to station.

Stylized Fact 2. *Some consumers know all prices (app users), whereas others need to search for prices sequentially.*

How well informed consumers are about prices often correlates with the fuel type they purchase. Frequent drivers often prefer diesel cars. Accordingly, diesel passenger vehicles drive 19,200 kilometers per year on average, compared to 10,800 kilometers for gasoline passenger vehicles.[15] Although diesel cars are more expensive to buy, the cost of fuel is usually lower, making diesel cars a fixed-cost investment to lower the marginal cost of driving. Therefore, drivers who select diesel engines have a higher incentive to save on fuel costs, so they are more likely to use price comparison apps.

To assess differences in search intensity across fuel types, we use data on search queries in 2015 from a major German price comparison smartphone app. Normalizing the number of app users by the number of registered vehicles, we find that the distinct number of users who search for diesel prices is approximately 50% higher than the number of users who search for gasoline prices. This is in line with the hypothesis that, on average, the share of diesel drivers who are well informed about prices is higher than the corresponding share of drivers of gasoline cars. Further details are presented in Appendix C.1.3.

Commercial vehicles often run on diesel. If drivers of commercial vehicles do not pay for their own fuel, they may be less sensitive to prices. It is therefore worth discussing why commercial vehicles are not a concern for our analysis. First, as of 1 January 2020, there were 15.1 million passenger vehicles with a diesel engine, but, including those with a gasoline engine, there were only 5.2 million commercial passenger vehicles (Kraftfahrt-Bundesamt, 2021). Hence, at least 66% of passenger cars with a diesel engine are owned by private individuals. Second, some commercial drivers, such as those receiving a fuel allowance or those who are self-employed, also have an incentive to reduce fuel costs. Therefore, the fact that many commercial vehicles run on diesel does not undermine our finding that drivers of diesel vehicles are, on average, more price sensitive than drivers of gasoline vehicles.

In addition to differences between diesel and gasoline, there are differences in price sensitivity between buyers of *E5* and *E10* in Germany. These differences are likely driven by unwarranted concerns about potential damage to the engine caused by biofuels, which arose around the introduction of *E10* in 2011 and help us further segment

[15]Based on 2019 data from the German Ministry of Transportation's *Verkehr in Zahlen 2020/2021*.

consumers according to how well informed they are.[16] Although *E5* is more expensive, the majority of gasoline drivers in Germany purchase *E5*. According to the German Automobile Association (ADAC), *E10* is approximately 1.5% less efficient than *E5*.[17] However, this efficiency gap accounts for only a fraction of the observed difference in prices between *E5* and *E10*, since *E10* is usually 4-6 eurocents cheaper. A survey conducted by the ADAC in 2020 suggests that the difference in price sensitivity between *E5* and *E10* is due to preferences and a lack of information. The survey found that, among respondents fuelling with *E10*, the most cited reason for doing so is lower prices (72%), followed by environmental concerns (37%). Among those not fuelling with *E10*, the most cited reasons are technical concerns (51%) and uncertainty about the cost and benefits (23%).[18] Since the octane rating is the same for *E5* and *E10*, there are no quality differences between the two fuel types.

This evidence strongly suggests that, among drivers of gasoline cars, more buyers of *E10* choose to become informed about prices in Germany. Again, we confirm this hypothesis with our search data in Appendix C.1.3. In particular, we find that, adjusted for the relative market shares of *E5* and *E10* within the gasoline market, search intensity is substantially higher among consumers buying *E10* than those purchasing *E5*. In France, in contrast, there was no controversy regarding *E10* similar to the one in Germany. Therefore, drivers of gasoline vehicles in France predominantly buy *E10*.

Stylized Fact 3. *The share of well-informed consumers (app users) differs by fuel type. In Germany, it is higher for diesel than for gasoline and for E10 than for E5.*

3.3 Theoretical Model

Motivated by these stylized facts, we adapt the Stahl (1989) model to analyse tax pass-through in a setting where firms sell a homogeneous good to consumers who either are fully informed or can search for lower prices. We use the model to derive predictions that guide our empirical analysis.

[16] Although biofuels can pose a significant threat to the engine of a vehicle not compatible with *E10*, approximately 90% of gasoline-run vehicles, including all vehicles produced after 2012, are compatible. A full list of compatible vehicles can be found at `https://www.dat.de/e10/` (last accessed: 19 February 2024).

[17] See `https://www.adac.de/verkehr/tanken-kraftstoff-antrieb/benzin-und-diesel/e10-tanken/` (last accessed: 19 February 2024).

[18] The full survey results can be found at `https://www.adac.de/news/umfrage-e10-tanken/` (last accessed: 19 February 2024).

3.3.1 Set-up

On the demand side, there is a mass M of consumers, each with the same valuation v and inelastic unit demand for a homogeneous good. Consumers in the market can be divided into two groups. On the one hand, there are fully informed shoppers, who know the prices of all sellers and always buy from the lowest-price seller. On the other hand, there are uninformed non-shoppers, who draw a first price for free, know the distribution of prices, and can decide to sequentially search for prices at an incremental search cost s until they find a price weakly below their reservation price p_r. The model assumes that a fraction ϕ of consumers are fully informed shoppers and that the remaining fraction $1 - \phi$ are non-shoppers.

On the supply side, there is an exogenous number of sellers denoted by N, which produce at a constant marginal cost of c.[19] Sellers are indexed by i. Sales are subject to an ad valorem tax τ.

Sellers first choose their price, and consumers then make search and purchase decisions. We search for the subgame perfect Nash equilibrium of the game via backward induction.

Before we proceed, we introduce some additional notation. Whenever mentioning prices, we refer to the gross price paid by consumers. We assume that sellers bear the initial incidence of the tax and then (partially) 'pass through' the cost of the tax to consumers. It is well established in the theoretical literature that equilibrium prices are equivalent regardless of whether the initial tax incidence is on buyers or sellers. We denote the pass-through rate of marginal costs as $\rho_c = \frac{\partial p}{\partial c}$. The pass-through rate of a per unit tax is equivalent to the pass-through rate of marginal costs. The pass-through rate of the ad valorem tax is denoted as

$$\rho_\tau = \frac{\partial p}{\partial \tau} \cdot \frac{1+\tau}{p}.$$

We focus on the determinants of the pass-through rate of the ad valorem tax. In Appendix C.2.3, we show that the main mechanisms are the same for a per unit tax.

The model differs from traditional models of pass-through in its notion of price sensitivity. Many models capture consumers' sensitivity to price changes through the price elasticity of aggregate demand or the closeness of substitution of differentiated products. In contrast, we consider the share of shoppers ϕ and the incremental search cost of non-shoppers s to be the primary determinants of price sensitivity. A larger share of shoppers results in more consumers purchasing from the lowest-price seller, thus reducing the expected profit from setting a price above the market minimum.

[19] We endogenize entry in Appendix C.2.2.

Similarly, lower search costs for non-shoppers incentivize them to search for lower prices, leading to lower reservation prices and lower prices overall. In section 3.7, we discuss why the price elasticity of aggregate demand or product differentiation in a full information model cannot explain our empirical findings.

3.3.2 Equilibrium Price Distribution

In the following, we characterize the equilibrium, while the analysis of the model is relegated to Appendix C.2.1. There exists no pure strategy equilibrium in prices. There is a unique symmetric mixed strategy equilibrium where all sellers draw a price from the interval $[\underline{p}, p_r]$ according to the distribution $F(p_i)$, where p_r is the reservation price of non-shoppers and $\underline{p}$ is the minimum price that a seller charges. Shoppers always buy from the lowest-price seller, whereas non-shoppers draw a single price and buy at this price. In equilibrium, non-shoppers do not search sequentially because any price they draw is below their reservation price.

The symmetric equilibrium pricing strategy is characterized by the equilibrium objects p_r, $\underline{p}$, and $F(p_i)$. The reservation price of non-shoppers is

$$p_r = \min \{E[p] + s, \upsilon\}.$$

If searching sequentially is sufficiently cheap, the reservation price of non-shoppers is the sum of the expected price at the next draw and the search cost s. With relatively high search costs, the reservation price of non-shoppers is simply the valuation of the good υ, and the model boils down to the well-known Varian (1980) setting.

The minimum element of the support from which sellers draw prices in equilibrium is

$$\underline{p} = \frac{p_r}{\frac{\phi N}{1-\phi} + 1} + c\frac{1+\tau}{1 + \frac{1-\phi}{\phi N}}.$$

The cumulative density function of the equilibrium pricing strategy is

$$F(p_i) = 1 - \left(\frac{p_r - p_i}{p_i - c(1+\tau)} \frac{1-\phi}{N\phi}\right)^{\frac{1}{N-1}}.$$

The expected profit of a seller is

$$E[\pi_i] = \left(\frac{p_r}{1+\tau} - c\right) \frac{1-\phi}{N} M.$$

In equilibrium, non-shoppers buy at the first price they draw, making the expected price equal to the average price paid by non-shoppers. Shoppers buy from the lowest-price seller, making the expected minimum price equal to the average price paid by shoppers.[20]

The expected price is

$$E[p] = \underline{p} + \left(\frac{1-\phi}{N\phi}\right)^{\frac{1}{N-1}} \int_{\underline{p}}^{p_r} \left(\frac{p_r - p}{p - c(1+\tau)}\right)^{\frac{1}{N-1}} dp.$$

The expected minimum price is

$$E[p_{min}] = \frac{1-\phi}{\phi}\left[p_r - E[p] + (p_r - c(1+\tau))c(1+\tau)\int_{\underline{p}}^{p_r} \frac{1}{(p-c(1+\tau))^2} F(p)dp\right].$$

3.3.3 Pass-Through of an Ad Valorem Tax

To analyse how ad valorem taxes are passed through to consumers, we first examine the impact of an increase in the ad valorem tax τ on the equilibrium pricing strategy. We assume that the search cost s is sufficiently high that the reservation price p_r is equal to the consumers' valuation υ, simplifying the framework to the Varian (1980) setting. In section 3.3.5, we relax this assumption and use numerical examples to show that our results hold even when search costs are low, which is the Stahl (1989) setting with sequential search.[21]

Since the reservation price now corresponds to the valuation of the good, only the minimum element of the support and the density of the pricing strategy are affected by a change in ad valorem taxes.[22]

Proposition 1. *With $0 < \phi < 1$, for any $\hat{\tau} > \tau$, the minimum element of the support of the equilibrium pricing strategy $\hat{\underline{p}} > \underline{p}$ and the Nash equilibrium pricing strategy with τ first-order stochastically dominates (FOSD) the pricing strategy with $\hat{\tau}$, i.e., $\hat{F}(p) \leq F(p) \quad \forall p$.*

When the share of shoppers is strictly positive, increasing the ad valorem tax τ leads to a shift in the support of prices from which sellers draw in equilibrium toward higher prices. Additionally, for each price on this support, the likelihood of a drawn price being lower than that price decreases with an increase in the ad valorem tax rate to $\hat{\tau}$. As the share of shoppers converges to zero, the Nash equilibrium converges toward

[20] The average minimum price refers to the average price paid by shoppers if this game is repeated often across time or space. At a given time and location, there are only one minimum price and N prices.

[21] An alternative simplification would be setting $N = 2$, which is less desirable for studying the effect of competition.

[22] The proofs of Proposition 1 and all following propositions can be found in Appendix C.2.4.

a degenerate distribution at the monopoly price, the classical result from Diamond (1971). The monopoly price corresponds to the consumers' valuation v.

Since the minimum element of the support of prices and the density function monotonically move toward higher prices, other moments of interest, such as the expected price $E[p]$ and the expected minimum price $E[p_{min}]$, also increase.

3.3.4 Effect of the Price Sensitivity on the Pass-Through Rate

We now turn to analysing how the pass-through rate of an ad valorem tax τ varies with the price sensitivity of consumers.

Proposition 2. *If the share of shoppers $\phi = 0$, pass-through of the ad valorem tax $\rho_\tau = 0$. If $\phi = 1$, there is full pass-through, i.e., $\rho_\tau = 1$. As $\phi \to 1$, the pass-through rate $\rho_\tau \to 1$.*

Let us begin by examining the two extreme cases. If there are no shoppers, the Nash equilibrium is a degenerate distribution at the monopoly price, which is unaffected by the ad valorem tax, and pass-through is zero. However, if the share of shoppers approaches one, the Nash equilibrium approaches the classical Bertrand equilibrium, where the Nash equilibrium is a degenerate distribution at $c(1 + \tau)$, and there is full pass-through.

As the share of shoppers ϕ increases from zero to one, the pass-through rate of the ad valorem tax to the lower bound of the equilibrium price strategy strictly increases. Furthermore, the rate at which an increase in the tax from τ to $\hat{\tau}$ reduces the probability of drawing a price below a certain price p, i.e., from $F(p)$ to $\hat{F}(p)$, also strictly increases as the share of shoppers increases. Therefore, the pass-through rate increases with the share of shoppers, and it reaches full pass-through as the share of shoppers approaches one.

3.3.5 Effect of the Number of Sellers on the Pass-Through Rate

In addition to the share of informed consumers, the number of active sellers is another important dimension of competition, often more salient in empirical applications.

Proposition 3. *With $0 < \phi < 1$, as $N \to \infty$, the pass-through of τ to the minimum element of the equilibrium price support converges to full pass-through, i.e., $\rho_{\tau,\underline{p}} \to 1$.*

With more sellers, competition for shoppers becomes more intense, leading to convergence of the minimum price that sellers consider charging in the symmetric Nash equilibrium toward $c(1+\tau)$. As a result, the pass-through rate of the ad valorem tax to $\underline{p}$ increases.

Proposition 4. *With $0 < \phi < 1$, as $N \to \infty$, the pass-through of τ to the expected price is non-monotonic.*

Finally, we can show that the relationship between the number of sellers and pass-through of an ad valorem tax to the expected price is non-monotonic. The idea of the proof is straightforward. Facing inelastic unit demand, the monopolist charges consumers their valuation and does not adjust the price when taxes change. When $N \to \infty$, $E[p]$ converges to the monopoly price, which, again, implies no pass-through. For intermediate numbers of sellers, pass-through is strictly positive. This is a striking result, as this is not the case in models with perfect information, where the rate of pass-through to $E[p]$ monotonically increases in the number of sellers.[23]

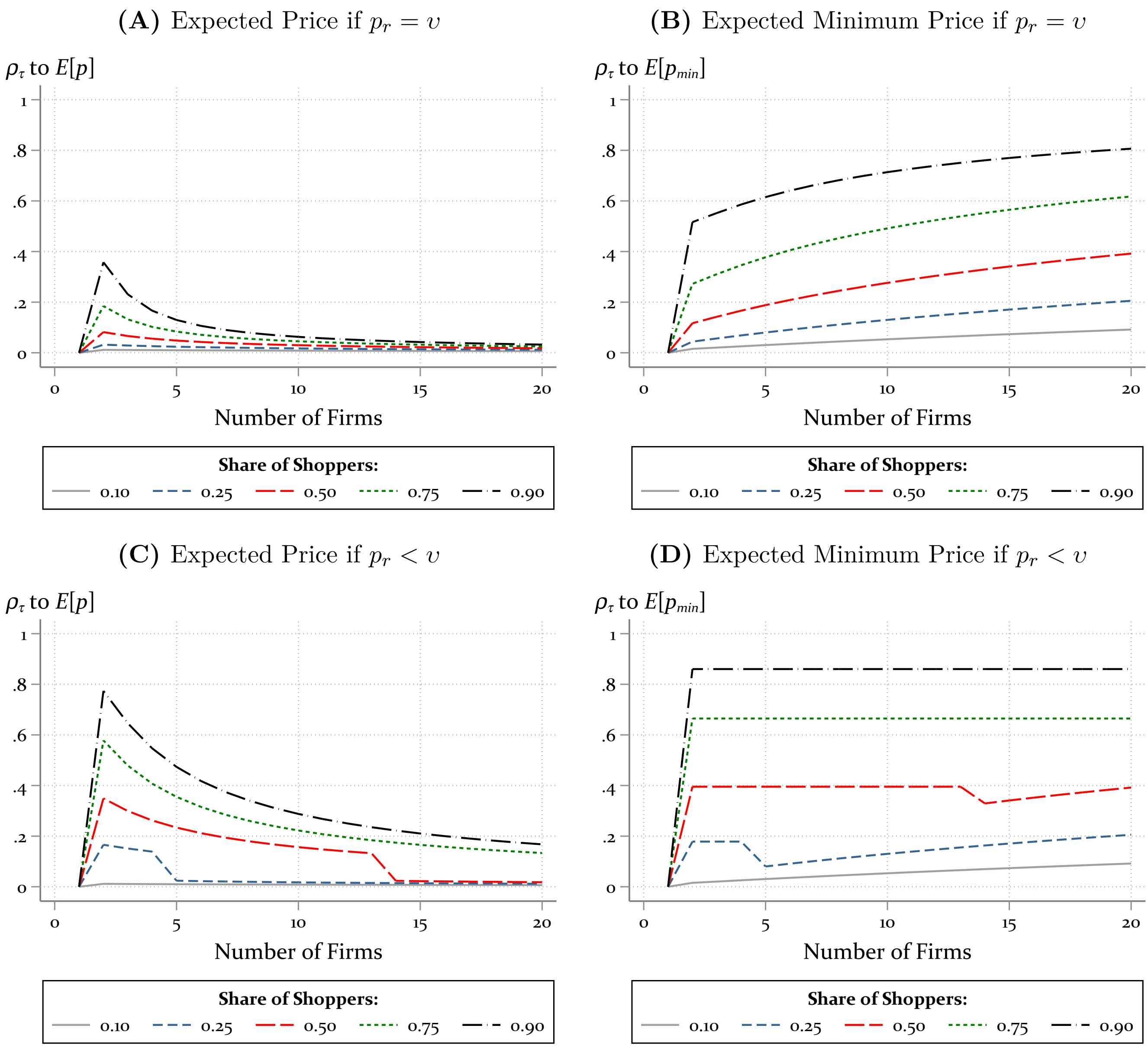
Although we can show that there is a non-monotonic relationship between the number of sellers and pass-through to $E[p]$, we cannot say anything about its exact shape. Similarly, we cannot prove how $E[p_{min}]$ varies with the number of sellers. To illustrate these relationships, we resort to numerical simulations of how a change in the tax affects $E[p]$ and $E[p_{min}]$ for a given set of parameters and varying the number of sellers N.

We show the numerical results for a particular choice of parameter values in Figure 3.2. Panels (A) and (B) illustrate how pass-through of an ad valorem tax to $E[p]$ and $E[p_{min}]$ varies with the number of sellers in a Varian (1980) setting, where the sequential search cost of non-shoppers s is so high that their reservation price is equal to their valuation of the good, i.e., $p_r = \upsilon$. Panels (C) and (D) show how pass-through of an ad valorem tax to $E[p]$ and $E[p_{min}]$ varies with the number of sellers in the Stahl (1989) setting, where the sequential search cost of non-shoppers is sufficiently low that their reservation price depends on the price they expect to draw if they were to search, i.e., $p_r < \upsilon$.

The simulations show that the relationship between the number of sellers and pass-through to $E[p]$ can take different shapes but is always non-monotonic. They also show that, with inelastic unit demand, there is always a peak at $N = 2$, after which pass-through falls. When deriving deriving empirically testable predictions, we discuss how product differentiation might affect this shape and shift the peak.

[23] This is analogous to the finding of Stahl (1989) that there is a non-monotonic relationship between the number of sellers and the expected price.

Figure 3.2. Numbers of Sellers and Tax Pass-Through



Notes: The figure shows simulation results of how the pass-through rate of the ad valorem tax τ varies with the number of sellers. Panels (A) and (B), respectively, show how the rate of pass-through to the expected price ($E[p]$) and to the expected minimum price ($E[p_{min}]$) varies with the number of sellers if the reservation price is exogenous. Panels (C) and (D) show the same if the reservation price of non-shoppers p_r is endogenous. In all panels, the different lines correspond to different values of the share of shoppers ϕ. Parameter values: $\upsilon = 4.5$, $c = 0.4$, $\tau = 0.2$, $\hat{\tau} = 0.22$, $s = \infty$ (without sequential search) and $s = 0.75$ (with sequential search).

For pass-through to $E[p_{min}]$, the results are more nuanced. In a Varian (1980) setting, more sellers decrease $E[p_{min}]$. The more $E[p_{min}]$ converges to marginal costs, the lower the margin that sellers could use to absorb a tax increase. Thus, pass-through to $E[p_{min}]$ monotonically increases in N. In the Stahl (1989) setting, there is an additional countervailing effect. If the reservation price is endogenous, this is a function of the expected price. When the number of sellers increases, $E[p]$ increases, and p_r increases. This decreases the incentive for sellers to set lower prices, increasing the expected minimum price and decreasing pass-through to $E[p_{min}]$. Depending on the relative strength of these two effects, when the reservation price is endogenous, pass-through to the expected minimum price can increase or decrease in the number of sellers.

Since there is no clear prediction about the relationship between the number of sellers and the expected minimum price, the key testable implication of Proposition 4 and the numerical exercise is that, if there is imperfect information, the relationship between the number of sellers and pass-through to $E[p]$ is non-monotonic.

3.3.6 Deriving Empirically Testable Predictions

Our empirical setting deviates from the theoretical model in several ways. Incorporating these features into the model is difficult, so we qualitatively discuss how they affect the testable predictions. In section 3.7, we discuss how well alternative hypotheses can explain the empirical results.

In the theoretical model, players have expectations about the average price and the minimum price in a market. These are the expected price and the expected minimum price, respectively. In the empirical application, we do not observe these expectations. Instead, we observe many different local markets. The sample equivalents to these theoretical objects are the average price and the minimum price in a market.

The first prediction is based on Propositions 1 and 2. Since close to all stations in Germany sell all three fuel types under consideration, station-level product differentiation should not affect the relative pass-through between fuel types. Propositions 1 and 2 deal with the full distribution of equilibrium prices. Therefore, Prediction 1 should hold for any moment of the price distribution.

Prediction 1. *Pass-through is higher when the share of well-informed consumers is higher. In Germany, we expect pass-through to be highest for diesel and lowest for E5.*

In the model, consumers purchase at most one unit of the good. They cannot choose to buy different quantities in response to different prices. If the price elasticity of aggregate demand is highest for diesel and lowest for *E5*, this works in the opposite direction as the effect of consumer information. Finding higher pass-through for diesel would thus further underscore the importance of imperfect information. If the price elasticity of aggregate demand is lowest for diesel and highest for *E5*, the price elasticity of aggregate demand would be an observationally equivalent explanation with respect to Prediction 1. A finding supporting Predictions 2 and 3 continues to require imperfect information, however.

The next prediction is based on results from the numerical exercise. Figure 3.2 shows that pass-through to the minimum price, paid by well-informed consumers, is predicted to be higher than pass-through to the average price, paid by uninformed consumers. This holds for any $N \geq 2$ and remains the case with horizontal differentiation.

Prediction 2. *Pass-through to the minimum price is higher than pass-through to the average posted price in a market.*

The final prediction is based on Proposition 4, which states that there is a non-monotonic relationship between the number of sellers and pass-through. In practice, there is some degree of horizontal product differentiation, which is not reflected in the model. Travelling between stations comes at a cost, and the degree of substitution between stations decreases with travel time. Horizontal differentiation, through the distance between stations, increases a station's market power. The closer the competing stations are to each other, the lower their market power becomes. In contrast to perfectly homogeneous fuel stations, the presence of only two rival stations may not be enough to achieve perfect competition, even with full information. With imperfect information, this effect works in the opposite direction as the increase in the pass-through to the average price observed when $N > 2$. Empirically, we should therefore continue to expect a non-monotonic relationship, but this can take various shapes, and the pass-through peak may occur at a higher number of competitors than $N = 2$.

Prediction 3. *The relationship between the number of competitors and pass-through to the average price is non-monotonic.*

3.4 Policy Changes and Descriptive Evidence

We analyse multiple tax changes in the German and French retail fuel markets from 2020 to 2023 to verify whether pass-through can be explained by competition under imperfect consumer information. In this section, we provide an overview of the tax changes and then present descriptive evidence on the pass-through of these interventions.

3.4.1 Tax Changes in the Retail Fuel Market

Taxes account for the largest share of fuel prices in Germany and France. In 2019, a lump-sum energy tax of 65.45 eurocents per liter was levied on gasoline and 47.04 eurocents per liter on diesel in Germany. In France, the lump-sum fuel tax varied by region, ranging from 67 to 70 eurocents per liter for gasoline (and approximately 61 eurocents per liter for diesel). In addition, Germany and France have a VAT of 19% and 20%, respectively, that is levied on the tax-inclusive price of fuel.

The retail fuel markets in Germany and France experienced several tax changes between 2020 and 2023. These changes include a temporary VAT reduction in Germany to combat the economic impact of the Covid-19 pandemic, the introduction of a carbon tax in the German fuel market, and temporary reductions in the energy tax in both countries in 2022/23 to address price increases resulting from the Russian invasion of Ukraine.

The first tax change was a temporary VAT reduction in Germany from 19% to 16% between July and December 2020. On 1 January 2021, at the same time as the VAT was raised back to 19%, the German federal government also introduced a carbon price of 25 euros per emitted tonne of CO_2 on oil, gas, and fuel. For *E5* and *E10*, this translates into a per unit tax of 6.00 eurocents per liter (7.14 eurocents including VAT). For diesel, the per unit tax is 6.69 eurocents per liter (7.96 eurocents including VAT).

We cannot separately identify the pass-through of the simultaneous increase in the VAT and the introduction of the carbon emissions price in Germany on 1 January 2021. Instead, we jointly estimate their pass-through rate. This does not raise concerns regarding the theoretical predictions, since we show in Appendix C.2.3 that the mechanisms that determine pass-through of an ad valorem tax and a per unit tax are the same.

In 2022, several tax changes occurred as a response to the Russian invasion of Ukraine and the resulting surge in energy prices. In France, between 1 April and 31 August 2022, there was a decrease in the fuel tax on gasoline and diesel of 18 eurocents per liter. This rebate increased to 30 eurocents between 1 September and 15 November 2022. It then dropped to 10 eurocents between 16 November and 31 December 2022, before being completely phased out on 1 January 2023. Instead, the government introduced a lump-sum transfer to poorer households depending on their use of a car to commute to work.

Germany also implemented a temporary tax rebate, but this tax change is not studied in our analysis because of the intense public scrutiny and concurrent market investigation by the Federal Cartel Office. To appease the public, the vertically integrated oligopolists heavily advertised that they would pass through the tax change fully and quickly.[24]

[24]In Appendix C.3, we present additional descriptive evidence showing that the 2022 tax changes in Germany are not suitable for our analysis.

3.4.2 Descriptive Evidence on Heterogeneous Pass-Through

Before turning to the econometric analysis, we descriptively study the pass-through of the 2020/21 tax changes in Germany by comparing fuel prices over time in Germany and France. This allows us to observe whether the pass-through differs between markets with a higher share of informed consumers (diesel) and those with fewer informed consumers (*E5*).

Panel (A) of Figure 3.3 displays non-parametric estimates of the pass-through rate by fuel type for the 2020 temporary VAT reduction in Germany. Prices before the tax reduction evolve similarly for the three fuel types, suggesting that post-reduction differences in pass-through rates are not driven by pre-trends. The pass-through rate is highest for diesel and lowest for *E5*, consistent with the theoretical prediction that pass-through is higher when more price-sensitive consumers are present in the market. Pass-through is relatively fast and stabilizes after approximately two weeks.

Panel (B) of Figure 3.3 presents non-parametric estimates of the pass-through rate by fuel type for the subsequent tax increase in winter 2020/21. Unlike in the case of the tax decrease, there appear to be anticipatory effects in the pass-through of the tax increase in the last two weeks of December. Therefore, we drop the second half of December 2020 from the econometric analysis. The sharp increase in the implied pass-through rate around 1 January 2021 stabilizes afterwards. Again, pass-through is highest for diesel, which is consistent with the theoretical predictions. Differences in pass-through between *E5* and *E10* appear less pronounced.

3.5 Empirical Strategy

Next, we estimate pass-through rates of the different tax changes separately by fuel type using a difference-in-differences (DiD) strategy.

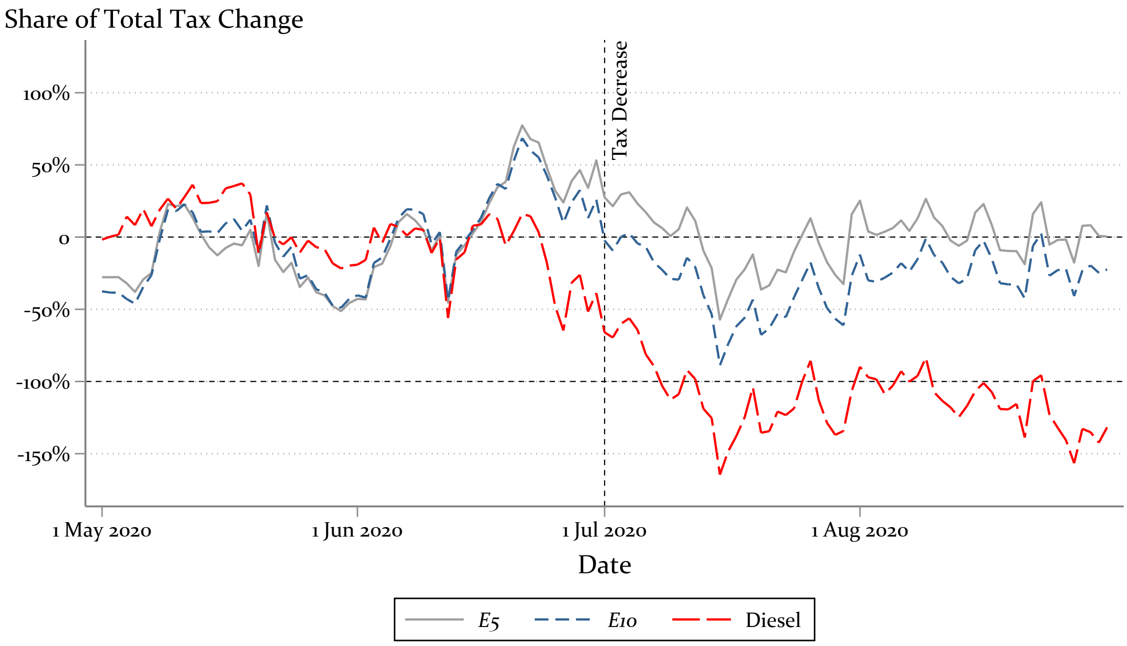
3.5.1 Difference-in-Differences Strategy

To estimate the average rate of pass-through of the tax changes to fuel prices, we compare stations in Germany and France, before and after the tax change. We use French fuel prices as the control group to estimate pass-through of the 2020/21 tax changes in Germany.[25] The treatment effect is the change in the difference between average fuel prices in Germany and France between the pre- and post-treatment periods.

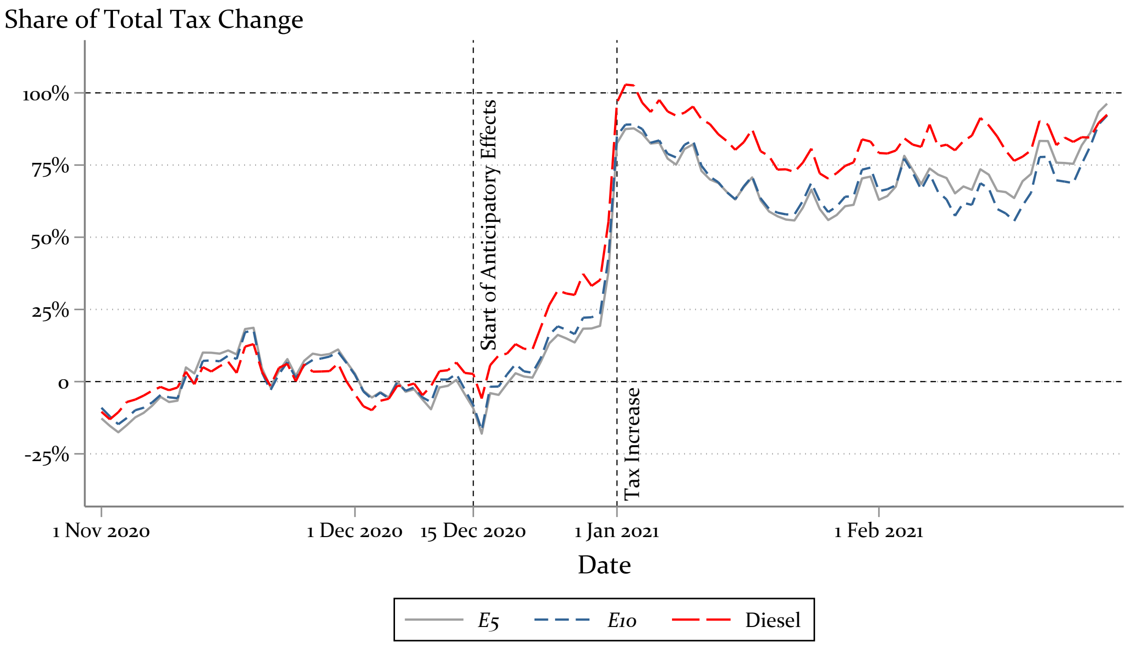
[25] Conversely, we use German stations as the control group to estimate pass-through of the 2022/23 tax changes in France. For simplicity, we explain the DiD strategy for the baseline tax changes in Germany.

Figure 3.3. Price Change as Share of Total Tax Change

(A) Tax Decrease



(B) Tax Increase



Notes: The figure depicts the price change as a share of the total tax change for the tax decrease in July 2020 and the tax increase in January 2021 in panels (A) and (B), respectively. The solid line shows the non-parametric estimate of the daily average pass-through rate to prices for *E5*. The short-dashed and long-dashed lines show analogous estimates for *E10* and diesel, respectively. To estimate pass-through, we first subtract the average pre-period price in Germany (France) from the daily average price in Germany (France). The pre-period is from 1 May until 30 June 2020 for the tax decrease in panel (A) and from 1 November until 15 December 2020 for the tax increase in panel (B). Next, we compute the difference between de-meaned average prices in Germany and France. Finally, we divide this difference by the difference under full pass-through. For the tax decrease, full pass-through would correspond to a price drop of 2.52%. Using average absolute prices from 24 June until 30 June (i.e., in the week prior to the tax change), this translates to a price decrease of 3.24 eurocents for *E5*, 3.15 eurocents for *E10*, and 2.72 eurocents for diesel under full pass-through. For the tax increase, full pass-through would correspond to a price increase of 2.59% due to the VAT increase plus the newly introduced carbon price. Using absolute prices in the week from 9 December until 15 December 2020 (i.e., in the week prior to the appearance of anticipatory effects), this translates to a price increase of 10.37 eurocents for *E5*, 10.24 eurocents for *E10*, and 10.75 eurocents for diesel under full pass-through. The vertical solid line marks the starting date of the tax change. The horizontal dashed line indicates full pass-through.

For our baseline pass-through estimation, we estimate the following DiD model:

$$Y_{i,t} = \beta \cdot \text{Tax}_{i,t} + \gamma X_{i,t} + \alpha_i + \pi_t + \epsilon_{i,t}, \tag{3.1}$$

where $Y_{i,t}$ is the logarithm of the weighted average price of gasoline or diesel at a fuel station i at date t. $\text{Tax}_{i,t}$ is a dummy variable that equals one for stations affected by the tax change at date t. For the analysis of the tax reduction, these are fuel stations in Germany from 1 July 2020 onward. For the analysis of the subsequent tax increase, these are fuel stations in Germany from 1 January 2021 onward. For the analysis of the French tax changes in 2022/23, $\text{Tax}_{i,t}$ equals one for French stations in the respective post-treatment periods. $X_{i,t}$ is a vector of controls, which in the baseline specification includes only an interaction term of the crude oil price with an indicator for stations in Germany. This allows for differential pass-through of the crude oil price in Germany and France. Finally, the variables α_i and π_t correspond to fuel station and date fixed effects, respectively. We cluster standard errors at the market level.

3.5.2 Stations in Neighbouring Country as a Control Group

Two assumptions must be met for us to identify the impact of the tax changes on fuel prices. First, there should be no temporary shocks that differentially affect fuel stations in Germany and France before and after the tax change, other than the policy change itself. Second, there should be no spillover effects from the tax changes onto the fuel market in the neighbouring country. Both of these assumptions are likely to be satisfied for the tax changes in 2020/21, but this is less likely for the tax changes in 2022/23.

Station fixed effects account for time-invariant differences between fuel stations in Germany and France, while date fixed effects control for transitory shocks that identically affect German and French stations. The two countries are similar in their geographic location, size, and wealth. Moreover, we restrict our analysis to relatively narrow time windows around the reforms, which should alleviate concerns about transitory shocks that may differently affect German and French fuel stations in 2020/21.

To strengthen our claim that the effects are not influenced by transitory shocks, we consider the most obvious threats to identification. Public and school holidays in Germany and France are highly correlated. Travel restrictions due to the Covid-19 pandemic were lifted simultaneously in both countries and the rest of the Schengen Area, starting on 15 June 2020. As most holidaymakers within Europe typically travel across several EU countries and France and Germany are popular travel destinations in close proximity, it is likely that demand shocks affected fuel stations in both countries similarly.

Transitory supply shocks should also affect German and French fuel stations in a similar way. Because of their geographic proximity, fuel stations in Germany and France procure most of their refined oil from similar sources. The two countries are also members of the European single market, which implies harmonized border checks, common customs policy, and identical regulatory procedures on the movement of goods within the EU.

No major reforms concerning the fuel market were implemented in Germany and France during our analysis period other than the tax changes discussed in section 3.4. In general, there is no fuel price-setting regulation in Germany and France, and both countries have mandatory disclosure of fuel prices, which reaffirms our choice of France as a suitable control group.

Focusing on tax changes in opposite directions reduces concerns about confounding factors driving our results. If we find similar heterogeneities in pass-through for the tax increase in January 2021 as for the tax decrease in July 2020, a transitory shock confounding our estimates in July 2020 would also have to be present in January 2021, but it would have to go in the opposite direction. For instance, if we overestimated the pass-through rate for diesel in July 2020 because of a positive demand shock in France, then overestimating pass-through for diesel in January 2021 would require France to experience a negative demand shock. This scenario is unlikely, and our finding of consistent heterogeneities between the two tax changes suggests that we are robustly estimating actual differences in pass-through.

For 2022/23, the picture is different. First, multiple tax changes occurred in Germany and France, sometimes simultaneously, making it impossible to identify these separately. Second, the tax changes are so large that they may change the opportunity cost of selling fuel in the other country. This may lead to spillover effects and break the stable unit treatment value assumption (SUTVA). Third, gasoline and diesel markets were hit differently by Russia's invasion of Ukraine, as diesel is a close substitute for heating oil whereas gasoline is not. Fourth, Germany and France were affected differently by the shocks to the global oil market in 2022.[26] As a consequence, we should refrain from interpreting the magnitude of pass-through for the 2022 tax changes and the differences between fuel types. However, analysing these tax changes can be help-

[26]We present evidence in support of these points in Appendix C.3. We show that the diesel and gasoline markets in Germany and France started developing differently right after the start of Russia's invasion and before any tax change was announced. We also show that margins increased in France immediately after Germany introduced a large fuel tax cut on 1 June 2022, suggesting spillover effects.

ful to understand the difference in pass-through to the average posted price and the minimum price within a given fuel type.

3.5.3 Testing the Theoretical Predictions Empirically

The aim of our empirical exercise is to test the theoretical predictions in section 3.3 empirically. To test Prediction 1, we estimate the baseline DiD model in equation (3.1) separately for the three different fuel types for each of the 2020/21 tax changes in Germany. We then compare the pass-through rates *across* fuel types. According to the theoretical model and the specificities of the German industry, we expect pass-through to be highest for diesel and lowest for *E5*. This should be the case for both the tax decrease and the increase. We do not use the tax changes in 2022/23 to assess Prediction 1, because the caveats discussed above make any comparison across fuel types unreasonable.

To test Prediction 2, we estimate tax pass-through to the average price and the minimum price in a market for the 2020/21 tax changes and the 2022/23 French tax changes. We now include the French tax changes because we compare pass-through to the average price and the minimum price *within* each fuel type. To remain as close as possible to the expected price and the expected minimum price in the theoretical model, we compute pass-through rates for the average posted price and the minimum price within non-overlapping geographic markets in Germany and France. We then estimate the following triple DiD variation of our baseline model:

$$Y_{j,t,s} = \beta_1 \cdot \text{Tax}_{j,t} + \beta_2 \cdot \text{Tax}_{j,t} \cdot \text{Min}_s + \alpha_{j,s} + \pi_{t,s} + \epsilon_{j,t,s}, \tag{3.2}$$

where $Y_{j,t,s}$ is the logarithm of the price of gasoline or diesel in market j at date t. Note that the regression is now at the market level (instead of the station level). The market-level price is either the average posted price or the minimum price, as indicated by the 'price type' s. Accordingly, Min_s is an indicator that equals one when the price type is the minimum price. As before, $\text{Tax}_{j,t}$ is a dummy variable that equals one for markets affected by the tax change at date t. The variables $\alpha_{j,s}$ and $\pi_{t,s}$ correspond to market × price type and date × price type fixed effects. In this model, β_1 allows us to derive the average rate of pass-through to the average posted price (i.e., when $\text{Min}_s = 0$). In contrast, β_2 allows us to derive the difference in the pass-through rate between the minimum price and average posted price in a local market. As the theoretical model predicts pass-through to the minimum price to be higher than that to the average price, we expect $\beta_2 > 0$ when estimating the parameters of the model in equation (3.2). This prediction holds for all fuel types and for all tax changes.

To test Prediction 3, we estimate tax pass-through for the 2020/21 tax changes at the station level. We begin by estimating a pass-through rate for every station in Germany for each fuel type. That is, we estimate the model in equation (3.1) adding an interaction between the treatment period and the station fixed effect. The station-specific treatment effect is then the sum of the average treatment effect and this additional interaction. Finally, we group stations by the number of competing price setters in a local non-overlapping market and calculate the average pass-through rate for each group. An important feature of our setting is that we can then compare pass-through rates *within* fuel type and thus hold an important source of variation in price sensitivity fixed. In a perfect information model, we expect the relationship between the number of competitors in a market and the average pass-through rate to be monotonically increasing. Instead, our model predicts a hump-shaped relationship between the number of competitors and average tax pass-through. As we do not have information on brands or price setters for France, we can test Prediction 3 only for tax changes in Germany.

3.5.4 Robustness Checks

We run several additional analyses to verify that our empirical results are robust to alternative model specifications.

In Appendix C.4.1, we present DiD estimates of the baseline pass-through rates when we include additional control variables in our regression model. In particular, we directly account for demand-related shocks by including regional information on daily mobility to work and to retail and recreational places from the Google Mobility Report. Our results are robust to including these control variables.

In Appendix C.4.2, we further show that our results are robust to estimating the baseline DiD model in a restricted (balanced) sample of fuel stations for which we have a price observation for every day. This allows us to rule out that our findings are driven by temporary closures (e.g., on weekends or holidays) of some fuel stations.

Based on the descriptive evidence in Figure 3.3, our preferred specification accounts for anticipatory effects in winter (tax increase) but not in summer (tax decrease). In Appendix C.4.3, we show that our empirical findings are robust to changing this assumption. In Appendix C.2.5, we also provide a brief theoretical discussion of the emergence of anticipatory price increases before a tax increase and a tax decrease.

In Appendix C.4.4, we employ a synthetic difference-in-differences (SDiD) strategy (Arkhangelsky et al., 2021). SDiD is a variation of DiD that aims to match pre-treatment trends between the treatment and control groups using weights. Our results are robust to using this alternative empirical strategy.

In Appendix C.4.5, we assess the robustness of our results regarding the differential pass-through to the average price and the minimum price. One potential concern is that the minimum price in a local market may reflect outlier prices valid for only a short period of time. We address this concern by replacing the minimum price with the price at the 10th percentile of the distribution of all hourly prices across all stations in a market on a given day. We find that our results on the difference in pass-through between the minimum price and the average price are robust to this specification and, hence, are not driven by outliers.

Finally, Appendices C.4.6 and C.4.7 repeat part of our analysis with a different definition of local markets. Instead of using non-overlapping markets, we define overlapping markets centered on each station and including all competing price setters within a driving distance of 5km around the centroid station. Again, our results are largely unaffected by this alternative approach.

3.6 Results

This section presents the results from the empirical analysis.

3.6.1 Consumer Information and Tax Pass-Through

Table 3.3 presents the estimated average treatment effect of the 2020/21 tax changes on fuel prices for *E5*, *E10*, and diesel, using the DiD model in equation (3.1). The outcome variable in all columns is the logarithm of price, including taxes and duties, for a given station and date.

Columns (1) to (3) show the effect of the tax decrease. The tax reduction caused prices for all fuel types to decrease. Under full pass-through, we expect prices for each fuel product to decrease by approximately 2.52%.[27] We estimate that 95% of the tax decrease is passed on to diesel consumers while the pass-through rates for *E10* and *E5* are 46% and 28%, respectively.

Columns (4) to (6) show that the tax increase raised prices for all fuel products. Under full pass-through, we expect an increase in prices by 8.31% for *E5*, 8.54% for

[27] With a decrease in the VAT rate from 19% before the VAT decrease to 16% after the policy change, this is $\frac{1.16-1.19}{1.19} \cdot 100 \approx -2.52\%$.

Table 3.3. Effect of the Tax Change on Log Prices

	Tax Decrease			Tax Increase		
	E5	*E10*	Diesel	*E5*	*E10*	Diesel
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Change	-0.0071***	-0.0116***	-0.0239***	0.0561***	0.0611***	0.0834***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Pass-Through Rate	28%	46%	95%	68%	72%	84%
	[25%, 31%]	[43%, 49%]	[92%, 97%]	[67%, 68%]	[71%, 72%]	[83%, 84%]
Date FE	✓	✓	✓	✓	✓	✓
Station FE	✓	✓	✓	✓	✓	✓
DE × Oil Price	✓	✓	✓	✓	✓	✓
Observations	2,133,377	2,324,131	2,703,604	1,804,703	1,988,649	2,318,672

Notes: The table presents DiD estimates using the model in equation (3.1). Columns (1) to (3) present average treatment effect estimates of the German VAT reduction on 1 July 2020 on *E5*, *E10*, and diesel log prices, respectively. Columns (1) to (3) use data from 1 May to 31 August 2020. Columns (4) to (6) present average treatment effect estimates of the German VAT increase and introduction of a carbon price on 1 January 2021 on *E5*, *E10*, and diesel log prices, respectively. Columns (4) to (6) use data from 1 November to 15 December 2020 for the pre-treatment period and from 1 January to 28 February 2021 for the post-treatment period. Standard errors clustered at the market level are shown in parentheses. We also compute the pass-through rates corresponding to the point estimates and report their 95% confidence intervals in brackets. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

E10, and 9.97% for diesel.[28] We find a joint pass-through rate of the tax increases of 68% for *E5*, 72% for *E10*, and 84% for diesel.

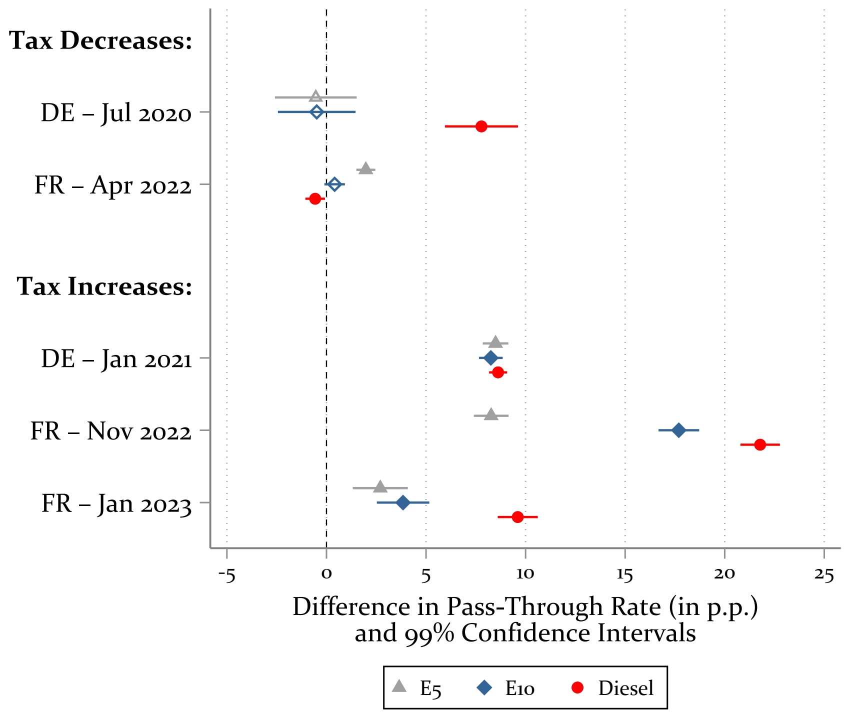
Overall, our findings are consistent with Prediction 1 that the pass-through rate is higher when there are more price-sensitive consumers. For both tax changes, pass-through is significantly higher for diesel than for gasoline. Within gasoline, pass-through is significantly higher for *E10* than for *E5*. Thus, the ordering of all the point estimates in Table 3.3 is consistent with our prediction. Since we observe that almost all fuel stations in Germany sell all three types of fuel, the differences in the pass-through rates cannot be explained by supply-side factors. In section 3.7, we discuss why differences in the price elasticity of aggregate demand between fuel types are also unlikely to explain the results.

3.6.2 Pass-Through to the Average and Minimum Price

Figure 3.4 shows the difference in the pass-through rate between the minimum price and the average posted price for different tax changes in the German and French retail fuel markets between July 2020 and January 2023. As previously noted, the 2022 tax changes are inadequate for measuring the relative and absolute pass-through

[28]Under full pass-through, a change in the VAT rate from 16% to 19% would increase the fuel price by $\frac{1.19-1.16}{1.16} \cdot 100 \approx 2.59\%$. To estimate by what percentage the fuel price would increase if the carbon emissions price was fully passed through, we divide the gross price per liter on carbon emissions for each fuel type by the average fuel price in Germany in the week from 9 December until 15 December 2020 (i.e., before we start seeing anticipatory effects).

Figure 3.4. Pass-Through to Market-Level Minimum vs. Average Prices



Notes: The figure shows the difference in the market-level pass-through rate (in percentage points) between the minimum price and the average posted price. The average posted price is the average daily price within a non-overlapping market, calculated by weighting the price at every full hour of the day between 6am and 10pm equally. The minimum price is the minimum price within a non-overlapping market at any point in time during the day. The figure depicts the pass-through rates implied by the DiD estimate β_2 in equation (3.2) along with 99% confidence intervals, based on standard errors clustered at the market level. Regressions are estimated separately for each tax change and fuel type. For most tax changes, we use data for the two months before and after every tax change. Exceptions include the German tax increase on 1 January 2021, where we exclude the second half of December to account for anticipatory effects. For the tax increase in France on 16 November 2022, we use only the period until 31 December 2022 as the post-treatment period. Similarly, for the French tax increase on 1 January 2023, we use the period from 16 November until 31 December 2022 as the pre-treatment period. Solid (hollow) symbols indicate point estimates that are statistically significant (insignificant) at the 1% level.

across different fuel types. Nonetheless, they allow us to compare pass-through to the minimum price and average posted prices within a particular fuel type.[29]

For the tax increases, pass-through rates to the minimum price are always significantly higher than those to the average posted price. This supports our theoretical prediction. For the tax decrease, this is true in two cases, while there are three cases where the difference in pass-through rates is statistically indistinguishable from zero and one case where pass-through to the average posted price is slightly higher. Over-

[29]Before estimating the triple DiD model in equation (3.2), we verify that market-level minimum and average prices evolved similarly in Germany and France prior to the tax changes. To this end, we repeat the non-parametric pass-through estimation as in section 3.4, but we now use the minimum and average posted price in a local non-overlapping market. The results are shown in Appendix C.3.1 and look very similar to those in Figure 3.3, which uses station-level prices.

all, we observe a statistically significantly higher rate of pass-through to the minimum price than to the average price in 11 out of the 15 cases depicted in Figure 3.4.

The findings presented in Figure 3.4 support Prediction 2 that the pass-through to the expected minimum price is higher than that to the expected price. Informed consumers, who typically buy fuel at prices closer to the within-market minimum, bear more of the cost of a tax increase (and gain more from a tax cut) than uninformed consumers, who buy fuel at the average posted price.

3.6.3 Number of Sellers and Tax Pass-Through

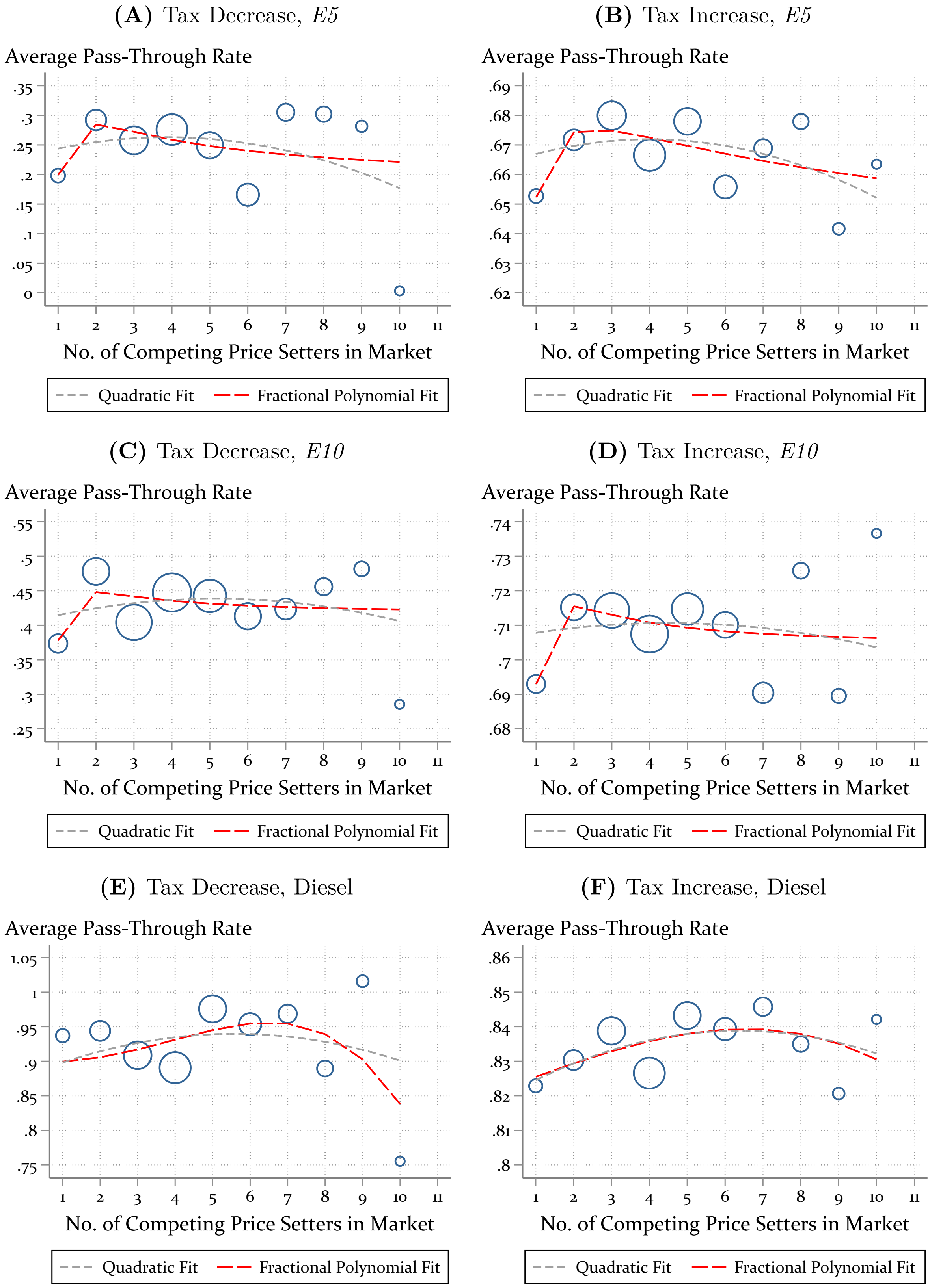
Finally, we study how the pass-through rate varies with the number of sellers in the market. Figure 3.5 shows the relationship between the pass-through rate and the number of competitors of a focal station for the 2020/21 German tax changes and the three fuel types. Each circle corresponds to the average pass-through rate for stations with a particular number of competing price setters within a non-overlapping local market. The size of a circle is proportional to the total number of stations with a given number of competitors. We also plot the curves of a fractional polynomial fit and a quadratic fit.

Panels (A), (C), and (E) depict the pass-through rates for the tax decrease in summer 2020 for *E5*, *E10*, and diesel, respectively. Panel (A) shows that the average pass-through rate for *E5* is relatively low for local monopolists. It is higher for markets with two competing price setters and then tends to decrease in the number of competitors. We observe a similar non-monotonic relationship between the number of sellers and the average pass-through rate for *E10*. This pattern looks strikingly similar to the numerically simulated patterns in Figure 3.2. For diesel, the relationship is flatter for smaller markets, but then the average pass-through rate also declines with the number of competitors.

In panels (B), (D), and (F) of Figure 3.5, we repeat this analysis for the tax increase in winter 2020/21. For all fuel types, we find relationships similar to those observed for the tax decrease. For *E5* and *E10*, pass-through is again relatively low for local monopolists and tends to decrease in the number of competitors when there are at least two competing price setters. For diesel, the pass-through rate is mildly increasing up to approximately six or seven competing price setters and then decreases in the number of sellers.

Overall, the results in Figure 3.5 indicate that pass-through to the average price is not monotonically increasing in the number of sellers, in line with our theoretical prediction under imperfect information. The fractional polynomial fits for *E5* and *E10* closely resemble our simulations in Figure 3.2 with a peak at $N = 2$. For diesel, the

Figure 3.5. Average Pass-Through by Number of Competitors



Notes: The figure shows how the rate of pass-through to the average price varies with the number of competing price setters in a market. Panels (A), (C), and (E) depict the pass-through rates for the German VAT decrease on 1 July 2020 for *E5*, *E10*, and diesel, respectively. Panels (B), (D), and (F) depict the pass-through rates for the German VAT increase and introduction of a carbon price on 1 January 2021 for *E5*, *E10*, and diesel, respectively. In every panel, each circle plots the average pass-through rate for a group of stations with a particular number of competing price setters within a nonoverlapping local market. The size of a circle is proportional to the total number of stations with a given number of competitors. The long-dashed line shows a fractional polynomial fit. The short-dashed line shows a quadratic fit. The number of competitor stations is trimmed at the 97.5th percentile.

relationship between the number of sellers and the average pass-through rate has an inverted-U shape with a peak at a higher number of sellers than in the case of *E5* and *E10*.

In Appendix C.4.7, we formally test for non-monotonicity by applying the U-test of Lind and Mehlum (2010). This allows us to test the null hypothesis of a monotone or U-shaped relationship against the alternative hypothesis of an inverted-U shape. We reject the null hypothesis at the 10% level for both tax changes for diesel and *E5*. For *E10*, the estimates still speak in favour of a hump-shaped relationship, but we cannot formally reject that the pass-through rate is increasing in the number of competitors. Overall, we conclude that our empirical findings are consistent with Prediction 3 that the relationship between the number of sellers and pass-through to the expected price is non-monotonic.

3.7 Alternative Hypotheses

While our analysis shows that the imperfect consumer information model proposed by Stahl (1989) effectively accounts for the relationship between competition and tax pass-through that we observe empirically, this does not prove that there is no observationally equivalent alternative explanation for the empirical findings. In the following, we collect alternative candidate explanations and discuss how well they can explain the empirical findings.

3.7.1 Alternative Hypotheses with Full Information

Before we discuss alternative hypotheses, it is worth briefly reviewing the full-information conduct parameter approach, which is a widely used approach to modelling imperfect competition in the tax pass-through literature. A natural way of modelling competition in the retail fuel market with full information is as symmetrically differentiated Nash-in-prices. Sellers have a symmetric cost structure, offer a homogeneous good, and are located in different places, with pricing as their primary decision variable. This is a special case of the analysis presented in Weyl and Fabinger (2013).

Most of the literature that uses pass-through as a sufficient statistic for welfare results assumes that markets are perfectly competitive (Weyl and Fabinger, 2013).[30] To extend this analysis to oligopolistic markets, Weyl and Fabinger (2013) use the con-

[30] Sumner (1981) notes that the price elasticity of residual demand is a critical factor in tax pass-through for oligopolistic markets. Bulow and Pfleiderer (1983) demonstrate that the degree of pass-through depends on the functional form of demand and that, by measuring it, the curvature of demand can be determined.

duct parameter approach, which encompasses most models of oligopolistic competition with symmetric sellers and perfect information.[31] Additionally, Weyl and Fabinger (2013) extend their analysis to cases of asymmetric competition, including homogeneous product oligopoly, differentiated Nash-in-prices, and monopolistic competition with perfect information.

The conduct parameter approach features a parameter θ, which varies from 0 for perfect competition, 1 for monopoly, and $1/N$ for Cournot competition with N symmetric competitors. For full information models with symmetric competitors, including symmetrically differentiated Nash-in-prices, pass-through of a per unit tax can be expressed as

$$\rho = \frac{1}{1 + \frac{\theta}{\epsilon_\theta} + \frac{\epsilon_D - \theta}{\epsilon_S} + \frac{\theta}{\epsilon_{ms}}},$$

where ϵ_D is the price elasticity of aggregate demand, ϵ_S is the price elasticity of supply, and ϵ_{ms} is the price elasticity of marginal surplus, i.e., the curvature of demand.

Genakos and Pagliero (2022) argue that, for the retail fuel market, it is reasonable to assume that marginal costs are constant and that the conduct parameter does not vary with quantities. In case of the former, $\frac{\epsilon_D - \theta}{\epsilon_S} = 0$. In case of the latter, $\frac{\theta}{\epsilon_\theta} = 0$. In the case of symmetric competition with constant marginal costs and conduct invariant to quantities, the pass-through therefore simplifies to

$$\rho = \frac{1}{1 + \frac{\theta}{\epsilon_{ms}}}.$$

The relationship between pass-through and competition is unclear since the curvature of demand may either increase or decrease with competition.

Next, we discuss alternative hypotheses of what could cause differences in pass-through rates for retail fuel, which do not rely on imperfectly informed consumers. Some of these are nested by the conduct parameter approach, whereas others are not. We contrast their predictions with the empirical evidence.

Curvature of demand. If we analyse the pass-through equation used by Genakos and Pagliero (2022), the simplest conclusion would be that differences in pass-through between different fuel types are based on differences in the curvature of aggregate demand. Although this *may* explain differences in pass-through between fuel types, it explains neither differences in pass-through between the average price and the minimum price in a market nor a non-monotonic relationship between the number of competitors and pass-through. In fact, when we hold the curvature of demand fixed and assume that θ decreases in the number of competitors, pass-through should increase monoton-

[31] This was first introduced by Bresnahan (1989) and Genesove and Mullin (1998).

ically with the number of competitors. Furthermore, with symmetric firms and full information, there should not be any within-market price dispersion.

Vertical differentiation. One reason for within-market differences in prices and pass-through that is unrelated to imperfect information could be vertical differentiation. One could argue that some stations are just better than others (e.g., because of their location). These stations may face a lower price elasticity of residual demand even under perfect information and therefore also have lower pass-through. Combined with a mechanism of why we should expect differences in pass-through between fuel types, this *may* explain some of the empirical patterns. However, this would not explain the non-monotonic relationship between the number of competitors and pass-through or the existence of random price dispersion. Most importantly, a story of vertical differentiation would require that the ranking of stations by price within a market remain mostly unchanged. This is at odds with what we observe empirically.

Price elasticity of aggregate demand. Similarly to the curvature of demand, we can analyse the price elasticity of aggregate demand through the lens of the symmetric competition conduct parameter approach. If we relax the assumption that marginal cost is constant and allow the price elasticity of supply to be positive but below infinity, all other things being equal, a higher price elasticity of aggregate demand leads to lower pass-through. If the price elasticity of aggregate demand is lowest for diesel and highest for *E5*, this *may* explain differences in pass-through between fuel types without requiring imperfect information. However, as with the curvature of demand, it does not explain differences in pass-through between the average and the minimum price, random price dispersion, or a non-monotonic relationship between the number of competitors and pass-through.

Price sensitivity unrelated to information. An alternative approach would be to let consumers differ in how strongly they react to lower prices (i.e., their price sensitivity) while assuming that they are perfectly informed. Even with perfect information, consumers may react to a lower price only if it is worth their time (e.g., because the lower-price station is further away). Purchasers of different fuel types may differ in their price sensitivity, which may affect competitive conduct. This could be modelled, for example, as the price coefficient in a logit demand model. Combined with firms competing in prices, the higher the price sensitivity of consumers, the higher is the intensity of competition and the tax pass-through. Without imperfect information, however, frequent changes in the ranking of stations by price, random price dispersion, and a non-monotonic relationship between the number of competitors and pass-through remain to be explained.

Overall, although the empirical findings can often be partially explained by alternative hypotheses that do not rely on imperfect information, the results in their entirety are inconsistent with any of these alternative hypotheses.

3.7.2 Alternative Hypotheses with Imperfect Information

Another set of alternative hypotheses relies on modelling competition under imperfect information differently. We briefly discuss what we consider to be the most obvious alternative modelling choices.

Rockets and feathers. Tappata (2009) proposes a dynamic model to explain the 'rockets and feathers' phenomenon – prices rising faster than they fall – via consumer search and cost uncertainty. In this model, atomistic consumers have unit demand and value the good at v. Consumers have the option to purchase access to an information clearinghouse, which allows them to observe all market prices, or they can choose to draw a single price at random. Some consumers have zero access costs and are always perfectly informed, while others draw access costs from a continuous distribution. This model is a variant of Varian (1980) in which the decision to become informed is endogenized.

Marginal costs in this model can be high or low, and they follow a first-order Markov process. Firms use mixed strategies to set prices. When production costs are high, the gap between marginal cost and v is narrow, resulting in low price dispersion and limited search gains. Conversely, if production costs are low, the gap between marginal cost and v is large, leading to high price dispersion and greater search gains. As cost decreases are possible only when marginal costs are high, they occur during periods of low search, resulting in slow pass-through. In contrast, cost increases are quickly passed through when marginal costs are low, prices are low, and search is high.

For this mechanism to explain the empirical findings, price dispersion, a measure for the gains from search, needs to be higher when prices are low. However, in Appendix C.1.4, we show that this is not the case. Since the basic premise of the model does not apply in our empirical application, modelling competition under imperfect information in this way is unlikely to better explain the empirical findings.

Reference price dependence. In a duopoly market where costs are unobservable to consumers, Lewis (2011) finds that whether consumers search depends on how the first price they draw compares to a reference price (e.g., the previous period's price). A positive cost shock increases the probability of the first price exceeding the reference price, inducing more search and higher pass-through. Conversely, a negative cost shock reduces search and lowers pass-through.

The model's main finding is that pass-through is faster for cost increases and slower for cost decreases. However, there are several drawbacks to analysing our empirical application through the lens of this model. First, since it is a duopoly model, it does not allow us to analyse the relationship between pass-through and the number of competitors. Second, the consumer search protocol is suboptimal, as it is unrelated to the *actual* gains from search, and only applicable if cost shocks – tax changes in our application – are unobservable and a surprise to consumers. Finally, unlike the German fuel market, where price cycles occur intraday and are unrelated to cost, price cycles result from cost shocks, similarly to what happens in Tappata (2009).

Unobservable input costs. Whereas we treat input costs (i.e., mainly the price of oil and tax rates) as observable to consumers, an alternative would be to treat these costs as unobservable. Janssen et al. (2011) extend the Stahl (1989) model to include unobservable input prices while assuming an exogenous share of shoppers. Janssen and Shelegia (2015) extend this to a vertical market, where an upstream manufacturer sets the input price. They find that lower sequential search costs for non-shoppers lead to less elastic upstream demand, incentivizing the manufacturer to reduce downstream retailer profits and resulting in higher prices than those under a vertically integrated monopolist.[32]

Our empirical application differs from this setting in several ways. First, vertical integration is prevalent in the industry. Second, upstream prices, represented by the oil price, are more transparent than prices in other industries. Even if the oil price is not observed by consumers on a daily basis, the cost shocks analysed in this chapter (i.e., the sizeable tax changes) were broadly publicized and salient for consumers. In addition, the timing, direction, and magnitude of the tax changes were well-known. It therefore seems unlikely that treating input costs as unobservable would increase the explanatory power of the model.

Although one can make the case for modelling imperfect information in the retail fuel market differently, none of the most obvious alternative ways of modelling competition better explain our empirical results. Therefore, we conclude that our empirical findings in their entirety are best explained by the imperfect consumer information model of Stahl (1989).

[32] Janssen et al. (2011) and Janssen and Shelegia (2015) both find that, in the presence of unobservable input costs or upstream prices, downstream prices are higher. However, neither has predictions about pass-through. Janssen and Shelegia (2020) study pass-through with imperfect information and differentiated products.

3.8 Conclusion

In this chapter, we highlight the role of imperfect information in explaining heterogeneities in tax pass-through. We show that when consumers do not know all prices, firms have market power, and this affects tax pass-through. While our approach imposes more structure on conduct and demand than that of Weyl and Fabinger (2013), this allows us to be more flexible in modelling consumer information.

Three results stand out and set our analysis apart from one with perfect information. First, the more well-informed consumers there are, the higher is tax pass-through. Second, taxes (and tax cuts) are passed through more to the price paid by well-informed consumers than to the price paid by uninformed consumers. Third, there is no monotonic relationship between the number of sellers and pass-through.

The results of this study have important implications for policy. When markets are imperfectly competitive, pass-through affects the effectiveness of unconventional fiscal policies (D'Acunto et al., 2018, 2022), how strongly prices react to Pigouvian taxes, and the distributional consequences of such policies. In the case of unconventional fiscal policy, its effectiveness hinges on consumers expecting that a future tax increase will be passed through to them. Similarly, in the case of Pigouvian taxes, consumers react by reducing quantities only if the tax is passed through. These considerations, as well as uncertainty about the level of consumer information, can affect the relative benefits of regulation versus Pigouvian taxes or subsidies.

We shed light on a novel explanation for what determines tax pass-through, which is relevant for any market where it is costly for consumers to learn about prices. Future research should aim to develop tools to estimate the parameters of demand and supply models with imperfectly informed consumers. This would allow researchers to estimate how a proposed tax or subsidy would affect prices and quantities even in the presence of firms with market power and consumers who are imperfectly informed about prices.

Appendix A

Appendix to Chapter 1

A.1 Supplementary Results for Class-Level Approach

This appendix contains supplementary results for my main class-level approach to estimate the effect of compulsory licensing of Xerox's patents on cumulative innovation. First, I present additional descriptives of the dataset. Then, I show robustness checks for my main results and heterogeneity analyses.

A.1.1 Data Description

Table A.1 presents additional information on Xerox's portfolio of compulsorily licensed patents by showing the top ten four-digit CPC technology classes. Additional summary statistics for the outcome variables used in the main part of the chapter are reported in Table A.2.

Table A.3 reports additional summary statistics for the treatment variable (Share_s) and the measure of closeness to Xerox (Closeness_i) defined in equation (1.4). Summary statistics for Share_s are shown for the 313 treated six-digit subclasses in the main sample, whereas those for Closeness_i are shown for the 1,635 firms in the sample of established firms. In the average treated subclass, 2.7% of the unexpired patents (as of 1975) were subject to compulsory licensing. In contrast, in the subclass most exposed to the antitrust case, 58.5% of the unexpired patents became available for licensing. As Share_s is defined in percentage terms, both variables can, in theory, take on values between 0 and 100. For the closeness measure, a value of 100 would indicate that all of a firm's unexpired patents were in a subclass in which all patents were compulsorily licensed. As no such technology class exists, the closeness measure has a maximum of

Table A.1. Xerox's Patent Portfolio Subject to Compulsory Licensing

4-Digit Classes		6-Digit Subclasses			
Top 10		Largest		Other	
Code	Title	Code	Weight	Number	Weight
G03G	ELECTROGRAPHY; ELECTROPHOTOGRAPHY; MAGNETOGRAPHY	G03G 15	40.7%	10	28.8%
G03B	APPARATUS OR ARRANGEMENTS FOR TAKING PHOTOGRAPHS OR FOR PROJECTING OR VIEWING THEM	G03B 27	3.0%	6	0.4%
B65H	HANDLING THIN OR FILAMENTARY MATERIAL, e.g. SHEETS, WEBS, CABLES	B65H 3	0.6%	12	1.6%
H04N	PICTORIAL COMMUNICATION, e.g. TELEVISION	H04N 1	1.3%	3	0.4%
H01J	ELECTRIC DISCHARGE TUBES OR DISCHARGE LAMPS	H01J 29	0.4%	10	1.0%
G02B	OPTICAL ELEMENTS, SYSTEMS OR APPARATUS	G02B 27	0.3%	9	1.0%
G06K	GRAPHICAL DATA READING	G06K 15	0.5%	4	0.4%
H01L	SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES NOT OTHERWISE PROVIDED FOR	H01L 31	0.4%	6	0.5%
G03C	PHOTOSENSITIVE MATERIALS FOR PHOTOGRAPHIC PURPOSES	G03C 1	0.5%	2	0.2%
B41M	PRINTING, DUPLICATING, MARKING, OR COPYING PROCESSES; COLOUR PRINTING	B41M 5	0.5%	3	0.3%

Notes: The table presents additional information on the top ten four-digit CPC technology classes in Xerox's portfolio of compulsorily licensed patents. The first and second column indicate the code and title of the ten largest four-digit classes, measured by the number of patents. The third column provides the code of the largest six-digit subclass and the fourth column shows the weight of this subclass in Xerox's patent portfolio. The last two columns indicate the number and weight of the remaining six-digit subclasses within the given four-digit class.

Table A.2. Class-Level Analysis: Summary Statistics

	Mean	SD
(A) Main Outcome		
# Patents	15.132	28.153
(B) Citations to Xerox		
Distance = 1	0.213	2.909
Distance $\leq$ 2	0.582	5.071
Distance $\leq$ 3	1.124	6.834
Unconnected to Xerox	14.008	25.348
(C) Applicant Country		
USA	8.934	16.889
Non-USA	5.742	12.252
Japan	2.251	7.708
Others	3.491	6.576
(D) Firm Sample		
Established Firms	8.572	19.249
Other Patentees	6.670	12.064
(E) Closeness to Xerox		
Top Decile of Closeness	1.607	7.568
75-90th Percentile	2.167	6.476
Below 75th Percentile	4.736	11.394
(F) Quality		
Top 10% of Forward Citations	1.603	4.649
Top 10% of Quality (KPST)	2.000	8.921
Mean of Forward Citations	14.144	12.392
Mean of Quality (KPST)	-0.005	0.114
(G) Active Subgroups		
All	4.790	6.196
USA	3.576	4.920
Non-USA	2.618	3.804
Japan	1.137	2.443
Others	1.964	2.811

Notes: The table presents summary statistics for the outcome variables used in the main part of the chapter. Panel (A) refers to the baseline specification, where the outcome is the number of patent applications by firms other than Xerox per six-digit CPC subclass per year. The variables depicted in panels (B) to (G) correspond to the results from Figure 1.4, Table 1.2, Figure 1.7, Table 1.3, and Table 1.4, respectively. There are 35,360 observations for each variable (i.e., 2,210 six-digit subclasses observed in 16 years).

33.5. The mean value is only around 0.3, indicating that most firms in the sample of established firms had very little exposure to copier technologies.

Table A.3. Class-Level Analysis: Summary Statistics for Share$_s$ and Closeness$_i$

	Observations	Mean	SD	Min	Max
Share$_s$	313	2.746	7.440	0.027	58.519
Closeness$_i$	1,635	0.339	1.904	0.000	33.483

Notes: The table presents summary statistics for the main treatment variable (Share$_s$) and the measure of closeness to Xerox (Closeness$_i$) defined in equation (1.4). As Share$_s$ is defined in percentage terms (i.e., $\times$ 100), both variables can, in theory, take on values between 0 and 100.

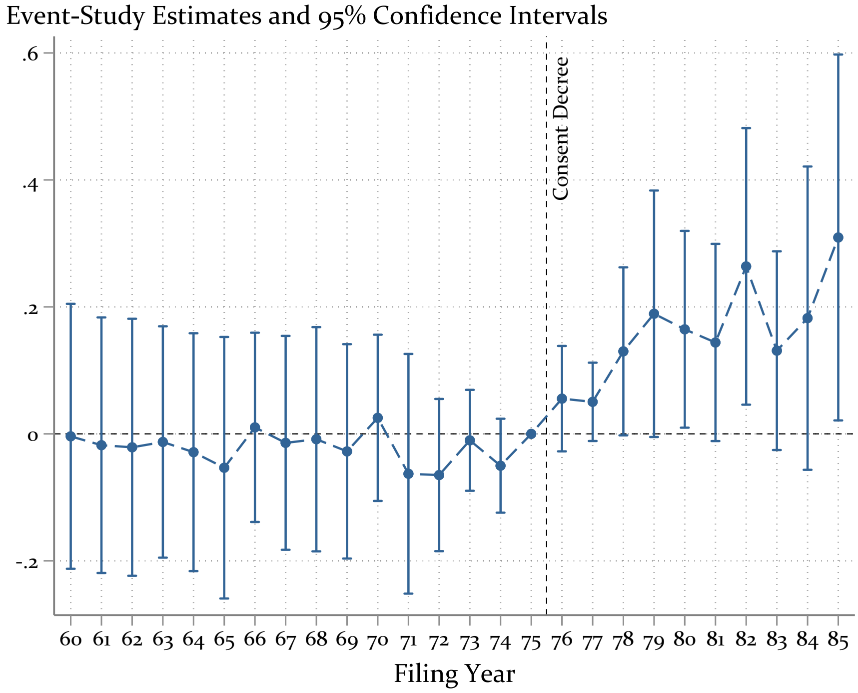
A.1.2 Extended Pre-Treatment Period

Figure A.1 reproduces the event-study analysis from my main class-level approach (see Figure 1.3 in the main part of the chapter), but it presents annual estimates and covers an extended pre-treatment period back to 1960. The estimates show that, on average, there were no significant differences in patenting across differentially exposed six-digit subclass in the 15 years prior to the consent decree. This further supports the identifying assumption underlying my class-level approach. After compulsory licensing of Xerox's patents in 1975, however, there was a disproportionate increase in patenting in technologies where most of Xerox's patents became available for licensing.

A.1.3 Model Specification

In Table A.4, I analyse the robustness of my baseline estimate by employing alternative model specifications. Column (1) repeats the baseline estimate from the main part of the chapter. In column (2), I estimate the model without using the coarsened exact matching (CEM) weights by Iacus et al. (2012), which does not change the point estimate by much. In column (3), I estimate a Poisson pseudo-likelihood regression instead of an ordinary least squares model, using the algorithm by Correia et al. (2020). This alternative non-linear estimation model takes into account that patent applications represent count data, but it comes at the sacrifice of the point estimate's simple (linear) interpretation. Reassuringly, the point estimate is still positive and highly statistically significant. Columns (4) to (6) employ alternative treatment definitions. In column (4), I apply the inverse hyperbolic sine (IHS) transformation to the share of compulsorily licensed patents per subclass. Again, the estimate remains positive and statistically significant. This addresses potential concerns that the underlying relationship between the treatment variable and the number of patent applications may be non-linear. The estimates in columns (5) and (6) further show that my results are robust to using a binary treatment measure or defining treatment based on the number (as opposed to the share) of compulsorily licensed patents. These alternative treatment measures are further discussed below in appendix A.1.4. Finally, in column (7) of Table A.4, I exclude

Figure A.1. Class-Level Analysis: Event-Study Estimates with Extended Pre-Period



Notes: The figure depicts point estimates and 95% confidence intervals from the event-study analysis in equation (1.2), using an extended pre-treatment period that covers years since 1960. The regression uses the weights by Iacus et al. (2012). Standard errors are clustered at the four-digit CPC technology class level.

the top 5% of subclasses with the greatest share of compulsorily licensed patents (i.e., the treatment variable) from the sample. This is to verify that the positive baseline estimate is not purely driven by patenting in these subclasses that were highly exposed to compulsory licensing. Again, the point estimate remains statistically significant at the 5% level. Additional robustness checks on the sample definition are reported below in appendix A.1.5.

A.1.4 Treatment Definition

Next, I turn to assessing the robustness of the treatment definition. In the first step, I employ a simple binary specification and show the corresponding event-study estimates in Panel (A) of Figure A.2. The estimates reveal no statistically significant pre-trends and indicate that there was a disproportionate increase in patenting after 1975 in technologies where at least one Xerox patent became available for licensing. The DiD estimate in column (5) of Table A.4 confirms this visual impression. The results are equally robust to defining an alternative binary treatment that considers all subclasses as treated that contain at least two (as opposed to one) compulsorily licensed Xerox patents. This is not surprising in light of the pattern shown in panel (B) of Figure 1.2 in

Table A.4. Class-Level Analysis: Alternative Model Specifications

	Model			Treatment			Outliers
	Baseline	No CEM Weights	Poisson	IHS of Share	Binary	IHS of Abs. No.	Excl. Top 5% Subcl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\text{Share}_s \cdot \text{Post}_t$	0.189**	0.169*	0.033***				0.666**
	(0.094)	(0.087)	(0.011)				(0.318)
$\text{IHS}(\text{Share}_s) \cdot \text{Post}_t$				1.760**			
				(0.806)			
$\mathbb{1}[\text{Lic}_s \geq 1] \cdot \text{Post}_t$					2.797*		
					(1.527)		
$\text{IHS}(\text{Lic}_s) \cdot \text{Post}_t$						2.974**	
						(1.430)	
Subclass FE	✓	✓	✓	✓	✓	✓	✓
Year × Class FE	✓	✓	✓	✓	✓	✓	✓
CEM Weights	✓			✓	✓	✓	✓
Mean of Outcome	15.13	15.13	15.13	15.13	15.13	15.13	15.06
4-Digit CPC Classes	141	141	141	141	141	141	140
Observations	35,360	35,360	34,585	35,360	35,360	35,360	35,088

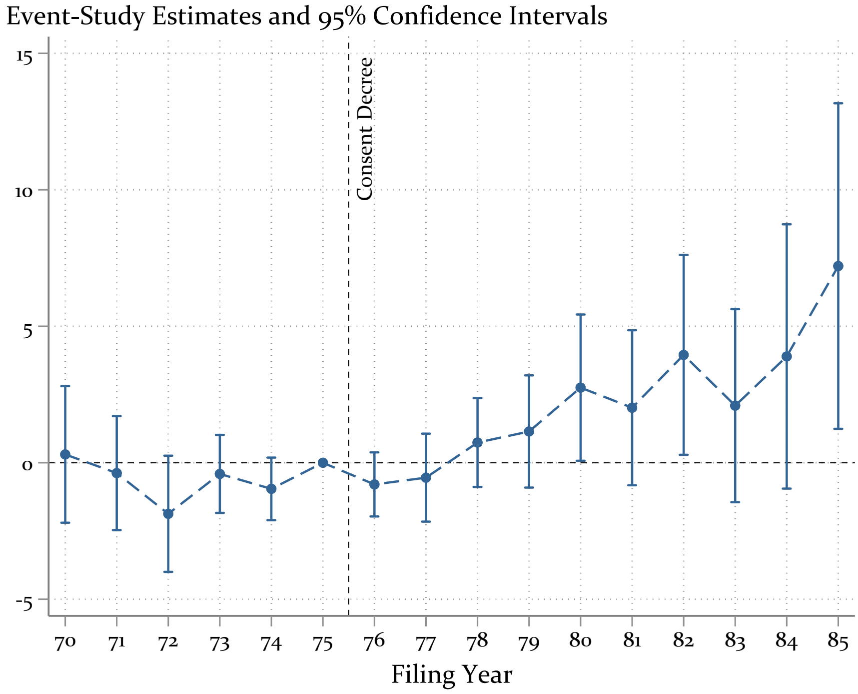
Notes: The table shows the results from difference-in-differences regressions following equation (1.1). The outcome variable in all regressions is the number of patent applications by firms other than Xerox in a given six-digit CPC subclass and year. Column (1) replicates the baseline estimate from Table 1.1 in the main part of the chapter. In column (2), the model is estimating without using the weights by Iacus et al. (2012). In column (3), I estimate a Poisson pseudo-likelihood regression instead of a linear ordinary least squares model. In column (4), I apply the inverse hyperbolic sine (IHS) transformation to the share of compulsorily licensed patents per class (i.e., the treatment variable). Column (5) employs a binary treatment indicator that equals one for subclasses with at least one licensable Xerox patent. In column (6), I use a an alternative continuous treatment variable defined as the inverse hyperbolic sine of the number of compulsorily licensed Xerox patents, denoted by the variable Lic_s. Finally, column (7) employs the baseline model specification but excludes the top 5% of six-digit subclasses with the highest share of compulsorily licensed patents from the sample. Standard errors clustered at the four-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

the main part of the chapter, indicating that subclasses with exactly one compulsorily licensed Xerox patent experienced a similar patenting trend as subclasses with no exposure to compulsorily licensing. However, one drawback of the alternative binary specification is that the number of treated subclasses is reduced to 122 and, as a consequence, there are only 66 clusters on the four-digit class level.

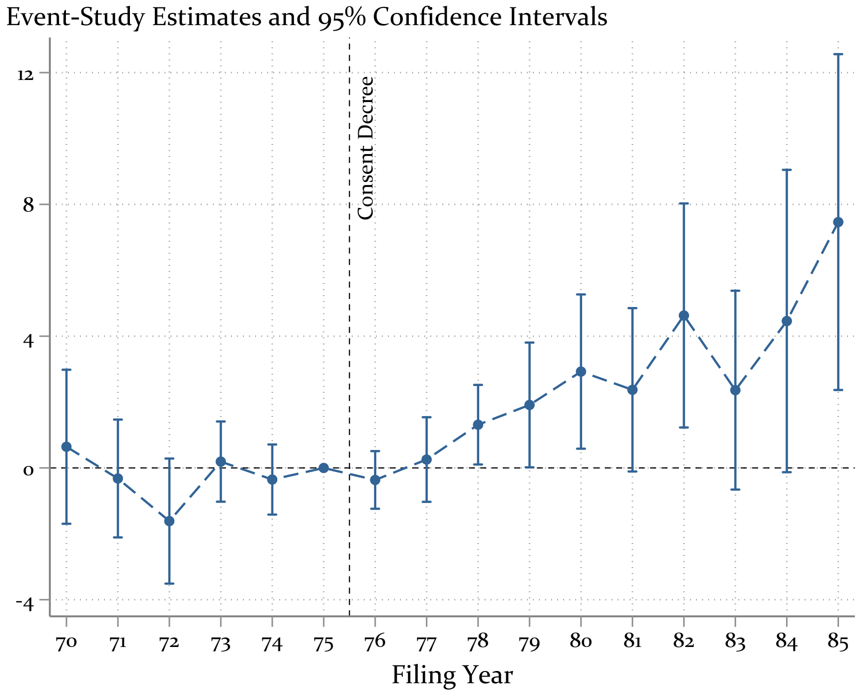
To avoid making an arbitrary choice which subclasses should be considered as treated, my main empirical approach employs a continuous treatment specification. As explained in the main part of the chapter, I use the variable Share_s as treatment measure, which captures the share of unexpired patents per subclass (as of 1975) that were subject to compulsorily licensing. In Panel (B) of Figure A.2, I further show that my results do not hinge on the specific definition of the continuous treatment variable. I now define the treatment variable as the inverse hyperbolic sine of the absolute number of compulsorily licensed Xerox patents per subclass (Bellemare and Wichman, 2020). This non-linear transformation is necessary to ensure that the DiD estimate is

Figure A.2. Class-Level Analysis: Event-Study Estimates with Alternative Treatments

(A) Binary Treatment

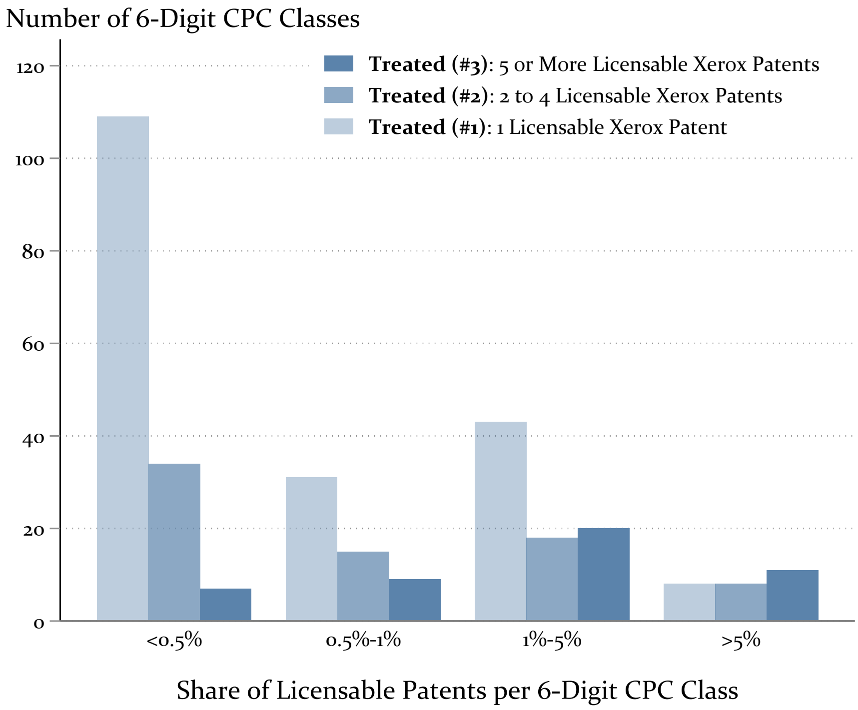


(B) Number of Licensable Patents



Notes: The figure depicts point estimates and 95% confidence intervals from two variations of the event-study analysis in equation (1.2). Panel (A) employs a binary treatment specification, where a six-digit subclass is treated if it contains at least one compulsorily licensed Xerox patents. Panel (B) uses an alternative continuous treatment specification, where the treatment variable is defined as the inverse hyperbolic sine of the number of compulsorily licensed Xerox patents per subclass. All regressions use the weights by Iacus et al. (2012). Standard errors are clustered at the four-digit CPC technology class level.

Figure A.3. Class-Level Analysis: Share vs. Number of Licensable Patents



Notes: The figure depicts the relationship between the share of unexpired patents per subclass that were subject to compulsorily licensing (i.e., the treatment variable) and the absolute number of licensable patents per subclass. It shows the number of subclasses in the sample (on the vertical axis) that have a given share variable (on the horizontal axis), which is discretized into four groups. The figure presents separate bars for subclasses with (#1) one, (#2) two to four, and (#3) five of more compulsorily licensed Xerox patents. Example: there are approximately 110 subclasses with one compulsorily licensed Xerox patent whose share variable is below 0.5%.

not driven by the single subclass G03G 15, which represents an extreme outlier in terms of the number (but not in terms of the share) of licensable Xerox patents. As can be seen in Panel (B) of Figure A.2, the pattern of increased patenting in those subclasses that were more strongly exposed to compulsory licensing of Xerox's patents is robust to this alternative treatment definition. Similarly, the point estimate in column (6) of Table A.4 remains positive and statistically significant.

Figure A.3 shows the relationship between the main treatment variable and the absolute number of compulsorily licensed Xerox patents per subclass. The figure uses the same subgroups of subclasses as in Table 1.1 or panel (B) of Figure 1.2 in the main part of the chapter. According to Figure A.3, on average, subclasses with a larger absolute number of licensable Xerox patents also have a higher share of compulsorily licensed patents. However, there is important variation in the share variable even within a subgroup, which stems from differences in the total number of unexpired patents across subclasses (i.e., the denominator of the share variable). Arguably, these differences con-

Table A.5. Class-Level Analysis: IPC Classes

	(1)	(2)	(3)
$\text{Share}_s \cdot \text{Post}_t$	0.156**	0.204***	
	(0.063)	(0.059)	
$\text{Share}_s \cdot \text{Post}_t \cdot \mathbb{1}[\text{Lic}_s = 1]$			0.009
			(0.029)
$\text{Share}_s \cdot \text{Post}_t \cdot \mathbb{1}[2 \leq \text{Lic}_s \leq 4]$			0.052
			(0.032)
$\text{Share}_s \cdot \text{Post}_t \cdot \mathbb{1}[5 \leq \text{Lic}_s]$			0.288***
			(0.063)
Additional Patents per Year	140	182	141
Relative Increase	1.2%	1.5%	1.2%
Subclass FE	✓	✓	✓
Year FE	✓		
Year × Class FE		✓	✓
Mean of Outcome	16.28	16.28	16.28
No. of 6-Digit IPC Classes	2,152	2,152	2,152
No. of 4-Digit IPC Classes	137	137	137
Observations	34,432	34,432	34,432

Notes: The table shows the results from difference-in-differences regressions following equation (1.1). The outcome variable in all regressions is the number of patent applications by firms other than Xerox in a given six-digit IPC (instead of CPC) subclass and year. In column (3), the treatment variable is interacted with indicators for subclasses with one, two to four, and five or more compulsorily licensed Xerox patents, as indicated by the variable Lic_s. Standard errors clustered at the four-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

tain meaningful information about the exposure of a subclass to compulsory licensing. Therefore, I employ the variable Share_s as my main treatment measure.

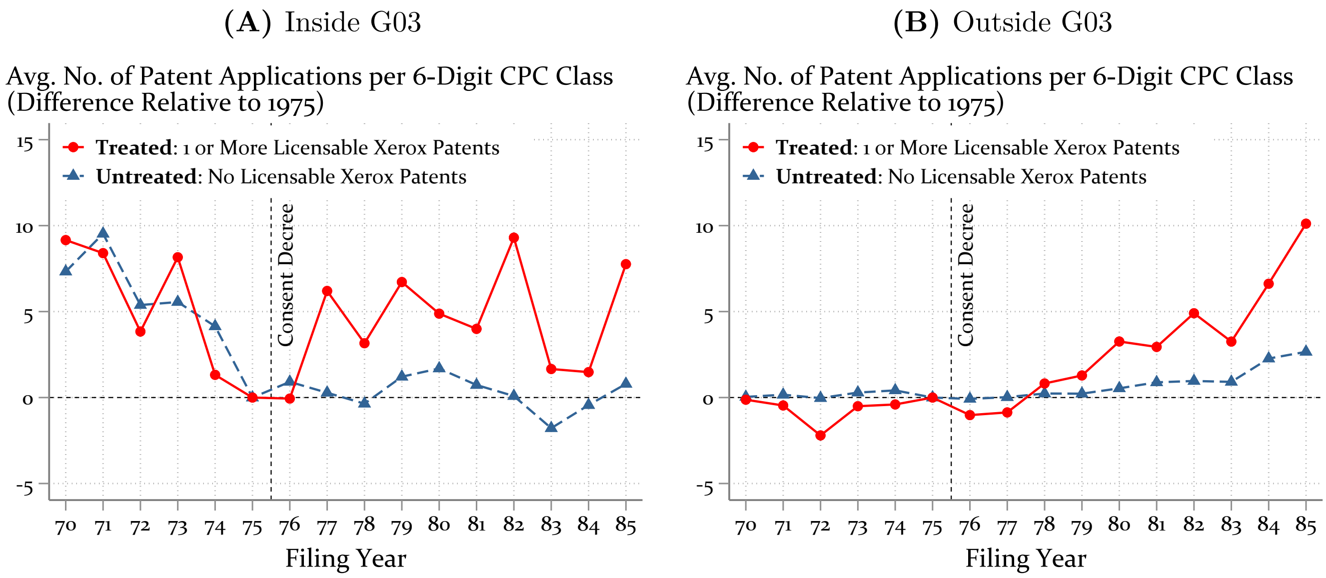
Finally, Table A.5 presents estimation results when using technology classes based on the International Patent Classification (IPC). The table shows that using IPC (instead of CPC) classes yields very similar estimates both in terms of magnitude and statistical significance.

A.1.5 Sample Definition

I now analyse whether my estimates are robust to different sample definitions. As noted above in appendix A.1.3, the estimate in column (7) of Table A.4 indicates that my estimates remain positive and statistically significant even after excluding the top ten subclasses with the greatest share of compulsorily licensed patents. That is, the positive effect is not driven by outliers with high values of the treatment variable.

In the next step, I more closely examine the role of the three-digit class G03, which covers photography and electrography among other technologies. As shown

Figure A.4. Class-Level Analysis: Patenting Inside vs. Outside Class G03

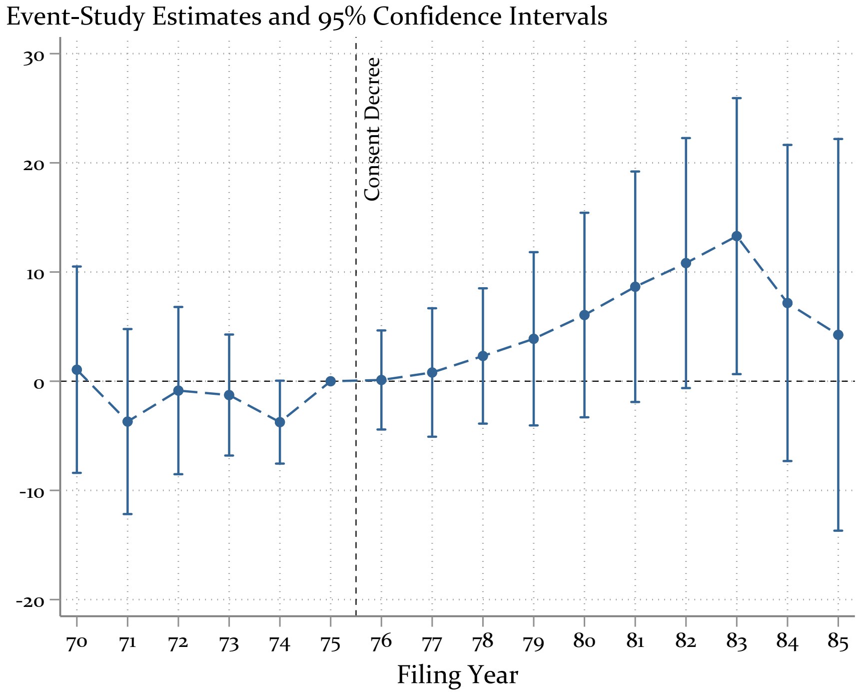


Notes: The figure depicts the average number of patent applications per six-digit subclass relative to 1975. Averages are computed separately for treated and untreated subclasses, where a subclass is defined as treated if it contained at least one compulsorily licensed Xerox patent. Panel (A) only includes subclasses that are part of the three-digit CPC class G03. Panel (B) includes all remaining subclasses in the sample. In both panels, the subclasses are aggregated using the weights by Iacus et al. (2012).

in Table A.1, a large fraction of Xerox's compulsorily licensed patents is clustered in this technology class. Figure A.4 presents descriptive evidence indicating that the post-1975 increase in patenting in subclasses exposed to compulsory licensing is present both inside and outside the three-digit class G03. Panel (A) of Figure A.4 depicts the average number of patent applications separately for treated and untreated subclasses inside class G03, while Panel (B) reports the averages for all remaining technologies. As is evident from the figure, even inside class G03, there was a relative increase in patenting in subclasses exposed to compulsory licensing of Xerox's patents. However, the averages in Panel (A) are relatively noisy due to the small number of observations. As shown in Panel (B), patenting also increased after 1975 in treated subclasses belonging to three-digit classes other than G03. However, it is noteworthy that the increase inside G03 is quite pronounced as soon as 1977, whereas the effect in the remaining technology fields evolves more progressively. This may point to the presence of spillovers across technologies over time.

In a second robustness check, I follow Watzinger et al. (2020) and aggregate subclasses into one treated and one untreated subclass per four-digit class. The advantage of this aggregation is that the resulting regression model gives equal weight to every four-digit class. In contrast, in my baseline approach, by using the weights of Iacus

Figure A.5. Class-Level Analysis: Event-Study Estimates with Aggregated Subclasses



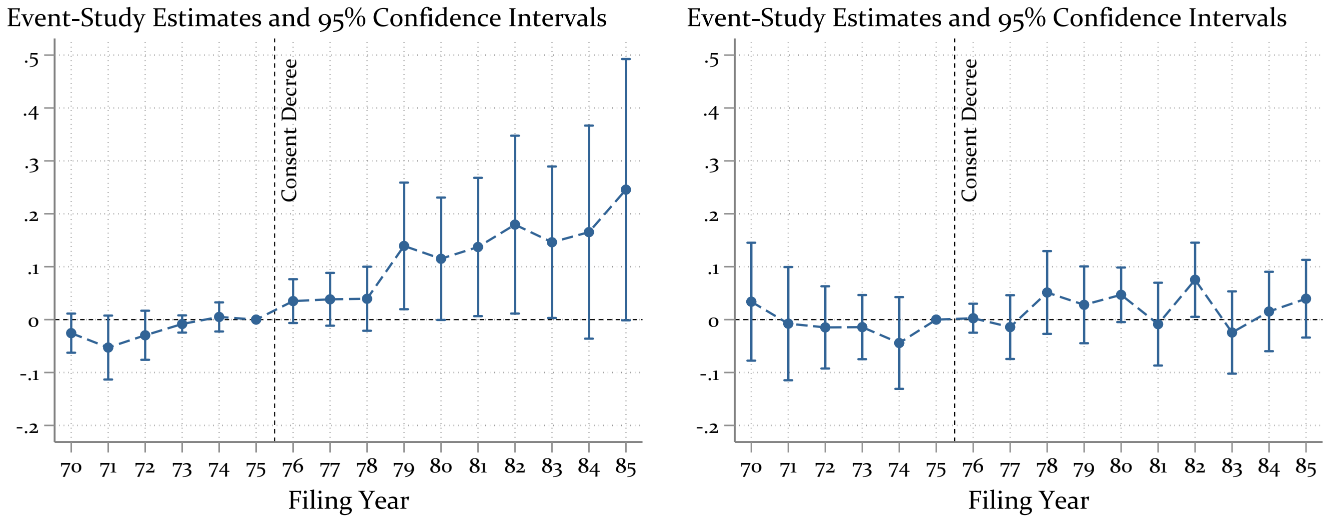
Notes: The figure depicts point estimates and 95% confidence intervals from an event-study analysis akin to equation (1.2), but where subclasses are aggregated into one treated and one untreated subclass per class. A six-digit subclass is defined as treated if it contained at least one compulsorily licensed Xerox patent. Standard errors are clustered at the four-digit CPC technology class level.

et al. (2012), the regression gives equal weight to every treated six-digit subclass, implying that classes with a larger number of treated subclasses get a higher weight.

Figure A.5 depicts the event-study estimates with aggregated subclasses. The estimates are based on a binary treatment specification, where a subclass is treated if it contained at least one compulsorily licensed Xerox patent. Consistent with my baseline results, there was a disproportionate increase in the number of patent application after 1975 in (aggregated) subclasses where Xerox patents became available for licensing. However, this positive effect is statistically significant at the 5% level only in 1983 and then fades out again. One reason for the lower precision of these estimates could be the binary treatment specification, which does not exploit all the available information in terms of the subclasses' exposure to compulsory licensing. Yet, reassuringly, the figure shows no meaningful pre-trends, which gives further credibility to my identification strategy.

Figure A.6. Class-Level Analysis: Event-Study Estimates by Applicant Country

(A) Japan **(B)** US



Notes: The figure depicts point estimates and 95% confidence intervals from the event-study analysis in equation (1.2). In panels (A) and (B), the outcome variable is restricted to patent applications by assignees from Japan and the US, respectively. All regressions use the weights by Iacus et al. (2012). Standard errors are clustered at the four-digit CPC technology class level.

A.1.6 Heterogeneity by Applicant Country

This section presents additional results of my heterogeneity analysis by applicant country. Figure A.6 depicts event-study estimates separately for US and Japanese patent applicants. These estimates correspond to the simple DiD estimates in columns (2) and (4) of Table 1.2. The figure shows that there were no significant pre-trends in either country, which again supports the identification strategy underlying my empirical approach. After 1975, technology classes with a greater exposure to compulsory licensing of Xerox's patents experienced an increase in the number of patent applications only by Japanese applicants. The point estimates for Japan are larger than those for the US throughout the post-treatment period. However, standard errors are also larger, because the overall number of Japanese patent applications was lower.

In Table A.6, I show results from a robustness check that uses data on patent applications at the Japanese Patent Office (JPO), using data from the Japanese Institute of Intellectual Property (Goto and Motohashi, 2007). In contrast, my main results are based on patenting at the USPTO. This robustness check addresses the concern that increased patenting by Japanese firms in the US may not necessarily represent novel innovation; it may also reflect Japanese copier producers seeking protection for existing technologies abroad. The set-up of Table A.6 is analogous to that of Table 1.2 in the main part of the chapter. The only difference is that the analysis with JPO data uses

Table A.6. Class-Level Analysis: Data from Japanese Patent Office

		Applicant Country			
	Baseline	USA	Non-USA	Among Non-USA	
				Japan	Others
	(1)	(2)	(3)	(4)	(5)
$\text{Share}_s \cdot \text{Post}_t$	0.195*	-0.004	0.235*	0.234*	0.002
	(0.105)	(0.008)	(0.136)	(0.136)	(0.002)
Mean of Outcome	24.5	1.6	22.9	21.1	1.8
No. of 6-Digit IPC Classes	1,392	1,392	1,392	1,392	1,392
No. of 4-Digit IPC Classes	117	117	117	117	117
Observations	22,272	22,272	22,272	22,272	22,272

Notes: The table shows the results from difference-in-differences regressions following equation (1.1). The underlying data in this table are patent applications at the Japanese Patent Office (Goto and Motohashi, 2007). The treatment definition is based on IPC (instead of CPC) classes. Column (1) reports the baseline estimates. In columns (2) to (5), the outcome variable is restricted to patent applications filed by assignees from selected countries. All regressions include subclass and year × class fixed effects. Standard errors clustered at the four-digit IPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

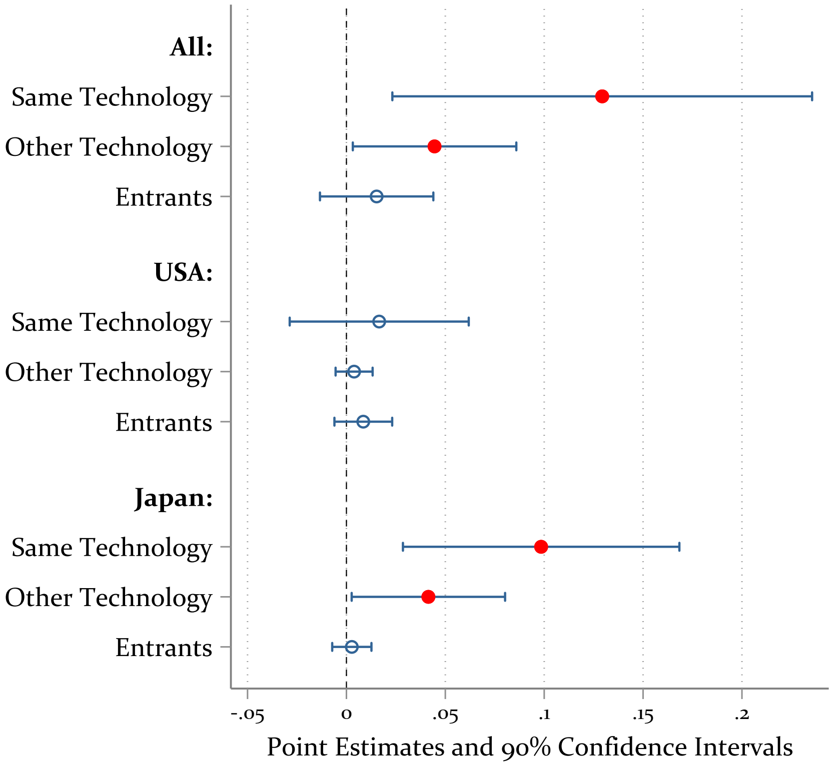
a treatment definition based on IPC (instead of CPC) classes, as the Japanese patent data only report IPC classes. However, as shown above in appendix A.1.4, my main results are also robust to using IPC classes.

Overall, the DiD estimates in Table A.6 support the interpretation that the increase in Japanese patenting after 1975 represents novel innovation. The baseline estimate in column (1) is positive and statistically significant at the 10% level. Again, this positive effect is almost entirely driven by increased patenting by Japanese firms, as indicated by column (4). In contrast, the estimates for patenting by applicants from the US or other countries are small in magnitude and statistically indistinguishable from zero.

A.1.7 Heterogeneity by Prior Patenting Experience

Figure 1.7 in the main part of the chapter shows that the observed increase in patenting after 1975 is primarily driven by firms with previous patenting experience in technologies related to Xerox. Figure A.7 shows estimates from a complementary heterogeneity analysis that considers whether firms had prior patenting experience and in which technologies they patented. The first row shows the effect on patenting by firms with previous experience in the same technology. More specifically, in the first row, the outcome variable only counts those patents whose assignee had filed a patent until 1970 in the same-three digit technology class as the focal subclass. In contrast, the second and third row depict estimates of the effect on patenting by firms with previous experience in other fields and by firms without any patenting experience (i.e., entrants), respectively.

Figure A.7. Class-Level Analysis: Alternative Heterogeneity by Patenting Experience



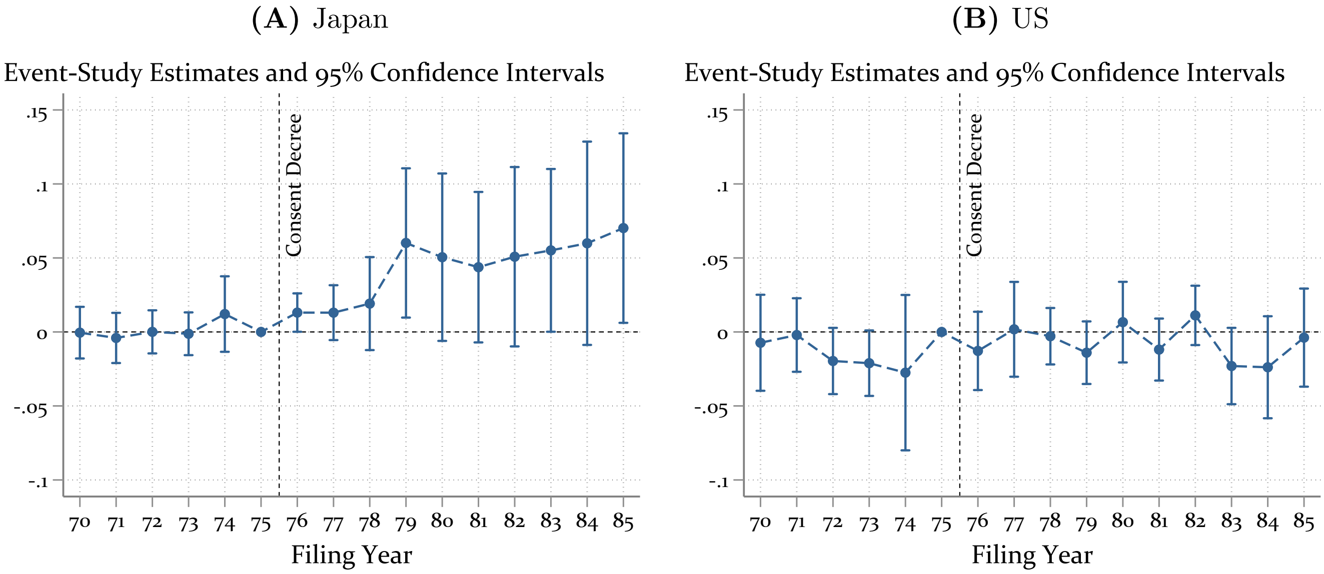
Notes: The figure depicts point estimates and 90% confidence intervals from estimating the regression model in equation (1.1). The outcome variable (i.e., the number of patent application) is split by applicant country as well as by whether the assignee had already filed a patent by 1970 and to which technology class this patent belonged. A previous patent is considered to be of the 'same' ('other') technology if it belongs to the same (another) three-digit CPC class as the focal subclass. Patents by assignees without any patenting experience until 1970 are labelled as 'entrant' patents. Red dots (blue circles) indicate statistically significant (insignificant) point estimates. Standard errors are clustered at the four-digit CPC technology class level.

The results again indicate that the observed increase in the number of patent applications after 1975 is due to firms with previous patenting experience in technologies related to Xerox. The remainder of Figure 1.7 repeats the analysis separately for applicants from the US and Japan. The estimates highlight that, even within Japan, the positive innovation effect is driven by patent applications by firms that previously patented in treated technologies. Overall, these results are consistent with those from Figure 1.7 in the main part of the chapter, where patents are split by firms.

A.1.8 Diversity and Direction of Innovation

This section presents additional results on the diversity and direction of innovation. In the main part of the chapter, I follow Watzinger and Schnitzer (2022) and look at the number of active technology subgroups to study the effect of compulsory licensing

Figure A.8. Class-Level Analysis: Event-Study Estimates for Active Subgroups



Notes: The figure depicts point estimates and 95% confidence intervals from a variation of the event-study analysis in equation (1.2). Unlike in the main approach, the outcome variable now is the number of 'active' subgroups (aggregated to two dots) within a six-digit technology class. In panels (A) and (B), only patent applications by assignees from Japan and the US, respectively, are counted to determine whether a subgroup was 'active'. All regressions use the weights by Iacus et al. (2012). Standard errors are clustered at the four-digit CPC technology class level.

on the diversity of innovation. Figure A.8 presents the corresponding event-study estimates separately for patents by applicants from Japan and the US in panels (A) and (B), respectively. The figure shows that the number of active subgroups with at least one Japanese patent application increased after 1975, but there were no significant pre-trends.

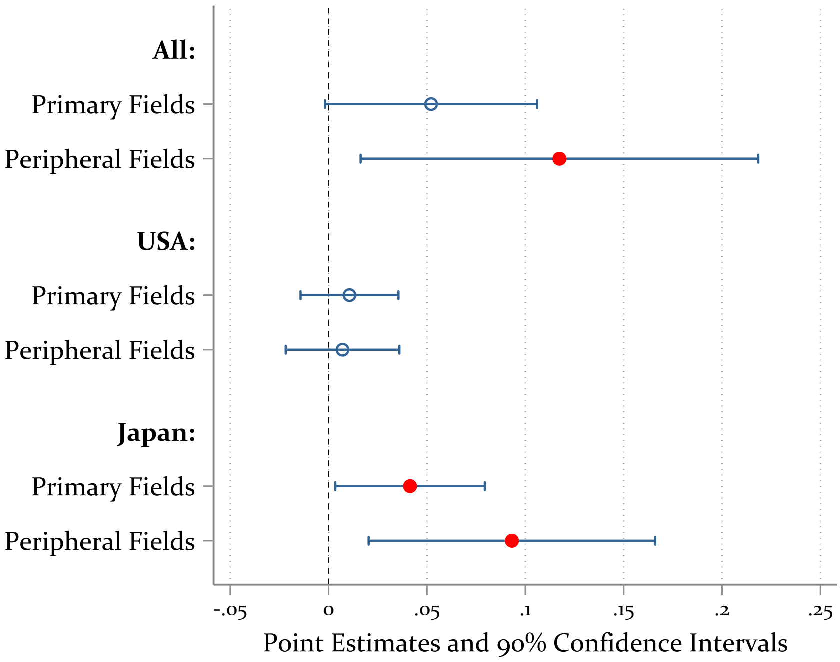
In the next step, I investigate whether firms also changed the direction of their innovation. To this end, similar to Kang (2021), I identify the primary and peripheral technology fields of every firm in my sample of established firms. I define a firm's primary field as its top four-digit CPC technology classes, where the firm filed at least 50% of its patents in the pre-treatment period from 1970 until 1975. All remaining four-digit technology classes represent a firm's peripheral technology fields.[1] This approach has the advantage that the number of patent applications is roughly equal across primary and peripheral technology fields in the pre-treatment period. I then use my main class-level approach to estimate how compulsory licensing of Xerox's patents affected patenting across firms' technology fields.

Figure A.9 shows that the increase in innovation after 1975 is driven disproportionately by patenting in firms' peripheral technology fields. Panel (A) shows the estimates

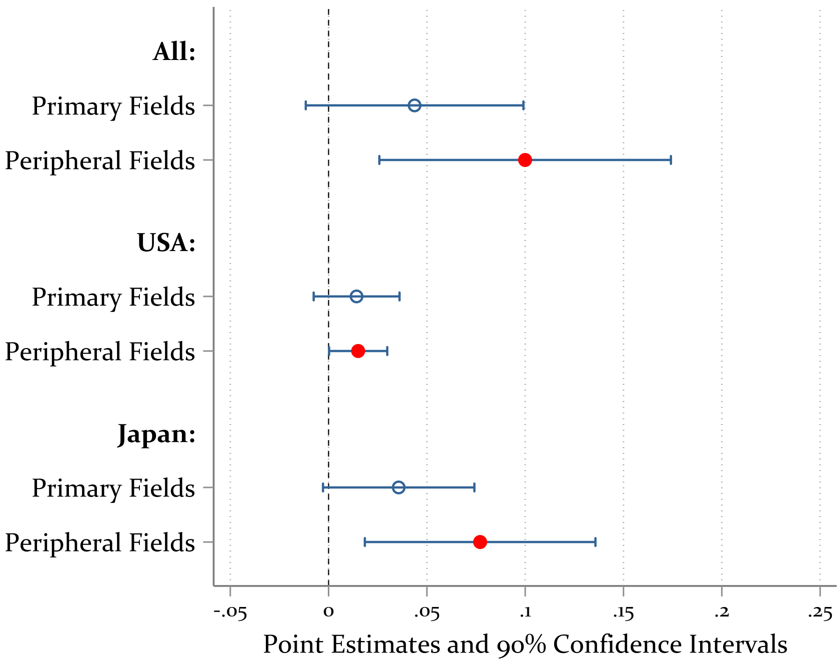
[1]In contrast, Kang (2021) defines a firm's primary technology field as the firm's top three technology classes. The results reported below remain robust to this alternative approach.

Figure A.9. Class-Level Analysis: Patenting in Firms' Primary vs. Peripheral Fields

(A) All Established Firms



(B) Firms in Top Decile of Closeness



Notes: The figure depicts point estimates and 90% confidence intervals from estimating the regression model in equation (1.1). In both panels, the outcome variable (i.e., the number of patent applications) is split by applicant country as well as by whether the focal patent belongs to its assignee's primary or peripheral technology field. A firm's primary (peripheral) field is defined as those four-digit CPC technology classes where that firm filed at least 50% (the remainder) of its patent applications from 1970 until 1975. Panel (A) considers patent applications by all 'established firms', whereas panel (B) restricts attention to patents filed by firms located in the top decile of the distribution of the closeness measure defined in equation (1.4). Red dots (blue circles) indicate statistically significant (insignificant) point estimates. Standard errors are clustered at the four-digit CPC technology class level.

for all established firms, whereas panel (B) only considers patents filed by firms in the top decile of the distribution of the closeness measure from equation (1.4). Both panels confirm the pattern from previous figures and tables that the overall positive effect of the antitrust case is primarily driven by increased innovation by Japanese applicants. When focusing on Japanese firms in the top decile of the distribution of the closeness measure, the DiD estimates are positive and statistically significant both for patenting in primary and peripheral fields. Yet, the point estimate for peripheral fields is about double the size of that for primary fields. This suggests that Japanese competitors shifted the focus of their innovation activities towards different technologies. That is, their research and development became more explorative (and less exploitative). In contrast, if firms had continued innovating in the same technology fields as before, one would expect the DiD estimates to be of roughly equal size.

A.2 Supplementary Results for Patent-Level Approach

This appendix contains supplementary results for my complementary analysis on the patent level to estimate the effect of compulsory licensing on direct follow-on innovation to Xerox's patents. Column (1) of Table A.7 presents the baseline DiD estimate from the patent-level analysis. It indicates that, on average, every compulsorily licensed Xerox patent received an additional 0.02 citations per year after 1975 relative to the matched control patents. In the remaining columns of Table A.7, I split the number of forward citations by the applicant country of the citing patent. Unlike in the results of the class-level analysis in Table 1.2, the estimate for the US is now statistically significant. It accounts for around half of the baseline estimate, implying that part of the increase in direct citations to compulsorily licensed Xerox patents came from patents filed by American firms. However, citations by non-US and especially Japanese applicants again play a key role in explaining the increase in citations to Xerox's patents. The estimate in column (4) is not only highly statistically significant; it is also quantitatively large compared to the average number of forward citations from Japanese patents. Overall, therefore, the estimates from Table A.7 confirm the finding from my main approach that Japanese competitors particularly benefited from access to Xerox's technology.

Table A.7. Patent-Level Analysis: Heterogeneity by Citing Country

	Baseline	Citing Country			
		USA	Non-USA	Among Non-USA	
				Japan	Others
	(1)	(2)	(3)	(4)	(5)
$\text{Xerox}_i \cdot \text{Post}_t$	0.024***	0.017**	0.009**	0.013***	-0.003
	(0.008)	(0.008)	(0.004)	(0.003)	(0.003)
Mean of Outcome	0.20	0.12	0.08	0.03	0.04
4-Digit CPC Classes	108	108	108	108	108
Observations	409,050	409,050	409,050	409,050	409,050

Notes: The table shows the results from difference-in-differences regressions following equation (1.3). Column (1) repeats the baseline estimates. In columns (2) to (5), the outcome variable is restricted to forward citations from citing patents filed by assignees from selected countries. All regressions include patent and year fixed effects. Standard errors clustered at the four-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.3 Supplementary Results for Firm-Level Approach

As discussed in section 1.5, I also construct a sample of established firms that filed at least ten patent applications from 1970 until 1975. This appendix contains additional information on the firm sample and presents results from an auxiliary analysis on the firm level.

A.3.1 Japanese Firms in Top Decile of Closeness Measure

Figure 1.7 in the main part of the chapter indicates that the positive innovation effect is primarily driven by Japanese firms in the top decile of the distribution of the closeness measure. Table A.8 presents the names of this set of Japanese firms. Reassuringly, the main Japanese competitors that, ex post, entered the US copier market (i.e., Canon, Konica, Minolta, Ricoh, Sharp, and Toshiba) are all included in this list. The list also contains a number of relatively unknown companies whose name suggests that their business may be related to printing or paper manufacturing. It is also important to note that other well-known Japanese companies in the high-technology sector (e.g., Hitachi, NEC, Sony), which were extremely successful at the time, are not included in the list.

A.3.2 Auxiliary Firm-Level Analysis

I now present an auxiliary analysis on the firm level that allows me to rule out alternative explanations for the heterogeneous effect across countries. In particular, I use the

Table A.8. Firm-Level Analysis: List of Japanese Firms in Top Decile of Closeness

Name
CANON
CASIO COMPUTER COMPANY
DENKI ONKYO COMPANY
DIC CORPORATION
DNP (DAINIPPON PRINTING COMPANY)
FUJIFILM
FUJITSU
HITACHI METALS
ISE ELECTRONICS CORP,JA
IWASAKI TSUSHINKI
IWATSU ELECTRIC CO LTD,JA
IWATSU ELECTRIC COMPANY
JSR CORPORATION (JAPAN SYNTHETIC RUBBER)
KANSAI PAINT COMPANY
KANZAKI PAPER MANUFACTURING COMPANY
KATSURAGAWA DENKI KK,JA
KONICA CORPORATION
MINOLTA CAMERA COMPANY
MITA INDUSTRIAL COMPANY
MITSUBISHI PAPER MILLS
NTT (NIPPON TELEGRAPH AND TELEPHONE CORP
PANASONIC CORPORATION
RICOH COMPANY
SHARP CORPORATION
TOSHIBA CORPORATION
WEST ELECTRIC COMPANY

Notes: The table reports the names of Japanese companies in the sample of established firms that are in the top decile in terms of the distribution of the closeness measure defined in equation (1.4).

firm sample to test whether the increase in patenting by Japanese competitors may be explained by either of the following possibilities.

First, it could be that differences in observable firm characteristics across countries explain the heterogeneity. For example, Japanese firms may have had more previous experience in copier technologies than American firms, hence giving them higher values of the closeness measure. I address this possibility by running simple firm-level DiD regressions, as shown in columns (1) and (2) of Table A.9. Specifically, in column (1), I regress the number of patent applications per firm and year on the interaction of the closeness measure (Closenss_i) and a post-1975 indicator (Post_t). I add a second interaction that uses a firm's stock of unexpired patents (Stock_i) as an additional firm-level covariate. The regression also includes firm and year × country fixed effects,

Table A.9. Firm-Level Analysis: Regression Estimates

	(1)	(2)	(3)	(4)
$\text{Closeness}_i \cdot \text{Post}_t$	0.003			-0.002
	(0.003)			(0.004)
$\text{Stock}_i \cdot \text{Post}_t$	0.001			-0.001
	(0.008)			(0.007)
$\mathbb{1}[\text{Closeness}_i \geq p90] \cdot \text{Post}_t$		3.592	-0.401	0.299
		(2.389)	(1.605)	(1.865)
$\mathbb{1}[\text{Closeness}_i \geq p90] \cdot \text{Post}_t \cdot \text{Japan}_i$			29.005**	29.548**
			(13.413)	(13.521)
$\mathbb{1}[\text{Closeness}_i \geq p90] \cdot \text{Post}_t \cdot \text{Other}_i$			-1.672	-1.901
			(2.336)	(2.326)
Firm FE	✓	✓	✓	✓
Year × Country FE	✓	✓	✓	✓
Mean of Outcome	11.50	11.50	11.50	11.50
Firms	1,635	1,635	1,635	1,635
Observations	26,160	26,160	26,160	26,160

Notes: The table shows the results from several firm-level difference-in-differences regressions, where the outcome is the number of patent applications per year and firm. Closeness_i is the measure of closeness to Xerox defined in equation (1.4) and $\mathbb{1}[\text{Closeness}_i \geq p90]$ denotes firms in the top decile of the distribution of the closeness measure. The variable Stock_i represents the number of unexpired patents that a firm held as of 1975. The indicator Post_t equals one in years after 1975. The variables Japan_i and Other_i are country indicators that equal one for firms from Japan and countries other than the US and Japan, respectively. Standard errors clustered at the firm level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

which may absorb any country-specific patenting trends. If the explanation above were correct and the heterogeneity across countries were driven by differences in firm observables, one would expect a significantly positive relationship between either of the interaction terms and the outcome variable. However, as shown in column (1), this is not the case: both estimates are quantitatively small and very imprecisely estimated. In column (2), I convert the closeness measure into a binary indicator that takes on the value of one for firms in the top decile of the distribution of the closeness measure ($\mathbb{1}[\text{Closeness}_i \geq p90]$). Although the estimate is much larger now, it is still statistically insignificant. This result indicates that, on aggregate, differences in the closeness to Xerox cannot explain the increase in Japanese patenting after 1975.

Second, one may be worried that the disproportionate increase in the number of patent applications among Japanese competitors could reflect a more aggregate country-specific pattern of increasing Japanese patenting. In the regression estimates in Table A.9, such an effect should be absorbed by the year × country fixed effects. I address this second possibility in columns (3) and (4) of Table A.9 by adding a triple interaction of an indicator for firms in the top decile of the distribution of the closeness measure ($\mathbb{1}[\text{Closeness}_i \geq p90]$), the post-treatment indicator (Post_t), and a country indicator for Japan (Japan_i). I include a similar interaction term with an indicator for all other countries other than the US and Japan (Other_i). If the increase in the number of patent applications were indeed driven by an aggregate increase in Japanese patenting, then the triple interaction with the Japan indicator should not be statistically significant, as the entire effect would be absorbed by the year × country fixed effects. Columns (3) shows that, again, this is not the case. The coefficient on the interaction with the Japan indicator is large and statistically significant at the 5% level. The point estimate implies that, following the consent decree, Japanese firms in the top decile of the distribution of the closeness measure filed an additional 29 patents per year on average, relative to American firms in the top decile and conditional on firm and year × country fixed effects. This estimate remains largely unchanged when controlling for the continuous closeness measure and the stock of a firm's patents, as shown in column (4). Intuitively, this result reflects the pattern visible in panel (A) of Figure 1.8 in the main part of the chapter – that is, there was a disproportionate increase in patenting among Japanese competitors, relative to their American counterparts but also relative to other Japanese firms.

Overall, this auxiliary analysis on the firm level suggests that the increase in patenting among Japanese competitors can neither be explained by differences in terms of observable firm characteristics across countries nor by an aggregate patenting trend that is specific to Japan. Yet, an important caveat regarding the first result is that my analysis is limited to firm characteristics related to firms' patenting, because I do not have any other firm-level data that cover all firms in the sample. However, if the heterogeneity were explained by other common firm characteristics, one would expect such differences to be reflected (at least partly) in the patent data as well.

A.4 Effect on Trade Values

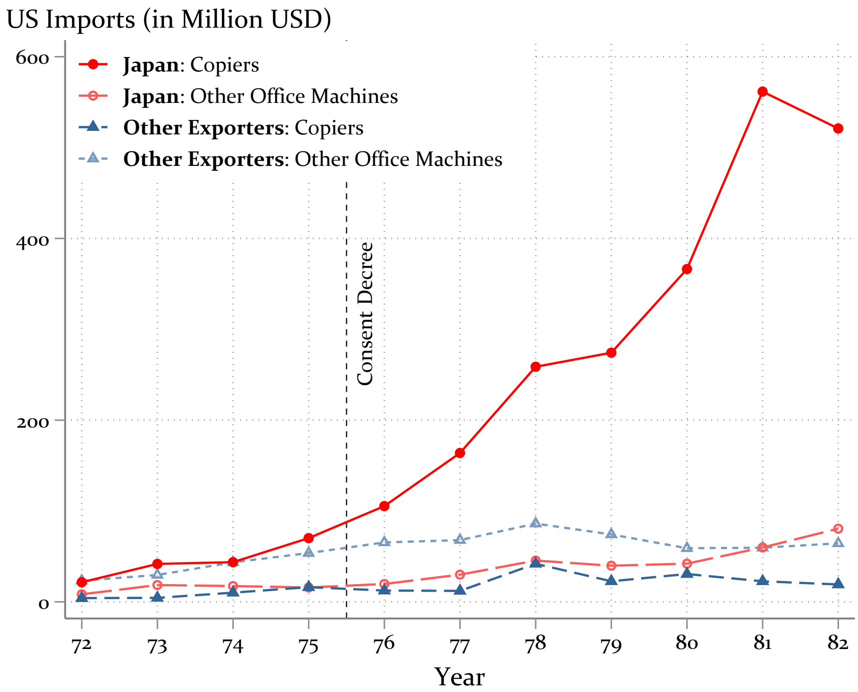
The empirical results of this chapter show that compulsory licensing of Xerox's patents promoted innovation particularly among Japanese competitors. Yet, this finding does not imply that Japanese firms also increased their revenues or profits. Therefore, this appendix analyses trade data to assess whether Japanese competitors benefited from the antitrust case in terms of subsequent exports to the US.

I use data from Feenstra (1996) that report the annual value of US imports by commodity and exporting country.[2] US imports were originally classified based on the Tariff Schedule of the United States Annotated (TSUSA). The data from Feenstra (1996) re-

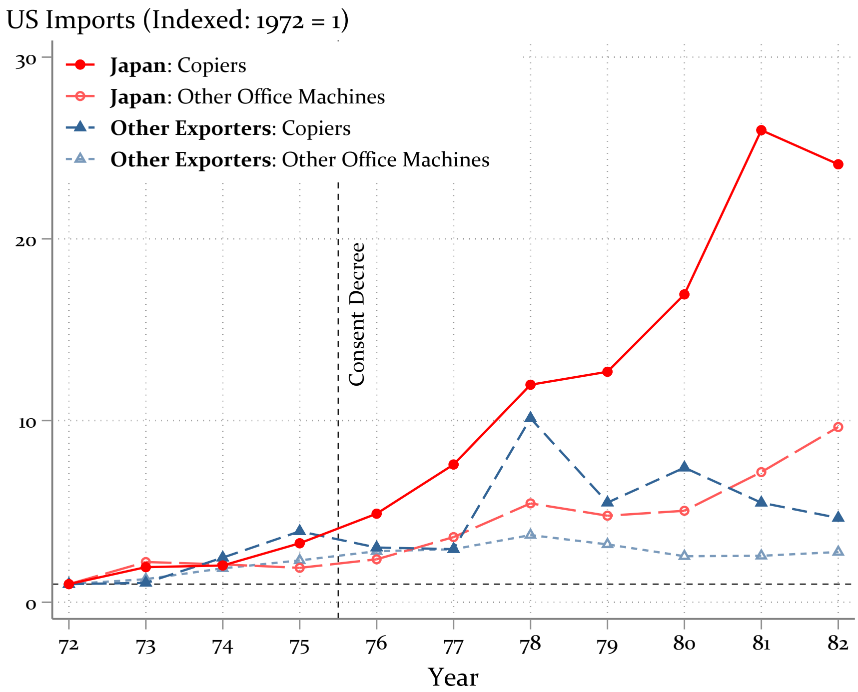
[2] These data are publicly available at `https://cid.ucdavis.edu/data` (last accessed: 11 February 2024).

Figure A.10. Effect on Trade Values: US Imports of Copiers and Other Office Machines

(A) Dollar Value



(B) Indexed



Notes: The figure depicts the value of US imports in the four-digit SITC class 7518, using data from Feenstra (1996). Trade values are shown separately for copiers and the remaining office machines in SITC class 7518. Copier imports are identified based on the more disaggregated TSUSA classes, which are also reported in the data. The figure also distinguishes trade values by exporting country. Panel (A) shows absolute dollar values. In panel (B), the time series are indexed to reflect growth rates relative to 1972.

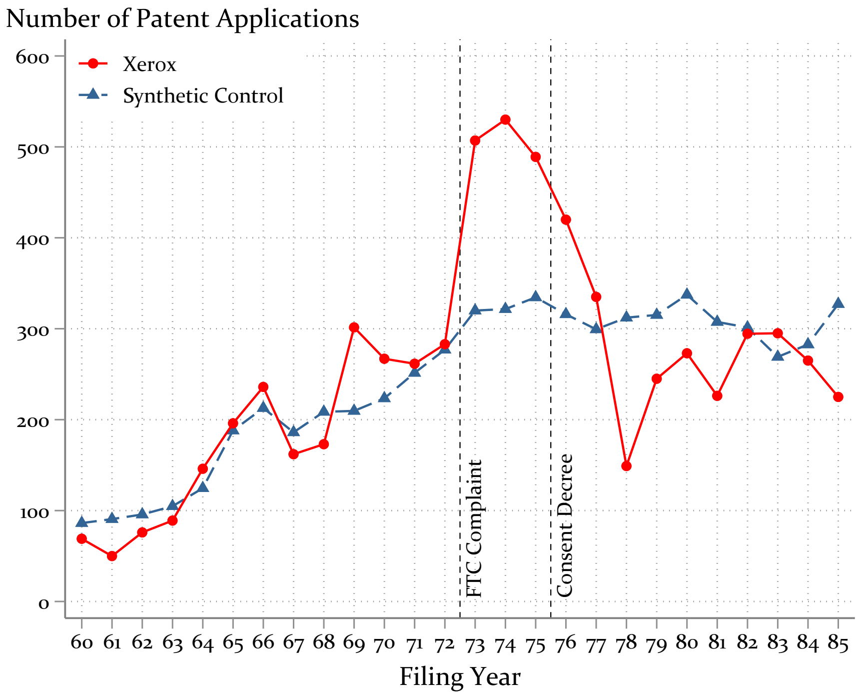
port these highly disaggregated TSUSA classes, which allows me to precisely identify the value of US copier imports. The data also contain a concordance to industry classes based on the Standard International Trade Classification (SITC, Rev. 2). Copiers are classified under the four-digit SITC class 7518, which covers 'office machines, n.e.s.' (i.e., not elsewhere specified). However, this class also contains goods other than copiers such as 'duplicating machines' or 'number, dating, and check-writing machines'.[3] One drawback of the data from Feenstra (1996) is that their coverage only starts in 1972, which is why I focus on the ten-year period until 1982.

Figure A.10 depicts the value of US imports in the four-digit SITC class 7518. Panel (A) shows that Japanese copier exports to the US skyrocketed in the late 1970s and amounted to a value of $560 million (in current dollars) in 1981. In contrast, all other countries jointly exported copiers worth less than $60 million to the US in the same year. As further shown in panel (B), Japanese copier exports to the US grew at similar rates as the aggregate copier exports by all other countries from 1972 until 1975. Then, however, copier exports from Japan increased disproportionately and reached up to 26 times of their 1972 value. Figure A.10 also shows trade values of other office machines that are classified in the same four-digit SITC class as copiers. These other office machines represent a natural control group for copiers. As is evident from the figure, Japanese exports of other office machines also increased after 1975, but this increase was less pronounced than that of Japanese copier exports. In addition, the total annual value of US imports of other office machines remained at a much lower value of around $140 million even in the early 1980s.

In summary, there are two key takeaways from my analysis of trade values. First, Japanese copier exports to the US increased after 1975, relative to the aggregate copier exports by all other countries. Second, there was no such disproportionate increase in Japanese exports for other office machines that are classified in the same four-digit SITC class. This descriptive evidence is consistent with my findings regarding the innovation effect of the antitrust case against Xerox. Therefore, the results on trade values support the conclusion that Japanese copier producers did not only benefit from access to Xerox's technology in terms of innovation; they also generated revenues in the product market by exporting copiers to the US.

[3]For more information on SITC, Rev. 2, see `https://unstats.un.org/unsd/publication/SeriesM/SeriesM_34rev2E.pdf` (last accessed: 11 February 2024).

Figure A.11. Effect on Xerox: Patenting by Xerox vs. Alternative Synthetic Control



Notes: The figure depicts the number of patent applications per year for Xerox and its subsidiaries (in red) and an alternative synthetic control group (in blue). The synthetic control group is computed using the algorithm by Abadie et al. (2010, 2015) and consists of 71.4% Siemens, 21.9% Westinghouse, 4.9% General Electric, and 1.8% Ciba-Geigy.

A.5 Supplementary Results for Effect on Xerox

In this appendix, I report supplementary results on the effect on Xerox by computing an alternative synthetic control group. As discussed in the main part of the chapter, including Bell into the synthetic control may not represent a good counterfactual for Xerox's patenting, as Bell itself faced an antitrust lawsuit in the late 1970s and was broken up in 1984 (Watzinger and Schnitzer, 2022). Therefore, I compute an alternative synthetic control, where Bell is excluded from the donor pool. The resulting control group consists of 71.4% Siemens, 21.9% Westinghouse, 4.9% General Electric, and 1.8% Ciba-Geigy.

The patenting trends depicted in Figure A.11 show an overall similar picture to Figure 1.10 in the main part of the chapter. However, the synthetic control group now has a slightly higher average number of patent applications after 1975. As a consequence, the hand-computed DiD estimate now indicates a decline in Xerox's patenting by around 30 patents per year on average.

A.6 Conceptual Framework

Why, in theory, should the removal of patent protection on Xerox's technology affect innovation by other firms? In this appendix, I introduce a brief conceptual framework explaining my key findings. In principle, patent rights should not hinder follow-on innovation, because efficient bargaining between the owner of an upstream technology and downstream innovators leads to ex-ante licensing (Green and Scotchmer, 1995). Consequently, any surplus-enhancing downstream innovation should be developed, irrespective of whether the upstream technology has patent protection or not.

However, patents may exert a blocking effect on follow-on innovation if upstream and downstream parties fail to reach a licensing agreement.[4] The economics literature has identified several reasons that may explain the absence of ex-ante licensing. On the one hand, there may be bargaining failure between the parties. This may arise, for example, due to asymmetric information (Bessen and Maskin, 2009) or coordination failure among downstream innovators (Galasso and Schankerman, 2015). On the other hand, rent dissipation may make licensing unprofitable for the upstream firm (Arora and Fosfuri, 2003). This is the case if the upstream firm's revenues from licensing are lower than its expected loss in profits due to increased product market competition.

In the case of Xerox, I argue that rent dissipation is the most likely reason for the absence of ex-ante licensing. Following the theoretical framework by Gaessler et al. (2019), the prevalence of rent dissipation versus bargaining failure can be assessed by considering (i) the degree of market overlap between upstream and downstream firms, (ii) the number of firms that control relevant technologies, and (iii) the size of the parties. In the case of Xerox, first, market overlap between Xerox and potential licensees was high. If Xerox had granted other firms unrestricted licenses, they may have become its direct competitors in the plain-paper copier market. Consequently, Xerox's monopoly rents would likely have dissipated. Second, Xerox was the single owner of the relevant technology, as it held almost all xerography patents. There was no need for downstream innovators to negotiate with several upstream parties, which speaks against bargaining failure. Third, both Xerox and potential licensees (e.g., Canon, IBM, Kodak) were large companies. This should have allowed face-to-face negotiations and facilitated cross-licensing agreements, again making bargaining failure unlikely (Gaessler et al., 2019).

[4]This blocking effect has also been identified empirically in several studies (e.g., Williams, 2013; Galasso and Schankerman, 2015; Gaessler et al., 2019). However, the evidence is still mixed, as other studies find no effect of patent protection on follow-on innovation (e.g., Sampat and Williams, 2019). For a broader overview of how patents affect research investments, see Williams (2017).

The rent dissipation theory is also most consistent with narrative evidence about Xerox's patenting strategy. Xerox executives believed that, by giving a license to one competitor, they would also have to license everybody else (Jacobson and Hillkirk, 1986). However, this would likely have created product market competition by reducing Xerox's technological advantage. In hindsight, this concern seems justified, given that I find that compulsory licensing increased subsequent innovation primarily in the copier industry. A further aspect that speaks against bargaining failure is that, in fact, Xerox did enter a number of licensing agreements with other firms. However, all licenses were restricted to manufacturing products other than plain-paper copiers.

Overall, a rent dissipation effect can plausibly explain the absence of ex-ante licensing between Xerox and potential follow-on innovators. Consequently, Xerox's patents could block follow-on innovation during the monopoly period. This changed in 1975 when antitrust intervention gave other firms access to Xerox's technology and, therefore, enabled subsequent innovation.

Appendix B

Appendix to Chapter 2

B.1 Supplementary Results for Aggregate Effect

This appendix contains supplementary results for my main approach on the technology class level to estimate the aggregate effect of RCA's technology transfer on Japanese innovation.

B.1.1 Data Description

Table B.1 presents additional information on RCA's patent portfolio by showing the top ten four-digit CPC technology classes. As can be seen in the table, RCA held patents in a broad range of different technologies. Additional summary statistics for the outcome variables used in the main part of the chapter are reported in Table B.2.

B.1.2 Computation of Semi-Elasticities for Transformed Outcome

In the DiD analysis in section 2.4, the outcome variable is subject to the inverse hyperbolic sine (IHS) transformation. This transformation has been frequently used in economics, because it approximates the natural logarithm but allows for zeros in the outcome variable (Bellemare and Wichman, 2020). Yet, a drawback of this approach is that the regression estimates cannot be interpreted quantitatively.[1] Therefore, I also compute how the overall number of Japanese patent applications changed both in absolute and relative terms.

[1] As recently pointed out by Chen and Roth (forthcoming), another potential issue with 'log-like' transformations such as the IHS is that the magnitude of the estimated treatment effects may depend on the units of the outcome variable, because the treatment may affect both the intensive and the extensive margin. I address this concern below in appendix B.1.4.

Table B.1. RCA's Top 10 4-Digit CPC Classes

Code	Title	Weight
H01J	ELECTRIC DISCHARGE TUBES OR DISCHARGE LAMPS	14.2%
H04N	PICTORIAL COMMUNICATION, e.g. TELEVISION	12.2%
G01S	RADIO DIRECTION-FINDING; RADIO NAVIGATION	6.3%
H03K	PULSE TECHNIQUE	6.1%
G11B	INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER	3.5%
H01Q	ANTENNAS, i.e. RADIO AERIALS	3.1%
H03F	AMPLIFIERS	3.0%
H03C	MODULATION	2.9%
H04B	TRANSMISSION	2.9%
H04L	TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION	2.7%

Notes: The table presents additional information on the top ten four-digit CPC technology classes in RCA's patent portfolio. The columns indicate the code and title of the ten largest four-digit classes along with their respective weights in RCA's patent portfolio.

Table B.2. Aggregate Effect: Summary Statistics

	Mean	SD
(A) Main Outcome		
# Japanese Patents	1.911	5.718
(B) Quality and Diversity		
Top 10% of Forward Citations	0.127	0.485
Top 10% of Quality (KPST)	0.277	1.376
Active CPC Subgroups	1.040	2.317
(C) Triple DiD		
Germany	5.403	11.402
(D) Citations to RCA		
Distance = 1	0.106	0.710
Distance = 2	0.221	1.180
Distance = 3	0.185	0.783
Unconnected to RCA	1.288	4.421
(E) Licensee Status		
License from RCA	0.416	2.125
No License	1.495	4.857

Notes: The table presents summary statistics for the outcome variables used in the main part of the chapter. Panel (A) refers to the baseline specification, where the outcome is the number of Japanese patent applications in the US per four-digit CPC subclass per year. The variables in panel (B) to (E) correspond to the results from Table 2.2, Figure 2.4, and Figure 2.5, respectively. There are 5,600 observations for each variable (i.e., 350 four-digit subclasses observed in 16 years).

I follow Bellemare and Wichman (2020) to derive (semi-)elasticities when using an IHS-transformed outcome variable. For the case with a binary treatment, the elasticity is undefined, but one can compute the percentage change in the outcome (i.e., the number of Japanese patent applications) associated with a change in the binary treatment variable (i.e. whether RCA has at least one patent in the given subclass) from zero to one. After estimating the model in equation (2.1), this proportional effect is given by

$$\hat{P} = \frac{\sinh(\hat{\beta} + \hat{\alpha}_s + \hat{\lambda}_{c,t} + \hat{\epsilon}_{c,s,t}) - \sinh(\hat{\alpha}_s + \hat{\lambda}_{c,t} + \hat{\epsilon}_{c,s,t})}{\sinh(\hat{\alpha}_s + \hat{\lambda}_{c,t} + \hat{\epsilon}_{c,s,t})}, \tag{B.1}$$

where $\sinh(\bullet)$ is the hyperbolic sine transformation (Bellemare and Wichman, 2020). I evaluate this expression at the mean of the treated subclasses in years after 1958 to approximate the counterfactual of how Japanese patenting in these fields would have evolved in the absence of RCA's technology transfer.

For the intensity specification, where $\text{Treat}_s = \text{RCA}_s$ (i.e., the number of RCA patents per subclass), I compute the semi-elasticity

$$\xi = \frac{\partial \text{Patents}_{c,s,t}}{\partial \text{RCA}_s} \cdot \frac{1}{\text{Patents}_{c,s,t}}, \tag{B.2}$$

which describes the percentage change in the number of Japanese patent applications associated with one additional RCA patent per subclass. Following Bellemare and Wichman (2020), after estimating the DiD model in equation (2.1), this semi-elasticity is given by

$$\hat{\xi} = \hat{\beta} \cdot \sqrt{1 + \frac{1}{\text{Patents}^2_{c,s,t}}}. \tag{B.3}$$

Again, I evaluate this expression at the mean of the treated subclasses in years after 1958.

In the next step, I attempt to derive the number of additional Japanese patents that is associated with the treatment. Intuitively, I use the estimated semi-elasticity to approximate the number of Japanese patent applications in the counterfactual scenario absent RCA's technology transfer. To do this, I first calculate the total relative increase in Japanese patenting for each treated subclass, given the estimated semi-elasticity and the number of RCA patents in that subclass.[2] Then, I can estimate the counterfactual Japanese patenting by dividing the observed number of Japanese patent applications by the total relative increase plus one (i.e., by the growth factor). Finally, the change in Japanese patenting associated with the treatment is the difference between the

[2]Since the regression model is IHS-linear, the underlying assumption is that every additional RCA patent has the same proportional effect on Japanese patenting.

observed and the counterfactual number of Japanese patent applications. Aggregating this number across all treated subclasses yields the overall number of additional Japanese patents, as reported in the regression tables. This procedure can be used analogously for the binary treatment specification. The only minor change is that the total relative increase in Japanese patenting is identical across all treated subclasses and is directly given by the estimated proportional effect ($\hat{P}$) in equation (B.1).

B.1.3 Treatment Definition

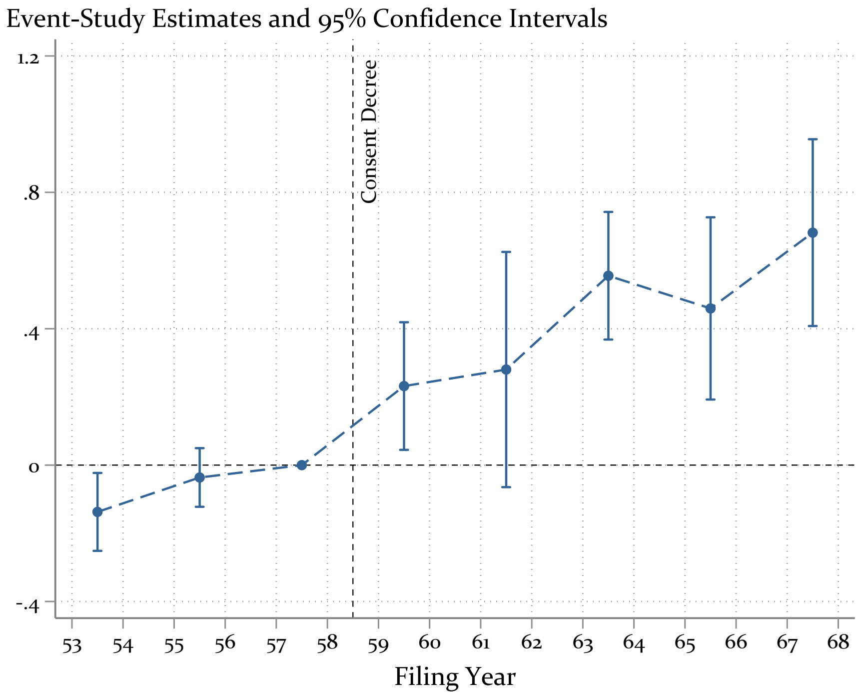
In this section, I analyse whether my estimates are robust to using alternative treatment definitions. In the main part of the chapter, I focus on using the intensity treatment that captures differences in the likelihood of licensing by RCA across subclasses.

Figure B.1 shows the event-study estimates from running the model in equation (2.2) with the binary treatment definition, where all subclasses with at least one RCA patent are equally treated. The figure looks similar to Figure 2.3 in that it shows a disproportionate increase in Japanese patenting in treated technologies after 1958. However, there seems to be a slight pre-trend around five to six years prior to the antitrust settlement. This indicates that, even before RCA began large-scale licensing in Japan, Japanese patenting was growing more in technologies where RCA was an active innovator. I will get back to discussing these pre-trends in appendix B.1.5. In particular, I show below that German patenting evolved similarly prior to 1958, whereas the subsequent increase in the number of Japanese patent applications represents an effect that is specific to Japan. Therefore, the potential pre-trends in Figure B.1 do not represent a concern for the validity of my empirical strategy.

In Table B.3, I present baseline DiD estimates using other treatment specifications. The baseline estimates from Table 2.3 are repeated in columns (1) and (4). Columns (2) and (3) present alternative specification with a continuous treatment variable, whereas columns (5) to (7) use different cut-offs to define a binary treatment variable.

In column (2) of Table B.3, I use the intensity treatment as in the baseline, but I also include the square of the number of RCA patents per subclass as an explanatory variable. This is motivated by the potentially restrictive assumption underlying the simple intensity treatment that the relationship between the number of RCA patents in a subclass and the proportional change in Japanese patenting is linear. Additionally including the square of the number of RCA patents allows for a non-linear relationship. Indeed, the estimated coefficient on the linear interaction term increases in magnitude, whereas the coefficient on the interaction with the squared number of RCA patents is negative. This indicates that the proportional effect of one additional RCA patent in a subclass on Japanese innovation decreases in the number of RCA patents. This pattern

Figure B.1. Aggregate Effect: Event-Study Estimates with Binary Treatment



Notes: The figure depicts point estimates and 95% confidence intervals from the event-study analysis in equation (2.2). Japanese patent applications are binned in two-year groups to reduce noise in the estimates. The figure uses the binary treatment with $\text{Treat}_s = \mathbb{1}[\text{RCA}_s > 0]$, where RCA_s is the number of unexpired RCA patents in subclass s. The regression uses the weights by Iacus et al. (2012). Standard errors are clustered at the three-digit CPC technology class level.

is intuitive and consistent with estimates in the literature (e.g., Moser and Voena, 2012). Interestingly, allowing for a non-linear relationship also leads to a higher estimate of the overall number of additional Japanese patents. The estimate in column (2) suggests that Japanese patenting increased by around 28% following RCA's licensing in Japan, compared to 16% in the baseline.

One potential concern with the intensity treatment (and its square) is that it may be volatile to the design of the patent classification system. Suppose, for example, that there is an update to the classification such that a treated subclass is split into two distinct subclasses. This would reduce the absolute number of RCA patents (i.e., the treatment variable) in each of the new subclasses. Yet, this would also re-distribute all Japanese patents (i.e., the outcome variable) to the new subclasses. As my empirical specification studies percentage changes in Japanese patenting, changes to the classification system do not present a threat for the validity of my identification strategy. Moreover, it is worth noting that the patent data from PATSTAT contain a consistent and potentially updated assignment of patents to CPC classes, rather than the assignment at the time of patent filing.

Table B.3. Aggregate Effect: Different Treatment Definitions

	Intensity			Binary			
	Number		Share	Baseline	Min. 2	Min. 3	Min. 5
	Baseline	w/ Square					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\text{Treat}_s \cdot \text{Post}_t$	0.003***	0.007***	0.025	0.500***	0.662***	0.693***	0.676***
	(0.001)	(0.002)	(0.021)	(0.073)	(0.104)	(0.128)	(0.115)
$\text{Treat}_s^2 \cdot \text{Post}_t$		-0.000**					
		(0.000)					
Implied Semi-Elasticity	0.3%	0.6%	2.6%	n/a	n/a	n/a	n/a
Additional Patents per Year	111	173	65	261	328	295	197
Relative Increase	16.3%	27.9%	8.9%	89.5%	122.1%	125.5%	119.5%
Mean of Outcome (w/o IHS)	1.9	1.9	1.9	1.7	2.1	2.2	2.1
No. of 4-Digit CPC Classes	350	350	350	307	243	212	170
No. of 3-Digit CPC Classes	63	63	63	54	39	33	26
Observations	5,600	5,600	5,600	4,912	3,888	3,392	2,720

Notes: The table shows the results from difference-in-differences regressions following equation (2.1). The baseline estimates from columns (1) and (2) of Table 2.1 are repeated in columns (4) and (1), respectively. Column (2) also uses the intensity treatment with $\text{Treat}_s = \text{RCA}_s$, where RCA_s is the number of unexpired RCA patents in subclass s, but it additionally includes its square. In column (3), the treatment variable reflects the share (as opposed to the number) of unexpired RCA patents in a subclass. Columns (5) to (7) use alternative binary treatment specifications with $\text{Treat}_s = \mathbb{1}[\text{RCA}_s \geq i]$, where the cut-offs i = 2, 3, and 5, respectively. All regressions include subclass and year × class fixed effects. Standard errors clustered at the three-digit CPC technology class level are in parentheses. Significance levels: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Nevertheless, column (3) of Table B.3 presents the DiD estimate using an alternative continuous treatment specification that relies on the share (as opposed to the number) of RCA patents in a subclass. The point estimate implies a smaller increase in Japanese patenting by around 9% and it is not statistically significant. This approach follows chapter 1 (Mamrak, 2023), where I study the impact of compulsory licensing of Xerox's patents on subsequent innovation. However, it is important to note that, in the case of Xerox, other firms did not actually obtain a formal license for Xerox's copier patents. Therefore, the treatment variable in chapter 1 does not aim to capture the likelihood of licensing. Instead, it approximates the propensity of a given technology class for containing copier-related patents, for which the overall number of patent applications in a given technology class should also be taken into account.[3] In contrast, in the present chapter, the objective of the treatment variable is to approximate the likelihood of licensing by RCA. Arguably, this likelihood is most closely associated with the number (rather than the share) of RCA patents in a given technology class, irrespective of the overall size of that class. Therefore, the smaller and insignificant estimate in column (3) should not be a huge concern, given that the estimates are robust across many other specifications.

[3]Moreover, in chapter 1, I consider absolute changes in the outcome variable, which further motivates using the share specification due to the potential concern above.

Finally, columns (5) to (7) of Table B.3 present estimates with different binary treatment variables. In the baseline binary treatment specification in column (4), I consider all subclasses as treated where RCA held at least one unexpired patent. In the remaining three columns of Table B.3, I only assign subclasses to the treatment group if they contain a minimum of two, three, and five RCA patents, respectively. The overall result of a large and statistically significant increase in Japanese patenting in RCA's technologies remains unchanged. Depending on the specification, the number of additional Japanese patents that is associated with the treatment varies between 197 and 328. That is, alternative definitions of a binary treatment variable also yield a larger increase in Japanese patenting than my baseline specification with the intensity treatment. The drawback of all the binary treatment specifications is that is unclear which cut-off is most appropriate. Hence, using a binary specification requires an arbitrary choice by the researcher. This is another argument for using the intensity treatment, as I do in the main part of the chapter.

Overall, the various alternative treatment specifications in Table B.3 suggest that my baseline result of a disproportionate increase in Japanese patenting in RCA's fields does not hinge on any specific definition of the treatment variable.

B.1.4 Model Specification

In Table B.4, I assess the robustness of my DiD estimates to alternative model specifications. Column (1) repeats the baseline estimate with the intensity treatment from the main part of the chapter. In column (2), I estimate a variation of the DiD model that only includes year fixed effects but not year $\times$ class fixed effects. This increases the point estimate, hence indicating that the year $\times$ class fixed effects play an important role by controlling for technology-specific shocks over time. In column (3), I estimate the model without using the coarsened exact matching (CEM) weights by Iacus et al. (2012). This hardly changes the point estimate relative to the baseline.

Columns (4) and (5) address recent concerns about using an IHS-transformed outcome variable when the non-transformed outcome can also equal zero. Chen and Roth (forthcoming) show that a potential issue with 'log-like' transformations (such as the IHS) is that the magnitude of the estimated treatment effect may depend on the units of the outcome variable, because the treatment may affect both the intensive and the extensive margin. They illustrate this issue by replicating several empirical studies from the *American Economic Review*, showing that multiplying the outcome variable by 100 may substantially change the estimated treatment effects. This issue is particularly prevalent if extensive-margin effects are important. In column (4), I show that the interpretation of my baseline estimate is robust to rescaling the outcome vari-

Table B.4. Aggregate Effect: Alternative Model Specifications

	Baseline	Other FE	No CEM Weights	Outcome × 100	Outcome w/o IHS	Excl. Top 5% Subcl.
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Treat}_s \cdot \text{Post}_t$	0.003***	0.004***	0.003***	0.003*	0.027***	0.010***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.003)	(0.002)
Implied Semi-Elasticity	0.3%	0.4%	0.3%	0.3%	n/a	1.0%
Additional Patents per Year	111	127	104	10,737	156	120
Relative Increase	16.3%	19.0%	15.2%	15.7%	24.6%	21.0%
Subclass FE	✓	✓	✓	✓	✓	✓
Year FE		✓				
Year × Class FE	✓		✓	✓	✓	✓
CEM Weights	✓	✓		✓	✓	✓
Mean of Outcome (w/o IHS)	1.9	1.9	1.9	192.1	1.9	1.9
No. of 4-Digit CPC Classes	350	350	350	350	350	340
No. of 3-Digit CPC Classes	63	63	63	63	63	62
Observations	5,600	5,600	5,600	5,600	5,600	5,440

Notes: The table shows the results from difference-in-differences regressions following variations of equation (2.1). All regressions use the intensity treatment with $\text{Treat}_s = \text{RCA}_s$, where RCA_s is the number of unexpired RCA patents in subclass s. Column (1) repeats the baseline estimated from column (2) of Table 2.1 in the main part of the chapter. In column (2), the model is estimated with year fixed effects instead of year × class fixed effects. In column (3), the model is estimating without using the weights by Iacus et al. (2012). In column (4), the outcome variable is multiplied by 100 before applying the inverse hyperbolic sine (IHS) transformation. In column (4), the outcome variable is the number of Japanese patent applications without applying the IHS transformation. Finally, column (6) excludes the top 5% of four-digit subclasses with the highest number of unexpired RCA patents from the sample. Standard errors clustered at the three-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

able. When multiplying the number of Japanese patent applications by 100 before applying the IHS transformation (as in Chen and Roth, forthcoming), the estimated relative increase in Japanese patenting is 15.7%, which is quantitatively very close to the baseline of 16.3%. This result indicates that the extensive margin (i.e., a change in the number of Japanese patents from zero to one in a given subclass in response to the treatment) plays a minor role in my application. In addition, column (5) shows the estimate from a regression that uses the absolute number of Japanese patents without the IHS transformation as outcome variable. The magnitude of this estimate implies that one additional RCA patent in a subclass is associated with 0.03 additional Japanese patents in that subclass per year after 1958. This corresponds to a surge in Japanese patenting by close to 25%, which is larger than the baseline estimate.

Finally, in column (6), I exclude the top 5% of subclasses with the most RCA patents (i.e., the highest values of the treatment variable) from the sample. This is to verify that the positive baseline estimate is not purely driven by patenting in subclasses where RCA held a very large number of patents. Again, the magnitude of the estimate is slightly higher than in the baseline, which is consistent with a potential non-linear relationship between the treatment and the outcome, as discussed above in appendix B.1.3.

In summary, all point estimates in Table B.4 remain positive and statistically significant. They indicate that the baseline result showing a disproportionate increase in Japanese patenting in RCA's fields is not driven by a particular model or sample specification.

B.1.5 Triple Difference-in-Differences Strategy

This section contains supplementary results for the triple DiD analysis introduced in section 2.4.3 in the main part of the chapter. Column (2) of Table B.5 presents the point estimates from running the regression model in equation (2.3). Consistent with Figure 2.4 in the main part of the chapter, the simple interaction that estimates the effect on Germany is very small in magnitude and statistically insignificant. Conversely, the triple interaction term that estimates the *additional* effect on Japan is positive and statistically significant at the 1% level. The magnitude of the triple DiD estimate implies that Japanese applicants filed 118 additional patents per year after 1958, which corresponds to a relative increase by 18%. These numbers are very close in magnitude to my baseline DiD estimate that does not include Germany as a comparison country. This indicates that the entire increase in Japanese patenting represents an effect that is specific to Japan and cannot be explained by broader patenting trends.

Column (1) of Table B.5 shows the corresponding triple DiD estimate when using the binary treatment specification. Again, the estimate of the effect on Germany is small and statistically insignificant, whereas the estimate of the additional effect on Japan is positive and highly significant. Figure B.2 presents the event-study estimates corresponding to the triple DiD model with the binary treatment. The figure looks very similar to Figure 2.4 in the main part of the chapter. It shows that German patenting did not change differentially in RCA's main technologies after 1958, while Japanese patenting increased disproportionately in technologies where RCA held more patents. Moroever, Figure B.2 highlights that, relative to Germany, Japanese patenting did not follow any different trend before 1958 across technologies with different numbers of RCA patents. Therefore, the slightly diverging pre-trends in Figure B.1 likely reflect a development that also affected countries other than Japan, hence supporting the identifying assumption underlying my empirical approach.

Finally, I check that the results from my triple DiD analysis do not hinge on using Germany as a comparison country. Therefore, in columns (3) and (4), I use all non-US countries other than Japan as a comparison category. Although the aggregate number of patent applications filed by applicants from all other non-US countries is

Table B.5. Aggregate Effect: Triple DiD Estimates

	Germany		Other Non-US	
	Binary	Intensity	Binary	Intensity
	(1)	(2)	(3)	(4)
$\text{Treat}_s \cdot \text{Post}_t$	0.022	-0.000	-0.062	-0.001
	(0.093)	(0.001)	(0.183)	(0.001)
$\text{Treat}_s \cdot \text{Post}_t \cdot \text{Japan}_i$	0.478***	0.003***	0.562**	0.004***
	(0.110)	(0.000)	(0.216)	(0.001)
Implied Semi-Elasticity	n/a	0.3%	n/a	0.4%
Additional Patents per Year	251	118	287	136
Relative Increase	83.5%	17.5%	108.2%	20.8%
Mean of Outcome (w/o IHS)	3.4	3.7	10.4	10.7
No. of 4-Digit CPC Classes	307	350	307	350
No. of 3-Digit CPC Classes	54	63	54	63
Observations	9,824	11,200	9,824	11,200

Notes: The table shows the results from triple difference-in-differences regressions following equation (2.3). Columns (1) and (3) use the binary treatment with $\text{Treat}_s = \mathbb{1}[\text{RCA}_s > 0]$, where RCA_s is the number of unexpired RCA patents in subclass s. Columns (2) and (4) use the intensity treatment with $\text{Treat}_s = \text{RCA}_s$. Columns (1) and (2) use Germany as a comparison country, whereas columns (3) and (4) consider the aggregate patenting in all non-US countries other than Japan. All regressions include subclass × country and year × class × country fixed effects. Standard errors clustered at the three-digit CPC technology class level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

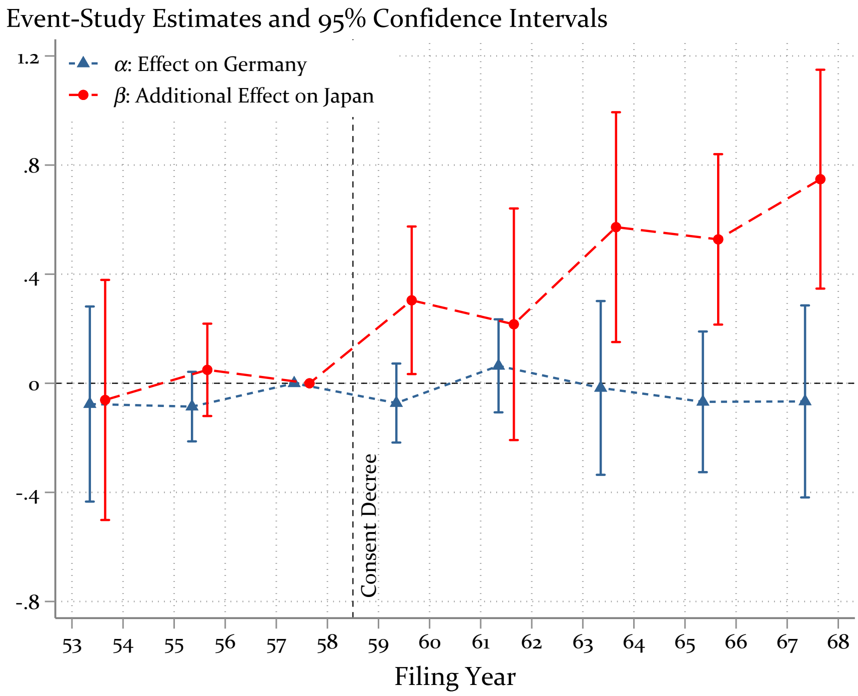
much higher than that filed by Japanese applicants, this should not be a concern for the empirical strategy since I focus on percentage changes in patenting. The estimates in columns (3) and (4) of Table B.5 confirm that the observed increase in the number of Japanese patent applications in RCA's technologies represents an effect that is specific to Japan. Even the magnitude of the implied increase in Japanese patenting that is associated with the treatment only changes slightly.

In summary, these robustness checks further corroborate my conclusion that the increase in Japanese patenting cannot be explained by more general patenting trends across technologies. This supports my interpretation that the estimates in this chapter likely capture the impact of RCA's technology transfer on Japanese innovation.

B.1.6 Heterogeneity by Distance to RCA and Licensee Status

Table B.6 presents DiD estimates corresponding to the graphical disaggregation of the additional Japanese patents filed after 1958 in Figure 2.5 in the main part of the chapter. Column (1) repeats the baseline estimate with the intensity treatment. In columns (2) to (4), the outcome variable is restricted to Japanese patent applications with a given citation distance to RCA, following the framework by Ahmadpoor and

Figure B.2. Aggregate Effect: Triple DiD with Binary Treatment



Notes: The figure depicts point estimates and 95% confidence intervals from an event-study variation of the triple DiD model in equation (2.3). Patent applications are binned in two-year groups to reduce noise in the estimates. The figure uses the intensity treatment with $\text{Treat}_s = \mathbb{1}[\text{RCA}_s > 0]$, where RCA_s is the number of unexpired RCA patents in subclass s. The regression uses the weights by Iacus et al. (2012). Standard errors are clustered at the three-digit CPC technology class level.

Jones (2017). Although the point estimates are statistically significant in all specifications, the magnitude of the estimated increase in Japanese patenting is higher when considering patents that are more directly linked to RCA. In particular, the relative increase in the number of Japanese patents that directly cite RCA was more than 150%, whereas it was only 3.6% for patents unconnected to RCA (i.e., with distance ≥ 4). Finally, columns (6) and (7) show the point estimates when splitting the number of Japanese patent applications by whether the applicant obtained a license from RCA or not. In line with the visual interpretation of panel (B) of Figure 2.5, RCA's licensees account for approximately two thirds of the overall increase in Japanese patenting after 1958.

B.2 Supplementary Results for Direct Effect

This appendix contains supplementary results for the firm-level analysis of the direct effect of receiving a patent license from RCA on follow-on innovation by the licensees.

Table B.6. Aggregate Effect: Distance to RCA and Licensee Status

	Baseline	Distance to RCA				Licensee Status	
		$D = 1$	$D = 2$	$D = 3$	$D \geq 4$	Licensees	Others
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\text{Treat}_s \cdot \text{Post}_t$	0.003***	0.002***	0.003***	0.002***	0.001***	0.003***	0.002***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Implied Semi-Elasticity	0.3%	0.8%	0.6%	0.4%	0.1%	0.4%	0.2%
Additional Patents per Year	111	34	48	23	18	73	40
Relative Increase	16.3%	159.1%	73.9%	33.5%	3.6%	53.2%	7.3%
Mean of Outcome (w/o IHS)	1.9	0.1	0.2	0.2	1.4	0.5	1.5
No. of 4-Digit CPC Classes	350	350	350	350	350	350	350
No. of 3-Digit CPC Classes	63	63	63	63	63	63	63
Observations	5,600	5,600	5,600	5,600	5,600	5,600	5,600

Notes: The table shows the results from difference-in-differences regressions following variations of equation (2.1). Column (1) repeats the baseline estimate from column (2) of Table 2.1. In columns (2) to (5), the outcome variable is restricted to patent applications with different citation distances to RCA, following the framework by Ahmadpoor and Jones (2017). In column (6), the outcome variable is restricted to patent applications by Japanese firms that received a patent license from RCA, whereas the outcome in column (7) includes all remaining Japanese patent applications. The table uses the intensity treatment with $\text{Treat}_s = \text{RCA}_s$, where RCA_s is the number of unexpired RCA patents in subclass s. All regressions include subclass and year × class fixed effects. Standard errors are clustered at the three-digit CPC technology class level. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B.2.1 Alternative Control Group

As discussed above, the interpretation of my estimates in section 2.5 hinges upon the selection of an appropriate control group. Therefore, I now present the results of a robustness check that does not rely on the synthetic control method.

I employ a matching strategy to construct an alternative control group for RCA's licensees. This strategy considers three factors: (i) a firm's aggregate number of patent applications prior to the licensing agreement, (ii) the year of its first patent, and (iii) its primary technology field. I group these variables into 'cells'. First, I assign the number of prior patent applications to one of eight bins. Second, for the year of the first patent, I compute a binary variable that indicates whether a firm already filed at least one patent in the US prior to 1960. Finally, I identify a firm's primary technology field as the most frequent one-digit technology class (based on the Cooperative Patent Classification) among its prior patent applications. This grouping ensures that all licensees can be matched to at least one Japanese firm that did not receive a license from RCA. I re-weight the control firms so that there effectively is one control firm for each licensee. Overall, this cell matching strategy allows me to identify Japanese firms that were similar to the licensees in terms of the scope, history, and field of their prior patenting.

Column (1) of Table B.7 presents the baseline estimate from running the DiD model in equation (2.4) with the alternative control group. The point estimate indicates

Table B.7. Direct Effect: Regression Estimates with Alternative Control Group

	Baseline	Distance to RCA		Patents in Top 10%		Diversity
		$D \leq 3$	$D \geq 4$	Forward Citations	Quality (KPST)	Active Subgroups
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Treat}_i \cdot \text{Post}_t$	2.862*	2.258**	0.604	0.139**	0.679*	2.569*
	(1.688)	(1.012)	(0.792)	(0.054)	(0.397)	(1.399)
Additional Patents per Year	120.2	94.9	25.4	5.8	28.5	107.9
Relative Increase	74.3%	129.7%	28.6%	80.8%	89.0%	81.0%
Mean of Outcome	0.2	0.1	0.2	0.0	0.0	0.2
No. of Treated Firms	42	42	42	42	42	42
Clusters	924	924	924	924	924	924
Observations	145,692	145,692	145,692	145,692	145,692	145,692

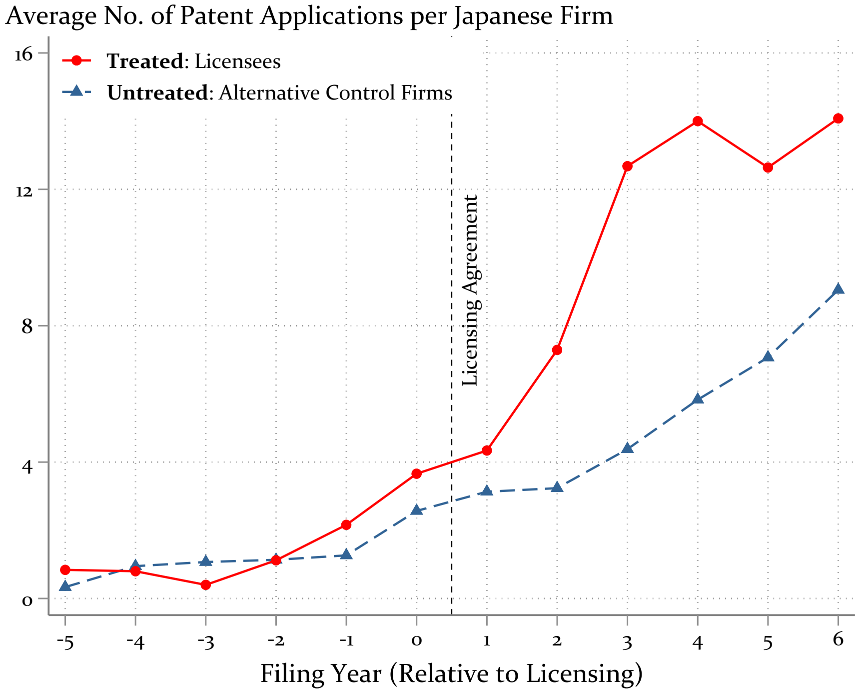
Notes: The table shows the results from difference-in-differences regressions following variations of equation (2.4). In column (1), the outcome variable is the number of patent applications in the US per firm. In columns (2) and (3), the outcome variable is restricted to patent applications that are connected (with distance ≤ 3) and unconnected (with distance ≥ 4) to RCA through citations, respectively, following the distance framework by Ahmadpoor and Jones (2017). In columns (4) and (5), the outcome variable is restricted to patent applications in the top 10% of the distribution of forward citations and the quality measure by Kelly et al. (2021, KPST), respectively. In column (6), the outcome variable counts the number of 'active' technology subgroups (based on the Cooperative Patent Classification) per firm. Therefore, in that column, the number in the row 'Additional Patents per Year' refers to the number of additional subgroups (as opposed to patents). All regressions include firm and absolute year fixed effects. Standard errors clustered at the firm level are in parentheses. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

that, on average, receiving a license from RCA was associated with an increase in the licensee's patenting by 2.9 patents per year. This estimate corresponds to around 120 additional patents per year.

Figure B.3 graphically shows how the average number of patent applications in the US developed over time for RCA's licensees and the alternative control firms. Prior to the licensing agreement, patenting by the control firms roughly followed that of the (future) licensees, but the matching strategy performs worse in producing parallel pre-trends than the synthetic control method (see Figure 2.6). This is because the matching strategy only takes into account the aggregate (as opposed to the annual) number of patent applications prior to licensing. After the treated firms received a license from RCA, the control firms also increased their patenting. This development stands in contrast to the evolution of the number of patent applications by the synthetic control firms, as shown in Figure 2.6. One potential explanation for this difference is that the matching strategy also takes into account a firm's primary technology field, which may control for possibly diverging patenting trends across industries. This pattern also explains why the point estimate in column (1) of Table B.7 is smaller in magnitude than the corresponding estimate in Table 2.3.

Despite the difference in magnitude, the results using the alternative control group are qualitatively similar to those using the synthetic control method. This is true also for the remainder of the estimates in Table B.7. Columns (2) and (3) indicate that

Figure B.3. Direct Effect: Patenting by Licensees vs. Alternative Control Firms



Notes: The figure depicts the average number of patent applications in the US separately for RCA's Japanese licensees (in red) and their alternative control firms (in blue) in the six years before and after the licensing agreement. The control group is obtained via cell matching based on a firm's aggregate number of patent applications prior to the licensing agreement, the year of its first patent, and its primary technology field.

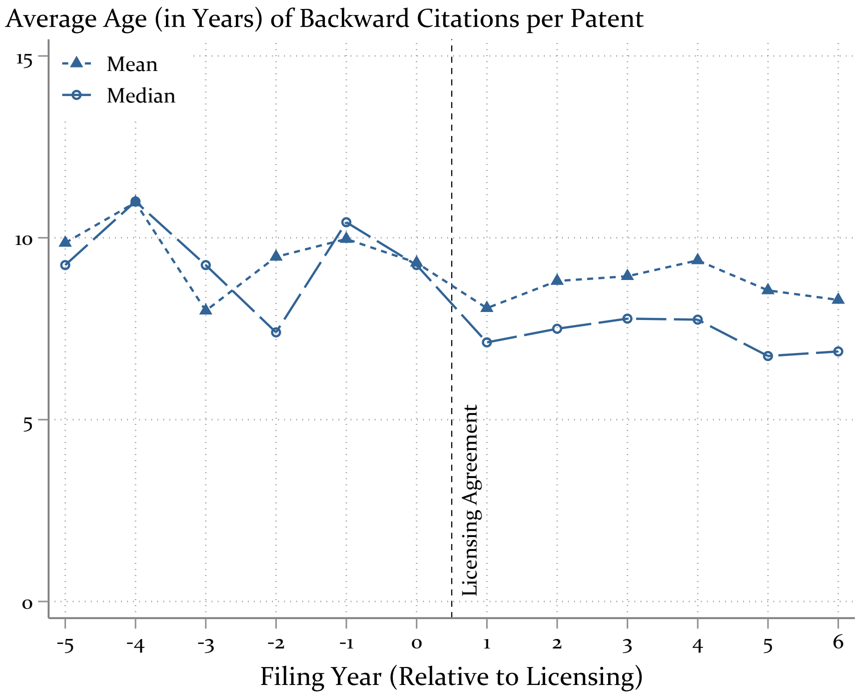
the increase in the licensees' patenting is driven by patents that are connected to RCA through citations, whereas there was no statistically significant change in the number of patents that do not build on RCA. Finally, as shown in columns (3) to (5), there was also an increase in the number of high-quality patents filed by RCA's licensees, and their innovation activity became more diverse.

Overall, the matching analysis shows that the finding that Japanese firms increased their patenting after receiving a license from RCA is robust to using an alternative control group. As discussed in the main part of the chapter, it is challenging to precisely quantify this effect, though. This is because the firms that entered into licensing agreements with RCA naturally represent a non-random subset of all Japanese firms. Therefore, I view the present analysis primarily as descriptive and the magnitude of the estimates should be interpreted with caution.

B.2.2 Age of Backward Citations

As discussed in section 2.5, one concern with the rise in the licensees' patenting is that it may not necessarily represent novel innovation. In principle, it could also be the case

Figure B.4. Direct Effect: Age of Backward Citations



Notes: The figure depict the average age of the backward citations among the patents filed by RCA's Japanese licensees. The age refers to the time (in years) between the filings of the cited patent and the citing patent. The figure shows both the mean and the median citation age across all patents filed in the same year, relative to the year in which the applicant received the (first) license from RCA.

that Japanese firms started to file patents in the US for existing technologies, which they could not previously use in the product market without a license from RCA.

To address this concern, I compute the average age of all backward citations for each US patent filed by a licensee. The age refers to the time between the filing of the cited patent and the filing of the focal (citing) patent by the licensee. Figure B.4 shows how the mean and median citation age developed over time, relative to the year of the licensing agreement. The figure shows that, if anything, the average age of backward citations decreased slightly, meaning that the patents filed after licensing built on more recent prior art. Thus, the figure does not support the concern that Japanese licensees started filing patents for previously existing inventions, which should be reflected by an increase in the mean or median citation age.

Appendix C

Appendix to Chapter 3

C.1 Appendix to Section 3.2: Data, Prices, and Search

In this appendix, we provide additional details on the construction of our price dataset and of non-overlapping local markets. We also present supplementary descriptive evidence on search and price dispersion in the retail fuel market.

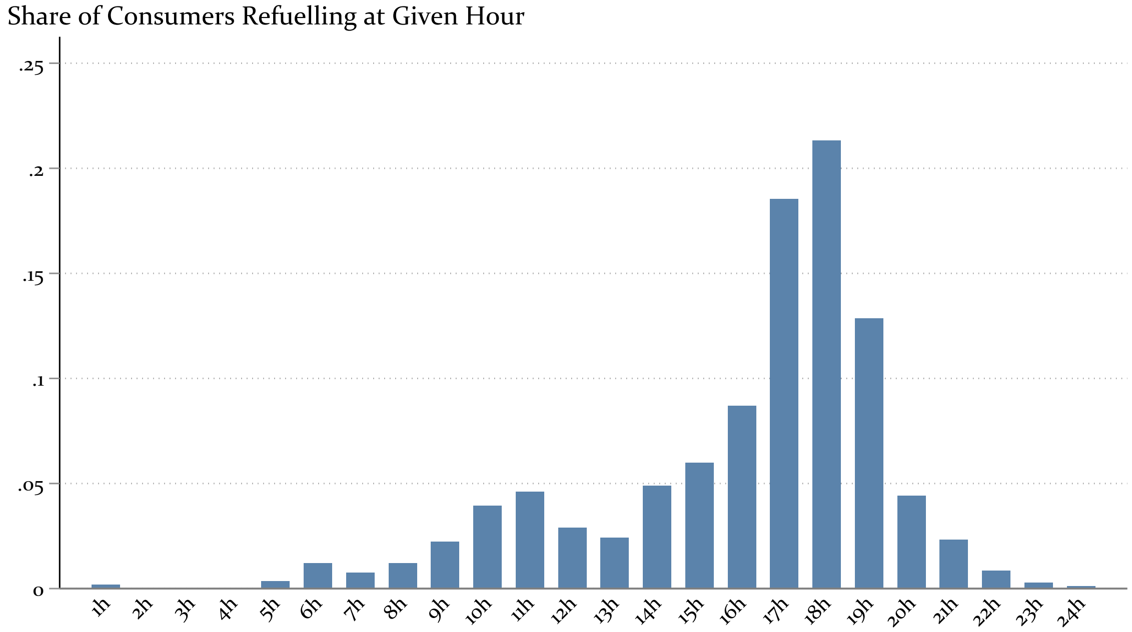
C.1.1 Construction of Price Dataset

We construct the station-level price panel for Germany and France as follows. For each fuel station in our dataset, we observe a price every time it changes, along with the precise time and date stamp for every change. On average, in 2019, fuel stations in Germany changed prices 14 times per day, whereas there was typically one price change per day at French fuel stations. Based on the distribution of price changes, we construct hourly fuel prices from 6am until 10pm for every fuel station in Germany and France.

In the next step, we compute daily weighted average prices from the hourly distribution of price changes. To construct the weights, we use data on hourly fuelling patterns reported in a representative survey among drivers for the German Federal Ministry of Economic Affairs. Figure C.1 shows the share of motorists in Germany who fuel at a particular time of day. We further reweight the hourly shares to produce weights for the hours between 6am and 10pm.

In Table 3.1 in the main part of the chapter, we also compute prices net of taxes and duties for both Germany and France. In Germany, taxes and duties consist of VAT, a lump-sum energy tax, and a fee for oil storage. The lump-sum energy tax is 65.45 eurocents per liter for *E5* and *E10* gasoline, and 47.04 eurocents per liter for diesel.

Figure C.1. Daily Fuelling Patterns in Germany



Notes: The figure shows shares of drivers in Germany who fuel at a given hour of the day. Data are based on a representative survey of motorists in Germany, commissioned by the German Federal Ministry of Economic Affairs.

The fee for oil storage is 0.27 eurocents per liter for *E5* and *E10* and 0.30 eurocents per liter for diesel.[1] Before the temporary VAT reduction in 2020, the German VAT rate on retail fuel was 19%. In mainland France, fuel products are subject to a lump-sum tax of 60 to 70 eurocents per liter, depending on the metropolitan region and fuel type.[2] In addition, the French VAT rate on retail fuel is 20%.

We impose a few restrictions on which fuel stations we include in our analysis. In Germany, we drop stations located on highways (i.e., 'Autobahn') because prices at these stations are typically around 20 to 30 eurocents higher than those at regular fuel stations. We identify highway stations based on their address and manual checks. In France, we focus only on stations in mainland France (i.e., we exclude stations on the island of Corsica and overseas).

C.1.2 Non-Overlapping Markets

To group fuel stations into non-overlapping local markets, we use an agglomerative hierarchical clustering algorithm based on the driving time between stations. This

[1] See `https://www.avd.de/kraftstoff/staatlicher-anteil-an-den-krafstoffkosten/` (last accessed: 19 February 2024).

[2] See `http://www.financespubliques.fr/glossaire/terme/TICPE/` (last accessed: 19 February 2024).

approach follows Carranza et al. (2015), Lemus and Luco (2021), and Assad et al. (2020).

In the first step, we compute the driving time between all pairs of fuel stations in each country. To do this, we use the `osrmtime` Stata package by Huber and Rust (2016), which relies on OpenStreetMap data using the Open Source Routing Machine (OSRM).

Next, we implement the hierarchical clustering algorithm separately for stations in Germany and France. The algorithm begins with each station in a separate cluster. Then, iteratively, the algorithm combines the closest two clusters into a larger cluster and records the additional driving time required to link the clusters. We use average linkage, implying that two clusters are linked based on the average driving time between the stations in the two clusters. As this procedure moves on, the algorithm builds a clustering tree that indicates which clusters have been linked at which iteration and how much additional driving time is required to link two clusters.[3] Eventually, all stations are combined into a single cluster.

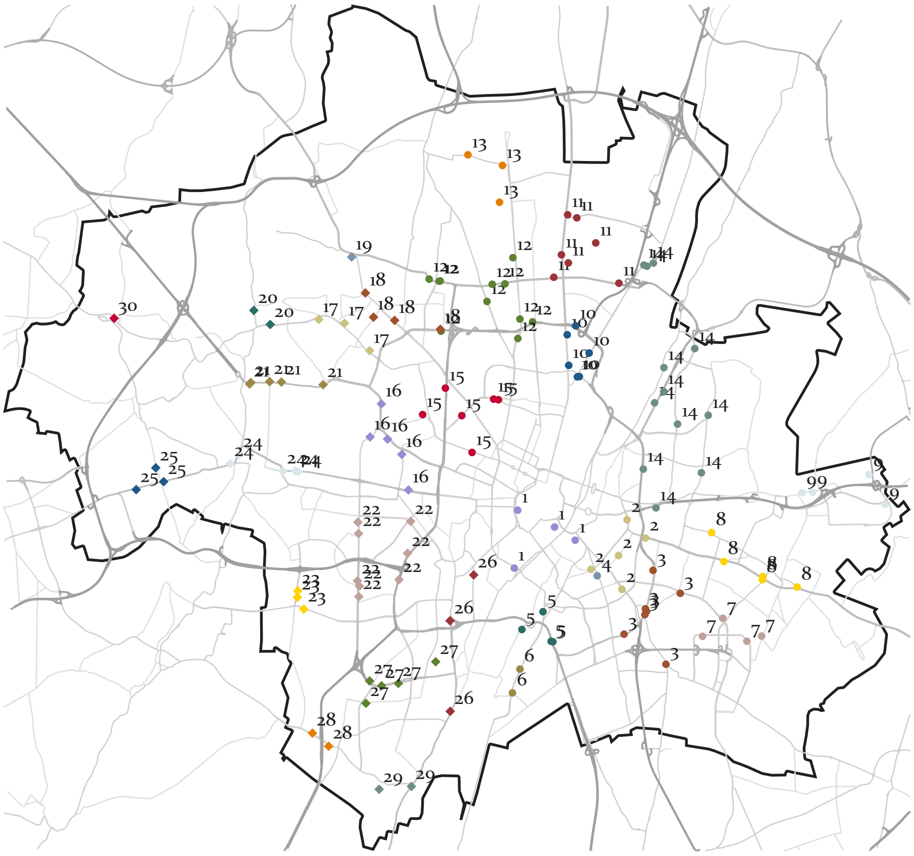
The objective of the clustering exercise is to find clusters of stations that are naturally separated from each other. The 'height' of each link (i.e., the average driving time needed to link one cluster and another) is informative about such natural separations. Formally, we compute an inconsistency coefficient for each link, which captures the height of the current link relative to the heights of previous links.[4] A high inconsistency coefficient indicates that two clusters are far apart from each other (i.e., there is an inconsistency in the linking of the two clusters). The idea underlying this inconsistency measure is twofold. First, two clusters linked at a low additional driving time are more likely to belong to one local geographic market than two clusters linked at a much higher additional driving time. This is true irrespective of whether the original clusters are individual stations or groups of stations linked in a previous iteration. Second, if the driving time required to link two clusters is similar to the driving time required to link clusters (or individual stations) in previous iterations, then there is unlikely to be a natural border between this group of stations. In contrast, if the driving time required to link two clusters is much higher than the time needed to drive from station to station *within* these two clusters, then the two clusters are likely to represent separate local markets.

Finally, based on the clustering tree and the inconsistency coefficients for each link, we group stations into non-overlapping markets. We do this by pruning the tree at a selected threshold in the distribution of the inconsistency coefficients. We choose to

[3] See Appendix C in Lemus and Luco (2021) for an example and illustration.

[4] For additional details, see `http://cda.psych.uiuc.edu/multivariate_fall_2012/matlab_help/cluster_analysis.pdf` (last accessed: 19 February 2024).

Figure C.2. Example of Local Markets in the City of Munich

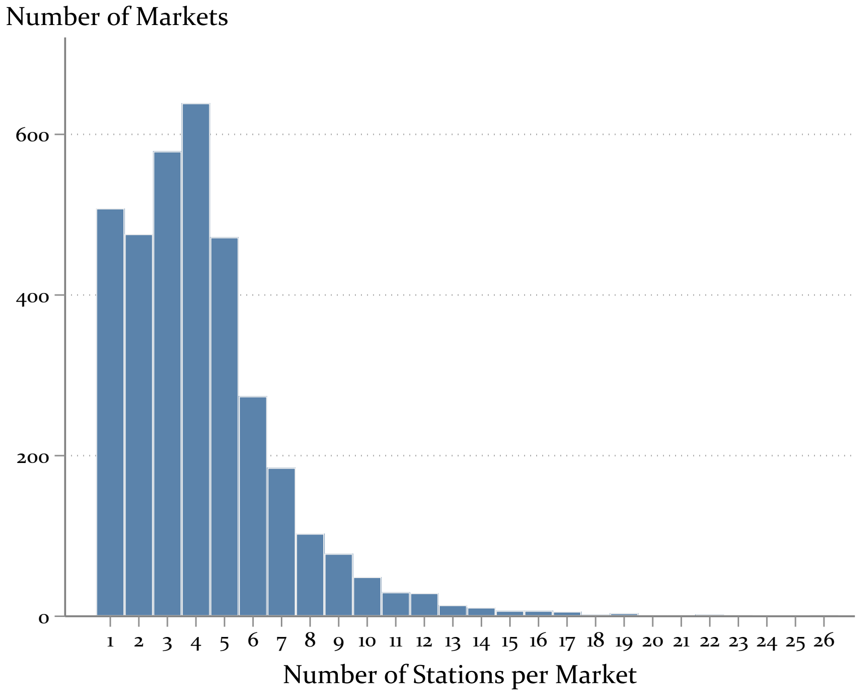


Notes: The figure shows a map of Munich where all fuel stations (indicated by solid squares or diamonds) are grouped into non-overlapping markets by means of the agglomerative hierarchical clustering algorithm. The solid black line indicates the city boundary. The gray lines represent the road network, with thinker and darker lines indicating larger roads.

prune the clustering tree at the 85th percentile of this distribution. This threshold is in line with choices in prior literature on retail fuel markets, which uses the 80th (Assad et al., 2020) or 90th percentile (Lemus and Luco, 2021). We verify that our results do not hinge on the specific choice of the threshold.

As pointed out by Lemus and Luco (2021) and Assad et al. (2020), the advantage of this agglomerative hierarchical clustering approach is that researchers do not need to specify the number or size of markets. Instead, in the entire procedure outlined above, we decide only where to prune the clustering tree. One potential drawback of the approach is that the clustering algorithm tends to group stations in rural areas together, although they may be far away from each other in terms of absolute driving time. Therefore, we make one more explicit choice and define all fuel stations with no

Figure C.3. Distribution of Market Sizes in Germany



Notes: The figure shows the distribution of market sizes in Germany when we apply the hierarchical clustering algorithm to identify nonoverlapping local markets. The histogram depicts the number of markets (on the y-axis) by the number of stations per market (on the x-axis).

competitor within a 10-minute drive as monopoly markets, without including them in the clustering procedure.[5]

To illustrate the outcome of applying the clustering algorithm, the map in Figure C.2 shows how we group stations into non-overlapping markets, using the city of Munich as an example. The figure highlights that nearby stations are usually grouped together into one local market. It also shows that the algorithm often identifies 'natural' clusters of stations. In Figure C.3, we also present the distribution of market sizes in Germany. The median market consists of four stations, and the 90th percentile is at seven stations. That is, the vast majority of the local markets defined by the clustering algorithm are of a very reasonable size.

[5]The European Commission also frequently uses market definitions based on 10-minute driving time. An example is the Commission's assessment of the recent takeover of OMV stations by EG Group in the German market (see Case M.10134 – EG Group/OMV Germany Business).

C.1.3 Search Intensity by Fuel Type

In this section, we use data on search queries in 2015 from a major German price comparison smartphone app to confirm that the share of well-informed consumers is higher for diesel than for gasoline and higher for *E10* than *E5*.

Panel (A) of Figure C.4 shows the daily number of distinct users searching for fuel prices by fuel type. Normalizing the number of users by the number of registered vehicles, we see that the ratio of searchers to the number of vehicles in circulation is approximately 50% higher for diesel than for gasoline. We report the number of distinct searchers rather than the total number of searches to adjust for the higher mileage of diesel drivers.

Panel (B) of Figure C.4 shows the number of distinct searchers for *E5* and *E10*, divided by the number of gasoline vehicles in circulation and adjusted for the relative market shares of *E5* and *E10* within the gasoline market. This shows that the search intensity is substantially higher among consumers buying *E10* than those purchasing *E5*.

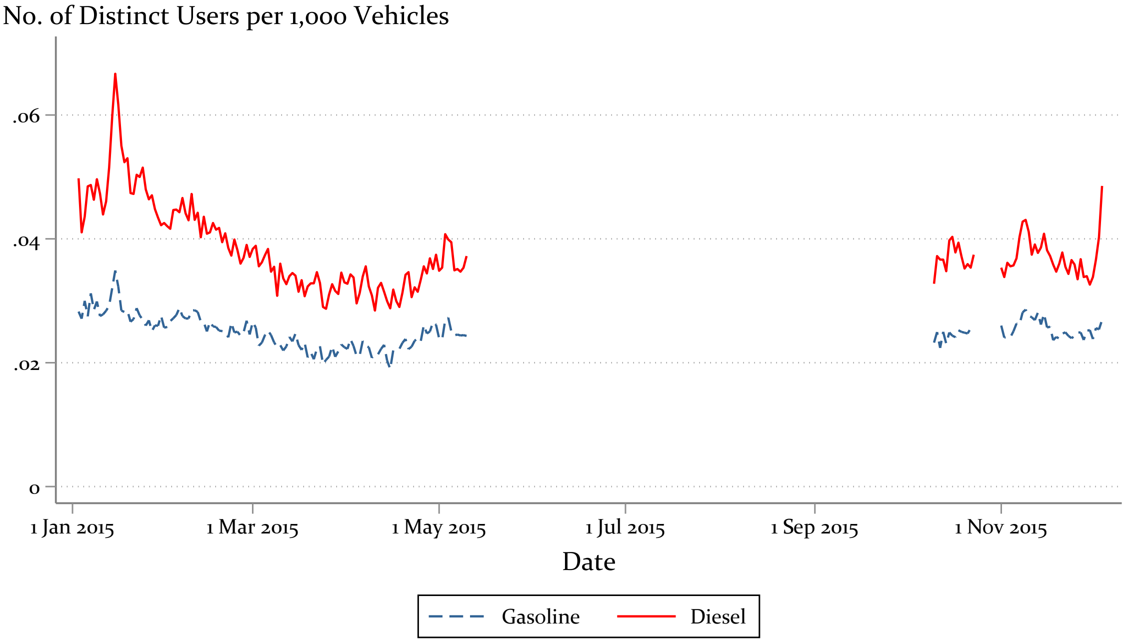
C.1.4 Additional Evidence on Search and Price Dispersion

In this section, we present additional evidence on search intensity and price dispersion. Figure 3.1 in the main part of the chapter shows average daily price cycles for *E10* in Germany in 2019. We now present price cycles at a more disaggregated level to show that these pricing patterns do not merely result from averaging over time and across stations. Therefore, in Figure C.5, we zoom in on one local market in the city of Munich (see market no. 24 in the map in Figure C.2) and present the stations' raw prices on four consecutive Mondays in the fall of 2019. Several things are noteworthy in the figure. First, on each of the four days, the stations' pricing follows a similar pattern, in line with that shown in Figure 3.1. Price increases typically occur at the same time, whereas the timing of price decreases is more idiosyncratic. Second, there are persistent differences in the average price level across stations, consistent with some degree of product differentiation (e.g., due to station amenities). For example, in Figure C.5, Mr. Wash typically sets the lowest price, as this is the only station that does not belong to a vertically integrated brand. Finally, even at a particular time, the order of the stations' prices may vary across different days. This indicates that there is a substantial amount of price variation that is unpredictable to consumers, which is consistent with the mixed strategy equilibrium in our theoretical model.

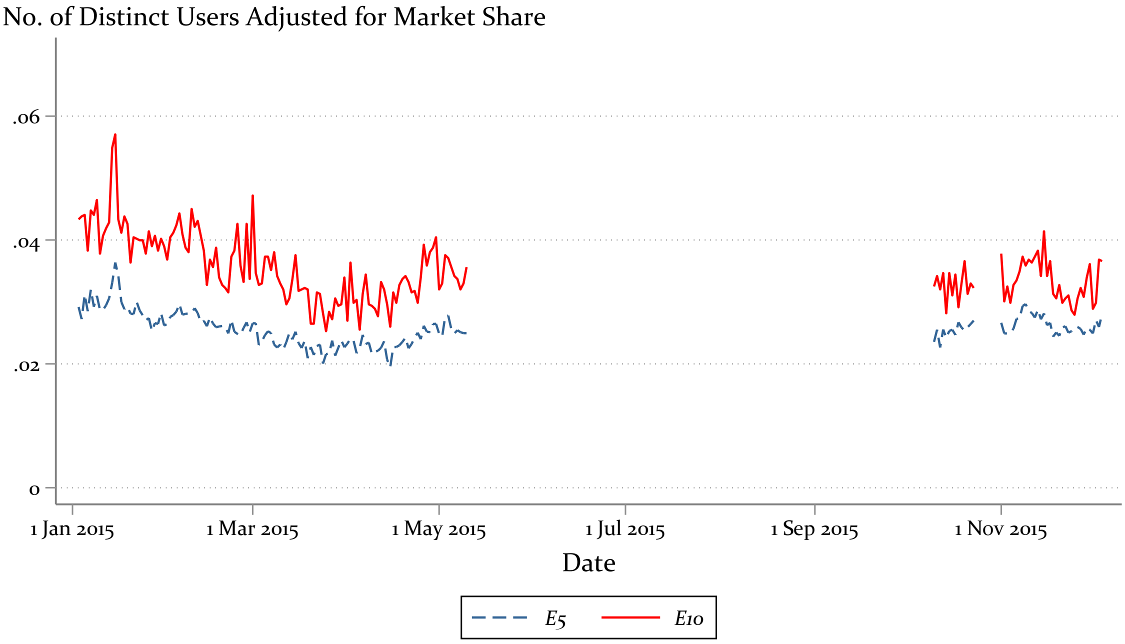
In Table 3.2 in the main part of the chapter, we analyse price dispersion more systematically by computing within-market price residuals for 5pm prices. Panel (A)

Figure C.4. Consumer Search Patterns in Germany

(A) Diesel vs. Gasoline

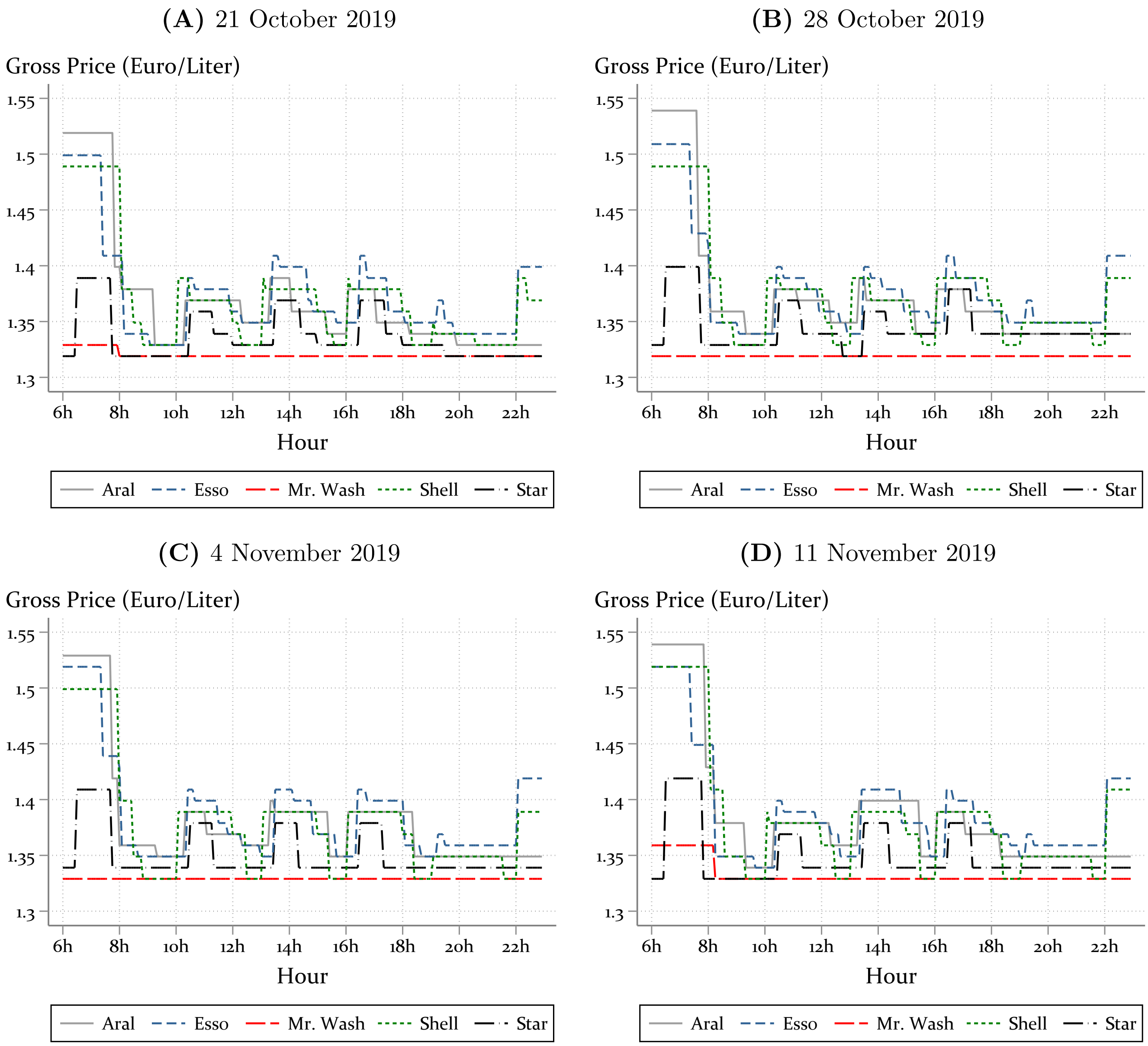


(B) *E5* vs. *E10*



Notes: The figure shows the daily number of searchers by fuel type on a major German smartphone app. The data are available for January to mid-May and mid-October to early December 2015. Panel (A) shows the number of distinct users who search for diesel vs. gasoline prices per 1,000 diesel or gasoline vehicles in circulation. The solid line corresponds to the search intensity for diesel, whereas the dashed line corresponds to the search intensity for gasoline. Panel (B) shows the number of distinct users who search for *E5* vs. *E10* per 1,000 gasoline vehicles in circulation and adjusted for the relative market shares of *E5* and *E10*. The solid line corresponds to the search intensity for *E10*, whereas the dashed line corresponds to the search intensity for *E5*.

Figure C.5. Daily Price Cycles for *E10* on Selected Mondays in One Local Market

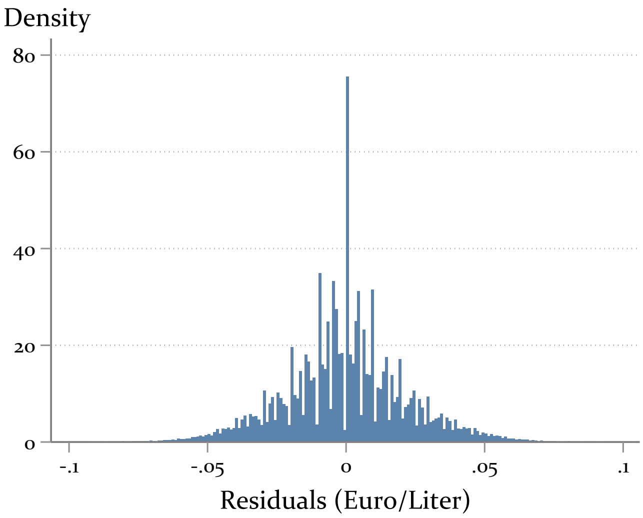


Notes: The figure shows prices of *E10* for five different stations in one local market in the city of Munich at different times of a specific day. Fuel prices are updated at five-minute intervals. Panels (A), (B), (C), and (D) depict prices on 21 October, 28 October, 4 November, and 11 November (all Mondays), respectively.

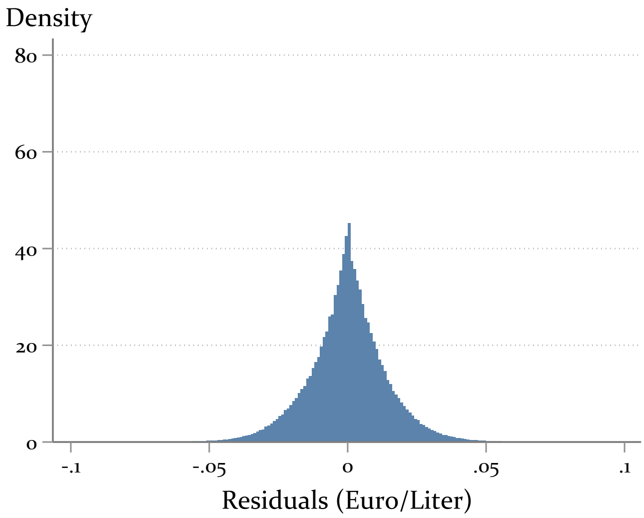
of Figure C.6 graphically illustrates the distribution of these residuals. Panel (B) shows the corresponding residuals when we additionally include station fixed effects to absorb any time-invariant price differences across stations. These residuals correspond to variation in prices that is unpredictable even to the most sophisticated consumers. Consistent with our stylized fact and the numbers in Table 3.2 in the main part of the chapter, Figure C.6 shows that this unpredictable price dispersion is substantial.

Figure C.6. Within-Market Price Residuals at 5pm, 2019

(A) Market × Date FE



(B) Market × Date FE and Station FE



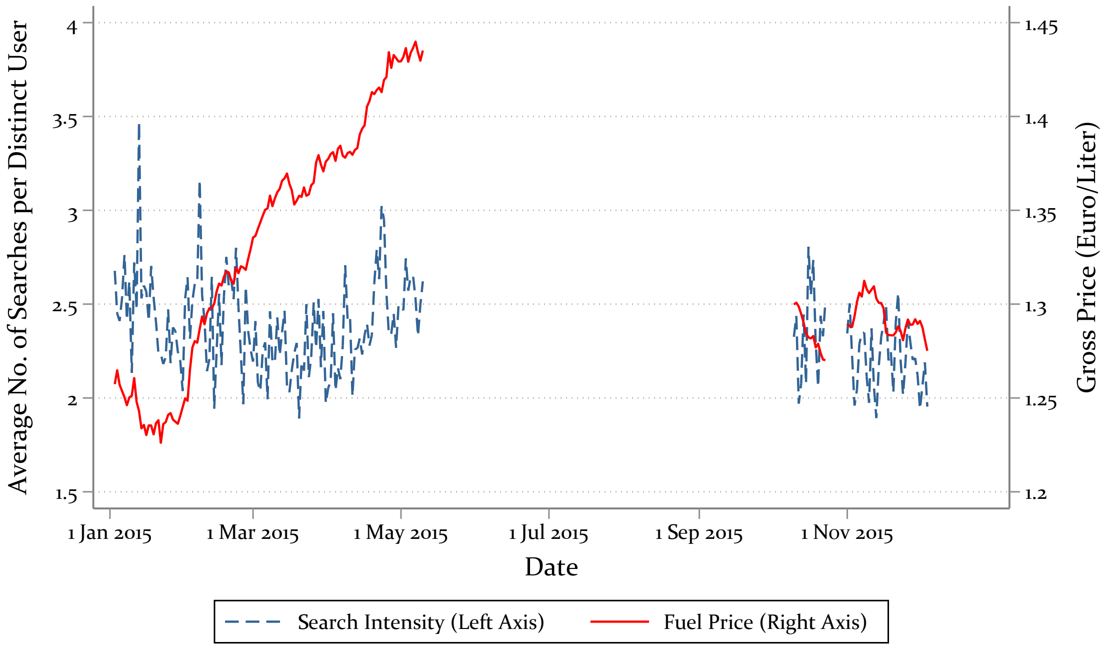
Notes: The figure shows the distribution of the deviation of a fuel station's price from the average price in the same market (i.e., within-market residuals) on the same day at 5pm for *E10* and for all stations that are not local monopolists. We use data for all weekdays in 2019. Panel (A) shows residuals when we include only market × date fixed effects. In Panel (B), we additionally include station fixed effects.

Next, we investigate whether and how search and price dispersion are correlated with the absolute price level. As outlined in section 3.7, the model with endogenous search by Tappata (2009) predicts that consumers search more when prices are low. Similarly, the model predicts that price dispersion (i.e., a measure of the gains from search) is high when prices are low.

Figure C.7 shows the average number of searches per app user in 2015 for *E10*, along with the development of the gross price of *E10*. As can be seen in the figure, search intensity and the price level are almost entirely uncorrelated. That is, in our empirical application, there is no evidence that consumers change their search behaviour on the intensive margin in response to changes in the price level.

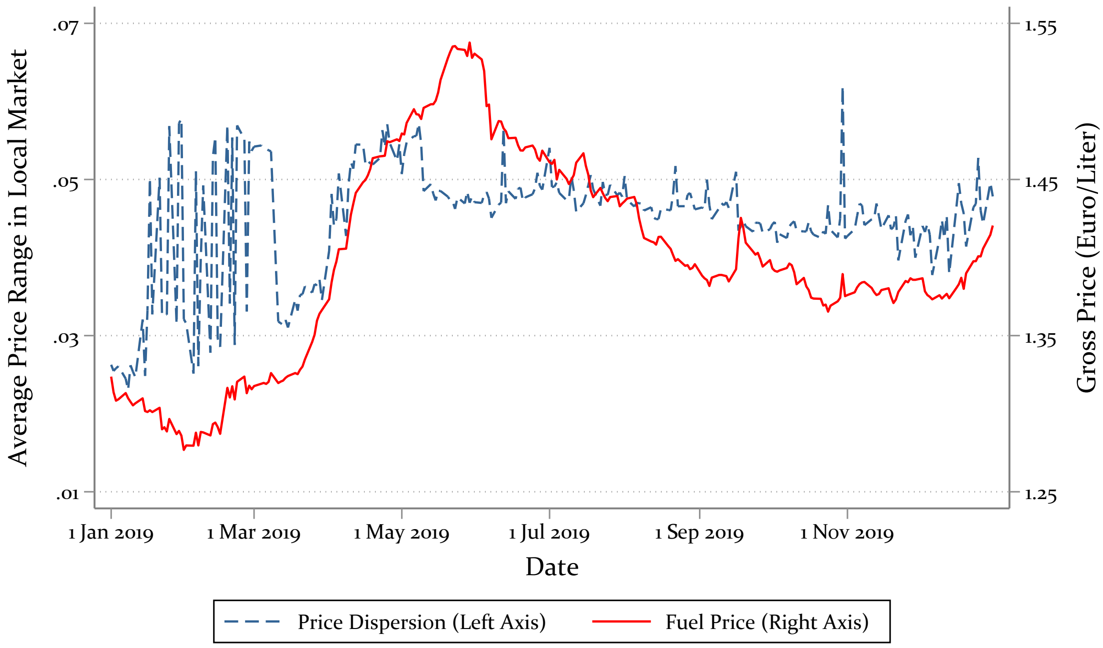
Figure C.8 depicts the relationship between price dispersion and the price level for *E10* in 2019. We compute price dispersion as the difference between the maximum and the minimum price within a local market, using daily prices at 5pm. Thus, the figure corresponds to the specification with only market × date fixed effects in the last column of panel (B) in Table 3.2 in the main part of the chapter. If anything, Figure C.8 points to a positive correlation between price dispersion and the price level. This would imply that gains from search are higher when prices are higher, which is at odds with the predictions of Tappata (2009).

Figure C.7. Search per User and Price Level of *E10*, 2015



Notes: The figure shows the development of daily search intensity (on the left axis, dashed line) and the daily weighted average fuel price (on the right axis, solid line) for *E10* in Germany in 2015. Search intensity is measured by the average number of searches per distinct user on a major German smartphone app. The search data are available for January to mid-May and mid-October to early December 2015.

Figure C.8. Price Dispersion and Price Level of *E10*, 2019



Notes: The figure shows the development of daily price dispersion (on the left axis, dashed line) and the daily weighted average fuel price (on the right axis, solid line) for *E10* in Germany in 2019. We compute price dispersion as the difference between the maximum and the minimum price within a local market, using prices at 5pm.

C.2 Appendix to Section 3.3: Theoretical Model

This appendix complements the theoretical model in section 3.3. Here, we formally solve the model, prove our propositions, and consider extensions such as endogenous entry or pass-through of marginal costs.

C.2.1 Equilibrium Price Distribution

Lemma 1. *There is no pure strategy Nash equilibrium in prices if $N \geq 2$.*

Proof. Suppose that all N sellers choose to set the same price strictly above the constant marginal cost c. Then, all sellers receive a share $\frac{1}{N}$ of shoppers and non-shoppers. This cannot be a stable equilibrium because all sellers have an incentive to marginally undercut the common price and attract all shoppers. All sellers setting the price at the constant marginal cost c also cannot be a stable equilibrium because sellers can profitably deviate by setting a higher price and serving only uninformed consumers.

Finally, suppose that sellers play pure strategies in which at least one seller chooses a lower price than the other sellers. This seller then serves all shoppers and its share of uninformed consumers. This cannot be an equilibrium because the lowest-price seller can always marginally increase its price without losing the shoppers to another seller. □

Lemma 2. *There are no mass points in the equilibrium pricing strategies.*

Proof. Suppose that any price is played with positive probability. This means that there is a positive probability of a tie for shoppers at that price. This cannot be an equilibrium because a seller could profitably deviate from that strategy by charging a marginally lower price with the same probability and capture all shoppers in that case.[6] □

Lemma 3. *There is a unique symmetric mixed strategy Nash equilibrium where all sellers draw a price from the distribution $F(p_i)$ on the interval $[\underline{p}, p_r]$, where*

$$\underline{p} = \frac{p_r}{\frac{\phi N}{1-\phi} + 1} + c\frac{1+\tau}{1+\frac{1-\phi}{\phi N}},$$

$$p_r = \begin{cases} E[p] + s & \textit{if } E[p] + s < v \\ v & \textit{otherwise} \end{cases}, \ \textit{and}$$

[6] For a more detailed proof, see Varian (1980).

$$F(p_i) = 1 - \left(\frac{p_r - p_i}{p_i - c(1+\tau)} \frac{1-\phi}{N\phi} \right)^{\frac{1}{N-1}}.$$

The expected profits of a seller are

$$E[\pi_i] = \left(\frac{p_r}{1+\tau} - c \right) \frac{1-\phi}{N} M.$$

The expected price is

$$E[p] = \underline{p} + \left(\frac{1-\phi}{N\phi} \right)^{\frac{1}{N-1}} \int_{\underline{p}}^{p_r} \left(\frac{p_r - p}{p - c(1+\tau)} \right)^{\frac{1}{N-1}} dp.$$

The expected minimum price is

$$E[p_{min}] = \frac{1-\phi}{N\phi} \left[p_r - E[p] + (p_r - c(1+\tau))c(1+\tau) \int_{\underline{p}}^{p_r} (p - c(1+\tau))^2 F(p) dp \right].$$

Proof. We begin by deriving the reservation price of non-shoppers p_r. Non-shoppers can search sequentially at an incremental search cost s. A necessary condition for search to occur, irrespective of the price initially drawn, is that the sum of the expected price at the next draw and the sequential search cost not exceed the valuation of the good. If this is fulfilled, non-shoppers with a particular first draw of p search as long as the expected gain from searching is greater than s. Thus, search occurs as long as

$$s < p - \int_{\underline{p}}^{p_{max}} p f(p) dp. \tag{C.1}$$

The reservation price of non-shoppers is such that they are exactly indifferent between continuing to search and buying at that price. No consumer buys at a price above the reservation price of non-shoppers. At the same time, sellers who do not sell to shoppers want to charge non-shoppers their reservation price. The maximum of the support of prices from which sellers draw in equilibrium is therefore $p_{max} = p_r$. Following Stahl (1989), a consistent reservation price $p_r \leq \upsilon$ must therefore satisfy

$$H(p_r; \phi, N, s) \equiv p_r - \int_{\underline{p}}^{p_r} p f(p) dp - s = 0. \tag{C.2}$$

Stahl (1989) shows that H has a unique root or none at all for a general class of demand functions that include linear demand. Thus, in this case, there is no other symmetric mixed strategy Nash equilibrium of the pricing game.

As explained before, if the sum of the expected price at the next draw and the sequential search cost exceed the valuation υ, search never occurs. In this case, the reservation price is simply the valuation of the good. The equilibrium reservation price of non-shoppers is thus

$$p_r = \begin{cases} E[p] + s & \text{if } E[p] + s < v \\ v & \text{otherwise} \end{cases}. \tag{C.3}$$

Since it is never an equilibrium strategy for any seller to choose a price above the reservation price of non-shoppers, there is no sequential search in equilibrium.

Next, we turn to finding the lowest price that sellers may draw in equilibrium, which is denoted by $\underline{p}$. Any price drawn with positive probability in equilibrium should yield the same expected profit. The expected profit from setting the price at $\underline{p}$ therefore has to equal the expected profit from setting the reservation price, thus,

$$E[\pi(\underline{p})] = E[\pi(p_r)]. \tag{C.4}$$

Since we established that there are no mass points in the equilibrium pricing strategies, the probability of a tie is zero. A seller setting its price at $\underline{p}$ will therefore attract all shoppers and its share of non-shoppers that randomly visit its store. A seller setting its price at p_r will never attract any shoppers and will serve its share of non-shoppers. We can therefore rewrite the expected profits as

$$\left(\frac{\underline{p}}{1+\tau} - c\right)\left(\phi + \frac{1-\phi}{N}\right) M = \left(\frac{p_r}{1+\tau} - c\right)\frac{1-\phi}{N} M. \tag{C.5}$$

We can simplify this expression and rearrange it to yield an expression for the lowest price that sellers may draw in equilibrium:

$$\underline{p} = \frac{p_r}{\frac{\phi N}{1-\phi} + 1} + c\frac{1+\tau}{1 + \frac{1-\phi}{\phi N}}. \tag{C.6}$$

The last ingredient necessary to characterize the distribution from which sellers draw prices in equilibrium is the density function of the distribution. To derive the density function, we can again exploit the equiprofit condition that

$$E[\pi(p_i)] = E[\pi(p_r)] \qquad \forall \quad p_i \in [\underline{p}, p_r]. \tag{C.7}$$

With probability $(1 - F(p_i))^{N-1}$, a seller choosing price p_i has the lowest price of all N sellers and will thus sell to all shoppers and its share of non-shoppers. With probability

$1-(1-F(p_i))^{N-1}$, there is another seller charging a lower price, and thus seller i sells only to its share of non-shoppers. Expected profits can be written as

$$\begin{aligned}&\left(\frac{p_i}{1+\tau}-c\right)\left(\phi+\frac{1-\phi}{N}\right)(1-F(p_i))^{N-1}M\\&+\left(\frac{p_i}{1+\tau}-c\right)\left(\frac{1-\phi}{N}\right)(1-(1-F(p_i))^{N-1})M=\left(\frac{p_r}{1+\tau}-c\right)\frac{1-\phi}{N}M.\end{aligned} \tag{C.8}$$

We can solve this equation for the equilibrium density function according to which each seller i draws its prices from the support $[\underline{p}, p_r]$:

$$F(p_i)=1-\left(\frac{p_r-p_i}{p_i-c(1+\tau)}\frac{1-\phi}{N\phi}\right)^{\frac{1}{N-1}}. \tag{C.9}$$

For a given number of entrants N and a given set of exogenous parameters, equations (C.3), (C.6), and (C.9) uniquely identify the symmetric mixed strategy Nash equilibrium in prices.

We can derive the expected profit of each seller i in this equilibrium. Since the expected profit of each seller in the symmetric equilibrium is the same for any price chosen with positive probability, the expected profit of seller i drawing a price from the equilibrium price distribution is

$$E[\pi_i]=E[\pi(p_r)]=\left(\frac{p_r}{1+\tau}-c\right)\frac{1-\phi}{N}M. \tag{C.10}$$

Finally, we can derive the expected prices paid by non-shoppers and shoppers, namely, the expected price and the expected minimum price. The expected price is

$$E[p]=\int_{\underline{p}}^{p_r}pf(p)dp=p_r-\int_{\underline{p}}^{p_r}F(p)dp \tag{C.11}$$

after integrating by parts. We can then insert the equilibrium price distribution and simplify the expression, which yields

$$E[p]=\underline{p}+\left(\frac{1-\phi}{N\phi}\right)^{\frac{1}{N-1}}\int_{\underline{p}}^{p_r}\left(\frac{p_r-p}{p-c(1+\tau)}\right)^{\frac{1}{N-1}}dp.$$

To derive the expected minimum price, we begin by setting up the probability density function of the minimum price. This can be written as

$$f_{min}(p)=N(1-F(p))^{N-1}f(p). \tag{C.12}$$

After we insert $F(p)$ and simplify the expression, this yields

$$f_{min}(p) = \frac{p_r - p}{p - c(1+\tau)} \frac{1-\phi}{\phi} f(p). \tag{C.13}$$

The expected minimum price is then

$$E[p_{min}] = \int_{\underline{p}}^{p_r} p f_{min}(p) dp = \int_{\underline{p}}^{p_r} p \frac{p_r - p}{p - c(1+\tau)} \frac{1-\phi}{N\phi} f(p) dp. \tag{C.14}$$

After adding and subtracting $c(1+\tau)$ in the numerator of the first fraction and further simplifications, we obtain

$$E[p_{min}] = \frac{1-\phi}{\phi} \left[\int_{\underline{p}}^{p_r} p \frac{p_r - c(1+\tau)}{p - c(1+\tau)} f(p) dp - E[p] \right].$$

Finally, we can use integration by parts and rearrange terms to obtain the following expression for the expected minimum price:

$$E[p_{min}] = \frac{1-\phi}{\phi} \left[p_r - E[p] + (p_r - c(1+\tau)) c(1+\tau) \int_{\underline{p}}^{p_r} \frac{1}{(p - c(1+\tau))^2} F(p) dp \right].$$

□

C.2.2 Endogenous Entry

To consider endogenous entry, we assume that there is an infinite number of symmetric firms that can potentially enter the market. Each firm can enter the market for a fixed and sunk cost F.

In this case, the game proceeds in two stages. In the first stage, firms decide whether to enter the market. Entry occurs as long as the expected second-stage profits of the entrant are greater than or equal to the fixed and sunk cost of entry F. No further entry occurs if the next potential entrant cannot expect to recoup her entry costs.

In the main analysis, we assume that there is no entry and treat the number of sellers as exogenous. This is because our empirical study is concerned with a short-term tax adjustment during which entry seems unlikely. In other applications, it will make sense to endogenize the number of active sellers for the analysis of pass-through as well. Unless otherwise stated, we focus on the case where $N^* \geq 2$ since there need to be at least two sellers active in the market for the informedness of consumers to matter.

Lemma 4. *Under free entry and with a number of symmetric potential entrants large enough that the number of potential entrants always exceeds the number of firms that* *can be supported by the market, in equilibrium an integer number of N^* firms enter the market, such that*

$$\left(\frac{p_r}{1+\tau}-c\right)\frac{1-\phi}{F}M-1<N^*\leq\left(\frac{p_r}{1+\tau}-c\right)\frac{1-\phi}{F}M.$$

Proof. Suppose that there is a large number of symmetric firms that are sequentially asked whether they want to enter the market at the fixed and sunk cost F, knowing how many firms decided to enter before them. Firms are going to decide to enter the market as long as their expected second-stage profits are at least as high as the fixed and sunk cost F. In equilibrium, the first N firms asked to enter will accept, and firm $N+1$ and all firms following thereafter will reject if, and only if, the expected second-stage profits of firms $1,...,N$ are equal to F or higher and the expected second-stage profits of firm $N+1$ are lower than F.

To derive the condition for the equilibrium number of firms entering the market, we use the expression for the expected second-stage profit of firm i in equation (C.10). We calculate the expected second-stage profits with N and $N+1$ entrants and rearrange these to yield a condition on the equilibrium number of entrants. In equilibrium, an integer number of N firms enter the market, such that

$$\left(\frac{p_r}{1+\tau}-c\right)\frac{1-\phi}{F}M-1<N^*\leq\left(\frac{p_r}{1+\tau}-c\right)\frac{1-\phi}{F}M. \tag{C.15}$$

□

C.2.3 Pass-Through of Marginal Costs

Next, we analyse how marginal costs or per unit taxes are passed through to consumers. Many of the results and intuitions regarding ad valorem taxes directly translate to marginal costs (or per unit taxes).

Proposition 5. *With $0<\phi<1$, for any $\hat{c}>c$, the minimum element of the support of the equilibrium pricing strategy $\hat{\underline{p}}>\underline{p}$ and the Nash equilibrium pricing strategy with c first-order stochastically dominates (FOSD) the pricing strategy with $\hat{c}$, i.e., $\hat{F}(p)\leq F(p)\quad\forall p$.*

Analogously to the explanation for ad valorem taxes, this means that if the share of shoppers is strictly positive, an increase in c leads to a shift in the support of the prices from which sellers draw in equilibrium toward higher prices. Furthermore, for each price on the equilibrium pricing support, the likelihood that a drawn price is below said price decreases if marginal costs increase from c to $\hat{c}$.

As for the pass-through of ad valorem taxes, the pass-through of marginal costs converges to zero as the share of shoppers converges to zero. Since the minimum element of the support of prices and the density function monotonically move toward higher prices, other moments of interest, such as the expected price $E[p]$ and the expected minimum price $E[p_{min}]$, also increase.

We now turn to analysing how the pass-through rate of marginal costs or per unit taxes varies with the price sensitivity of consumers and the number of active sellers.

Proposition 6. *If the share of shoppers $\phi = 0$, marginal cost pass-through $\rho_c = 0$. If $\phi = 1$, there is full pass-through, i.e., $\rho_c = 1 + \tau$. As $\phi \to 1$, the pass-through rate $\rho_c \to 1 + \tau$.*

We can begin by looking at the cases when there are no shoppers and when there are only shoppers. If there are no shoppers, all sellers choose the monopoly price, and pass-through of marginal costs is zero. If all consumers are shoppers, there is full pass-through of marginal costs or per unit taxes.[7]

For all values of ϕ between zero and one, we can show that the pass-through rate of marginal costs to the lower bound of the equilibrium price strategy is strictly increasing in the share of shoppers. We can also show that the rate at which an increase in marginal costs from c to $\hat{c}$ reduces the probability that a drawn price is below a particular price p, i.e., from $F(p)$ to $\hat{F}(p)$, strictly increases in the share of shoppers. Thus, the pass-through rate of marginal costs increases in the share of shoppers.

Let us now consider how pass-through of marginal costs varies with the number of active sellers. As we will see, all of our results and intuitions with respect to ad valorem tax pass-through extend to marginal cost pass-through.

Proposition 7. *With $0 < \phi < 1$, as $N \to \infty$, the pass-through of c to the minimum element of the equilibrium price support converges to full pass-through, i.e., $\rho_{c,\underline{p}} \to 1+\tau$.*

As the number of sellers increases, competition for shoppers becomes fiercer, and the pass-through rate of marginal costs to $\underline{p}$ increases. Furthermore, we expect pass-through of marginal costs to $E[p]$ to first increase and then decrease, whereas pass-through to $E[p_{min}]$ should always increase as $N \to \infty$. The same reasoning as that laid out for ad valorem taxes applies.

Proposition 8. *With $0 < \phi < 1$, as $N \to \infty$, the pass-through of c to the expected price is non-monotonic.*

[7]Although an increase in the marginal cost from c to $\hat{c}$ leads to an increase of $(\hat{c} - c)(1 + \tau)$ to consumers, we would still classify this case as full pass-through (instead of overshifting), since the producer price increases by only $\hat{c} - c$.

As for the pass-through of an ad valorem tax, there is a non-monotonic relationship between the number of sellers and the pass-through of a per unit tax.

The results from the numerical simulation in Figure C.9 are very similar to those for ad valorem tax pass-through. As N increases, pass-through of c to the expected price first increases and then decreases. This is the case with and without sequential search. Pass-through to the expected minimum price always increases in the number of sellers if there is no sequential search. If non-shoppers can search sequentially, either the pass-through to the expected minimum price can be monotonically increasing in the number of sellers, or there can be a non-monotonic relationship between the number of sellers and the pass-through rate.

C.2.4 Proof of Propositions

Proof of Proposition 1. First, we assess the pass-through of τ to $\underline{p}$ if $0 < \phi < 1$.[8] Taking the first derivative with respect to τ, we find that

$$\frac{\partial \underline{p}}{\partial \tau} = c(1 + \frac{1-\phi}{\phi N})^{-1} > 0.$$

Thus, with $0 < \phi < 1$, pass-through of τ to the minimum element of the support of the equilibrium pricing strategy is strictly positive.

Next, we assess the pass-through of the ad valorem tax to $F(p)$ if $0 < \phi < 1$. Taking the first derivative with respect to τ, we find that

$$\frac{\partial F(p)}{\partial \tau} = -(\frac{1-\phi}{\phi N})^{\frac{1}{N-1}} \frac{1}{N-1} (\frac{p_r - p}{p - c(1+\tau)})^{\frac{1}{N-1}} \frac{c}{p - c(1+\tau)} < 0.$$

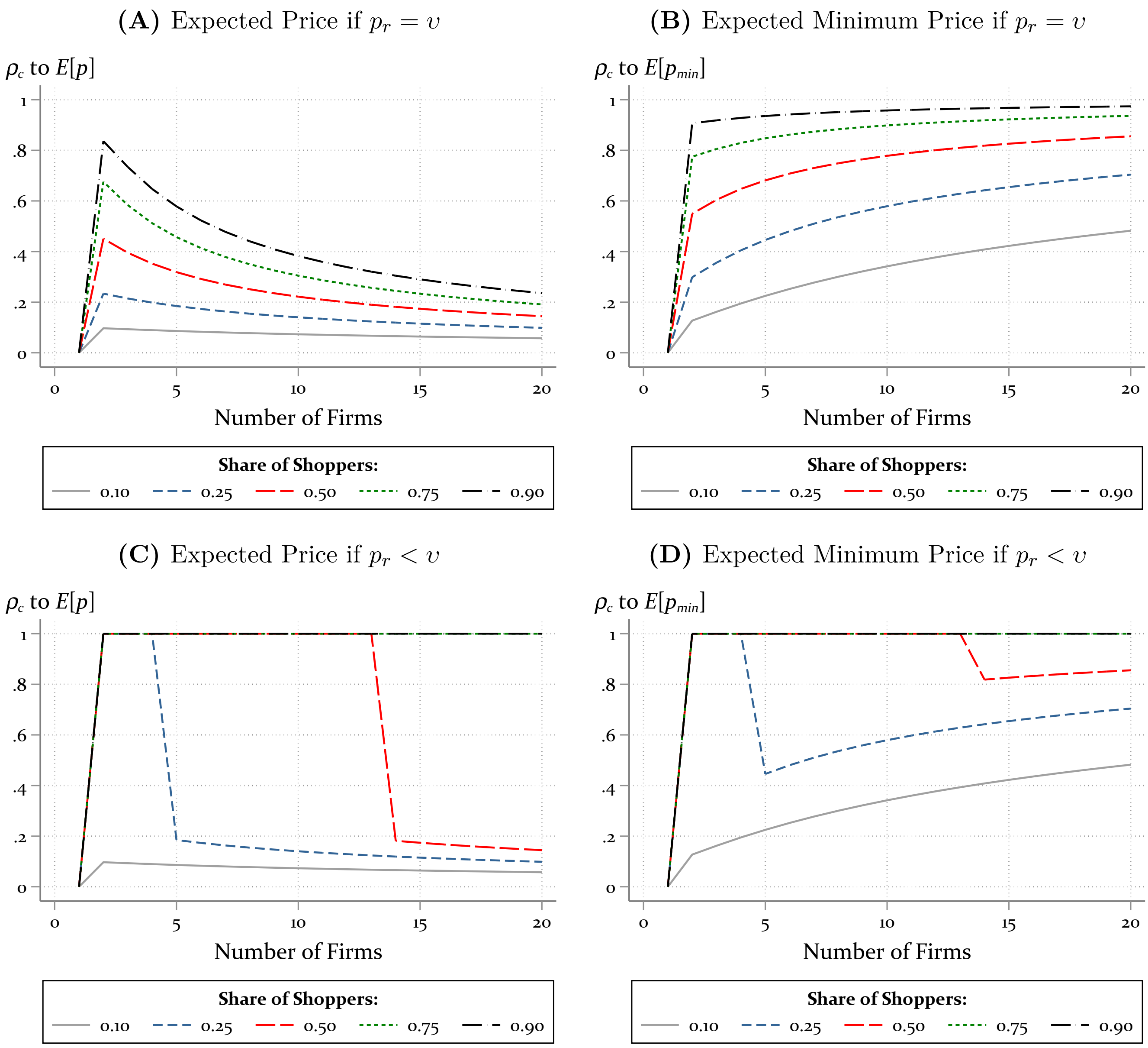
Thus, with $0 < \phi < 1$, for any $\hat{\tau} > \tau$, $\hat{F}(p) \leq F(p) \quad \forall p \in [\underline{p}, p_r]$. □

Proof of Proposition 2. Let us begin by examining the case where $\phi = 0$. In this case, the price equilibrium is a degenerate distribution at the monopoly price, with $\underline{p} = p_r = \upsilon$. An increase in τ is fully absorbed by sellers since these already fully extract the entire valuation from consumers.

Next, we examine the case where $\phi = 1$. In this case, the price equilibrium is a degenerate distribution at the perfectly competitive price, with $\underline{p} = p_r = c(1+\tau)$. An increase in the ad valorem tax τ is now fully passed through to consumers, as sellers already operate at zero profits and absorbing some of the marginal cost would mean that they would be making losses.

[8] $\underline{p}$ is not defined for $\phi = 0$ or $\phi = 1$.

Figure C.9. Numbers of Sellers and Marginal Cost Pass-Through



Notes: The figure shows simulation results of how the pass-through rate of marginal cost c varies with the number of sellers. Panels (A) and (B), respectively, show how the rate of pass-through to the expected price ($E[p]$) and to the expected minimum price ($E[p_{min}]$) varies with the number of sellers if the reservation price is exogenous. Panels (C) and (D) show the same if the reservation price of non-shoppers p_r is endogenous. In all panels, the different lines correspond to different values of the share of shoppers ϕ. Parameter values: $\upsilon = 4.5$, $\tau = 0.2$, $c = 0.4$, $\hat{c} = 0.44$, $s = \infty$ (without sequential search) and $s = 0.75$ (with sequential search).

Finally, we study the case where $0 < \phi < 1$. Let us begin by analysing how the pass-through rate changes with ϕ:

$$\frac{\partial^2 p}{\partial \tau \partial \phi} = c\left(1 + \frac{1-\phi}{\phi N}\right)^{-2} \frac{1}{\phi^2 N} > 0.$$

Thus, with $0 < \phi < 1$, the pass-through of τ to the minimum element of the support of the equilibrium pricing strategy strictly increases in ϕ.

Next, we consider how the effect of an increase from τ to $\hat{\tau}$ on the cumulative density function of the pricing strategy changes if ϕ increases:

$$\frac{\partial^2 F(p)}{\partial\tau\partial\phi} = (\frac{1}{N-1})^2(\frac{p_r - p}{p - c(1+\tau)})^{\frac{1}{N-1}} \frac{c}{p - c(1+\tau)}(\frac{1-\phi}{\phi N})^{\frac{1}{N-1}-1} \frac{1}{\phi^2 N} > 0.$$

Thus, for higher ϕ, an increase from τ to $\hat{\tau}$ decreases the probability that prices are below a certain p more strongly. $\square$

Proof of Proposition 3. To see how the pass-through rate of a value-added tax τ to the minimum element of the support varies with N, we study the limit to which the pass-through rate converges as $N \to \infty$. We find that

$$\lim_{N\to\infty} \rho_{\tau,\underline{p}} = \lim_{N\to\infty} \frac{\partial \underline{p}}{\partial\tau} \cdot \frac{1+\tau}{\underline{p}} = \frac{c(1+\tau)}{c(1+\tau)} = 1.$$

Thus, with $N \to \infty$, pass-through of a value-added tax to the minimum element of the support of the equilibrium pricing strategy converges to full pass-through. $\square$

Proof of Proposition 4. To show that pass-through of τ to $E[p]$ is non-monotonic in the number of sellers N, we show that pass-through is zero when $N = 1$, converges to zero when $N \to \infty$, and is strictly positive in between.

If $N = 1$, the monopolist faces unit demand and maximizes profits by setting $p = \upsilon$. It cannot achieve higher profits by charging a price above υ, as the quantity demanded drops to zero in this case. There is no reason to charge a price below υ, as this does not increase quantities. Thus, the monopoly price is independent of τ and therefore also does not vary with τ. Pass-through is zero.

To analyse pass-through when $N \to \infty$, let us begin by recalling the expression for the expected price $E[p]$, which is

$$E[p] = p_r - \int_{\underline{p}}^{p_r} F(p)dp.$$

The second key expression to recall is

$$F(p) = 1 - \left(\frac{p_r - p}{p - c(1+\tau)} \frac{1-\phi}{N\phi}\right)^{\frac{1}{N-1}}.$$

When $N \to \infty$, $\frac{1}{N-1}$ converges to zero, which means that the second term in $F(p)$ converges to 1, letting $F(p)$ converge to zero. When $F(p)$ converges to zero, it becomes immediately apparent that $E[p]$ converges to p_r. Since $E(p) = p_r$ and $p_r = E(p) + s$ cannot both be true when s is strictly positive, $p_r = \upsilon$ when $N \to \infty$. Thus, when $N \to \infty$, $E(p) \to \upsilon$, which, again, is independent of τ. Pass-through to $E(p)$ is thus zero when $N \to \infty$.

Let us now consider the case where $1 < N < \infty$ and $0 < \phi < 1$. As shown in the proof of Proposition 1, $\frac{\partial F(p)}{\partial \tau} < 0$ and $\frac{\partial \underline{p}}{\partial \tau} > 0$. In the case where $p_r = \upsilon$, the analysis is simple. In this case, $\frac{\partial p_r}{\partial \tau} = 0$. Since $F(p)$ decreases when τ increases and $\underline{p}$ increases, it must mean that $\int_{\underline{p}}^{p_r} F(p)dp$ decreases in τ. Thus, $\frac{\partial E[p]}{\partial \tau} > 0$, which means that pass-through of an ad valorem tax is strictly positive.

In the case where $p_r = E[p] + s$, let us first consider that after an increase in τ, it remains that $p_r = E[p] + s$. If this is true, then plugging p_r into $E[p]$ and simplifying leads to

$$s = \int_{\underline{p}}^{p_r} F(p)dp.$$

Since s is independent of τ and it remains true that $\frac{\partial F(p)}{\partial \tau} < 0$ and $\frac{\partial \underline{p}}{\partial \tau} > 0$, this equality can continue to hold after a tax increase only if $\frac{\partial p_r}{\partial \tau} > 0$. This is true if, and only if, $\frac{\partial E[p]}{\partial \tau} > 0$.

Finally, let us consider the case where p_r moves from $E[p] + s$ to υ after the tax increase. This can happen only if $E[p] + s < \upsilon$ before the tax increase and $E[p] + s \geq \upsilon$ after the tax increase, which requires $\frac{\partial E[p]}{\partial \tau} > 0$.

Since $\frac{\partial E[p]}{\partial \tau} = 0$ for $N = 1$, converges to 0 for $N \to \infty$, and is strictly positive for intermediate N, it therefore must be that there is a non-monotonic relationship between the number of sellers and the pass-through rate of an ad valorem tax τ when $0 < \phi < 1$. □

Proof of Proposition 5. We begin by assessing the pass-through of marginal costs to $\underline{p}$ if $0 < \phi < 1$. Taking the first derivative with respect to c, we find that

$$\frac{\partial \underline{p}}{\partial c} = (1 + \tau)(1 + \frac{1 - \phi}{\phi N})^{-1} > 0.$$

Thus, with $0 < \phi < 1$, pass-through of marginal costs to the minimum element of the support of the equilibrium pricing strategy is strictly positive.

Next, we assess the pass-through of marginal costs to $F(p)$ if $0 < \phi < 1$. Taking the first derivative with respect to c, we find that

$$\frac{\partial F(p)}{\partial c} = -(\frac{1 - \phi}{\phi N})^{\frac{1}{N-1}} \frac{1}{N-1} (\frac{p_r - p}{p - c(1 + \tau)})^{\frac{1}{N-1}} \frac{1 + \tau}{p - c(1 + \tau)} < 0.$$

Thus, with $0 < \phi < 1$, for any $\hat{c} > c$, $\hat{F}(p) \leq F(p) \quad \forall p \in [\underline{p}, p_r]$. □

Proof of Proposition 6. Again, we begin by examining the case where $\phi = 0$. In this case, the price equilibrium is a degenerate distribution at the monopoly price, with $\underline{p} = p_r = v$. An increase in marginal costs is fully absorbed by sellers since these already fully extract the entire valuation from consumers.

Next, we examine the case where $\phi = 1$. In this case, the price equilibrium is a degenerate distribution at the perfectly competitive price, with $\underline{p} = p_r = c(1+\tau)$. An increase in c is now fully passed through to consumers.[9]

Finally, we study the case where $0 < \phi < 1$. Let us begin by analysing how the pass-through rate changes with ϕ:

$$\frac{\partial^2 \underline{p}}{\partial c \partial \phi} = (1+\tau)(1+\frac{1-\phi}{\phi N})^{-2}\frac{1}{\phi^2 N} > 0.$$

Thus, with $0 < \phi < 1$, the pass-through of c to the minimum element of the support of the equilibrium pricing strategy strictly increases in ϕ.

Next, we consider how the effect of an increase from c to $\hat{c}$ on the cumulative density function of the pricing strategy changes if ϕ increases:

$$\frac{\partial^2 F(p)}{\partial c \partial \phi} = (\frac{1}{N-1})^2(\frac{p_r - p}{p - c(1+\tau)})^{\frac{1}{N-1}}\frac{1+\tau}{p-c(1+\tau)}(\frac{1-\phi}{\phi N})^{\frac{1}{N-1}-1}\frac{1}{\phi^2 N} > 0.$$

Thus, for higher ϕ, an increase from c to $\hat{c}$ decreases the probability that prices are below a certain p more strongly. $\square$

Proof of Proposition 7. To see how the pass-through rate of marginal costs to the minimum element of the support varies with N, we study the limit to which the pass-through rate converges as $N \to \infty$. We find that

$$\lim_{N\to\infty} \rho_{c,\underline{p}} = \lim_{N\to\infty} \rho_{c,\underline{p}}(1+\tau)(1+\frac{1-\phi}{\phi N})^{-1} = 1+\tau.$$

Thus, with $N \to \infty$, pass-through of marginal costs to the minimum element of the support of the equilibrium pricing strategy converges to full pass-through. $\square$

Proof of Proposition 8. To show that pass-through of c to $E[p]$ is non-monotonic in the number of sellers N, we show that pass-through is zero when $N = 1$, converges to zero when $N \to \infty$, and is strictly positive in between.

[9]Although an increase in the marginal cost from c to $\hat{c}$ leads to an increase of $(\hat{c} - c)(1+\tau)$ to consumers, we would still classify this case as full pass-through (instead of overshifting) since the producer price increases by only $\hat{c} - c$.

If $N = 1$, the monopolist faces unit demand and maximizes profits by setting $p = \upsilon$. As in the proof of Proposition 4, the monopoly price is independent of c, which means that pass-through is zero.

By the same reasoning as in the proof of Proposition 4, when $N \to \infty$, $F(p)$ converges to zero, and $E[p]$ converges to p_r. Since $E(p) = p_r$ and $p_r = E(p) + s$ cannot both be true when s is strictly positive, $p_r = \upsilon$ when $N \to \infty$. Thus, when $N \to \infty$, $E(p) \to \upsilon$, which, again, is independent of c. Pass-through to $E(p)$ is thus zero when $N \to \infty$.

Let us now consider the case where $1 < N < \infty$ and $0 < \phi < 1$. As shown in the proof of Proposition 5, $\frac{\partial F(p)}{\partial c} < 0$ and $\frac{\partial \underline{p}}{\partial c} > 0$. In the case where $p_r = \upsilon$, the analysis is simple. In this case, $\frac{\partial p_r}{\partial c} = 0$. Since $F(p)$ decreases when c increases and $\underline{p}$ increases, it must mean that $\int_{\underline{p}}^{p_r} F(p)dp$ decreases in c. Thus, $\frac{\partial E[p]}{\partial c} > 0$, which means that pass-through of a per unit tax is strictly positive.

In the case where $p_r = E[p] + s$, let us first consider that after an increase in c, it remains that $p_r = E[p] + s$. If this is true, then plugging p_r into $E[p]$ and simplifying leads to

$$s = \int_{\underline{p}}^{p_r} F(p)dp.$$

Since s is independent of c and it remains true that $\frac{\partial F(p)}{\partial c} < 0$ and $\frac{\partial \underline{p}}{\partial c} > 0$, this equality can continue to hold after a tax increase only if $\frac{\partial p_r}{\partial c} > 0$. This is true if, and only if, $\frac{\partial E[p]}{\partial c} > 0$.

Finally, let us consider the case where p_r moves from $E[p] + s$ to υ after the tax increase. This can happen only if $E[p] + s < \upsilon$ before the tax increase and $E[p] + s \geq \upsilon$ after the tax increase, which requires $\frac{\partial E[p]}{\partial c} > 0$.

Since $\frac{\partial E[p]}{\partial c} = 0$ for $N = 1$, converges to 0 for $N \to \infty$, and is strictly positive for intermediate N, it therefore must be that there is a non-monotonic relationship between the number of sellers and the pass-through rate of a per-unit tax c when $0 < \phi < 1$. □

C.2.5 Dynamics and Anticipatory Effects

Since we analyse pass-through in a static model, we abstract from how expectations about future prices affect current price setting. Nevertheless, we briefly discuss how expectations may lead to anticipatory effects if extended to a dynamic framework. In particular, anticipatory price increases before a tax increase and a tax decrease are not at odds with the longer-term relationship between price sensitivity, competition, and pass-through that we focus on in this chapter.

First, let us extend our model and consider a dynamic framework in which there are not only informed shoppers and uninformed non-shoppers but also, within both groups, patient consumers (who could buy before or after the tax change) and impatient consumers (who cannot or do not want to wait).

Let us now consider how an anticipatory price increase could occur before a large pre-announced tax decrease. In this case, all patient consumers wait until the next period. Sellers cannot compete for patient consumers before the tax decrease, and so they are left with impatient consumers who do not have the option to wait. Within the group of shoppers and non-shoppers, patient consumers are more price sensitive since they have the option to wait, including in the absence of a tax change. Before a large pre-announced tax decrease, the more price-sensitive consumer groups among shoppers and non-shoppers drop out. Compared to their counterparts in a situation without a tax change, equilibrium prices therefore increase, and quantities decrease.

Finally, let us consider how an anticipatory price increase could occur before a large, pre-announced tax increase. In this case, the option of waiting for another period becomes worse for patient consumers. Therefore, patient consumers become more likely to accept a particular price draw before the tax increase than if there were no pre-announced tax change. For impatient consumers, nothing changes. Patient consumers are willing to accept higher prices than they would without the large, pre-announced tax increase and are more likely to buy in the current period, whereas impatient consumers behave just as they would do without the pre-announced tax increase. Compared to their counterparts in a situation without a tax change, equilibrium prices therefore increase, and quantities also increase.

C.3 Appendix to Section 3.4: Descriptive Evidence

This appendix presents additional descriptive evidence on tax pass-through and provides support for the argument that we should be cautious in comparing pass-through rates across fuel types for the 2022/23 tax changes.

C.3.1 Additional Descriptive Evidence

In Figure 3.3 in the main part of the chapter, we present non-parametric estimates of the pass-through rate for the German 2020/21 tax changes by fuel type. The figure is based on station-level prices, which form the basis for our analysis of the difference in pass-through rates between fuel types. In section 3.6.2, we also investigate the difference in pass-through to the minimum price and the average posted price within local markets. That analysis, in contrast, is based on market-level (instead of station-level) prices. Therefore, we verify that the market-level minimum and average price also evolved similarly in Germany and France prior to the tax changes.

Figures C.10 and C.11 show the results from a non-parametric pass-through estimation similar to that in Figure 3.3, but we now use the average and minimum price in a local non-overlapping market, respectively. The results look very similar to those in Figure 3.3 in the main part of the chapter. Market-level average and minimum prices evolve similarly before the tax changes for the three fuel types, suggesting that differences in pass-through rates after the tax changes are not driven by pre-trends. Again, we see that pass-through rates are highest for diesel and lowest for *E5* in the case of the tax decrease. For the tax increase, pass-through is again highest for diesel, whereas the differences between *E5* and *E10* appear less pronounced.

Importantly, these interpretations hold for both the average posted price (see Figure C.10) and the minimum price (see Figure C.11) within a local market. Therefore, these additional results indicate that the non-parametric evidence shown in Figure 3.3 is robust to using market-level prices, as well as different moments of the price distribution.

C.3.2 The Need for Caution in Considering 2022/23 Tax Changes

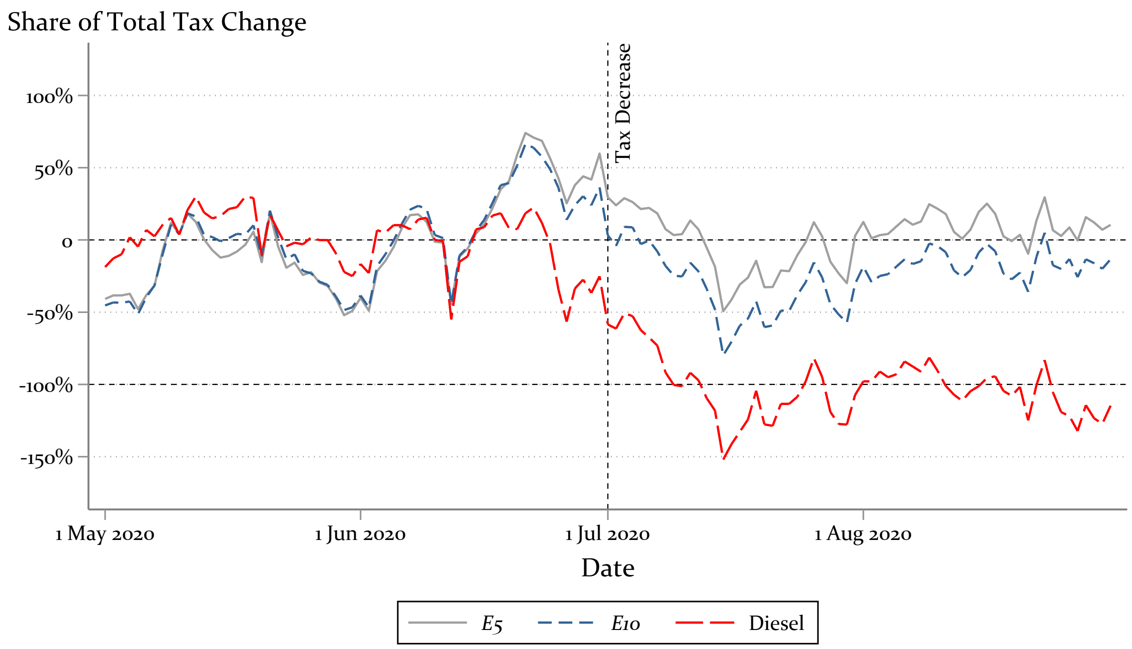
In our main empirical analysis, we focus on a temporary VAT reduction in Germany in the second half of 2020. The 2020/21 tax changes are the only recent policy shifts that allow us to study differences in pass-through across fuel types. Thus, they allow us to test our first theoretical prediction that pass-through increases in the price sensitivity of consumers. In addition, in section 3.6.2, we use tax changes in France in 2022/23 to study differences in pass-through to the average posted price and the minimum price *within* a given fuel type.

In this appendix, we present additional descriptive evidence suggesting that we should be cautious in using the 2022/23 tax changes for other analyses. First, we argue that comparisons across fuel types are problematic for the 2022/23 tax changes. Second, we show that there may have been spillover effects to France from the German tax cut in June 2022.

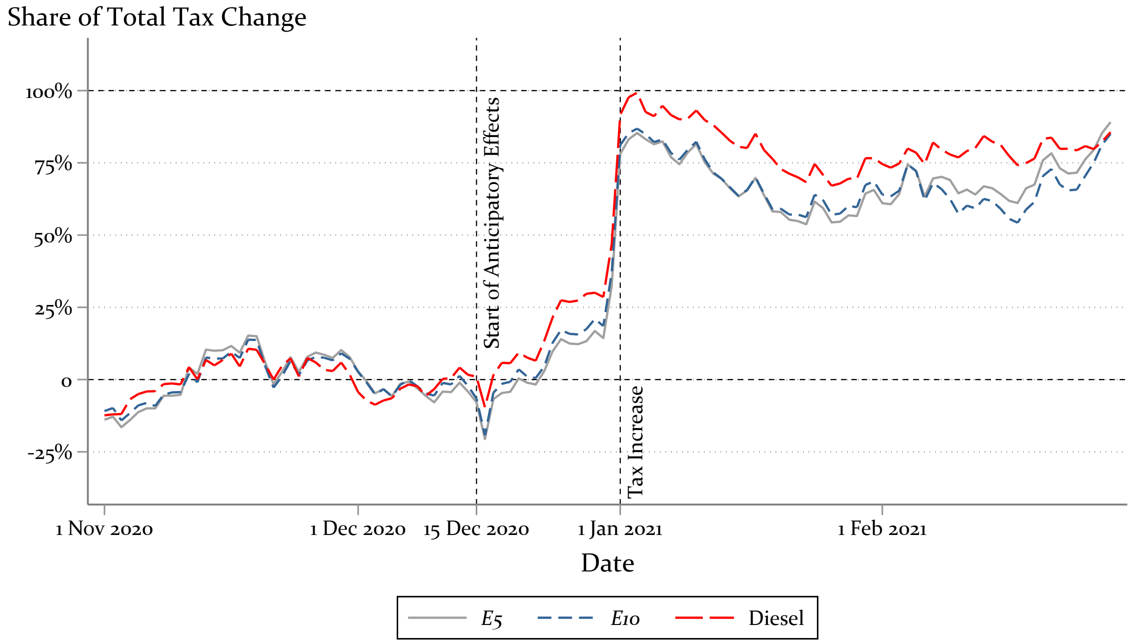
Figure C.12 shows the development of gross fuel prices in Germany and France in February and March of 2022. Panels (A) and (B) present German and French prices, respectively. All prices are normalized to one on 1 February 2022. Two findings emerge from this figure. First, there was a divergence between diesel and gasoline prices in

Figure C.10. Price Change as Share of Total Tax Change (Market-Level Avg. Prices)

(A) Tax Decrease



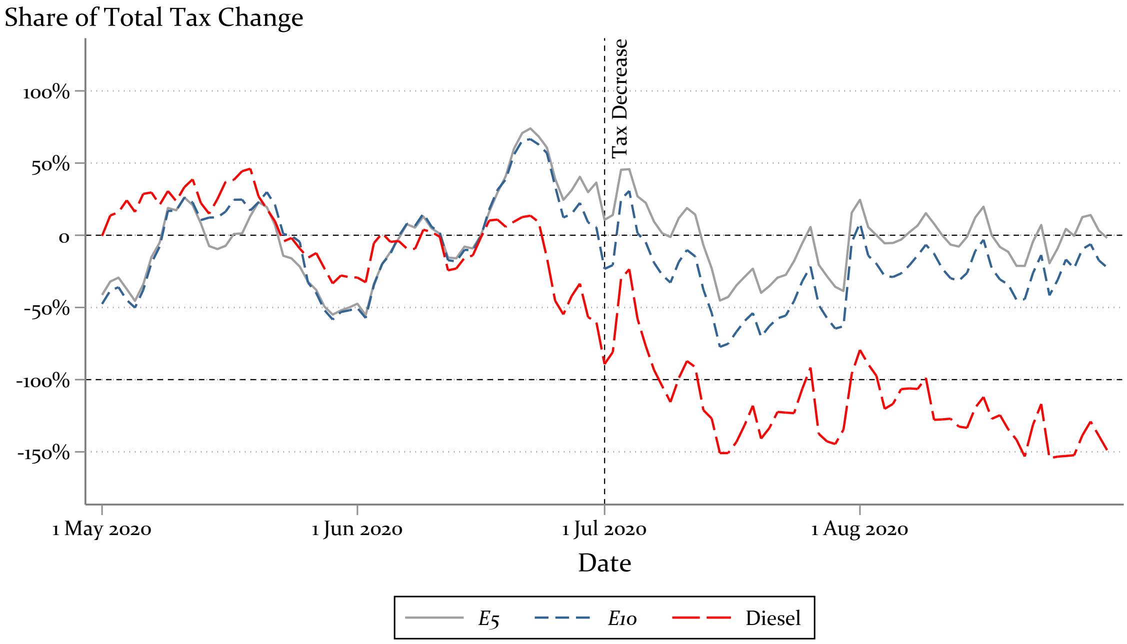
(B) Tax Increase



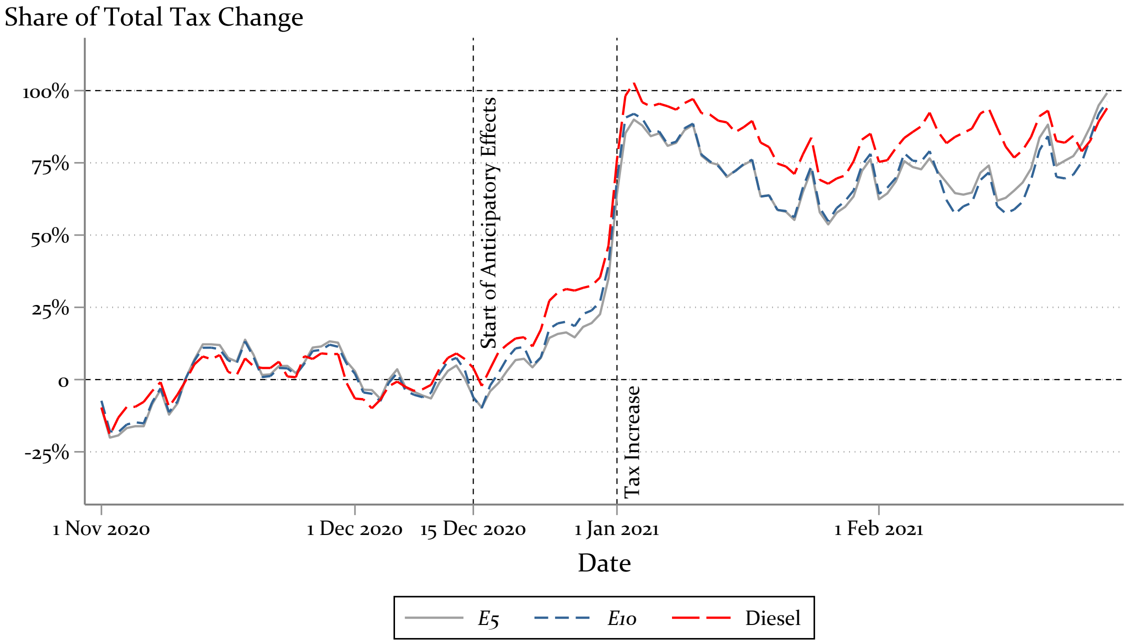
Notes: The figure depicts the price change as a share of the total tax change for the tax decrease in July 2020 and the tax increase in January 2021 in panels (A) and (B), respectively. The solid line shows the non-parametric estimate of the daily average rate of pass-through to prices for *E5*. The short-dashed and long-dashed lines show analogous estimates for *E10* and diesel, respectively. Unlike in Figure 3.3 in the main part of the chapter, a price observation here represents the average posted price in a non-overlapping local market on a given day. To estimate pass-through, we first subtract the average pre-period price in Germany (France) from the daily average price in Germany (France). The pre-period is from 1 May until 30 June 2020 for the tax decrease in panel (A) and from 1 November until 15 December 2020 for the tax increase in panel (B). Next, we compute the difference between de-meaned average prices in Germany and France. Finally, we divide this difference by the difference under full pass-through. For the tax decrease, full pass-through would correspond to a price drop of 2.52%. If we use average absolute prices from 24 June until 30 June (i.e., in the week prior to the tax change), this translates to a price decrease of 3.26 eurocents for *E5*, 3.18 eurocents for *E10*, and 2.75 eurocents for diesel under full pass-through. For the tax increase, full pass-through would correspond to a price increase of 2.59% because of the VAT increase, plus the newly introduced carbon price. If we use absolute prices in the week from 9 December until 15 December 2020 (i.e., in the week prior to the appearance of anticipatory effects), this translates to a price increase of 10.40 eurocents for *E5*, 10.27 eurocents for *E10*, and 10.78 eurocents for diesel under full pass-through. The vertical solid line marks the starting date of the tax change. The horizontal dashed line indicates full pass-through.

Figure C.11. Price Change as Share of Total Tax Change (Market-Level Min. Prices)

(A) Tax Decrease

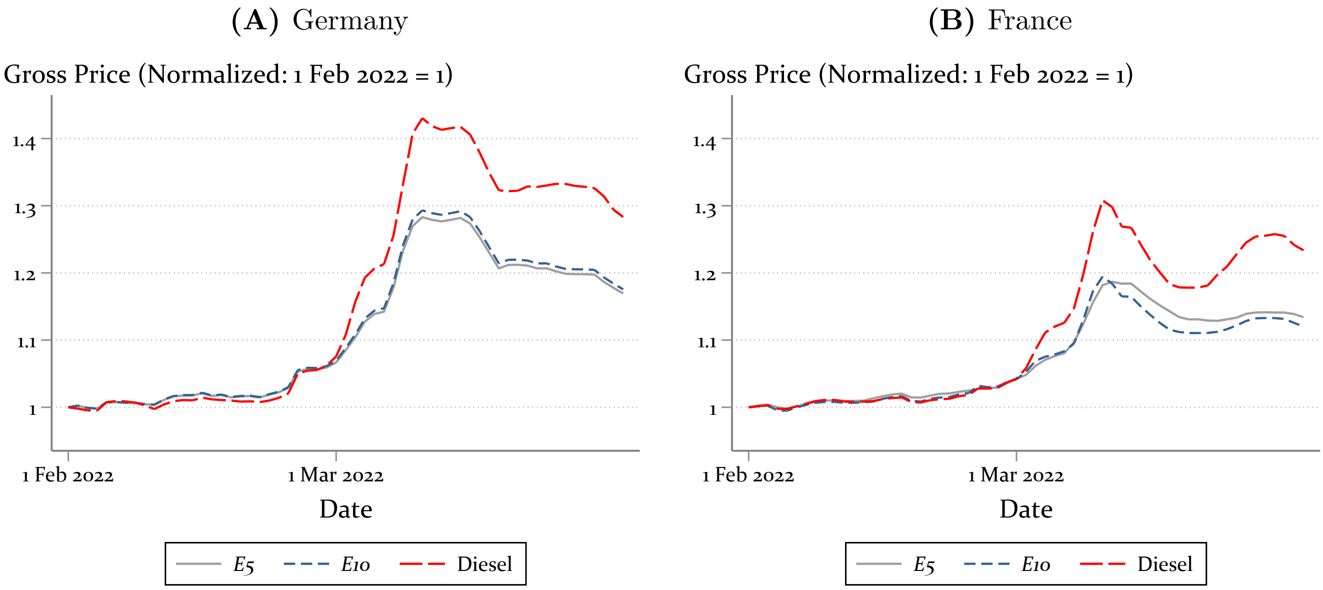


(B) Tax Increase



Notes: The figure depicts the price change as a share of the total tax change for the tax decrease in July 2020 and the tax increase in January 2021 in panels (A) and (B), respectively. The solid line shows the non-parametric estimate of the daily average rate of pass-through to prices for *E5*. The short-dashed and long-dashed lines show analogous estimates for *E10* and diesel, respectively. Unlike in Figure 3.3 in the main part of the chapter, a price observation here represents the minimum price in a non-overlapping local market on a given day. To estimate pass-through, we first subtract the average pre-period price in Germany (France) from the daily average price in Germany (France). The pre-period is from 1 May until 30 June 2020 for the tax decrease in panel (A) and from 1 November until 15 December 2020 for the tax increase in panel (B). Next, we compute the difference between de-meaned average prices in Germany and France. Finally, we divide this difference by the difference under full pass-through. For the tax decrease, full pass-through would correspond to a price drop of 2.52%. If we use average absolute prices from 24 June until 30 June (i.e., in the week prior to the tax change), this translates to a price decrease of 3.13 eurocents for *E5*, 3.04 eurocents for *E10*, and 2.62 eurocents for diesel under full pass-through. For the tax increase, full pass-through would correspond to a price increase of 2.59% because of the VAT increase, plus the newly introduced carbon price. Using absolute prices in the week from 9 December until 15 December 2020 (i.e., in the week prior to the appearance of anticipatory effects), this translates to a price increase of 10.27 eurocents for *E5*, 10.14 eurocents for *E10*, and 10.66 eurocents for diesel under full pass-through. The vertical solid line marks the starting date of the tax change. The horizontal dashed line indicates full pass-through.

Figure C.12. Evolution of Gross Prices in Early 2022



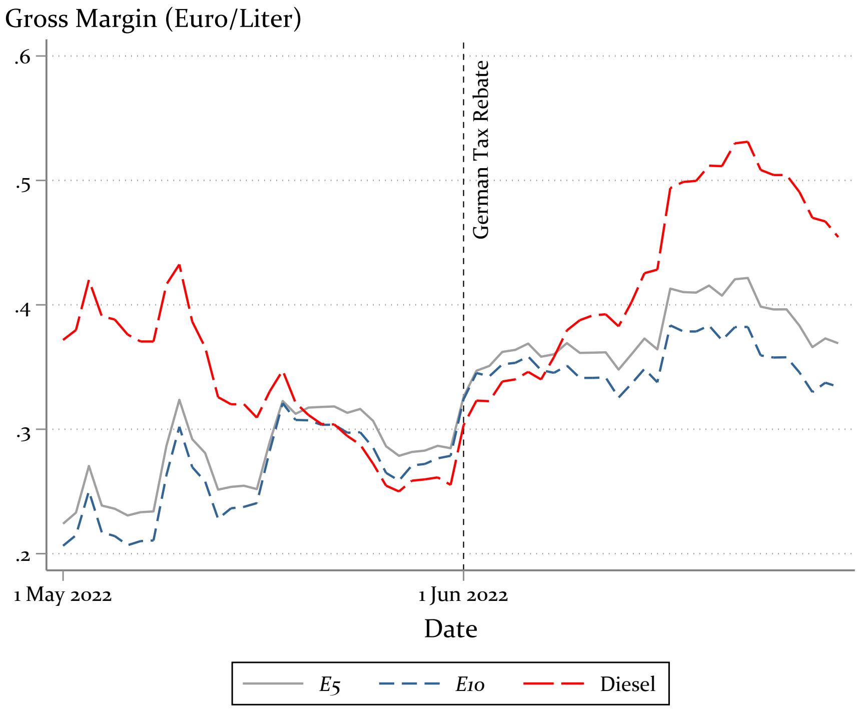
Notes: The figure shows the evolution of daily gross fuel prices in February and March of 2022. Panels (A) and (B) present German and French prices, respectively. All prices are normalized to one at 1 February 2022. The solid line shows prices for *E5*. The short-dashed and long-dashed lines show prices for *E10* and diesel, respectively.

March 2022 in both Germany and France. This is because gasoline and diesel markets were hit differently by Russia's invasion of Ukraine on 24 February 2022. As diesel is a close substitute for heating oil, demand for diesel increased relatively more than that for gasoline. As a consequence, diesel prices also increased disproportionately. Second, Figure C.12 shows that Germany and France were affected differently by the shocks to the global oil market. While diesel (gasoline) prices increased by up to 43% (29%) in Germany relative to those on 1 February 2022, they increased by only up to 31% (19%) in France. That is, fuel prices increased much more in Germany than in France following Russia's invasion of Ukraine.

On 1 April 2022 (i.e., right after the time window shown in Figure C.12), France introduced a fuel tax rebate of 18 eurocents per liter on both diesel and gasoline. Because of the divergence in diesel and gasoline prices prior to the French tax cut, we cannot use this tax change to compare pass-through *across* fuel types.

Similarly, Germany implemented a temporary tax rebate on diesel and gasoline starting on 1 June 2022. As discussed in section 3.2, we do not analyse this tax change because of the intense public scrutiny and concurrent market investigation by the Federal Cartel Office. An additional concern regarding the 2022 tax changes is that they were so large that they may have changed the opportunity cost of selling fuel across countries. That is, there may have been spillover effects from Germany to

Figure C.13. Margins in France Around 1 June 2022



Notes: The figure shows the evolution of daily retail margins at French stations in May and June of 2022. The solid line shows margins for *E5*. The short-dashed and long-dashed lines show margins for *E10* and diesel, respectively. The vertical solid line marks the starting date of the German tax rebate on 1 June 2022.

France (or vice versa), which would violate the stable unit treatment value assumption (SUTVA) underlying our empirical approach.

Figure C.13 presents evidence of such potential spillover effects, showing an increase in French retail margins immediately after the introduction of the German tax cut on 1 June 2022. To compute margins for *E5*, *E10*, and diesel, we subtract taxes, duties, and the estimated input cost of crude oil from the gross price.[10] The figure shows that margins at French stations increased by approximately 5 eurocents on the day when the German fuel rebate went into effect and continued increasing in subsequent weeks.

[10]To compute retail margins, we obtain daily data on the Brent price of crude oil at the port of Rotterdam from the US Energy Information Administration. On average, one barrel (42 gallons) of crude oil is refined into approximately 19 gallons of gasoline, 12 gallons of diesel, or 13 gallons of other products (e.g., jet fuel). See `https://www.eia.gov/energyexplained/oil-and-petroleum-products/refining-crude-oil.php` (last accessed: 19 February 2024). Assuming that, among the other products, only jet fuel is of high value, we split the price of one barrel into the cost of producing gasoline, diesel, and jet fuel to compute the share of the Brent price that corresponds to a particular fuel product. Approximately 54% of the Brent oil price per barrel corresponds to the production of 19 gallons of gasoline, while approximately 34% corresponds to the production of 12 gallons of diesel. Finally, we transform these values into the approximate input cost per liter of gasoline and diesel.

Therefore, estimating pass-through of the Germany tax cut on 1 June 2022 with France as the control group is problematic, and we do not do so in this chapter.

C.4 Appendix to Section 3.6: Empirical Results

In this appendix, we present additional results and several robustness checks of our empirical findings in section 3.6.

C.4.1 Robustness: Additional Controls

In Table C.1, we report results on the effect of the tax change on *E5*, *E10*, and diesel prices when we additionally control for regional mobility for retail and recreational purposes and to workplaces. To this end, we use data from the Google Mobility Report. Overall, the point estimates of the pass-through rates are similar to our main estimation results in Table 3.3 in the main part of the chapter.

The results in columns (1) to (3) show that the average price for *E5* decreased by 0.80% after the tax reduction in July 2020, while average prices for *E10* and diesel decreased by 1.25% and 2.34%, respectively. This implies pass-through rates of 32% for *E5*, 49% for *E10*, and 93% for diesel. The results in columns (4) to (6) show that, following the subsequent tax increase, the average price of *E5* increased by approximately 5.57%, whereas *E10* and diesel prices increased by approximately 6.00% and 8.14%, respectively. We compute a joint pass-through rate of the VAT increase and the carbon emissions price of 67% for *E5*, 70% for *E10*, and 82% for diesel.

Overall, the estimates in Table C.1 are close to our baseline estimates without controls. Therefore, they show that including controls in our regression model does not affect our main results. In particular, pass-through is still significantly higher for diesel than for gasoline, and it is significantly higher for *E10* than for *E5*, consistent with Prediction 1.

C.4.2 Robustness: Balanced Sample

In our baseline approach, we estimate the DiD model in equation (3.1) on an unbalanced panel of fuel stations. That is, we do not observe a price for every station on every day. This could be, for example, because some stations are closed on weekends or holidays or because of permanent station closures or new openings. To ensure that this does not drive our results, we now restrict the sample to fuel stations in Germany and France for which we have a price observation for every day in our sample period. For diesel, for example, this is the case for 83% of fuel stations in Germany and 62% in France

Table C.1. Effect of the Tax Change on Log Prices (with Additional Controls)

	Tax Decrease			Tax Increase		
	E5	*E10*	Diesel	*E5*	*E10*	Diesel
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Change	-0.0080***	-0.0125***	-0.0234***	0.0557***	0.0600***	0.0814***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)
Retail & Recreation	0.0025***	0.0032***	0.0046***	-0.0018***	-0.0043***	-0.0052***
	(0.0008)	(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0005)
Workplaces	0.0130***	0.0118***	-0.0018**	0.0000	-0.0001	-0.0039***
	(0.0009)	(0.0007)	(0.0007)	(0.0008)	(0.0007)	(0.0006)
DE $\times$ Oil Price	0.2282***	0.1931***	0.0470***	0.0790***	0.0202***	0.0755***
	(0.0066)	(0.0055)	(0.0052)	(0.0036)	(0.0036)	(0.0031)
Pass-Through Rate	32%	49%	93%	67%	70%	82%
	[29%, 35%]	[46%, 53%]	[90%, 96%]	[66%, 68%]	[69%, 71%]	[81%, 82%]
Date FE	✓	✓	✓	✓	✓	✓
Station FE	✓	✓	✓	✓	✓	✓
Observations	2,128,241	2,318,268	2,694,252	1,804,493	1,988,459	2,318,185

Notes: The table presents DiD estimates using the model in equation (3.1), but we now additionally control for regional mobility for retail and recreational purposes and to workplaces, using data from the Google Mobility Report. Columns (1) to (3) present estimates of the average treatment effect of the German VAT reduction on 1 July 2020 on *E5*, *E10*, and diesel log prices, respectively. Columns (1) to (3) use data from 1 May to 31 August 2020. Columns (4) to (6) present estimates of the average treatment effect of the German VAT increase and introduction of a carbon price on 1 January 2021 on *E5*, *E10*, and diesel log prices, respectively. Columns (4) to (6) use data from 1 November to 15 December 2020 for the pre-treatment period and from 1 January to 28 February 2021 for the post-treatment period. Standard errors clustered at the market level are shown in parentheses. We also compute the pass-through rates corresponding to the point estimates and report their 95% confidence intervals in brackets. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

for the analysis of the tax reduction, and for 81% of stations in Germany and 72% in France for the analysis of the tax increase.

Columns (1) to (3) show the effect of the tax decrease. We estimate that 90% of the tax decrease is passed on to diesel consumers, while the pass-through rates for *E10* and *E5* are 43% and 14%, respectively. Columns (4) to (6) show the estimates for the tax increase in January 2021. We find pass-through rates of 66% for *E5*, 70% for *E10*, and 84% for diesel. That is, most of the pass-through rates are slightly lower than those estimated from our baseline specification in Table 3.3 in the main part of the chapter, where we use the full unbalanced sample. Importantly, however, our main result regarding the heterogeneous pass-through across fuel types remains unchanged.

Overall, the estimates show that our results are robust to estimating the DiD model on a restricted (balanced) sample of fuel stations for which we have a price observation on every day.

Table C.2. Effect of the Tax Change on Log Prices (Balanced Sample)

	Tax Decrease			Tax Increase		
	E5	*E10*	Diesel	*E5*	*E10*	Diesel
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Change	-0.0035***	-0.0107***	-0.0226***	0.0544***	0.0601***	0.0833***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Pass-Through Rate	14%	43%	90%	66%	70%	84%
	[11%, 17%]	[39%, 46%]	[87%, 93%]	[65%, 66%]	[70%, 71%]	[83%, 84%]
Date FE	✓	✓	✓	✓	✓	✓
Station FE	✓	✓	✓	✓	✓	✓
DE × Oil Price	✓	✓	✓	✓	✓	✓
Observations	1,734,669	1,967,631	2,174,886	1,465,464	1,691,664	1,922,544

Notes: The table presents DiD estimates using the model in equation (3.1). Unlike in Table 3.3 in the main part of the chapter, we now estimate the model in a balanced sample of stations for which we have a price observation for every day. Columns (1) to (3) present estimates of the average treatment effect of the German VAT reduction on 1 July 2020 on *E5*, *E10*, and diesel log prices, respectively. Columns (1) to (3) use data from 1 May to 31 August 2020. Columns (4) to (6) present estimates of the average treatment effect of the German VAT increase and introduction of a carbon price on 1 January 2021 on *E5*, *E10*, and diesel log prices, respectively. Columns (4) to (6) use data from 1 November to 15 December 2020 for the pre-treatment period and from 1 January to 28 February 2021 for the post-treatment period. Standard errors clustered at the market level are shown in parentheses. We also compute the pass-through rates corresponding to the point estimates and report their 95% confidence intervals in brackets. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.4.3 Robustness: Anticipatory Effects

In Table C.3, we examine how our estimates of pass-through rates change when we alter the assumptions on anticipatory effects. In columns (1) to (3), we estimate the pass-through rate of the tax decrease when we drop the second half of June 2020 from the control period. In this case, the gap between pass-through rates among *E5*, *E10* and diesel widens, but the order of estimates in terms of magnitude remains the same. This is not our preferred estimation strategy since we do not think that there is sufficient evidence of anticipatory pass-through of the tax decrease in June 2020. We therefore treat these point estimates of the pass-through rate with caution. Reassuringly, however, our main result regarding the heterogeneity of pass-through with respect to the price sensitivity of consumers does not change.

In columns (4) to (6), we report the estimates of the pass-through rate for the tax increase when we include the second half of December 2020 in the control period. In this case, the point estimate of the pass-through rate decreases from 68% to 63% for *E5*, from 72% to 64% for *E10*, and from 84% to 70% for diesel. This is expected since Figure 3.3 in the main part of the chapter reveals important anticipatory effects of the tax pass-through in the second half of December 2020. Therefore, including this time period in the control period necessarily leads to an underestimation of the pass-through rate. Reassuringly, the differences across fuel types remain similar to those in our main results, although the difference in pass-through for *E5* and *E10* is very small. Although

Table C.3. Effect of the Tax Change on Log Prices (with Anticipatory Effects)

	Tax Decrease			Tax Increase		
	E5	*E10*	Diesel	*E5*	*E10*	Diesel
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Change	0.0006	-0.0058***	-0.0235***	0.0523***	0.0548***	0.0696***
	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0003)	(0.0003)
Pass-Through Rate	-2%	23%	93%	63%	64%	70%
	[-7%, 2%]	[19%, 27%]	[90%, 97%]	[62%, 64%]	[64%, 65%]	[69%, 70%]
Date FE	✓	✓	✓	✓	✓	✓
Station FE	✓	✓	✓	✓	✓	✓
DE × Oil Price	✓	✓	✓	✓	✓	✓
Observations	1,869,158	2,037,695	2,369,148	2,084,101	2,295,092	2,677,702

Notes: The table presents DiD estimates using the model in equation (3.1). Columns (1) to (3) present estimates of the average treatment effect of the German VAT reduction on 1 July 2020 on *E5*, *E10*, and diesel log prices, respectively. Columns (1) to (3) use data from 1 May to 15 June 2020 for the pre-treatment period and from 1 July to 31 August 2020 for the post-treatment period. Columns (4) to (6) present estimates of the average treatment effect of the German VAT increase and introduction of a carbon price on 1 January 2021 on *E5*, *E10*, and diesel log prices, respectively. Columns (4) to (6) use data from 1 November to 28 February 2021. Standard errors clustered at the market level are shown in parentheses. We also compute the pass-through rates corresponding to the point estimates and report their 95% confidence intervals in brackets. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

not accounting for anticipatory effects would slightly modify our estimates, the overall conclusions remain the same. However, the important anticipatory effects that are obvious in the data lead us to believe that excluding the second half of December 2020 from the analysis is preferable.

C.4.4 Robustness: Synthetic Difference-in-Differences Analysis

Next, we repeat our baseline pass-through estimation but employ a synthetic differences-in-differences (SDiD) approach instead of a standard DiD. SDiD is a variation of DiD that aims to match pre-treatment trends between the treatment and control groups using weights. In this sense, SDiD is similar to synthetic control methods (Arkhangelsky et al., 2021).

With the SDiD approach, we estimate pass-through using a two-step procedure. First, we calculate unit and time weights that minimize the difference in pre-treatment trends between treated and control groups and the difference in outcomes between pre- and post-treatment periods for the control group. In the second step, we estimate a DiD model similar to that in equation (3.1) using the weights from the first step. We use clustered bootstrapping with 300 replications and clustering at the station level to estimate standard errors.

Formally, to estimate the average pass-through rate of the tax changes to fuel prices, SDiD solves the following minimization problem:

$$(\hat{\beta}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\pi}) = \underset{\beta,\mu,\alpha,\pi}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{i,t} - \mu - \alpha_i - \pi_t - \text{Tax}_{i,t}\beta\right)^2 \hat{w}_i^{sdid} \hat{\lambda}_t^{sdid} \right\}, \quad \text{(C.16)}$$

where $\hat{\beta}^{sdid}$ is the estimated effect of the policy change and $\hat{w}_i^{sdid}$ and $\hat{\lambda}_t^{sdid}$ are the SDiD unit and time weights, respectively. As in our baseline DiD model in equation (3.1), $Y_{i,t}$ is the logarithm of the weighted average price of gasoline or diesel at fuel station i at date t, and $\text{Tax}_{i,t}$ is a dummy variable that equals one for stations affected by the tax change at date t. The variables α_i and π_t again correspond to fuel station and date fixed effects, respectively. Using SDiD requires a balanced panel, but we have shown above in Appendix C.4.2 that this sample restriction by itself does not affect our baseline estimates.

Using the SDiD approach, we first re-estimate the baseline pass-through rates of the German tax changes in 2020/21. In a second step, we also assess the robustness of our results regarding the relationship between the pass-through rate and the number of sellers in the market. We cannot use SDiD to estimate the difference in the pass-through rates between the minimum price and the average price, as SDiD does not allow a triple interaction term.

C.4.4.1 Baseline Pass-Through Estimation

Table C.4 shows the results when estimating the regression model in equation (C.16) for the analysis of the 2020/21 tax changes in Germany. Columns (1) to (3) show the effect of the tax decrease. For *E5*, the pass-through rate is 23%, while approximately 45% and 83% of the tax decrease is passed on to consumers who refuel with *E10* and diesel, respectively. Therefore, the ranking of pass-through rates with respect to fuel types is robust to using an SDiD strategy.

Columns (4) to (6) show that the tax increase raised prices for all fuel products. We find a joint pass-through rate of the VAT increase and the introduction of carbon price of 75% for *E5* and *E10* and 86% for diesel. That is, for the tax increase, pass-through is again significantly higher for diesel than for gasoline, whereas the pass-through rates for *E5* and *E10* are statistically indistinguishable.

Overall, the ranking of pass-through rates with respect to fuel types and their magnitude remain largely robust to using SDiD instead of a simple DiD approach.

Table C.4. Effect of the Tax Change on Log Prices (SDiD)

	Tax Decrease			Tax Increase		
	E5	*E10*	Diesel	*E5*	*E10*	Diesel
	(1)	(2)	(3)	(4)	(5)	(6)
Tax Change	-0.0058***	-0.0115***	-0.0209***	0.0625***	0.0636***	0.0859***
	(0.0007)	(0.0004)	(0.0005)	(0.0005)	(0.0004)	(0.0004)
Pass-Through Rate	23%	45%	83%	75%	75%	86%
	[17%, 28%]	[42%, 49%]	[79%, 86%]	[74%, 77%]	[74%, 75%]	[85%, 87%]
Date FE	✓	✓	✓	✓	✓	✓
Station FE	✓	✓	✓	✓	✓	✓
Observations	1,734,669	1,967,631	2,174,886	1,465,464	1,691,664	1,922,544

Notes: The table presents SDiD estimates using the model in equation (C.16). Columns (1) to (3) present estimates of the average treatment effect of the German VAT reduction on 1 July 2020 on *E5*, *E10*, and diesel log prices, respectively. Columns (1) to (3) use data from 1 May to 31 August 2020. Columns (4) to (6) present estimates of the average treatment effect of the German VAT increase and introduction of a carbon price on 1 January 2021 on *E5*, *E10*, and diesel log prices, respectively. Columns (4) to (6) use data from 1 November to 15 December 2020 for the pre-treatment period and from 1 January to 28 February 2021 for the post-treatment period. Standard errors obtained via clustered bootstrap with 300 replications are shown in parentheses. We also compute the pass-through rates corresponding to the point estimates and report their 95% confidence intervals in brackets. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.4.4.2 Number of Sellers and Tax Pass-Through

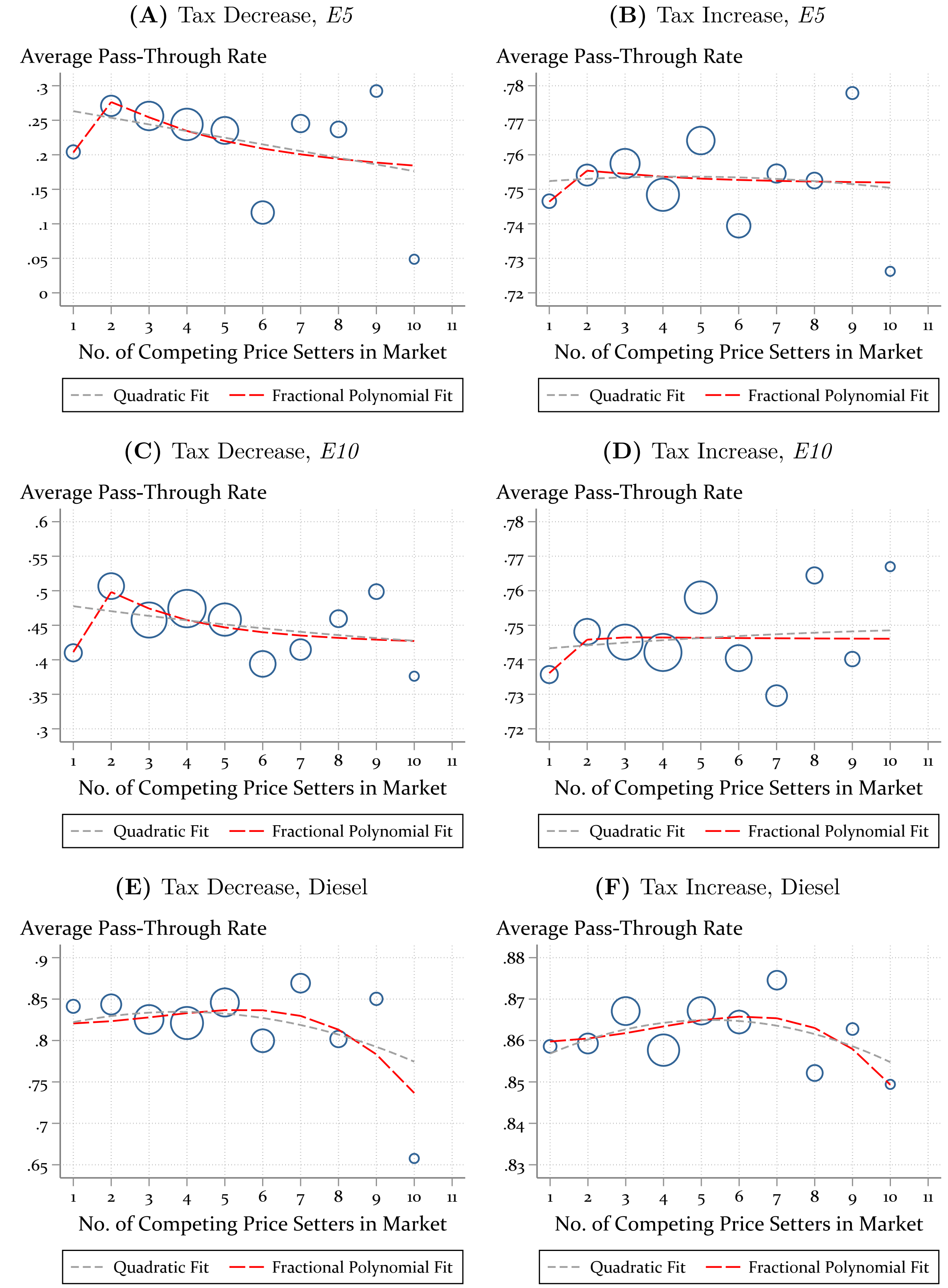
Figure C.14 shows the relationship between the pass-through rate and the number of competitors of a focal station when we estimate the station-level pass-through rates with SDiD (i.e., with additional weights) and a restricted balanced sample of stations. We estimate the time and unit weights only once and then use them to estimate each station-specific treatment effect.

The results look very similar to our main results in Figure 3.5 in the main part of the chapter. Consistent with Prediction 3, we still find a non-monotonic relationship between the number of competing price setters in a local market and the average pass-through rate.

For *E5* and *E10*, pass-through is again relatively low for local monopolists for both the tax decrease in summer 2020 and the tax increase in winter 2020/21. With at least two competing price setters in a local market, the average pass-through tends to decrease in the number of sellers in the case of the tax decrease and appears constant in the case of the tax increase. As before, for diesel, the relationship between the number of sellers and pass-through has an inverted-U shape with a peak at a higher number of sellers than in the case of *E5* and *E10*.

In summary, our analysis shows that the non-monotonic relationship between the number of sellers and pass-through is robust to using an SDiD strategy.

Figure C.14. Average Pass-Through by Number of Competitors (SDiD)



Notes: The figure shows how the rate of pass-through to the average price varies with the number of competing price setters in a market. Unlike in Figure 3.5 in the main part of the chapter, pass-through rates here are estimated with an SDiD strategy and a restricted balanced sample. Panels (A), (C), and (E) depict the pass-through rates for the German VAT decrease on 1 July 2020 for *E5*, *E10*, and diesel, respectively. Panels (B), (D), and (F) depict the pass-through rates for the German VAT increase and introduction of a carbon price on 1 January 2021 for *E5*, *E10*, and diesel, respectively. In every panel, each circle plots the average pass-through rate for a group of stations with a particular number of competing price setters within a non-overlapping local market. The size of a circle is proportional to the total number of stations with a given number of competitors. The long-dashed line shows a fractional polynomial fit. The short-dashed line shows a quadratic fit. The number of competitor stations is trimmed at the 97.5th percentile.

C.4.5 Robustness: Price at 10th Percentile

In Figure 3.4 in the main part of the chapter, we show that pass-through to the minimum price in a local market is generally higher than pass-through to the average posted price. One potential concern regarding the minimum price is that it may reflect outlier prices that are valid for only a short period of time. In that case, the minimum price that we identify would not actually be relevant for consumers.

We address this point in Figure C.15, where we re-estimate the difference in the pass-through rate between the minimum and the average price. This time, however, instead of the minimum price in a local market, we use the price at the 10th percentile of the distribution of all hourly prices across all stations in that market on a given day.

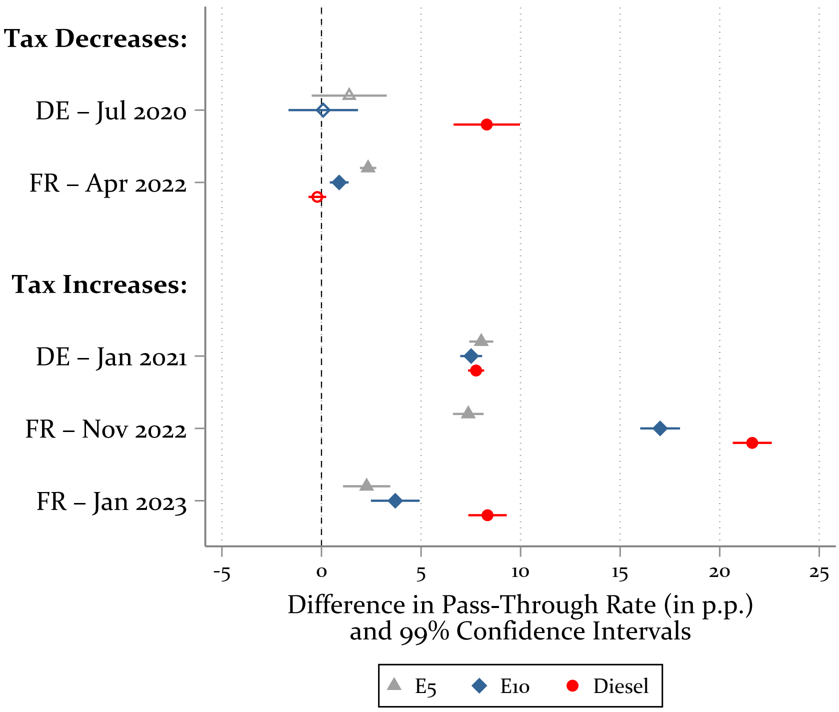
The results look very similar to those from our main specification in Figure 3.4. They indicate that our findings on the difference in pass-through between the minimum and the average price are not driven by outlier prices. If anything, the estimated difference between pass-through to the minimum price and the average price becomes slightly more pronounced for the tax decreases. Therefore, Figure C.15 provides even stronger support for Prediction 2 that pass-through to the price paid by well-informed consumers is higher than that to the price paid by uninformed consumers.

C.4.6 Robustness: Overlapping Markets

In the main part of the chapter, we group fuel stations into non-overlapping markets using a hierarchical clustering algorithm based on driving time. The details of the clustering method are explained in Appendix C.1.2. The clustering approach has been used in several previous studies (e.g., Carranza et al., 2015; Lemus and Luco, 2021; Assad et al., 2020). An alternative way of defining local markets is to center one market around each station and include all competitors within a predefined radius around the centroid station. The drawback of this approach is that it leads to overlapping markets, where some stations are assigned to multiple markets. This station-centered approach has also been used in the literature as well as by competition authorities, and it is easier to implement. Luco (2019) argues that markets centered on each station are more suitable when the researcher does not observe the universe of stations. This argument does not apply to our setting, since we do observe the universe of stations both in Germany and France. Therefore, we believe that the clustering approach with non-overlapping markets is preferable in our case.

Nevertheless, we analyse whether our results are robust to using a market definition with overlapping markets centered on each station. We implement this approach by including all stations within a driving distance of 5km as belonging to the same market

Figure C.15. Pass-Through to Market-Level Min. (10th Percentile) vs. Avg. Prices



Notes: The figure shows the difference in the market-level pass-through rate (in percentage points) between the minimum price and the average posted price. The average posted price is the average daily price within a non-overlapping market by weighting the price at every full hour of the day between 6am and 10pm equally. In contrast to Figure 3.4 in the main part of the chapter, we here replace the minimum price with the price at the 10th percentile of the distribution of all hourly prices within a non-overlapping market on a given day. The figure depicts the pass-through rates implied by the DiD estimate β_2 in equation (3.2) along with 99% confidence intervals, based on standard errors clustered at the market level. Regressions are estimated separately for each tax change and fuel type. For most tax changes, we use data for the two months before and after every tax change. Exceptions include the German tax increase on 1 January 2021, where we exclude the second half of December to account for anticipatory effects. For the tax increase in France on 16 November 2022, we use only the period until 31 December 2022 as the post-treatment period. Similarly, for the French tax increase on 1 January 2023, we use the period from 16 November until 31 December 2022 as the pre-treatment period. Solid (hollow) symbols indicate point estimates that are statistically significant (insignificant) at the 1% level.

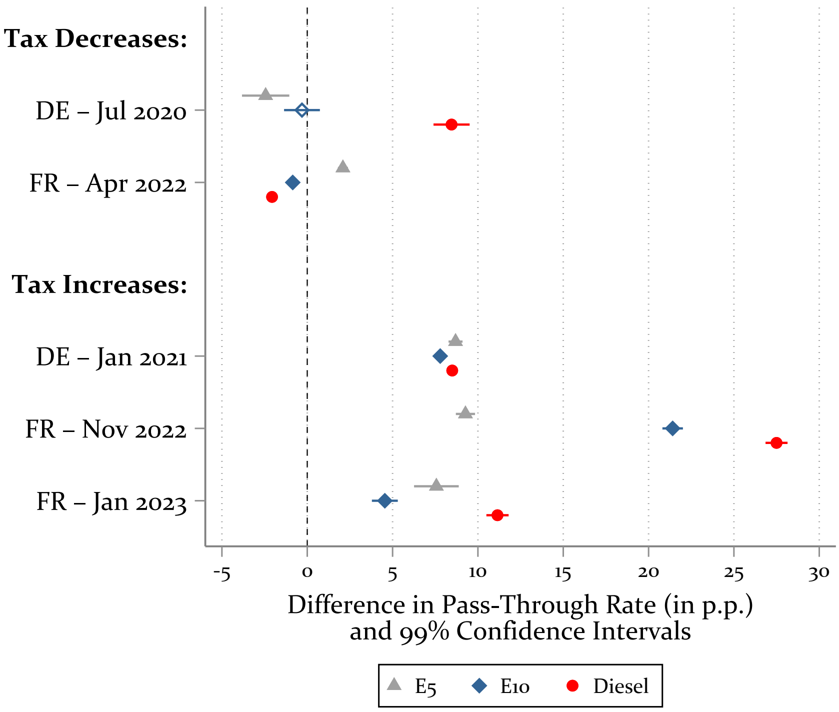
as the centroid station. Hence, the resulting number of local markets is equivalent to the number of stations, as shown in Table 3.1 in the main part of the chapter.

C.4.6.1 Pass-Through to the Average and Minimum Price

In Figure C.16, we re-estimate the difference in pass-through between the minimum price and the average posted price for all the tax changes in our sample period. Unlike in Figure 3.4 in the main part of the chapter, we now use overlapping markets.

When using overlapping markets, we still observe that the rate of pass-through to the minimum price is statistically significantly higher than that to the average posted price in 11 out of the 15 cases depicted in Figure C.16. For the tax increases,

Figure C.16. Pass-Through to Market-Level Min. vs. Avg. Prices (Overlapping Markets)



Notes: The figure shows the difference in the market-level pass-through rate (in percentage points) between the minimum price and the average posted price. The average posted price is the average daily price within a local market by weighting the price at every full hour of the day between 6am and 10pm equally. The minimum price is the minimum price within a local market at any point of time during the day. Unlike Figure 3.4 in the main part of the chapter, this figure uses overlapping markets centered on each station and including all competing price setters within a driving distance of 5km around the centroid station. The figure depicts the pass-through rates implied by the DiD estimate β_2 in equation (3.2) along with 99% confidence intervals, based on standard errors clustered at the market level. Regressions are estimates separately for each tax change and fuel type. For most tax changes, we use data for the two months before and after every tax change. Exceptions include the German tax increase on 1 January 2021, where we exclude the second half of December to account for anticipatory effects. For the tax increase in France on 16 November 2022, we only use the period until 31 December 2022 as post-treatment period. Similarly, for the French tax increase on 1 January 2023, we use the period from 16 November until 31 December 2022 as pre-treatment period. Solid (hollow) symbols indicate point estimates that are statistically significant (insignificant) at the 1% level.

pass-through rates to the minimum price remain significantly higher than those to the average posted price in all cases. For the tax decrease, the picture is more mixed. Pass-through to the minimum price is higher than pass-through to the average posted price in two cases, it is lower in three cases, and the difference is statistically indistinguishable from zero in the remaining case.

It is important to keep in mind that the results in Figure C.16 may differ from those in Figure 3.4 for two reasons. The first is that different stations are grouped together into local markets. The second reason is the overlapping nature of the alternative local markets, which can be problematic for the analysis. One potential issue is that, when we compute average posted prices in overlapping markets, the prices of stations in areas with a higher density of fuel stations receive a higher weight because they belong to multiple markets. Similarly, when we compute the minimum price in several markets with nearby centroid stations, it is possible that a single price notification by a single station represents the minimum price in all these markets. Note that independent stations typically set lower prices and, therefore, are more likely to set the minimum price in a market. They also often have less sophisticated pricing strategies or algorithms. As a consequence, the estimates in the robustness check in Figure C.16 may be overly sensitive to the prices set by individual stations and especially lower-price independent stations. Despite these potential concerns, the robustness check further supports Prediction 2 that the pass-through to the expected minimum price is higher than pass-through to the expected price.

C.4.6.2 Number of Sellers and Tax Pass-Through

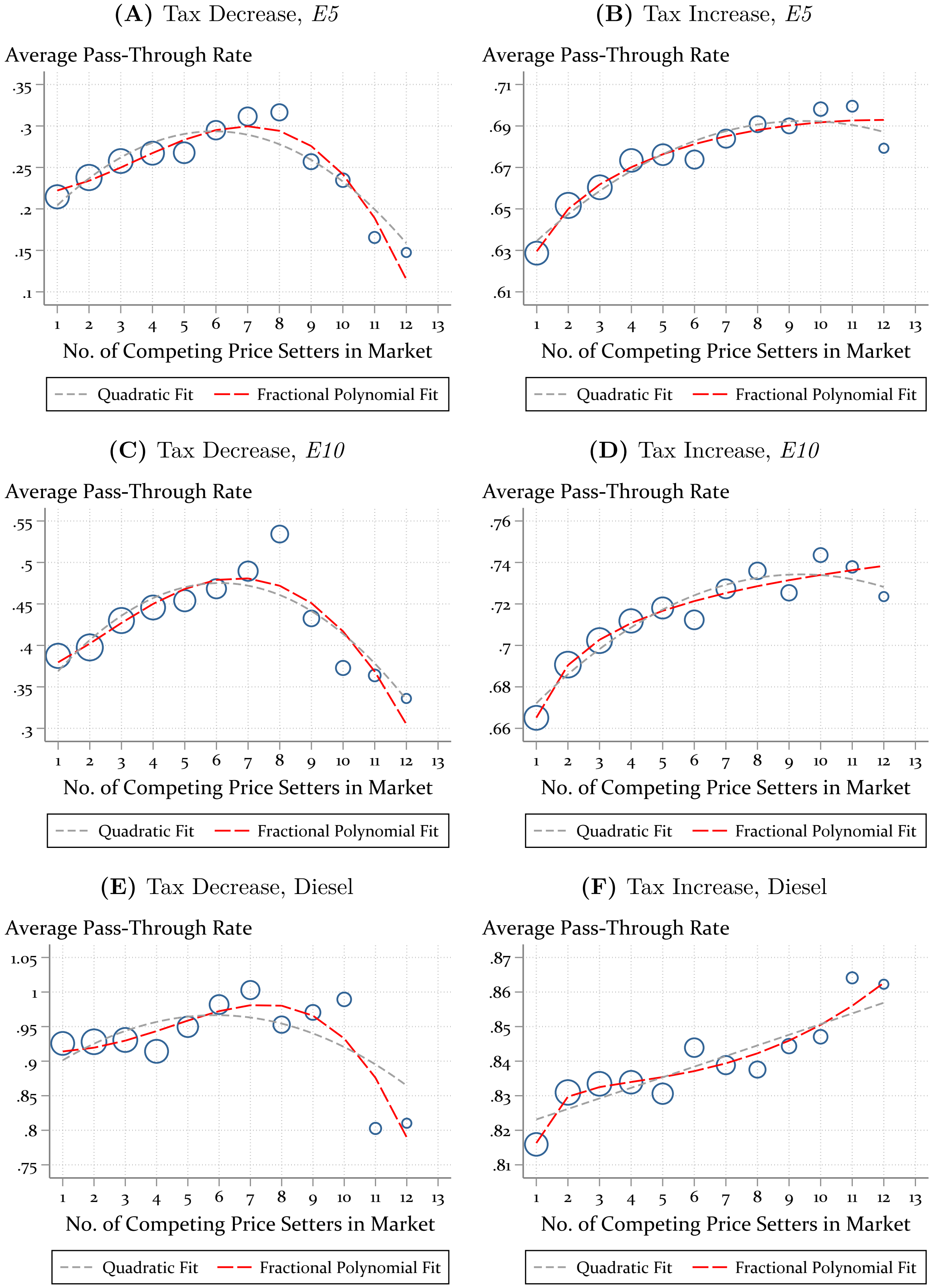
Next, we assess how the pass-through rate varies with the number of sellers when we use overlapping markets. Figure C.17 shows the relationship between the pass-through rate and the number of competing price setters in local markets for the 2020/21 German tax changes and the three fuel types.

Panels (A), (C), and (E) depict the pass-through rates for the tax decrease in summer 2020 for *E5*, *E10*, and diesel, respectively. For all three fuel types, there is a very pronounced inverse-U-shaped relationship between the average pass-through rate and the number of sellers. The highest pass-through rates are observed for stations in local markets with seven to eight competing price setters.

In panels (B), (D), and (F) of Figure C.17, we repeat this analysis for the tax increase in winter 2020/21. For *E5* and *E10*, pass-through is increasing in the number of sellers up to approximately 10 or 11 sellers but then tends to slightly decrease again. For diesel, the pass-through rate seems to be continuously increasing in the number of sellers. However, this result, which stands in contrast to our theoretical prediction, is the exception in Figure C.17.

Overall, Figure C.17 confirms our finding that pass-through to the average price is not monotonically increasing in the number of sellers. Therefore, our results are robust to using an alternative market definition that considers overlapping markets centered on each station and including all competing price setters within a driving distance of 5km around the centroid station.

Figure C.17. Average Pass-Through by Number of Competitors (Overlapping Markets)



Notes: The figure shows how the rate of pass-through to the average price varies with the number of competing price setters in a market. Panels (A), (C), and (E) depict the pass-through rates for the German VAT decrease on 1 July 2020 for *E5*, *E10*, and diesel, respectively. Panels (B), (D), and (F) depict the pass-through rates for the German VAT increase and introduction of a carbon price on 1 January 2021 for *E5*, *E10*, and diesel, respectively. In every panel, each circle plots the average pass-through rate for a group of stations with a particular number of competing price setters. Unlike Figure 3.5 in the main part of the chapter, this figure uses overlapping markets centered on each station and including all competing price setters within a driving distance of 5km around the centroid station. The size of a circle is proportional to the total number of stations with a given number of competitors. The long-dashed line shows a fractional polynomial fit. The short-dashed line shows a quadratic fit. The number of competitor stations is trimmed at the 97.5th percentile.

Table C.5. U-Test of Non-Monotonicity

	Tax Decrease			Tax Increase		
	E5	*E10*	Diesel	*E5*	*E10*	Diesel
	(1)	(2)	(3)	(4)	(5)	(6)
No. of Competing Price Setters	0.0178	0.0136	0.0221**	0.0043	0.0017	0.0064***
	(0.0124)	(0.0137)	(0.0110)	(0.0030)	(0.0032)	(0.0023)
(No. of Competing Price Setters)^2	-0.0023*	-0.0013	-0.0020*	-0.0005**	-0.0002	-0.0005**
	(0.0012)	(0.0014)	(0.0010)	(0.0003)	(0.0003)	(0.0002)
Constant	0.2283***	0.4030***	0.8779***	0.6632***	0.7070***	0.8186***
	(0.0294)	(0.0312)	(0.0262)	(0.0074)	(0.0079)	(0.0058)
Inverse-U Test						
t-value	1.3043	0.8777	1.6463	1.3138	0.4782	1.7083
p-value	0.0961	0.1900	0.0499	0.0945	0.3163	0.0438
Extreme Point	3.8900	5.1693	5.5786	3.9854	4.2498	6.3627
Lower Bound	1	1	1	1	1	1
Slope at Lower Bound	0.0132	0.0110	0.0182	0.0032	0.0013	0.0054
t-value	1.3043	0.9926	2.0243	1.3138	0.4782	2.7812
Upper Bound	10	10	10	10	10	10
Slope at Upper Bound	-0.0279	-0.0127	-0.0175	-0.0065	-0.0022	-0.0036
t-value	-2.3261	-0.8777	-1.6463	-2.4840	-0.7975	-1.7083
Observations	13,998	13,394	14,250	14,033	13,426	14,296

Notes: The table shows the estimates from regressing the station-level pass-through rate on the number of competing price setters in a local non-overlapping market, its square, and a constant. Columns (1) to (3) present estimates for the German VAT reduction on 1 July 2020, while columns (4) to (6) present estimates for the German VAT increase and introduction of a carbon price on 1 January 2021. The bottom part of the table shows the application of the U-test of Lind and Mehlum (2010). Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.4.7 Formal Test of Non-Monotonicity

The results in Figure 3.5 in the main part of the chapter indicate that pass-through to the average price is not monotonically increasing in the number of sellers. In this section, we formally test for non-monotonicity by applying the U-test of Lind and Mehlum (2010). This allows us to test the null hypothesis of a monotone or U-shaped relationship against the alternative hypothesis of an inverted-U shape.

To this end, we first regress the station-level pass-through rate on the number of competing price setters and its square. As the number of competing price setters in a local market is always positive, a necessary condition for a hump-shaped relationship is that the coefficient on the linear term is positive, that the coefficient on the quadratic term is negative, and that the extreme point lies within the data range. Lind and Mehlum (2010) argue, however, that these conditions are not sufficient to guarantee an inverse-U-shaped relationship. Therefore, they propose a formal statistical test for a U (or inverse-U) shape, taking into account the slope at the lower and upper bounds of the data range.

Table C.6. U-Test of Non-Monotonicity (Overlapping Markets)

	Tax Decrease			Tax Increase		
	E5	*E10*	Diesel	*E5*	*E10*	Diesel
	(1)	(2)	(3)	(4)	(5)	(6)
No. of Competing Price Setters	0.0435***	0.0496***	0.0322***	0.0153***	0.0166***	0.0030**
	(0.0080)	(0.0093)	(0.0068)	(0.0019)	(0.0021)	(0.0015)
(No. of Competing Price Setters)^2	-0.0037***	-0.0041***	-0.0027***	-0.0008***	-0.0009***	0.0000
	(0.0007)	(0.0008)	(0.0006)	(0.0002)	(0.0002)	(0.0001)
Constant	0.1644***	0.3234***	0.8718***	0.6200***	0.6564***	0.8201***
	(0.0193)	(0.0213)	(0.0165)	(0.0049)	(0.0053)	(0.0038)
Inverse-U Test						
t-value	5.1481	4.1823	4.6571	2.1582	2.2468	n/a
p-value	0.0000	0.0000	0.0000	0.0155	0.0123	n/a
Extreme Point	5.9354	6.1154	5.8892	9.4466	9.3822	-5.5e+02
Lower Bound	1	1	1	1	1	1
Slope at Lower Bound	0.0362	0.0415	0.0267	0.0137	0.0148	0.0030
t-value	5.3941	5.3960	4.6571	8.4753	8.4316	2.3983
Upper Bound	12	12	12	12	12	12
Slope at Upper Bound	-0.0445	-0.0478	-0.0334	-0.0041	-0.0046	0.0031
t-value	-5.1481	-4.1823	-4.6636	-2.1582	-2.2468	1.9540
Observations	14,082	13,515	14,350	14,112	13,541	14,387

Notes: The table shows the estimates from regressing the station-level pass-through rate on the number of competing price setters, its square, and a constant. Unlike in Table C.5, this table uses overlapping markets centered on each station and including all competing price setters within a driving distance of 5km around the centroid station. Columns (1) to (3) present estimates for the German VAT reduction on 1 July 2020, while columns (4) to (6) present estimates for the German VAT increase and introduction of a carbon price on 1 January 2021. The bottom part of the table shows the application of the U-test of Lind and Mehlum (2010). Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

We present our estimation results in Table C.5, following the exhibition in Pennerstorfer et al. (2020). The table shows that our estimated coefficients satisfy all three necessary conditions for each fuel type and tax changes, although not all the estimates are statistically significantly different from zero. The p-value on the inverse-U test indicates that we reject the null hypothesis of a monotonic or U-shaped relationship at the 10% level for both tax changes for diesel and *E5*. For *E10*, the pass-through rate is also increasing (decreasing) in the number of competing price setters for local monopolies (large markets), but the slopes at the bounds are imprecisely estimated. Therefore, we cannot formally reject a monotonic relationship between the number of competitors and the pass-through rate, but the estimates for *E10* still speak in favour of a hump-shaped relationship.

Overall, the U-test of Lind and Mehlum (2010) further supports Prediction 3 that the relationship between the number of sellers and pass-through to the expected price is non-monotonic.

In Table C.6, we repeat the U-test with a different definition of local markets. As in Appendix C.4.6, we use overlapping markets centered on each station and including all competing price setters within a driving distance of 5km around the centroid station. As the table shows, the estimated coefficients satisfy the necessary conditions for an inverse-U shape in most cases. The only exception is diesel in the case of the tax increase, where the estimates suggest an increasing relationship between the number of competitors and average pass-through. In the five other cases, the *p*-value on the inverse-U test allows us to reject the null hypothesis of a monotonic or U-shaped relationship, consistent with our theoretical prediction. Therefore, Table C.6 again shows that our conclusion on how average pass-through varies with the number of competitors is robust to using an alternative market definition.

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505.

——— (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510.

Acemoglu, D. and U. Akcigit (2012): "Intellectual Property Rights Policy, Competition and Innovation," *Journal of the European Economic Association*, 10, 1–42.

Adachi, T. and M. Fabinger (2022): "Pass-Through, Welfare, and Incidence Under Imperfect Competition," *Journal of Public Economics*, 211, 1045–1089.

Aghion, P., A. Dechezleprêtre, D. Hémous, R. Martin, and J. Van Reenen (2016): "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry," *Journal of Political Economy*, 124, 1–51.

Ahmadpoor, M. and B. F. Jones (2017): "The Dual Frontier: Patented Inventions and Prior Scientific Advance," *Science*, 357, 583–587.

Akcigit, U. and S. T. Ates (2023): "What Happened to U.S. Business Dynamism?" *Journal of Political Economy*, 131, 2059–2124.

Alcácer, J. and M. Gittelman (2006): "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations," *Review of Economics and Statistics*, 88, 774–779.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021): "Synthetic Difference-in-Differences," *American Economic Review*, 111, 4088–4118.

Arora, A. and A. Fosfuri (2003): "Licensing the Market for Technology," *Journal of Economic Behavior and Organization*, 52, 277–295.

Assad, S., R. Clark, D. Ershov, and L. Xu (2020): "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market," *CESifo Working Paper No. 8521.*

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 135, 645–709.

Bai, J., P. J. Barwick, S. Cao, and S. Li (2020): "Quid Pro Quo, Knowledge Spillover, and Industrial Quality Upgrades: Evidence from the Chinese Auto Industry," *NBER Working Paper No. 27644.*

Barkai, S. (2020): "Declining Labor and Capital Shares," *Journal of Finance*, 75, 2421–2463.

Basu, S. (2019): "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence," *Journal of Economic Perspectives*, 33, 3–22.

Baye, M. R., J. Morgan, and P. Scholten (2006): "Information, Search, and Price Dispersion," in *Handbook on Economics and Information Systems*, ed. by T. Hendershott, Elsevier, chap. 6, 323–376.

Bellemare, M. F. and C. J. Wichman (2020): "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 82, 50–61.

Benedek, D., R. A. De Mooij, M. Keen, and P. Wingender (2020): "Varieties of VAT Pass Through," *International Tax and Public Finance*, 27, 890–930.

Benzarti, Y., D. Carloni, J. Harju, and T. Kosonen (2020): "What Goes Up May Not Come Down: Asymmetric Incidence of Value-Added Taxes," *Journal of Political Economy*, 128, 4438–4474.

Bergeaud, A. and C. Verluise (2022): "A New Dataset to Study a Century of Innovation in Europe and in the US," *CEP Discussion Paper No. 1850.*

Berry, S. T., M. Gaynor, and F. Scott Morton (2019): "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization," *Journal of Economic Perspectives*, 33, 44–68.

Bessen, J. and E. Maskin (2009): "Sequential Innovation, Patents, and Imitation," *RAND Journal of Economics*, 40, 611–635.

Bilby, K. (1986): *The General: David Sarnoff and the Rise of the Communications Industry*, Harper & Row.

Borenstein, S., A. C. Cameron, and R. Gilbert (1997): "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?" *Quarterly Journal of Economics*, 112, 305–339.

Bresnahan, T. F. (1985a): "Post-Entry Competition in the Plain Paper Copier Market," *American Economic Review*, 75, 15–19.

——— (1985b): "The Transition to Competition in the Plain Paper Copier Market," Federal Trade Commission Report.

——— (1989): "Empirical Studies of Industries with Market Power," in *Handbook of Industrial Organization*, ed. by R. Schmalensee and R. Willig, Elsevier, vol. 2, chap. 17, 1011–1057.

Bulow, J. I. and P. Pfleiderer (1983): "A Note on the Effect of Cost Changes on Prices," *Journal of Political Economy*, 91, 182–185.

Burdett, K. and K. L. Judd (1983): "Equilibrium Price Dispersion," *Econometrica*, 51, 955–969.

Busse, M., J. Silva-Risso, and F. Zettelmeyer (2006): "$1,000 Cash Back: The Pass-Through of Auto Manufacturer Promotions," *American Economic Review*, 96, 1253–1270.

Büttner, T. and B. Madzharova (2021): "Unit Sales and Price Effects of Preannounced Consumption Tax Reforms: Micro-Level Evidence from European VAT," *American Economic Journal: Economic Policy*, 13, 103–134.

Byrne, D. P. and N. de Roos (2017): "Consumer Search in Retail Gasoline Markets," *Journal of Industrial Economics*, 65, 183–193.

Byrne, D. P. and L. A. Martin (2021): "Consumer Search and Income Inequality," *International Journal of Industrial Organization*, 79, 102716.

Cabral, L. M. B. (2018): "Standing on the Shoulders of Dwarfs: Dominant Firms and Innovation Incentives," *CEPR Discussion Paper No. 13115*.

Carranza, J. E., R. Clark, and J.-F. Houde (2015): "Price Controls and Market Structure: Evidence from Gasoline Retail Markets," *Journal of Industrial Economics*, 63, 152–198.

Carrier, M. A. (2002): "Unraveling the Patent-Antitrust Paradox," *University of Pennsylvania Law Review*, 150, 761–854.

Cason, T. N., D. Friedman, and E. Hopkins (2020): "Discrete Prices and the Incidence and Efficiency of Excise Taxes," *Journal of Political Economy*, 129, 790–841.

Chandler, A. D. (2005): *Inventing the Electronic Century: The Epic Story of the Consumer Electronics and Computer Industries*, Harvard University Press.

Chandra, A. and M. Tappata (2011): "Consumer Search and Dynamic Price Dispersion: An Application to Gasoline Markets," *RAND Journal of Economics*, 42, 681–704.

Chen, J. and J. Roth (forthcoming): "Logs with Zeros? Some Problems and Solutions," *Quarterly Journal of Economics*.

Chesbrough, H. and R. S. Rosenbloom (2002): "The Role of the Business Model in Capturing Value from Innovation: Evidence from Xerox Corporation's Technology Spin-Off Companies," *Industrial and Corporate Change*, 11, 529–555.

Chetty, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–488.

Chetty, R., A. Looney, and K. Kroft (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–1177.

Chien, C. (2003): "Cheap Drugs at What Price to Innovation: Does the Compulsory Licensing of Pharmaceuticals Hurt Innovation?" *Berkeley Technology Law Journal*, 18, 853–907.

Choi, H. (2008): "Technology Importation, Corporate Strategies, and the Rise of the Japanese Semiconductor Industry in the 1950s," *Comparative Technology Transfer and Society*, 6, 103–126.

Conlon, C. and N. L. Rao (2020): "Discrete Prices and the Incidence and Efficiency of Excise Taxes," *American Economic Journal: Economic Policy*, 12, 111–143.

——— (2023): "The Cost of Curbing Externalities with Market Power: Alcohol Regulations and Tax Alternatives," *NBER Working Paper No. 30896*.

Conlon, C., N. L. Rao, and Y. Wang (forthcoming): "Who Pays Sin Taxes? Understanding the Overlapping Burdens of Corrective Taxes," *Review of Economics and Statistics*.

Correia, S., P. Guimarães, and T. Zylkin (2020): "Fast Poisson Estimation with High-Dimensional Fixed Effects," *Stata Journal*, 20, 95–115.

Cunningham, C., F. Ederer, and S. Ma (2021): "Killer Acquisitions," *Journal of Political Economy*, 129, 649–702.

Curtis, P. J. (1994): *The Fall of the U.S. Consumer Electronics Industry: An American Trade Tragedy*, Quorum Books.

D'Acunto, F., D. Hoang, and M. Weber (2018): "Unconventional Fiscal Policy," *AEA Papers and Proceedings*, 108, 519–523.

——— (2022): "Managing Households' Expectations with Unconventional Policies," *Review of Financial Studies*, 35, 1597–1642.

De Loecker, J. and J. Eeckhout (2018): "Global Market Power," *NBER Working Paper No. 24768*.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561–644.

Delrahim, M. (2004): "Forcing Firms to Share the Sandbox: Compulsory Licensing of Intellectual Property Rights and Antitrust," *European Business Law Review*, 15, 1059–1069.

Deltas, G. and M. Polemis (2020): "Estimating Retail Gasoline Price Dynamics: The Effects of Sample Characteristics and Research Design," *Energy Economics*, 92, 1049–1076.

Diamond, P. A. (1971): "A Model of Price Adjustment," *Journal of Economic Theory*, 3, 156–168.

Dubois, P., R. Griffith, and M. O'Connell (2020): "How Well Targeted Are Soda Taxes?" *American Economic Review*, 110, 3661–3704.

Duso, T. and F. Szücs (2017): "Market Power and Heterogeneous Pass-Through in German Electricity Retail," *European Economic Review*, 98, 354–372.

Eckert, A. (2013): "Empirical Studies of Gasoline Retailing: A Guide to the Literature," *Journal of Economic Surveys*, 27, 140–166.

Eggertsson, G. B., J. A. Robbins, and E. G. Wold (2021): "Kaldor and Piketty's Facts: The Rise of Monopoly Power in the United States," *Journal of Monetary Economics*, 124, S19–S38.

Eizenberg, A., S. Lach, and M. Oren-Yiftach (2021): "Retail Prices in a City," *American Economic Journal: Economic Policy*, 13, 175–206.

Fabra, N. and M. Reguant (2014): "Pass-Through of Emissions Costs in Electricity Markets," *American Economic Review*, 104, 2872–2899.

Federico, G., F. Scott Morton, and C. Shapiro (2020): "Antitrust and Innovation: Welcoming and Protecting Disruption," *Innovation Policy and the Economy*, 20, 125–190.

Feenstra, R. C. (1996): "U.S. Imports, 1972-1994: Data and Concordances," *NBER Working Paper No. 5515.*

Fischer, K., S. Martin, and P. Schmidt-Dengler (2023): "The Heterogeneous Effects of Entry on Prices," *CEPR Discussion Paper No. 18297.*

Federal Trade Commission (1975): "Xerox Corporation: Consent Order, etc., in Regard to Alleged Violation of the Federal Trade Commission Act," in *Federal Trade Commission Decisions*, vol. 86, 364–386.

Furman, J. L., M. Nagler, and M. Watzinger (2021): "Disclosure and Subsequent Innovation: Evidence from the Patent Depository Library Program," *American Economic Journal: Economic Policy*, 13, 239–270.

Gaessler, F., D. Harhoff, and S. Sorg (2019): "Bargaining Failure and Freedom to Operate: Re-Evaluating the Effect of Patents on Cumulative Innovation," *CEPR Discussion Paper No. 13969.*

Galasso, A. and M. Schankerman (2015): "Patents and Cumulative Innovation: Causal Evidence from the Courts," *Quarterly Journal of Economics*, 130, 317–369.

Ganapati, S., J. S. Shapiro, and R. Walker (2020): "Energy Cost Pass-Through in US Manufacturing: Estimates and Implications for Carbon Taxes," *American Economic Journal: Applied Economics*, 12, 303–342.

Genakos, C. and M. Pagliero (2022): "Competition and Pass-Through: Evidence from Isolated Markets," *American Economic Journal: Applied Economics*, 14, 35–57.

Genesove, D. and W. P. Mullin (1998): "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890–1914," *RAND Journal of Economics*, 29, 355–377.

Gilbert, R. J. (2022): *Innovation Matters: Competition Policy for the High-Technology Economy*, MIT Press.

Giorcelli, M. (2019): "The Long-Term Effects of Management and Technology Transfers," *American Economic Review*, 109, 121–152.

Giorcelli, M. and B. Li (2021): "Technology Transfer and Early Industrial Development: Evidence from the Sino-Soviet Alliance," *NBER Working Paper No. 29455.*

Goldberg, P., A. Khandelwal, N. Pavcnik, and P. Topalova (2009): "Trade Liberalization and New Imported Inputs," *American Economic Review*, 99, 494–500.

Gomes-Casseres, B. and K. McQuade (1991): "Xerox and Fuji Xerox," *Harvard Business School Case 9-391-156.*

Goolsbee, A., S. Levitt, and C. Syverson (2016): *Microeconomics*, Worth Publishers, 2nd ed.

Goto, A. and K. Motohashi (2007): "Construction of a Japanese Patent Database and a First Look at Japanese Patenting Activities," *Research Policy*, 36, 1431–1442.

Graham, M. B. W. (1986): *The Business of Research: RCA and the VideoDisc*, Cambridge University Press.

Green, J. R. and S. Scotchmer (1995): "On the Division of Profit in Sequential Innovation," *RAND Journal of Economics*, 26, 20–33.

Hamilton, S. F. (1999): "Tax Incidence under Oligopoly: A Comparison of Policy Approaches," *Journal of Public Economics*, 71, 233–245.

Harding, M., E. Leibtag, and M. F. Lovenheim (2012): "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data," *American Economic Journal: Economic Policy*, 4, 169–198.

Harju, J., T. Kosonen, M. Laukkanen, and K. Palanne (2022): "The Heterogeneous Incidence of Fuel Carbon Taxes: Evidence from Station-Level Data," *Journal of Environmental Economics and Management*, 112, 102607.

Hegde, D., K. Herkenhoff, and C. Zhu (2023): "Patent Publication and Innovation," *Journal of Political Economy*, 131, 1845–1903.

Heim, S. (2021): "Asymmetric Cost Pass-Through and Consumer Search: Empirical Evidence from Online Platforms," *Quantitative Marketing and Economics*, 19, 227–260.

Higham, K. and S. Nagaoka (2023): "Language Barriers and the Speed of International Knowledge Diffusion," *SSRN Electronic Journal*.

Hindriks, J. and V. Serse (2019): "Heterogeneity in the Tax Pass-Through to Spirit Retail Prices: Evidence From Belgium," *Journal of Public Economics*, 176, 142–160.

Hollenbeck, B. and K. Uetake (2021): "Taxation and Market Power in the Legal Marijuana Industry," *RAND Journal of Economics*, 52, 559–595.

Holt, T., M. Igami, and S. Scheidegger (2023): "Detecting Edgeworth Cycles," *SSRN Electronic Journal*.

Huber, S. and C. Rust (2016): "Calculate Travel Time and Distance with OpenStreetMap Data Using the Open Source Routing Machine (OSRM)," *Stata Journal*, 16, 416–423.

Iacus, S. M., G. King, and G. Porro (2012): "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis*, 20, 1–24.

Jacobson, G. and J. Hillkirk (1986): *Xerox: American Samurai*, Macmillan.

Jaffe, A. B. and M. Trajtenberg (1996): "Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations over Time and Across Institutional and Geographic Boundaries," *Proceedings of the National Academy of Sciences*, 93, 12671–12677.

Janssen, M., P. Pichler, and S. Weidenholzer (2011): "Oligopolistic Markets with Sequential Search and Production Cost Uncertainty," *RAND Journal of Economics*, 42, 444–470.

Janssen, M. and S. Shelegia (2015): "Consumer Search and Double Marginalization," *American Economic Review*, 105, 1683–1710.

——— (2020): "Beliefs and Consumer Search in a Vertical Industry," *Journal of the European Economic Association*, 18, 2359–2393.

Johnson, C. (1982): *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925-1975*, Stanford University Press.

Johnson, R. N. (2002): "Search Costs, Lags and Prices at the Pump," *Review of Industrial Organization*, 20, 33–50.

Johnstone, B. (1999): *We Were Burning: Japanese Entrepreneurs and the Forging of the Electronic Age*, Basic Books.

Juhász, R. (2018): "Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade," *American Economic Review*, 108, 3339–3376.

Juhász, R., N. Lane, and D. Rodrik (2023): "The New Economics of Industrial Policy," *NBER Working Paper No. 31538*.

Kang, H. (2021): "How Does Competition Affect Innovation? Evidence from U.S. Antitrust Cases," *USC Marshall School of Business Research Paper*.

Kearns, D. T. and D. A. Nadler (1992): *Prophets in the Dark: How Xerox Reinvented Itself and Beat Back the Japanese*, Harper Business.

Keller, W. (2004): "International Technology Diffusion," *Journal of Economic Literature*, 42, 752–782.

Keller, W. and S. R. Yeaple (2009): "Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States," *Review of Economics and Statistics*, 91, 821–831.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021): "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, 3, 303–320.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017): "Technological Innovation, Resource Allocation, and Growth," *Quarterly Journal of Economics*, 132, 665–712.

Kopczuk, W., J. Marion, E. Muehlegger, and J. Slemrod (2016): "Does Tax-Collection Invariance Hold? Evasion and the Pass-Through of State Diesel Taxes," *American Economic Journal: Economic Policy*, 8, 251–286.

Kosonen, T. (2015): "More and Cheaper Haircuts after VAT Cut? On the Efficiency and Incidence of Service Sector Consumption Taxes," *Journal of Public Economics*, 131, 87–100.

Kraftfahrt-Bundesamt (2021): "Der Fahrzeugbestand im Überblick am 1. Januar 2020," `https://www.kba.de/DE/Statistik/Fahrzeuge/Bestand/Jahrebilanz_`

Bestand/2020/2020_b_ueberblick_pdf.pdf?__blob=publicationFile&v=1 (last accessed: 19 February 2024).

Kroft, K., J.-W. P. Laliberté, R. Leal-Vizcaíno, and M. J. Notowidigdo (2021): "Efficiency and Incidence of Taxation with Free Entry and Love-of-Variety Preferences," *NBER Working Paper No. 28838.*

——— (2024): "Salience and Taxation with Imperfect Competition," *Review of Economic Studies*, 91, 403–437.

Lane, N. (2021): "Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea," *SoDa Laboratories Working Paper No. 2021-10.*

Lemus, J. and F. Luco (2021): "Price Leadership and Uncertainty about Future Costs," *Journal of Industrial Economics*, 69, 305–337.

Levy, J. D. (1981): *Diffusion of Technology and Patterns of International Trade: The Case of Television Receivers*, Yale University ProQuest Dissertations Publishing.

Lewis, M. S. (2011): "Asymmetric Price Adjustment and Consumer Search: An Examination of the Retail Gasoline Market," *Journal of Economics & Management Strategy*, 20, 409–449.

Li, J. and J. H. Stock (2019): "Cost Pass-Through to Higher Ethanol Blends at the Pump: Evidence from Minnesota Gas Station Data," *Journal of Environmental Economics and Management*, 93, 1–19.

Lind, J. T. and H. Mehlum (2010): "With or Without U? The Appropriate Test for a U-Shaped Relationship," *Oxford Bulletin of Economics and Statistics*, 72, 109–118.

Liu, C. (2020): "The Effects of World War I on the Chinese Textile Industry: Was the World's Trouble China's Opportunity?" *Journal of Economic History*, 80, 246–285.

Luco, F. (2019): "Who Benefits from Information Disclosure? The Case of Retail Gasoline," *American Economic Journal: Microeconomics*, 11, 277–305.

Mamrak, R. (2023): "Antitrust and (Foreign) Innovation: Evidence from the Xerox Case," *CRC Discussion Paper No. 396.*

Marion, J. and E. Muehlegger (2011): "Fuel Tax Incidence and Supply Conditions," *Journal of Public Economics*, 95, 1202–1212.

Martin, S. (forthcoming): "Market Transparency and Consumer Search – Evidence from the German Retail Gasoline Market," *RAND Journal of Economics.*

Miller, N. H., M. Osborne, and G. Sheu (2017): "Pass-through in a Concentrated Industry: Empirical Evidence and Regulatory Implications," *RAND Journal of Economics*, 48, 69–93.

Miravete, E. J., K. Seim, and J. Thurk (2018): "Market Power and the Laffer Curve," *Econometrica*, 86, 1651–1687.

Montag, F., R. Mamrak, A. Sagimuldina, and M. Schnitzer (2023a): "Imperfect Price Information, Market Power, and Tax Pass-Through," *Stigler Center Working Paper No. 337*.

Montag, F., A. Sagimuldina, and C. Winter (2023b): "Whom to Inform About Prices? Evidence from the German Gasoline Market," *CRC Discussion Paper No. 415*.

Moser, P. (2012): "Innovation without Patents: Evidence from World's Fairs," *Journal of Law and Economics*, 55, 43–74.

Moser, P. and A. Voena (2012): "Compulsory Licensing: Evidence from the Trading with the Enemy Act," *American Economic Review*, 102, 396–427.

Moser, P., A. Voena, and F. Waldinger (2014): "German Jewish Émigrés and US Invention," *American Economic Review*, 104, 3222–3255.

Mostafa, R. and S. Klepper (2018): "Industrial Development Through Tacit Knowledge Seeding: Evidence from the Bangladesh Garment Industry," *Management Science*, 64, 613–632.

Nagler, M., M. Schnitzer, and M. Watzinger (2022): "Fostering the Diffusion of General Purpose Technologies: Evidence from the Licensing of the Transistor Patents," *Journal of Industrial Economics*, 70, 838–866.

Nakamura, E. and D. Zerom (2010): "Accounting for Incomplete Pass-Through," *Review of Economic Studies*, 77, 1192–1230.

Owen, D. (2005): *Copies in Seconds: How a Lone Inventor and an Unknown Company Created the Biggest Communication Breakthrough Since Gutenberg – Chester Carlson and the Birth of the Xerox Machine*, Simon & Schuster.

Paley, N. (2006): *The Manager's Guide to Competitive Marketing Strategies*, Thorogood Publishing, 3rd ed.

Pennerstorfer, D., P. Schmidt-Dengler, N. Schutz, C. Weiss, and B. Yontcheva (2020): "Information and Price Dispersion: Theory and Evidence," *International Economic Review*, 61, 871–899.

Petralia, S., P.-A. Balland, and D. Rigby (2016): "HistPat Dataset," Harvard Dataverse.

Poege, F. (2022): "Competition and Innovation: The Breakup of IG Farben," *Boston University School of Law Research Paper No. 22-24.*

Porter, M. E. (1985): *Competitive Advantage: Creating and Sustaining Superior Performance*, Free Press.

——— (1988): "Canon, Inc.: Worldwide Copier Strategy," *Harvard Business School Case 9-384-151.*

Sampat, B. and H. L. Williams (2019): "How Do Patents Affect Follow-on Innovation? Evidence from the Human Genome," *American Economic Review*, 109, 203–236.

Scherer, F. M. (2005): "The Role of Patents in Two US Monopolization Cases," *International Journal of the Economics of Business*, 12, 297–305.

Scherer, F. M. and J. Watal (2014): "Competition Policy and Intellectual Property: Insights from Developed Country Experience," *HKS Working Paper No. RWP14-013.*

Segal, I. and M. D. Whinston (2007): "Antitrust in Innovative Industries," *American Economic Review*, 97, 1703–1730.

Shih, W. C. and G. Dieterich (2014): "RCA: Color Television and the Department of Justice," *Harvard Business School Case 614-072.*

Stahl, D. O. (1989): "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review*, 79, 700–712.

Stern, N. (1987): "The Effects of Taxation, Price Control and Government Contracts in Oligopoly and Monopolistic Competition," *Journal of Public Economics*, 32, 133–158.

Sumner, D. A. (1981): "Measurement of Monopoly Behavior: An Application to the Cigarette Industry," *Journal of Political Economy*, 89, 1010–1019.

Syverson, C. (2019): "Macroeconomics and Market Power: Context, Implications, and Open Questions," *Journal of Economic Perspectives*, 33, 23–43.

Tappata, M. (2009): "Rockets and Feathers: Understanding Asymmetric Pricing," *RAND Journal of Economics*, 40, 673–687.

Tom, W. K. (2001): "The 1975 Xerox Consent Decree: Ancient Artifacts and Current Tensions," *Antitrust Law Journal*, 68, 967–990.

United States Congress (1970): "Electronics, Heavy Electrical Equipment, Item 807," in *Tariff and Trade Proposals, Hearings Before the Committee on Ways and Means*, vol. 10, 2827–3049.

United States Patent and Trademark Office (1975): "Xerox License Offer," in *Official Gazette of the United States Patent and Trademark Office*, vol. 939, 1665–1739.

United States v. Radio Corporation of America (1954): "Complaint, Civil No. 97-38," Southern District of New York.

——— (1958): "Final Judgement, Civil No. 97-38," Southern District of New York.

Varian, H. R. (1980): "A Model of Sales," *American Economic Review*, 70, 651–659.

Watzinger, M., T. A. Fackler, M. Nagler, and M. Schnitzer (2020): "How Antitrust Enforcement Can Spur Innovation: Bell Labs and the 1956 Consent Decree," *American Economic Journal: Economic Policy*, 12, 328–359.

Watzinger, M. and M. Schnitzer (2022): "The Breakup of the Bell System and its Impact on US Innovation," *CEPR Discussion Paper No. 17635.*

Weyl, E. G. and M. Fabinger (2013): "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition," *Journal of Political Economy*, 121, 528–583.

Williams, H. L. (2013): "Intellectual Property Rights and Innovation: Evidence from the Human Genome," *Journal of Political Economy*, 121, 1–27.

——— (2017): "How Do Patents Affect Research Investments?" *Annual Review of Economics*, 9, 441–469.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

München, 11. März 2024 Robin Mamrak