Four Essays on the Analysis of Market Power and its Implications

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades

an der Wirtschaftswissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

Eingereicht von: Hendrik Döpper, M.Sc.

Erstbetreuer: Prof. Dr. Alexander Rasch (DICE)

Zweitbetreuer: Prof. Dr. Joel Stiebale (DICE)

Abgabedatum: 21. November 2023

Acknowledgement

I am at the end of a long journey during which I have met many brilliant people. This is a mere attempt to acknowledge their impact on my career and personal life. Unfortunately, the nature of this type of list is to overlook someone or understate someone's contribution. I hope everyone will forgive me for such mistakes.

I will start my list with Susanne Thorwarth. When I decided to start my doctoral studies in 2017, I left the Heinreich Heine University to work in consulting. The idea was to pursue the doctoral studies part-time alongside my job and perhaps move back into an academic job after the doctorate. Susanne was my boss at Düsseldorf Competition Economics (formerly DICE Consult), and she taught me how academia works during numerous lunch breaks. She provided me with a unique, in-depth knowledge of the academic world and all its peculiarities. In doing so, she shaped many decisions that I made in the subsequent years. Most importantly, Susanne is the reason why I left consulting and returned to the university. I will admit that this decision and all the other decisions I made based on Susanne's advice were truly life changing. Without these decisions, I likely would not have shifted my focus to empirical work. I likely would not have started my work with my second supervisor, Joel Stiebale, and with my co-authors Alexander MacKay and Nathan Miller. I likely would not have done a research stay at Harvard University. (Many thanks to Elie Tamer for the invitation!) And all these things are just the tip of the iceberg of things I likely would not have done if I had not returned to the university. Thank you, Susanne! Thank you for your effort to bring me on track!

Next, I would like to thank my PhD supervisors, Alexander Rasch and Joel Stiebale. Both have contributed significantly to my career, although their contribution could not be more different. Alexander guided me through the early years, helping me start the first projects and gain a foothold in the academic world. Since my early days at DICE, he has been an integral part of my circle that I consult frequently. Joel, on the other hand, joined this circle later. He helped me shift my focus to empirical work and develop my empirical skills. In doing so, he has significantly shaped the last years of my time at DICE. I look forward to many great projects with them in the future!

In addition to my supervisors, my co-authors Alexander MacKay and Nathan Miller deserve special thanks. Both have significantly shaped my career. We met at the beginning of the COVID-19 crisis, and I am glad that they frequently challenged me intellectually during that difficult time. They have always been supportive and allowed me to develop my analytical and empirical skills. I would also like to especially thank Alexander for the numerous hours he spent with me at Harvard (and beyond) discussing my projects and job market opportunities.

I also would like to thank my other co-authors, Geza Sapi, Jannika Schad, Benjamin Schröder, Olaf Stypa, and Christian Wey. All have contributed to my academic career.

During my time at DICE, I have met a lot of great colleagues. I am afraid to name specific names here. There were way too many great colleagues, and my biggest fear is to miss someone if I try to name them all. The only person that I would like to highlight is my former office partner Jannika Schad. Thank you so much for the great more than 2 years we shared our office.

Finally, I would like to thank my family and friends who have supported me over the last few years. An underestimated aspect of academic success is the ability to stop working and take breaks. I am convinced that a cornerstone of brilliant work is to allow your brain to rest and process information in the background while you focus on completely unrelated activities. I am grateful to my family and my friends for regularly reminding me of this!

Coauthors and Publications

Chapter 1: Combinable Products, Price Discrimination, and Collusion

Coauthor: Alexander Rasch (Heinrich Heine University)

Status: The chapter is a revised version of a working paper published in the *DICE Discussion Paper Series* (No 377) and has been resubmitted to the *International Journal of Industrial Organization* ("first round R&R"). Appendix 1.A is not part of the resubmission, but documents findings from previous work.

Chapter 2: A Bargaining Perspective on Vertical Integration

Coauthors: Geza Sapi (DG Comp, European Commission) and Christian Wey (Heinrich Heine University)

Status: The chapter is accepted for publication at the *Canadian Journal of Economics*. Appendix 2.D is the online appendix in the accepted version and Appendix 2.E is not part of the accepted version, but was part of a response letter to a referee. An old version of the accepted article is published in the *DICE Discussion Paper Series* (No 389).

Chapter 3: Rising Markups and the Role of Consumer Preferences

Coauthors: Alexander MacKay (Harvard University), Nathan H. Miller (Georgetown University) and Joel Stiebale (Heinrich Heine University)

Status: The chapter is available as a working paper. The latest version is published in the *Kilts Center at Chicago Booth Marketing Data Center Paper Series*, the *Harvard Business School: Strategy Unit Working Paper Series* (No. 22-025) and the *Georgetown University McDonough School of Business Research Paper Series* (No. 3939126). It can be accessed on SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3939126` (last accessed on June 28, 2023). The chapter was submitted to the *Journal of Political Economy* and the journal has asked us to revise and resubmit the article ("first round R&R"). The chapter in its current form is the version that was originally submitted to the journal and does not contain any revisions made on the referees' requests.

Chapter 4: The Portfolio Power Theory Revisited: Evidence from Cross-Category Mergers in US Retailing

Coauthors: none

Status: The chapter is available as a working paper. The latest version is published in the *Kilts Center at Chicago Booth Marketing Data Center Paper Series* and can be accessed on SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4630353` (last accessed on November 14, 2023).

Contents

Introduction

The understanding of mechanisms driving market power and the implications of market power for market outcomes and consumers is a longstanding issue of Industrial Organization (IO). An entire subfield of IO, called competition or antitrust economics, is devoted to the questions under which conditions firms have too much market power, how they can abuse market power at the expense of rivals and consumers (thereby also defining what "too much" market power actually means), and how to prevent firms from gaining too much market power.

At the same time, market power has not remained a niche topic for researchers but has also attracted the interest of people outside of the academic economic profession. A great example is how competition economics has helped shape legislation in many jurisdictions, such as the European Union or the United States of America. For instance, laws have been implemented to prohibit cartels, prevent harmful mergers, or punish the abuse of dominant positions.

Despite a large body of IO literature on market power, there are still many open questions, and new questions arise frequently. This dissertation is my attempt to contribute to understanding recent developments related to market power in several ways.

The *first chapter* of my dissertation is the result of my work with Alexander Rasch and tries to shed light on undesired consequences of policy interventions that are meant to benefit consumers. We focus on markets that are characterized by "mixing." The idea is that in many markets, consumers are confronted with differentiated products and the likelihood that a product perfectly fits their preferences is rather small. However, consumers might be able to lower the disutility resulting from a mismatch of their preferences by buying multiple products. One example would be TV channels that are often specialized. For instance, some TV channels focus on sports events, while others concentrate on documentaries. However, a consumer may not like to watch either sports or documentaries exclusively, and doing so would result in a low utility. By switching the TV channels whenever (s)he feels attracted to another genre, (s)he can get a better fit with his/her preferences, and the utility level increases.[1]

Markets with mixing sometimes come with another characteristic that makes them particularly interesting for policymakers. In many markets, firms use rather complex pricing schemes. With complex, we primarily mean the number of available price instruments. Returning to the TV example, firms could use advertising breaks to generate revenues. The longer a consumer watches, the more ads (s)he will see; thus, advertising works similarly to a linear price that scales with the quantity consumed. In contrast, the TV operator could also charge a fixed fee, which is independent of the actual time watched. While both linear and fixed prices constitute simple pricing schemes, the operator can also use more complex tariffs. For instance, it can combine linear and fixed prices in a two-part tariff. Extending the list to even more complex examples like three-part tariffs is straightforward.

Firms can use these complex pricing schedules to price-discriminate among consumers.

[1] In the chapter, we focus on the banking and insurance industry. In the introduction, I refer to examples from the media and entertainment industry to highlight that the model we consider is not restricted to one narrow industry but applies to a more extensive set of industries.

For our work, we are particularly interested in second-degree price discrimination. Our baseline analysis focuses on two-part tariffs where firms can set a linear price along with a fixed fee and compares the market outcomes to the special cases of linear prices and fixed fees only.

There are many examples of price discrimination that attract the interest of policymakers who want to intervene and design regulations to force firms to use pricing schemes with fewer price components (examples are provided in the chapter). Prior literature indeed points to potential benefits for consumers these regulations can have (Anderson and Neven, 1989; Hoernig and Valletti, 2007, 2011). The contribution of our paper is to show that there can also be undesired consequences of such policy interventions in the sense that they can increase the likelihood of (tacit) collusion. More precisely, we show that banning either price component of a two-part tariff makes collusive agreements easier to sustain. Although our baseline analysis considers only three pricing schemes and collusion on profit-maximizing prices, we show that this result remains robust when moving to a more flexible pricing scheme. We also show that under partial collusion (i.e., collusion on prices below the profit-maximizing prices), consumers can be harmed by a ban on the fixed price component of a two-part tariff.

The *second chapter* of this dissertation is a joint work with Geza Sapi and Christian Wey. It deals with the question of what drives mergers, which is an essential question for the design of merger control policies and, thus, another important topic in competition economics. Traditionally, researchers and practitioners relate horizontal and vertical mergers to different driving forces. The commonly held view is that horizontal mergers are typically related to the elimination of competition (on the same market level), while vertical integration can give rise to efficiency gains, such as through the elimination of double marginalization, or allow the integrated firms to foreclose their competitors.

Our contribution is to challenge this widely held view by analyzing the impact of bargaining power[2] on merger decisions. We show that from a pure bargaining perspective, the incentives to merge horizontally and vertically are similar or, more precisely, that vertical merger incentives are a combination of horizontal merger incentives up- and downstream. Our analysis builds on the model of Inderst and Wey (2003), who analyze the impact of bargaining power on horizontal merger decisions. Their framework allows them to isolate the pure bargaining power effect, thereby ignoring other driving forces like those mentioned earlier. They show that horizontal merger incentives are driven by the shape of the (unit) cost function, while horizontal upstream merger incentives are driven by the complementarity or substitutability of the upstream firms' products. We extend their analysis to the case

[2]The definition of bargaining power that underlies our analysis in Chapter 2 differs somewhat from the definition of market power. Market power is usually defined as the ability of firms to set prices above marginal costs. Firms with larger market power can increase the markup on their marginal costs and thus earn larger profits. The model used in Chapter 2 considers efficient bargaining where changes in the bargaining power affect only the distribution of rents among up- and downstream firms but not the total profit generated. However, I suspect that most non-economists who are not familiar with these definitions are unlikely to be able to make this fine distinction, so this chapter still fits the overall theme of my thesis, at least to some extent.

of vertical integration and find that these two determinants also drive vertical mergers.

We complement our analysis with a formal comparison of the strength of merger incentives, which sheds light on the question of whether firms have an incentive to preempt a merger with a rival. The idea is that a firm might want to take over a potential target even though there are other firms in the market that would have larger gains from the merger. The reason could be that a merger with a rival is particularly harmful, so the firm has a large incentive to preempt it and carry out the merger on its own. To study this question, we use a simple auction model where an arbitrary firm is up for sale, and all other firms are allowed to bid for the target. We show that preemption never drives the outcome of the auction, which means that the firm that has the largest gain from a merger (i.e., the largest difference between pre- and post-merger profits) carries out the merger.

Like many other IO studies, the first two chapters of my dissertation focus on particular markets or market environments. However, other researchers have recently expressed interest in the evolution of market power over time across industries and countries. Most notable in this regard is the literature that arose following the seminal work of De Loecker et al. (2020). Since market power has always been a key topic in IO, IO economists would be perfectly qualified to participate in this debate and contribute with the expertise they have built up over the last decades. However, this would require extending many standard IO tools so that they can be applied across markets.

The *third* chapter of my dissertation tries to contribute to this domain. It is a joint work with Alexander MacKay, Nathan Miller, and Joel Stiebale and studies the evolution of market power in the US consumer packaged goods retail industry between 2006 and 2019. Our work differs from most of the prior literature in the approach used. While others typically take a firm perspective and use the so-called supply-side (or production function) approach, which builds on assumptions of cost-minimizing firms and on balance sheet data (or similar data sets), we take a consumer perspective and leverage the so-called demand-side approach. We use data on realized consumer purchases in the form of average prices and market shares to estimate demand elasticities as a measure of consumers' reactions to price changes and use a theoretical link between the substitution patterns and firms' price-setting behavior to calculate markups. This approach has various advantages. First, we can analyze the same trend of rising market power with a completely unrelated approach. Every approach comes with its own assumptions and other scholars have recently pointed out challenges associated with the assumptions of the supply-side approach (see, for instance, Bond et al., 2021; Raval, 2023). Therefore, it is reassuring that we find similar markup trends with a completely unrelated approach. Second, we can carry out the analysis at a much more detailed level by using product level information so that we can determine whether changes in markups are driven by particular groups of products or by all products. Third, since consumer behavior is an integral part of our model, we are able to gain additional insights regarding mechanisms driving changes in markups and implications for consumer surplus and welfare.

Our main finding is that markups have increased by about 30 percent from 2006 to 2019.

The increase in markups is primarily driven by falling marginal costs, which are not passed on to consumers because consumers became less price-sensitive. In this sense, there are two opposing channels. On the one hand, firms want to lower prices because of lower marginal costs, while on the other hand, they want to increase prices because consumers react less strongly to price increases. These two channels roughly offset each other, resulting in a relatively constant price level and increasing markups.

This also has implications for the consumer surplus. Because prices remain relatively stable over time and consumers care less about the prices (that is, they become less price-sensitive), the utility increases for the average consumer. However, there are substantial differences across consumers of different income groups. For instance, consumers from the lowest income quartile experience a strictly lower utility level in the middle of our sample period. Although their utility increases again until the end of our sample period, their utility level is roughly the same in 2019 as in 2006.

In addition, we find that the increase in markups happens "within products," meaning that the markup for the average product increased drastically and that a redistribution of sales from low-markup to high-markup products does not play an important role.

We complement our analysis with anecdotal evidence describing how consumers' behavior has changed over time. First, we look at the usage of coupons. Coupons are a tool for consumers that they can use to save money. However, consumers face a trade-off because saving money requires exerting a small amount of effort. This effort level should decrease over time since using coupons has arguably become easier due to technological progress. Our analysis shows that the usage of coupons declined drastically and that this trend started long before 2006. Second, we look at surveys on time usage and find that the amount of time consumers spend grocery shopping also declines over time. This could fit our result if, for example, consumers spent less time comparing prices and finding deals.

The *fourth chapter* of my dissertation is single-authored and deals with the portfolio power theory, which is sometimes discussed in the context of (mostly conglomerate) mergers. Quoting from the chapter, "[t]he idea is usually that if two firms sell their products to the same downstream firms, a merger can benefit them even if their product portfolios do not overlap before the merger. In other words, the increase in the sheer size of a firm's product portfolio can change market outcomes, leaving aside possible substitutability and complementarity considerations within the portfolio." The potential mechanisms underlying the portfolio power theory are often related to bargaining considerations and follow the idea that a merger between two (upstream) firms leads to change in bargaining positions when negotiations with downstream firms.

As discussed in more detail in the introduction of the chapter, the portfolio power theory caused heated debates around the turn of the millennium when the European Commission used it in the evaluation of multiple large merger cases. It has recently attracted the attention of economists and legal scholars after Lina Khan—Chair of the Federal Trade Commission—mentioned it in a speech to the International Competition Network. Despite the broad attention it has received in the past, surprisingly little empirical research has been conducted to assess its presence and potential effects in the practice. My chapter tries to help filling this gap by providing an analysis of 57 cross-category mergers in the US consumer packaged goods retail industry. Cross-category mergers are appealing for studying the portfolio power theory because the merging firms have (almost) no overlap in their product portfolio before the merger. Therefore, changes in market outcomes must arise from alternative channels like portfolio effects.

I exploit the fact that the pre-merger bargaining positions differ across targets and acquirers at the different retailers. I provide evidence that the merging party with a weaker pre-merger bargaining position at a given retailer benefits through an increase in revenues post-merger. This change in revenues is driven by an increase in quantities and not by a change in prices. To shed more light on potential channels, I use my work with Alexander MacKay, Nathan Miller, and Joel Stiebale (Chapter 3) to derive measures of marginal costs and perceived quality of the products. I find that changes in the perceived quality can help explain the patterns, while changes in marginal costs do not contribute to understanding the changes in revenues. Finally, I briefly discuss potential channels behind the documented patterns and discuss if and how they are related to the portfolio power theory.

Finally, I want to emphasize that each of these four chapters was originally written as a research paper. I have not attempted to adjust the wording, notation, and style. For instance, each chapter still refers to itself as a paper, and there are no cross-references between chapters. From an academic perspective, other researchers will most likely look at the papers rather than the dissertation, and these papers will likely be updated in the future. In other words, from an academic perspective, streamlining the wording, notation, and style seems to be of little benefit, and the biggest value is the contribution of each individual chapter and not the interplay between the different chapters. I also highly recommend that interested readers check for updated versions of the papers rather than reading the (likely outdated) versions in this dissertation.

Bibliography

Anderson, S. P. and Neven, D. J. (1989). Market efficiency with combinable products. *European Economic Review*, 33(4):707–719.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 121:1–14.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, 135(2):561–644.

Hoernig, S. H. and Valletti, T. M. (2007). Mixing goods with two-part tariffs. *European Economic Review*, 51(7):1733–1750.

Hoernig, S. H. and Valletti, T. M. (2011). When two-part tariffs are not enough: Mixing with nonlinear pricing. *B.E. Journal of Theoretical Economics*, 11(1).

Inderst, R. and Wey, C. (2003). Bargaining, mergers, and technology choice in bilaterally oligopolistic industries. *RAND Journal of Economics*, 34(1):1–19.

Raval, D. (2023). Testing the production approach to markup estimation. *Review of Economic Studies*, 90(5):2592–2611.

Chapter 1

Combinable Products, Price Discrimination, and Collusion

Coauthor: Alexander Rasch

Abstract:
We analyze the effect of different pricing schemes on horizontally differentiated firms' ability to sustain collusion when customers have the possibility to combine (or mix) products to achieve a better match of their preferences. To this end, we compare two-part tariffs with linear prices and quantity-independent fixed fees. We find that a ban of either price component of the two-part tariff makes it more difficult to sustain collusion at profit-maximizing prices. Moreover, linear pricing—as the most beneficial pricing schedule for customers in absence of collusion—harms customers most in presence of collusion.

Acknowledge:
We would like to thank Bernhard Kasberger, Nicholas Shunda, Achim Wambach, David Zeimentz, and participants at the CRESSE 2019, the EARIE 2019, the Competition and Innovation Summer School 2019, and the 4th DICE Winter School 2019 for very helpful comments and discussions. We also thank the Co-Editor of the International Journal of Industrial Organization, José Moraga, and two anonymous referees. Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf. Funding by the Deutsche Forschungsgemeinschaft (DFG) (project 235577387/GRK1974) is gratefully acknowledged.

1.1 Introduction

The present paper contributes to the ongoing debate in competition and customer protection policy that centers around the competitive or anti-competitive effects of different pricing schemes. In our analysis, we focus on the aspect of combining products (or mixing) and further explore the framework applied by Anderson and Neven (1989) and Hoernig and Valletti (2007, 2011). Under the possibility of mixing, customers can demand products from different firms, and in doing so, they create their own products that better fit their needs and preferences. The idea is that, from the customers' perspective, each product has advantages and disadvantages, and by combining different products, customers can enjoy the benefits of different products while offsetting the disadvantages of each product. In this sense, the characteristics of the new individualized "product" are a combination of the characteristics of the individual products. The possibility to combine products is a widespread phenomenon in many important industries (for example, banking and insurance; see our discussion below).

The scope of mixing (and, hence, the extent of product match) is crucial for welfare and policy considerations. In our framework, each customer buys one unit. The customers realize a basic utility from consumption, and prices are mere transfers of customer utility to the firms. The only source of disutility that may lower total welfare comes from a mismatch of customers' preferences that reduces customer surplus. Previous contributions analyze how different pricing schemes affect the scope of mixing and find that simple pricing schemes can be beneficial for customers. The best outcome is observed with linear prices. In this case, all customers optimally combine products, so that welfare is maximized. By contrast, the worst outcome is achieved with quantity-independent fixed fees that lead to a non-mixing result, that is, all customers buy exclusively from one firm. As a consequence, customers do not buy their preferred products and, thus, suffer from a disutility, which in turn leads to a welfare loss. With two-part tariffs and nonlinear pricing in general, some mixing occurs, so that welfare ranks second after the case of linear prices. Therefore, regulations of the pricing regimes, such as a ban of the fixed price component of a two-part tariff, can increase the number of customers who combine their products and, hence, increase customer surplus and total welfare.

Whereas the aforementioned contributions focus on a static environment, we extend the framework to a dynamic model to investigate firms' incentives to collude in an infinitely repeated game. By applying grim-trigger strategies, we derive the critical discount factors to compare the impact of linear prices, fixed fees, and two-part tariffs on firms' ability to collude. We show that firms' access to multi-part pricing schemes can reduce the scope for collusion in dynamic environments. More precisely, our main result is that firms' ability to use two-part tariffs can reduce collusion. In this sense, the effect of restrictions on pricing schemes are less clear-cut, and regulations that are explicitly imposed to benefit customers can backfire.

We investigate the robustness of this result with respect to two important assumptions. First, we demonstrate that the consequence of banning price discrimination does not only hold for two-part tariffs, but also extends to nonlinear pricing in general. Second, we allow firms to collude on prices below the profit-maximizing collusive prices (partial collusion). We find that firms are most likely to gain the largest profits when they partially collude on linear prices instead of fixed fees or two-part tariffs. By contrast, partial collusion on fixed prices does not yield any advantage. This finding complements the result of our baseline analysis in the sense that linear pricing schedules are most beneficial for customers in the absence of collusion, but they harm customers most in the presence of (partial) collusion.

Our analysis sheds light on the potential for anti-competitive practices in some of the most essential sectors of the economy. Prime examples for markets to which our results can be applied are the banking and the insurance sector and related services. These markets are indeed characterized by the four most important ingredients of our model: (i) mixing, (ii) two-part tariffs (or, more generally, nonlinear pricing), (iii) regulation of/ban on certain tariff components, and (iv) collusion. Having a closer look at the market for banking services shows that in the United States, for example, people have three credit cards on average.[1] Similarly, half of Americans use more than one bank, and it is recommended to have four accounts (at different banks).[2] Moreover, credit card and banking services usually come at various fees. As an example, credit card holders and bank account owners pay an annual fixed fee and linear fees for additional services (for instance, credit card usage, cash withdrawal).[3] Furthermore, the industry has seen various efforts by regulators to cap or ban certain fees.[4] Regulators have particularly focused on overdraft fees that come in different forms (fixed, linear, nonlinear).[5] In the UK, the Financial Conduct Authority (FCA) banned fixed fees for borrowing through an overdraft in 2019. Banks and building societies are now required to price overdrafts by a simple annual interest rate. A similar approach by the Consumer Financial Protection Bureau (CFPB) is currently under way in the United States, where the CFPB intends to intervene in overdraft practices. In this context, large banks such as Capital One and Bank of America decided to end or greatly reduce the use of these fees, and it has been argued that they had done so in anticipation of regulatory intervention.[6] With regard to regulating fees for credit cards, the Credit Card Accountability Responsibility and Disclosure Act of 2009 (CARD Act) stipulates that a

[1] See `https://www.forbes.com/advisor/credit-cards/credit-card-statistics/`.

[2] See `https://www.gobankingrates.com/banking/banks/how-many-bank-accounts-americans-have/`, `https://www.businessinsider.com/personal-finance/how-many-bank-accounts-should-i-have-tiffany-aliche`, and `https://www.gobankingrates.com/banking/checking-account/how-many-bank-accounts-can-have/`.

[3] Note that these fees may be negative, for example, because firms offer frequent-flyer points and insurance coverage as extra benefits.

[4] On a general note, the Biden-Harris Administration has recently launched a campaign against so-called junk fees and related pricing practices. One of the sectors targeted is banking (see `https://www.whitehouse.gov/briefing-room/blog/2022/10/26/the-presidents-initiative-on-junk-fees-and-related-pricing-practices/`). Other examples include airline pricing, event ticketing, and hotel booking.

[5] See `https://www.moneysavingexpert.com/news/2018/12/fixed-daily-and-monthly-overdraft-charges-to-be-banned/`.

[6] See `https://www.theregreview.org/2022/01/22/saturday-seminar-united-states-over-overdraft-fees/`.

credit card late payment fee must not surpass $30 for a first late payment.[7] On a related note, the prohibition of surcharges for credit cards above a certain percentage level in some US states and other jurisdictions (for example, Australia) can be viewed as an indirect form of regulating the linear pricing component. Because merchants' adoption decision is crucially affected by the credit card firms' interchange fee, credit card firms are restricted in their pricing strategy. In the European Union, there is a direct link between interchange regulation and the no-surcharge rule: "In all cases where the card charges imposed on merchants will be capped, in accordance with the complimentary multilateral interchange fees (MIF) Regulation (...), merchants will no longer be allowed to surcharge customers for using their payment card."[8] Turning to the fourth characteristic, the banking sector has seen quite a few cases of collusion. In a recent case in the United States, eight banks are being investigated because they are suspected of artificially blowing up interest rates that state and local governments must pay on a popular tax-exempt municipal bond.[9] Another case of anti-competitive behavior concerns initial public offerings. In 2017, the OECD concluded that banks' high fees were "akin to tacit collusion."[10]

With regard to the application of our set-up to the insurance sector, we point out that in the market for insurance services, mixing is also a widespread phenomenon. For example, a study for the German insurance market found that on average people have contracts with 2.7 insurance companies (Bain & Company, 2012). Such insurance contracts often include different price components; for example, a fixed annual fee and a deductible (either in fixed or in percentage terms). Furthermore, regulation is an important aspect in this sector. For example, in Germany, insurers and brokers cannot pass on commissions to customers.[11] Bans on upfront commissions are also being discussed in other parts of the world.[12] The Central Bank of Ireland recently introduced a ban on so-called "price walking" for private car and home insurance customers. Under this policy, discounts to attract new customers have to be made available also to those customers who remain with the same insurance provider (no loyalty penalty).[13]. As such, the policy hinders insurance firms to price-discriminate. Last, price-fixing practices are an issue in the industry as exemplified by the discussions at an OECD policy roundtable (OECD, 1998).

[7] See `https://www.bankrate.com/finance/credit-cards/late-fee-on-a-credit-card-late-fee/`.

[8] `https://ec.europa.eu/commission/presscorner/detail/de/MEMO_13_719`. In Europe, the interchange is capped at 0.3% for credit cards, such that merchants are estimated to be indifferent between accepting payment by card or in cash. In most of those US states where surcharging is legal, the surcharge is capped at 4% of the transaction total. In any case, it must not be used by merchants to make an extra profit.

[9] See `https://www.reuters.com/business/finance/judge-narrows-san-diego-baltimore-bond-collusion-cases-against-big-banks-2022-06-28/`.

[10] See `https://www.theglobeandmail.com/report-on-business/streetwise/the-tacit-collusion-of-big-bank-fee-setting/article35156863/`. Chen and Ritter (2000) provide evidence for this observation. Hansen (2001) argues that profits are not abnormal in the industry.

[11] See `https://www.lexology.com/commentary/insurance/germany/arnecke-sibeth-dabelstein/prohibition-on-passing-on-commission-in-reinsurance-context-exemption-uncertainty`.

[12] See, for example, `https://www.investmentexecutive.com/news/from-the-regulators/insurance-regulators-start-consultation-on-banning-upfront-commissions/`.

[13] See `https://www.centralbank.ie/news-media/press-releases/press-release-end-the-loyalty-penalty-for-private-car-and-home-insurance-21-July-2021`

The remainder of this paper is organized as follows. In the next section, we discuss the related literature. We describe the model and previous findings in Section 1.3. Section 1.4 analyzes collusion when firms can set profit-maximizing prices (full collusion). We investigate the robustness of our findings in Section 1.5. In doing so, we start with a general comment suggesting that the result with regard to the comparison of the stability of collusion with linear prices and two-part tariff can be expected to be robust to changes in many assumptions (Section 1.5.1). We discuss so-called fully nonlinear tariffs in Section 1.5.2 and partial collusion in Section 1.5.3. We summarize our findings and discuss possible limitations of our model in Section 1.6. All proofs are relegated to 1.B.

1.2 Related Literature

We first contribute to the literature on combinable products. Customers' possibility to mix different products was first analyzed by Anderson and Neven (1989) for the case with linear prices and was later adopted by, among others, Gal-Or and Dukes (2003) and Gabszewicz et al. (2004) to analyze media markets. Hoernig and Valletti (2007, 2011) investigate the impact of different pricing schemes by analyzing two-part tariffs and nonlinear pricing in general. As noted in the introduction, the contributions of Anderson and Neven (1989) and Hoernig and Valletti (2007, 2011) show that the scope of mixing crucially depends on the pricing policy. Whereas customers buy from one firm exclusively and, hence, do not mix at all if firms charge fixed prices, customers optimally mix in the sense that they get a perfect match of their preferences if firms charge linear prices. With two-part tariffs and nonlinear pricing in general, some mixing occurs: Only those customers whose preferences are met worst combine products to achieve a better fit. In a static environment, Hoernig and Valletti (2007) stress that the main and robust result is that firm profits are higher as the number of pricing instruments increases.[14] Because mixing benefits customers under competition, the contributions highlight that regulations of the pricing regimes can benefit customers. We add to this strand of literature by analyzing the impact that such regulations can cause in dynamic environments.

We contribute to the literature on the interplay between price discrimination and collusion.[15] Although price discrimination has been an important topic in the antitrust community, the literature on the effects of price discrimination on collusion is rather limited. Two-part tariffs are a classic tool to price-discriminate between customers. Thus, a ban of one of the two price components corresponds to a ban of price discrimination. Gössl and Rasch (2020) are the closest to us in that they use a Hotelling (1929) framework with

[14]This result is different from the related literature on mixed bundling, which uses mix-and-match models and shows that profits are lower with more instruments, particularly in the case in which firms practice mixed bundling compared to the situation in which products are sold separately (see Matutes and Regibeau, 1992).

[15]Although we focus on models with horizontal product differentiation, a related strand of literature has emerged in the context of vertical product differentiation. The process of customers self-selecting into their preferred quality levels is a form of second-degree price discrimination. In such a set-up, Häckner (1994), Symeonidis (1999), and Bos and Marini (2019) analyze the sustainability of collusive agreements. Bos et al. (2020) further explore the role of cartel formation.

linear transport costs and elastic demand to study how a ban of either the linear or fixed price component of a two-part tariff affects the ability of firms to sustain collusion. The underlying mechanisms driving the models are fundamentally different. In the traditional Hotelling (1929) framework, firms compete for the indifferent customer and the location of that customer alone determines from which firms customers buy. By contrast, firms face a more complex demand structure in our model in which each customer can decide whether the customer wants to buy from one firm exclusively or combine products from both firms. This possibility to mix can result in two indifferent customers and a share of customers who granularly adjust their demand at both firms when prices change. The different mechanisms at work correspond to different industries. Given the underlying difference in both set-ups, it is not surprising that the results diverge too: Among other things, Gössl and Rasch (2020) find that a ban on linear prices facilitates collusion, whereas a ban on fixed fees hampers collusion. This is in stark contrast to our finding that both types of ban indeed facilitate collusion.

Both Gössl and Rasch (2020) and our work consider second-degree price discrimination. Liu and Serfes (2007) use a model à la Hotelling (1929) with linear transport costs to analyze the impact of customer-specific information (that is, third-degree price discrimination) on tacit collusion. In their framework, information gives firms the ability to distinguish between different subintervals (market segments) of the linear city. A higher quality of information results in smaller subintervals and, hence, in more market segments. Firms can perfectly identify their customers by their segment and charge prices based on this information. The authors find that collusion becomes less likely as the quality of information (that is, the number of market segments) increases.[16] Liu and Serfes (2007) also show that this result is not clear ex ante because of two opposing channels. The more information firms have, the better they can target their customers. On the one hand, this leads to higher collusive profits and harsher punishment; but on the other hand, deviation profits increase as well. Our results are similar in the sense that multi-part pricing schemes allow firms to better target their customers and make collusion at profit-maximizing prices more difficult. Note, however, that we analyze a situation with second-degree price discrimination because firms cannot distinguish between customers who self-select into their preferred mixing choice.

Whereas Liu and Serfes (2007) assume that firms have perfect information in the sense that they can identify customers of each market segment with certainty, Colombo and Pignataro (2022) and Peiseler et al. (2022) investigate the role of imperfect information about customers' locations on firms' ability to collude. In both set-ups, firms try to distinguish between their loyal customers located in their own turf and disloyal customers located in their

[16] Colombo (2010) analyzes the impact of product differentiation on the stability of collusion in the limiting case in which the number of market segments approaches infinity (delivered pricing/perfect price discrimination). He picks up the prevalent finding in the literature that the relationship between the degree of product differentiation and firms' ability to collude is positive in the context of the Hotelling (1929) framework (see, for example, Chang, 1991, 1992 and Häckner, 1995; Ross, 1992 is a notable exception). He finds no relationship between product differentiation and the likelihood of collusion if firms collude on either customer-specific prices or on the decision whether to apply price discrimination (but not on the price level). In addition, he finds a negative relationship when firms collude on a uniform price. Further contributions that examine the effects of delivered prices on collusion include Jorge and Pires (2008) and Miklós-Thal (2008).

rival's turf. In Colombo and Pignataro (2022), the signal quality determines which share of loyal customers a firm can identify. It is, however, not possible to distinguish unidentified loyal customers from disloyal customers. This is different in Peiseler et al. (2022), where the signal quality determines the ability of firms to identify the brand loyalty of an arbitrary customer. This difference gives rise to different results: Whereas Peiseler et al. (2022) find that a ban of price discrimination hampers collusion for a sufficiently low signal quality, Colombo and Pignataro (2022) find that a ban of price discrimination hampers collusion for large degrees of product differentiation and a sufficiently large signal quality.

Finally, Helfrich and Herweg (2016) also analyze the effect of third-degree price discrimination on collusion in a linear city. The authors find that a ban of price discrimination raises the ability of firms to sustain collusion. The findings of Helfrich and Herweg (2016) are similar to ours in the sense that in our paper, a ban of price discrimination also facilitates collusion. However, in contrast to Helfrich and Herweg (2016), we analyze second-degree instead of third-degree price discrimination and consider various extensions that entail important consequences. For instance, we allow firms to collude on prices below the profit-maximizing prices (partial collusion).[17]

1.3 Model and Previous Findings

We adopt the models of Anderson and Neven (1989) and Hoernig and Valletti (2007) and consider two horizontally differentiated, symmetric firms that are located at the end points of a linear city of unit length (Hotelling, 1929). Fixed and marginal costs are normalized to zero. Firms discount future profits by the common discount factor δ per period. We analyze three different pricing regimes. Firms charge either linear prices $p_{i,L}$ per unit purchased, fixed fees $f_{i,F}$ that are independent of actual usage, or two-part tariffs $(p_{i,T}, f_{i,T})$ that include both a linear and a fixed price. The different scenarios are denoted by the subscripts L, F, and T.

Customers of mass one are uniformly distributed along the line. Each customer has a total demand of zero or one. Customers have a basic valuation of v for each product and incur transport costs. Transport costs reflect the fact that customers' preferences are not fully matched by the firms' products, that is, a customer located at x incurs quadratic transport costs of τx^2 or $\tau(1-x)^2$ when buying only from firm 1 or firm 2.

So far, the application of the transport costs follows the standard logic of the classic framework in Hotelling (1929). The modification of Anderson and Neven (1989) is to allow customers to save on these costs by splitting their demand across the two firms to purchase an optimal individual mix of both products. Let λ (with $0 \leq \lambda \leq 1$) denote the share of overall demand that a customer buys from firm 1; the remaining share of $1-\lambda$ is bought

[17] Horstmann and Krämer (2013) analyze the impact of third-degree price discrimination on collusive outcomes in an experimental setting. In contrast to theoretical predictions, the authors find that third-degree price discrimination leads to significantly higher prices and profits compared to uniform pricing.

from firm 2. Then, mixing leads to transport costs of

$$\tau(\lambda \cdot 0 + (1-\lambda) \cdot 1 - x)^2 = \tau(1-\lambda-x)^2.$$

The first part in brackets $\lambda \cdot 0 + (1-\lambda) \cdot 1$ is the location of the new product, which results from combining the products of both firms. It is a weighted combination of the locations of the individual products and, thus, depends on the decision of the customer about λ. The new location is then evaluated against the location of the customer, x, and the difference in the locations describes the distance that is used to calculate the transport costs.

Another way to think about the transport costs is to rewrite the above expression in the following way

$$\tau\left(\lambda \cdot (0-x) + (1-\lambda) \cdot (1-x)\right)^2.$$

Here, we have a weighted combination of the distances between the customer's location and both firm locations. Note that the distance to the product of firm 1 is (weakly) negative and the distance to the product of firm 2 is (weakly) positive. The important difference between the standard Hotelling (1929) framework and the model of Anderson and Neven (1989) is that we assign an interpretation to the sign of a distance. The idea is that each product comes with advantages and disadvantages (from the viewpoint of each customer). For instance, we can think of two credit cards that differ in their usage benefits for different customers. Both credit cards can be used both at home and abroad, but they differ in their acceptance rate: One credit card can be used more frequently in one country (or region), whereas the other card is more prominently accepted in another. In this case, those customers who mostly shop in their home country and only seldom go abroad can be expected to carry the credit card that is widely accepted at home. By contrast, customers who often go on business trips or on holiday abroad may want to contemplate whether to carry both cards to gain more flexibility. The model of Anderson and Neven (1989) captures this by allowing negative and positive distances to offset each other. Because both products are located at the extremes, λ can be chosen to achieve an arbitrary location of the new product on the line $[0, 1]$. If the customer wants to fully save on the transport costs, the customer has to select λ such that the location of the new product equals exactly his/her own location.

Even though (complete) savings on transport costs are possible, transport costs usually affect prices because they are a measure of firms' market power. Typically, the focus in models with horizontal product differentiation is on situations in which the market is covered, that is, the basic valuation v must be relatively large compared to the transport costs, so that the utility in equilibrium is not negative and customers disregard the outside option not to buy. We also focus on this case and make the following assumption:

Assumption 1. *Transport costs are not too high relative to the basic valuation from buying, that is,* $0 < \tau \leq 4v/5$.

Customer face three different potential utilities, depending on where they buy. For the case of two-part tariffs, the following utility functions U refer to the cases in which the customer located at x buys exclusively from firm 1, buys exclusively from firm 2, or combines the products of both firms (dropping subscripts for the different pricing scenarios for now):

$$\begin{aligned} U_1(x) &= v - f_1 - p_1 - \tau x^2, \\ U_2(x) &= v - f_2 - p_2 - \tau(1-x)^2, \\ U_m(x) &= v - f_1 - f_2 - \lambda p_1 - (1-\lambda)p_2 - \tau\left(1-\lambda-x\right)^2. \end{aligned}$$

A mixing customer will optimally choose share λ to maximize utility depending on the location, that is,

$$\frac{\partial U_m}{\partial \lambda} = 0 \quad \Leftrightarrow \quad \lambda\left(x\right) = 1 - x - \frac{p_1 - p_2}{2\tau}. \tag{1.1}$$[18]

Given the decision about the optimal share, we derive the customer who is indifferent between buying exclusively from firm 1 and mixing. Denote this customer's location by $\underline{x}$. Similarly, denote the location of the customer who is indifferent between mixing and buying exclusively from firm 2 by $\overline{x}$. Assume that $0 \leq \underline{x} \leq \overline{x} \leq 1$ holds. Then, the locations of the indifferent customers are given by

$$\begin{aligned} U_1\left(\underline{x}\right) = U_m\left(\underline{x}\right) &\quad \Leftrightarrow \quad \underline{x} = \sqrt{\frac{f_2}{\tau}} - \frac{p_1 - p_2}{2\tau}, \\ U_m\left(\overline{x}\right) = U_2\left(\overline{x}\right) &\quad \Leftrightarrow \quad \overline{x} = 1 - \sqrt{\frac{f_1}{\tau}} - \frac{p_1 - p_2}{2\tau}. \end{aligned}$$

For $0 \leq \underline{x} \leq \overline{x} \leq 1$, the profit function of each firm consists of three parts:

$$\pi_1\left(f_1, p_1; f_2, p_2\right) = f_1 \overline{x} + p_1 \underline{x} + p_1 \int_{\underline{x}}^{\overline{x}} \lambda\left(x\right) dx,$$

$$\pi_2\left(f_1, p_1; f_2, p_2\right) = f_2\left(1-\underline{x}\right) + p_2\left(1-\overline{x}\right) + p_2 \int_{\underline{x}}^{\overline{x}} \left(1-\lambda\left(x\right)\right) dx.$$

The first part consists of the fixed fee that is paid by both loyal and mixing customers. The second and third parts quantify the linear payments. Whereas loyal customers buy exclusively from one firm and, hence, pay the full linear price (second part), mixing customers buy their optimal shares that depend on their locations (third part).

If $\underline{x} > \overline{x}$, customers never mix. In this case, we are back in the classic Hotelling (1929) game with quadratic transport costs as analyzed by d'Aspremont et al. (1979).

Further note that the cases of linear prices and fixed fees are special cases of two-part

[18] Note that the second-order condition is satisfied, that is, $\partial^2 U_m / \partial \lambda^2 = -2\tau < 0$.

tariffs. All formulas for these cases follow immediately from setting the respective price component equal to zero.

Before we turn to analyzing the cases of collusion and deviation, we briefly recap the results in the competitive scenarios (denoted by an asterisk) that are derived in Anderson and Neven (1989) and Hoernig and Valletti (2007) in the static one-shot game:

Recap 1. *Competitive prices are given by*

$$p_L^* = f_F^* = \tau \qquad \text{(linear and fixed prices)},$$
$$(f_T^*, p_T^*) = \left(\frac{\left(7 - 3\sqrt{5}\right)\tau}{2}, \frac{\left(3\sqrt{5} - 5\right)\tau}{2} \right) \qquad \text{(two-part tariffs)}.$$

Competitive profits amount to

$$\pi_L^* = \pi_F^* = \frac{\tau}{2} \qquad \text{(linear and fixed prices)},$$
$$\pi_T^* = \frac{\left(13\sqrt{5} - 27\right)\tau}{4} \qquad \text{(two-part tariffs)}.$$

Customer surplus and welfare is given by

$$CS_L^* = v - \tau \quad \text{and} \quad W_L^* = v \qquad \text{(linear prices)},$$
$$CS_F^* = v - \frac{13\tau}{12} \quad \text{and} \quad W_F^* = v - \frac{\tau}{12} \qquad \text{(fixed prices)},$$
$$CS_T^* = v - \frac{\tau(23\sqrt{5} - 45)}{6} \quad \text{and} \quad W_T^* = v - \frac{\tau\left(3 - \sqrt{5}\right)^3}{12} \qquad \text{(two-part tariffs)}.$$

Although linear and fixed prices (and profits) are the same, they lead to remarkably different market outcomes. In the case of linear prices, all customers buy their optimal mix, such that transport costs are zero. As a result, welfare is maximized. By contrast, customers do not mix in the case of fixed fees, and, hence, the outcome is the same as in the classic game analyzed by d'Aspremont et al. (1979). The total welfare loss due to transport costs is $\tau/12$, and customers are worse off.

Two-part tariffs enable firms to segment customers. By setting a strictly positive fixed fee, firms extract additional surplus from their loyal customers. In turn, they are willing to lose extremely disloyal customers who are located close to their competitor and would only buy a small share anyway. In contrast to pure fixed pricing, the fixed-price component is lower, so that it is beneficial for customers around the center of the linear city to mix. These customers face higher transport costs compared to the loyal customers and, hence, are willing to pay the fixed fee twice to save on these costs.

Because some customers mix and others do not, the welfare loss through transport costs is lower than in the case of fixed fees only, but larger than in the case of linear pricing. Although the additional price component in the case of a two-part tariff allows firms to extract additional surplus from loyal customers, the decline in the overall transport costs is so large compared to the case of fixed pricing, such that customers benefit overall, that is, customer surplus is larger under two-part tariffs than under fixed fees only. However, customer surplus is largest under linear prices because customers face zero transport costs and firms extract less surplus.

1.4 Collusion at Maximum Prices

In our main analysis, we focus on three scenarios that relate to the situations discussed in the Introduction: (i) Firms can only set linear prices due to a regulation that does not allow the use (or setting) of fixed fees; (ii) firms can only choose fixed fees because linear prices are banned (or regulated); and (iii) firms are unrestricted in their price-setting and may choose linear and fixed prices (no ban or regulation).

Collusive Strategy

Throughout our analysis, we adopt the critical discount factor as a measure for the likelihood of (full) collusion. To this end, we focus on the standard grim-trigger strategies defined by Friedman (1971).[19] Denote the profits in the cases of collusion and deviation by π^c and π^d. Then, collusion is profitable as long as the discounted profits from collusion are higher than those from deviation and the ensuing punishment phase, that is,

$$\sum_{t=0}^{\infty} \delta^t \pi^c \geq \pi^d + \sum_{t=1}^{\infty} \delta^t \pi^*.$$

Hence, collusion can be sustained for any discount factor larger than the critical discount factor defined as

$$\overline{\delta} := \frac{\pi^d - \pi^c}{\pi^d - \pi^*}. \tag{1.2}$$

As a consequence, collusion is facilitated when the critical discount factor decreases because firms can sustain collusion for a larger range of discount factors.

Because we already discussed the competitive profits, we directly turn to the remaining cases of collusion and deviation.

Collusive outcomes

The following lemma summarizes the collusive outcomes:

[19] Apart from grim-trigger strategies, where firms return to Nash pricing after collusion is detected, other punishment strategies are possible. Most notable are optimal punishment strategies following the seminal work in Abreu (1986, 1988) and Abreu et al. (1986). In the context of the Hotelling (1929) framework, there is tentative evidence that optimal punishment strategies lead to similar results compared to grim-trigger strategies. For example, Häckner (1996) uses a standard set-up with quadratic transport costs and symmetric firms and shows that the impact of product differentiation on collusive prices is qualitatively similar with optimal punishment compared to the results achieved by Chang (1991) with grim-trigger strategies. Furthermore, in the context of price discrimination, Liu and Serfes (2007) report that their main result that is derived in a Hotelling (1929) set-up with linear transport costs is also robust when they move from grim-trigger to stick-and-carrot punishments. Because optimal punishment strategies come at the expense of less tractable models, we stick to grim-trigger strategies.

Lemma 1. *Collusive prices are given by*

$$p_L^c = v \qquad \textit{(linear prices)},$$
$$f_F^c = v - \frac{\tau}{4} \qquad \textit{(fixed prices)},$$
$$(p_T^c, f_T^c) = (v, 0) \qquad \textit{(two-part tariffs)}.$$

Collusive profits amount to

$$\pi_L^c = \pi_T^c = \frac{v}{2} \qquad \textit{(linear prices and two-part tariffs)},$$
$$\pi_F^c = \frac{v}{2} - \frac{\tau}{8} \qquad \textit{(fixed prices)}.$$

Customer surplus and welfare is given by

$$CS_L^* = 0 \quad \textit{and} \quad W_L^* = v \qquad \textit{(linear prices)},$$
$$CS_F^* = \frac{\tau}{6} \quad \textit{and} \quad W_F^* = v - \frac{\tau}{12} \qquad \textit{(fixed prices)},$$
$$CS_T^* = 0 \quad \textit{and} \quad W_T^* = v \qquad \textit{(two-part tariffs)}.$$

With linear prices, firms charge prices that are equal to the basic valuation. The reason for this behavior can be explained by two effects. First, by setting equal prices, all customers buy their optimal mix and, hence, do not incur transport costs. As a consequence, firms maximize customers' utility. Second, by setting the price level to the basic valuation, firms fully extract the maximized utility, such that producer surplus equals the maximized welfare and customer surplus is zero.

Because firms gain the highest possible profits with linear prices, firms cannot take advantage of the additional fixed fee in the case of two-part tariffs. A strictly positive fixed fee would lead to a share of customers who do not mix and, hence, suffer from a loss in utility due to strictly positive transport costs. As a consequence, firms set the fixed component equal to zero and charge the linear price equal to the basic valuation. Again, welfare is maximized and equals producer surplus; customer surplus is zero.

Finally, customers do not mix with fixed-fee pricing, and firms are in the same situation as in the classic set-up analyzed by Chang (1991). They set the optimal fixed fees, such that the indifferent customer at the center is indifferent between buying and not buying. As a result, all customers incur strictly positive transport costs, which leads to a welfare loss of $\tau/12$. However, in contrast to the other two pricing environments, customer surplus is strictly positive.

Deviation

Based on the collusive outcomes, we determine optimal prices and profits of a deviating firm:

Lemma 2. *Define $A := \sqrt{v^2 - 4v\tau + 28\tau^2}$. Optimal deviation prices are given by*

$$p_L^d = \frac{2v - 4\tau + A}{3} \qquad \textit{(linear prices)},$$

$$f_F^d = \begin{cases} v - \frac{5\tau}{4} & \textit{if } 0 < \tau \leq \frac{4v}{13} \\ \frac{v}{2} + \frac{3\tau}{8} & \textit{if } \frac{4v}{13} < \tau \leq \frac{4v}{5} \end{cases} \qquad \textit{(fixed prices)},$$

$$\left(f_T^d, p_T^d\right) = (\tau, v - 2\tau) \quad \textit{if } 0 < \tau \leq \frac{v}{4} \qquad \textit{(two-part tariffs)}.$$

For $\tau > v/4$, we use

$$\left(f_T^d, p_T^d\right) = \left(\frac{(v-\tau)^2}{9\tau}, \frac{v+2\tau}{3}\right)$$

to calculate a bound on the profit for two-part tariffs. Then, optimal deviation profits amount to

$$\pi_L^d = \frac{(-2v + 4\tau - A)(v^2 - vA - 4v\tau + 2\tau A - 20\tau^2)}{108\tau^2} \qquad \textit{(linear prices)},$$

$$\pi_F^d = \begin{cases} v - \frac{5\tau}{4} & \textit{if } 0 < \tau \leq \frac{4v}{13} \\ \frac{(4v+3\tau)^2}{128\tau} & \textit{if } \frac{4v}{13} < \tau \leq \frac{4v}{5} \end{cases} \qquad \textit{(fixed prices)},$$

$$\pi_T^d \begin{cases} = v - \tau & \textit{if } 0 < \tau \leq \frac{v}{4} \\ \geq \frac{-v^3 + 12v^2\tau + 6v\tau^2 + 10\tau^3}{54\tau^2} & \textit{if } \frac{v}{4} < \tau \leq \frac{4v}{5} \end{cases} \qquad \textit{(two-part tariffs)}.$$

Consider the case of linear prices first. The deviating firm sets its price, such that it serves a loyal customer base exclusively and sells shares of its product to disloyal customers. Thus, it never monopolizes the market, but leaves its competitor always with a strictly positive market share.

In contrast to linear prices, customers do not mix neither under competition nor under collusion in the case of fixed fees. This enables a deviating firm to monopolize the whole market if product differentiation is sufficiently low (that is, $0 < \tau \leq 4v/13$). The monopolization requires that the firm compensates the farthest customer for the transport costs. With an increasing degree of product differentiation (that is, $\tau > 4v/13$), this compensation becomes unattractive, so that the deviating firm leaves its competitor with a strictly positive market share.

The case of two-part tariffs poses some challenges. Although the mathematical problem is well-defined and it is possible to write down the optimization problem of the deviating firm, it is difficult to derive closed-form solutions for all cases because of the highly nonlinear nature of some equations. In principle, the deviating firm can set its prices in three different ways. First, it can set its prices to monopolize the market. Second, it can cover the entire market, but leave its rival with a strictly positive market share, that is, some customers close to the rival's location buy from both firms. The third option is to set prices, such that it does not cover the entire market. In this case, some customers who are located close to the rival's location choose the outside option and do not buy. The reason is that the rival charges a linear price of v which equals the basic utility. This means that after paying the price, a customer would have zero utility. All customers except the marginal customer at $x = 1$ would also have to pay transport costs on top, resulting in a negative utility. Therefore, these customers refrain from buying from any firm. In addition, there would be two other groups of customers, those who are located near the deviator's location and buy exclusively from that firm and those who are located towards the center and buy from both firms.

It is possible to derive closed-form solutions for the prices and profits in the first two cases. These two cases allow us to derive a lower bound for the profit. In addition, we will discuss in Section 1.5.2 that the profit in the case of fully nonlinear tariffs can serve as an upper bound. This allows us to provide an even more precise characterization for the case $\tau \leq v/4$ because the lower bound equals the upper bound. This allows us to conclude that in this case, the deviating firm monopolizes the market and serves all customers.

Table 1.1 provides a summary of our results so far to help to understand the mechanism behind Proposition 1 below.

Critical Discount Factors

The following lemma summarizes the critical discount factors that result from inserting the outcomes for the cases of collusion, deviation, and competition into Expression (1.2) for the critical discount factor. For the case of two-part tariffs, we use the lower bound on the deviation profit to obtain a lower bound for the corresponding critical discount factor. This makes sense because a larger deviation profit ceteris paribus renders deviation more attractive and thus leads to an increase in the critical discount factor.

Lemma 3. *Define $B := v^3 - v^2A - 6v^2\tau + 4v\tau A - 28\tau^2 A$. The critical discount factors are given by*

$$\bar{\delta}_L = \frac{B - 6v\tau^2 + 136\tau^3}{B - 60v\tau^2 + 190\tau^3} \qquad \textit{(linear prices)},$$

$$\bar{\delta}_F = \begin{cases} \frac{4v-9\tau}{2(4v-7\tau)} & \textit{if } 0 < \tau \leq \frac{4v}{13} \\ \frac{4v-5\tau}{4v+11\tau} & \textit{if } \frac{4v}{13} < \tau \leq \frac{4v}{5} \end{cases} \qquad \textit{(fixed prices)},$$

$$\bar{\delta}_T \begin{cases} = \frac{2(v-2\tau)}{4v+23\tau-13\sqrt{5}\tau} & \textit{if } 0 < \tau \leq \frac{v}{4} \\ \geq \frac{2(10\tau-v)(v-\tau)^2}{-2v^3+24v^2\tau+12v\tau^2+749\tau^3-351\sqrt{5}\tau^3} & \textit{if } \frac{v}{4} < \tau \leq \frac{4v}{5} \end{cases} \qquad \textit{(two-part tariff)}.$$

The comparison of the critical discount factors reveals how the different pricing regimes affect the sustainability of (full) collusion. Figure 1.1 plots the critical discount factors for the case in which $v = 1$ against the degree of product differentiation. The two bottom lines refer to the cases of linear prices and fixed fees. The shaded area depicts the area where the critical discount factor in the case of two-part tariffs is located. The lower bound of this area is the bound derived in Lemma 3. This bound is sufficient to compare the three

Panel A: Competitive profits				
Type	Linear price	Fixed fee	Profit	Customer surplus
Linear prices	$p_L^* = \tau$	–	$\pi_L^* = \frac{\tau}{2}$	$CS_L^* = v - \tau$
Fixed fees	–	$f_F^* = \tau$	$\pi_F^* = \frac{\tau}{2}$	$CS_F^* = v - \frac{13\tau}{12}$
Two-part tariffs	$p_T^* = \frac{\tau(3\sqrt{5}-5)}{2}$	$f_T^* = \frac{\tau(7-3\sqrt{5})}{2}$	$\pi_T^* = \frac{\tau(13\sqrt{5}-27)}{4}$	$CS_T^* = v - \frac{\tau(23\sqrt{5}-45)}{6}$
Relevant comparisons for the understanding of Proposition 1: $\pi_T^* > \pi_L^* > \pi_F^*$				

Panel B: Collusive profits				
Type	Linear price	Fixed fee	Profit	Customer surplus
Linear prices	$p_L^c = v$	–	$\pi_L^c = \frac{v}{2}$	$CS_L^c = 0$
Fixed fees	–	$f_F^c = v - \frac{\tau}{4}$	$\pi_F^c = \frac{v}{2} - \frac{\tau}{8}$	$CS_F^c = \frac{\tau}{6}$
Two-part tariffs	$p_T^c = v$	$f_T^c = 0$	$\pi_T^c = \frac{v}{2}$	$CS_T^c = 0$
Relevant comparisons for the understanding of Proposition 1: $\pi_L^c = \pi_T^c > \pi_F^c$				

Panel C: Deviation profits			
Type	Linear price	Fixed fee	Profit
Linear prices	$p_L^d = \frac{2v-4\tau+A}{3}$	–	$\pi_L^d = \frac{(-2v+4\tau-A)(v^2-vA-4v\tau+2\tau A-20\tau^2)}{108\tau^2}$
Fixed fees	–	$f_F^d = v - \frac{5\tau}{4}$ if $0 < \tau \leq \frac{4v}{13}$ $f_F^d = \frac{v}{2} + \frac{3\tau}{8}$ otherwise	$\pi_F^d = v - \frac{5\tau}{4}$ if $0 < \tau \leq \frac{4v}{13}$ $\pi_F^d = \frac{(4v+3\tau)^2}{128\tau}$ otherwise
Two-part tariffs	$p_T^d = v - 2\tau$ if $0 < \tau \leq \frac{v}{4}$ p_T^d undefined otherwise	$f_T^d = \tau$ if $0 < \tau \leq \frac{v}{4}$ f_T^d undefined otherwise	$\pi_T^d = v - \tau$ if $0 < \tau \leq \frac{v}{4}$ $\pi_T^d \geq \frac{-v^3+12v^2\tau+6v\tau^2+10\tau^3}{54\tau^2}$ otherwise
Relevant comparison for the understanding of Proposition 1: $\pi_T^d > \pi_F^d$			

Table 1.1: Summary of results.

critical discount factors derived above. The upper bound $\overline{\delta}_N$ will be of interest later when we discuss fully nonlinear tariffs in Section 1.5.2.

Figure 1.1 shows that collusion is most difficult to sustain under two-part tariffs and the comparison for linear prices and fixed fees is ambiguous. The following proposition states that this result is independent of the basic valuation v:

Proposition 1. *A comparison of the critical discount factors gives:*

1. *Collusion is less likely under two-part tariffs than under linear prices and fixed fees, that is,* $\overline{\delta}_T \geq \overline{\delta}_L$ *and* $\overline{\delta}_T \geq \overline{\delta}_F$.

2. *Collusion is less likely under fixed prices (linear prices) than under linear prices (fixed prices) for relatively low (high) degrees of product differentiation, that is,* $\tau^{(1)}$ *exists* ($\approx 0.5293500486v$), *such that* $\overline{\delta}_L(\tau) < \overline{\delta}_F(\tau)$ *for* $\tau < \tau^{(1)}$ *and* $\overline{\delta}_L(\tau) > \overline{\delta}_F(\tau)$ *for* $\tau > \tau^{(1)}$.

To understand this result, we compare the different profits that determine the critical discount factors (see Table 1.1). Comparing competitive and deviation profits, we find that competition is less harsh and deviation is more profitable with two-part tariffs than with linear or fixed prices. This makes it harder to sustain collusion in the case of two-part tariffs. With linear prices, the collusive profits are identical to those with two-part tariffs, and, hence, the critical discount factor is lower. For the case of fixed fees, the collusive profits are lower, which means that there is an opposing effect that makes it more difficult to sustain collusion. However, as Proposition 1 states, this destabilizing effect is strictly dominated by the aforementioned facilitating effects.

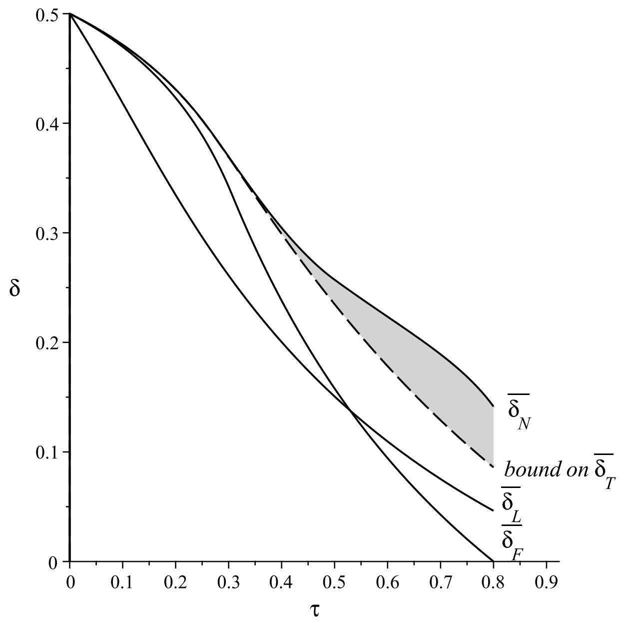


Figure 1.1: Comparison of the critical discount factors in the three scenarios (for $v = 1$ and $0 < \tau \leq 4/5$).

Comparing fixed to linear prices, we point out that both pricing schemes yield the same competitive profits. However, linear prices have two opposing effects with regard to the likelihood of collusion: On the one hand, collusive profits are higher, which makes collusion more attractive for firms. On the other hand, higher collusive prices make deviation easier, which results in higher deviation profits. For moderately differentiated products (that is, $\tau < \tau^{(1)}$), deviation proves attractive in the case of fixed fees, which means that the first effect dominates, and collusion is easier to sustain with linear prices. The reason is that with fixed fees, customers do not mix and, hence, a deviating firm has to compensate the customers for their transport costs. This compensation is relatively cheap when the degree of product differentiation is low. When product differentiation increases, compensating the customers becomes more costly and the second effect becomes more important (that is, for $\tau > \tau^{(1)}$).

1.5 Robustness Checks

In this section, we will investigate to what extent our results rely on the assumptions imposed in the two previous sections. In doing so, we will first focus on the structure of firms' pricing schedules and extend our model to capture more flexible pricing structures (so-called "fully nonlinear tariffs"). Second, we will relax the assumption that firms collude only on profit-maximizing prices. Before we discuss these two aspects in more detail, however, a more general remark with regard to the comparison of the likelihood of collusion under linear prices and two-part tariffs seems to be in order.

1.5.1 An Initial Remark on the Likelihood of Collusion under Linear Prices and Two-Part Tariffs

In our baseline analysis, we assume that firms use grim-trigger strategies, that is, a firm will set the competitive price in all future periods once it observes that its competitor deviates from a collusive agreement. Although grim-trigger strategies are themselves an assumption, their application may shed additional light on the comparison of the stability of collusive agreements under linear prices and two-part tariffs.

As shown in the previous section, the linear price component is an important tool for colluding firms to extract customers' surplus. Its importance is so large that colluding firms use only linear prices and dismiss the possibility to set a strictly positive fixed fee even if they may use two-part tariffs in general. It is easy to think of many modifications to our model that leave this result intact. For instance, if customers are not uniformly distributed, but are distributed according to any other distribution, the decision to focus only on linear prices is still optimal in the sense that it maximizes (joint) collusive profits. The reason is that by setting equal linear prices, firms allow each customer—regardless of the location—to obtain the optimal combination, which in turn maximizes welfare. At the same time, firms are able to extract the entire utility from each customer, so that they obtain the highest possible profit.

This has in turn implications for the optimal prices of a deviating firm. The deviation profits must be (weakly) larger if a firm has access to an additional pricing instrument. If the additional instrument is not useful, the firm will simply not use it and set it equal to zero. Therefore, deviation profits must be (weakly) larger with two-part tariffs than with linear prices whenever it is optimal for colluding firms to focus only on linear prices.

We believe that the above argumentation that collusive profits are the same under both pricing regimes and deviation profits are (weakly) larger under two-part tariffs applies to various modifications. With grim-trigger strategies, the critical discount factor depends on three different types of profits. The above results for the collusive and deviation profits point to lower stability of collusive agreements with two-part tariffs than with linear prices. The only opposing force that could lead to a lower likelihood of collusion under linear prices could stem from the competitive profits. However, the intuition behind how firms use the fixed fee component in the case of two-part tariffs suggests that competitive profits are unlikely to be higher under linear prices than under two-part tariffs. The reason is that by setting a moderate fixed fee, firms can exploit loyal customers. Customers located in the middle of the interval can save on their transport costs by buying from both firms, and as long as the magnitude of the fixed fee is moderate (relative to the transport costs), these customers will prefer to buy from both firms. The customers a firm loses are those who are located close to the competitor and would buy only small shares under linear prices anyway. Therefore, the fixed fee component appears to be a beneficial additional tool that is likely to increase rather than decrease competitive profits.

Following this line of reasoning, we conjecture that our result that collusion is more stable under linear prices than under two-part tariffs is robust to many modifications (for instance, to different distributional assumptions).

1.5.2 Fully Nonlinear Pricing Schemes

We now turn our focus to the structure of tariffs used in the baseline analysis. So far, we have focused on three pricing schemes that are common in the real world and often used in economic analyses. However, this limitation may raise the concern of whether our results extend to other pricing schemes. In particular, it is of interest how larger flexibility in the pricing structure might affect our results. To address this issue, we introduce fully nonlinear tariffs along the lines of Hoernig and Valletti (2011).

A fully nonlinear tariff is a function $T : [0,1] \to \mathbb{R}$ that assigns a price $p\,(q)$ to each possible quantity $q \in [0,1]$. This tariff structure incorporates all other possible pricing schemes. For example, a linear price p is equal to $T(q) = pq$, a fixed fee f to $T(q) = f$ and a two-part tariff (p, f) to $T(q) = pq + f$.

Hoernig and Valletti (2011) base their analysis on two sets of assumptions that we adopt as well. First, they require the tariffs to be differentiable on $(0, 1)$, but do not impose any assumption on the endpoints of the interval. Second, they allow each customer to buy more than one unit in total and restrict the information set of the retailers. More precisely, the authors assume that each firm can only observe the quantity that a customer buys from itself, but not the quantity bought from its competitor or the total quantity purchased. Thus, in combination with the assumption on customer behavior, it cannot expect that a customer who buys a share of λ from itself will automatically buy $1-\lambda$ from its competitor. Although each customer ends up buying only one unit in total, customers may buy more than one unit if it is cheaper (and throw away the additional quantity purchased). Therefore, this assumption rules out various forms of contractual relationship, including exclusive dealing, quantity forcing, and market-share contracts.

Based on these assumptions, Hoernig and Valletti (2011) analyze competitive outcomes and prove the existence of a unique Nash equilibrium:

Recap 2. *If firms compete in fully nonlinear tariffs, they set*

$$T_N^*(q) = \tau \cdot q + \frac{\tau \cdot q \cdot (1-q)}{3}$$

and earn profits of $\pi_N^* = 14\tau/27$.

With this result in hand, we only need to calculate the collusive and deviation profits to derive the critical discount factor. Based on our previous analysis, it is straightforward to derive the profit-maximizing collusive prices. As argued before, a linear collusive price of v leads to the largest possible profit because it maximizes welfare and allows firms to extract all rents, so that customer surplus is zero. Linear prices can be expressed as fully nonlinear tariffs, which leads to the following result:

Corollary 1. *Profit-maximizing collusive tariffs are given by* $T(q) = vq$, *and firms earn profits of* $\pi_N^c = v/2$.

This leaves us with the question of the optimal deviation strategy. The following lemma specifies the profit of a deviating firm and the resulting critical discount factor:

Lemma 4. *With fully nonlinear tariffs, the optimal deviation profits are given by:*

$$\pi_N^d = \begin{cases} v - \tau & \text{if } \tau \leq \frac{v}{4} \\ \frac{16\tau^3 + 12\tau v^2 - v^3}{48\tau^2} & \text{if } \frac{v}{4} < \tau < \frac{(3-\sqrt{6})v}{2} \\ \frac{99\sqrt{6}v^3 - 243v^3 + 8\tau^3 + 6v^2\tau}{24\tau^2} & \text{if } \tau \geq \frac{(3-\sqrt{6})v}{2}. \end{cases}$$

Let $C := -360\left(\frac{4\tau}{5} + v\right)\sqrt{2}\,v\sqrt{\tau^2 v\,(2\tau + v)} + 513\tau v^3 + 864\tau^2 v^2$. *The critical discount factor is*

$$\bar{\delta}_N = \begin{cases} \frac{54\tau - 27v}{82\tau - 54v} & \text{if } \tau \leq \frac{v}{4} \\ \frac{144\tau^3 + 108\tau v^2 - 216\tau^2 v - 9v^3}{108\tau v^2 - 80\tau^3 - 9v^3} & \text{if } \frac{v}{4} < \tau < \frac{v}{2} \\ \frac{C + 216\tau^3 v}{C + 224\tau^4} & \text{if } \tau \geq \frac{v}{2}. \end{cases}$$

As in the case of two-part tariffs, three cases characterize the optimal deviation strategy. If product differentiation is sufficiently low, a deviating firm prefers to monopolize the market because it is rather cheap to compensate customers located close to its competitor for their transport costs. With an increasing degree of product differentiation, this compensation becomes too costly. Therefore, an intermediate level of product differentiation enables customers close to the competitor to buy from both firms, thereby reducing their transport costs. At a high level of product differentiation, the deviating firm completely abandons these customers.

In Figure 1.1, the line for the critical discount factor in the case of fully nonlinear tariffs reveals that the critical discount factor is largest among all four pricing schemes. The following proposition states that this finding is independent of the parameterization used for the figure:

Proposition 2. *The critical discount factor is largest in the case of fully nonlinear tariffs, that is,* $\overline{\delta}_N > \overline{\delta}_t$ *with* $t \in \{L, F, T\}$.

Our previous analysis showed that collusion is least likely under two-part tariffs compared to linear prices or fixed fees. To understand Proposition 2, it is therefore sufficient to compare fully nonlinear tariffs to two-part tariffs. Our analysis reveals that under both pricing regimes, firms use the same collusive prices and obtain the same profits. A deviating firm can therefore only benefit from the more flexible pricing structure in the case of fully nonlinear tariffs. To see this, note that this pricing structure also incorporates two-part tariffs. Thus, a deviating firm could simply use the optimal two-part tariff if this was the best strategy overall. Our analysis indicates that this is indeed true for low degrees of product differentiation because it monopolizes the entire market under both pricing regimes and the costs of compensating the customer located at $x = 1$ are independent of the pricing schemes analyzed. However, once the deviating firm does not want to monopolize the entire market, the nonlinear tariffs make it easier to compensate customers located far away. The reason is that some customers combine products of both firms and, depending on their location, can reduce their transport costs to different extents. The fully nonlinear tariff allows the deviating firm to better adjust to these differences in utility levels and to better target individual customers (and, thus, better extract each customer's surplus). In summary, deviating becomes more profitable, which in turn destabilizes collusion. Moreover, we know that the competitive profits are larger under fully nonlinear tariffs than under two-part tariffs, which also leads to lower stability of collusion. We can thus conclude that the insights from the baseline set-up hold under fully nonlinear tariffs.

1.5.3 Partial Collusion

In our baseline analysis, we considered firms' ability to collude on profit-maximizing prices. In the context of the Hotelling (1929) framework, Chang (1991) shows that if collusion on these prices is not sustainable, firms can still collude by setting prices below the profit-maximizing collusive but above the competitive prices.[20] We refer to this behavior as partial collusion. In this section, we adopt idea of Chang (1991) and discuss whether and how our results change when firms collude on prices different from the profit-maximizing prices.

We start with the case of fixed fees. We already know from the previous analysis that firms do not have an incentive to set prices so low that (some) customers buy from both

[20] In Chang (1991), firms have only access to one price instrument (a per-unit price). With two-part tariffs, the framing that firms collude on "prices below the profit-maximizing prices" might be wrong. For instance, a firm might find it beneficial to decrease the linear price and to increase the fixed fee to stabilize a collusive agreement. Therefore, when we refer to two-part tariffs, we will talk about prices that are *different* from the profit-maximizing prices.

firms. So far, this was true for all cases (competition, collusion, and deviation). Given this result, it is not surprising that the same is true for partial collusion. If all customers decide to buy exclusively from one firm, we face the same game as in Chang (1991) and get the same result:

Corollary 2. *Assume that firms' pricing is restricted to fixed fees only and that collusion at profit-maximizing prices is not sustainable (that is, $\delta < \overline{\delta}_F$). Optimal collusive prices are given by*

$$f_F^c(\delta) = \begin{cases} \frac{\tau(2-3\delta)}{1-2\delta} & \text{if } \delta > \frac{1}{3} \\ \frac{\tau(1+3\delta)}{1-\delta} & \text{if } \delta \leq \frac{1}{3}, \end{cases}$$

and the profit is $\pi_F^c(\delta) = f_F^c(\delta)/2$.

Turning to linear prices, our previous analysis has shown that it is never profitable for the deviating firm to monopolize the market. This also remains true for the case of partial collusion because the deviating firm now faces an even lower collusive price set by the rival. Thus, it always leaves its rival with a strictly positive market share. We obtain the following characterization of the optimal linear prices:

Lemma 5. *Assume that firms' pricing is restricted to linear prices only and that collusion at profit-maximizing prices is not sustainable (that is, $\delta < \overline{\delta}_L$). The optimal collusive price $p_L^c(\delta)$ is implicitly defined by*

$$\delta = \frac{(-28\tau^2 + 4\tau p_L^c - {p_L^c}^2)\sqrt{28\tau^2 - 4\tau p_L^c + {p_L^c}^2} + 136\tau^3 - 6p_L^c\tau^2 - 6\tau\, {p_L^c}^2 + {p_L^c}^3}{(-28\tau^2 + 4\tau p_L^c - {p_L^c}^2)\sqrt{28\tau^2 - 4\tau p_L^c + {p_L^c}^2} + 190\tau^3 - 60p_L^c\tau^2 - 6\tau\, {p_L^c}^2 + {p_L^c}^3},$$

and the profit is $\pi_L^c(\delta) = p_L^c(\delta)/2$. The optimal collusive price $p_L^c(\delta)$ is strictly decreasing in the discount factor $\delta > 0$.

With the results for linear and fixed prices in hand, we can compare the profits from both pricing schemes for any given discount factor:

Proposition 3. *If firms collude, firms always gain larger profits with linear prices than with fixed fees, that is, $\pi_L^c(\delta) > \pi_F^c(\delta)$ for all $\delta > 0$.*

To understand this result, it is useful to focus on the impact of product differentiation on colluding firms' price setting. We start with a relatively small degree of product differentiation. We know from the previous analysis that in this case, collusion on profit-maximizing prices is sustainable for a larger range of discount factors with linear prices than with fixed prices. This is because with fixed prices, it is cheap for the deviating firm to monopolize the entire market, which results in relatively large deviation profits. When we extend this idea to the current set-up with partial collusion, this likely means that with fixed fees, firms must lower the collusive fee substantially in order to render deviation unprofitable. Combined with the fact that even if firms are able to collude on profit-maximizing prices, firms obtain larger collusive profits with linear prices than with fixed fees, it is not surprising that firms gain larger (partially) collusive profits with linear prices.

Turning the focus to relatively large degrees of product differentiation, deviation becomes less attractive in the case of fixed fees. The reason is that the deviating firm has to compensate the indifferent customer for the transport costs to induce this customer to switch, and this compensation comes increasingly costly as the degree of product differentiation increases. This also leads to the result that collusion on profit-maximizing prices is sustainable for a larger range of the discount factor with fixed fees than with linear prices. If deviation becomes less attractive with fixed prices, the extent to which (partially) colluding firms have to lower the collusive price to render deviation unprofitable decreases as well. This positively affects collusive profits with fixed prices. There is, however, an opposing channel that is caused by a large degree of product differentiation. As noted above, even if firms are able to collude on profit-maximizing prices, firms obtain larger collusive profits with linear prices than with fixed fees. This difference in profits increases in the degree of product differentiation. The reason is that with symmetric linear prices, customers always buy their optimal mix, so that transport costs are zero. Thus, the surplus that colluding firms can extract is independent of the degree of product differentiation. By contrast, with fixed fees, customers do not mix and, hence, the total surplus that firms can extract decreases in the degree of product differentiation. Proposition 3 shows that the first positive effect of higher transport costs does not outweigh this second effect.

In Section 1.4, where we consider collusion on profit-maximizing prices, we use a bound on the critical discount factor for two-part tariffs because it is difficult to derive a closed-form solution. Not surprisingly, we run into similar problems with the analysis of partial collusion. Therefore, we base the remaining investigation of partial collusion on numerical simulations.[21] In each simulation, we fix the set of exogenous parameters (that is, the basic valuation and the transport cost parameter) and identify the firms' optimal collusive behavior for discount factors between 0.01 and 0.5. For discount factors larger than 0.5, firms can collude on profit-maximizing prices in any of the three pricing scenarios (see also Figure 1.1).

For linear prices and fixed fees, we can calculate the outcomes using the results above, so our simulation concerns only the case of two-part tariffs. In the first step of our simulation, we run a "brute force" procedure and go through all possible collusive (linear and fixed) prices with precision 0.01. For a given collusive two-part tariff, we search for the optimal deviation response of the competitor.

This approach gives us information about the optimal deviation strategies for a given collusive two-part tariff. In the next step, we evaluate which collusive tariff is optimal for a given critical discount factor. The discount factors are arranged on a grid with precision 0.01. For each discount factor, we loop over all possible collusive tariffs. For each candidate tariff, we can calculate the collusive profit and extract the deviation profit from the "brute

[21] Note that we do not claim that a solution does not exist, but such a solution is simply too difficult to derive because of the highly nonlinear relationships between prices and the discount factor.

force" procedure. The competitive behavior is not affected by the candidate tariff, and the competitive profit results from the static Nash equilibrium. First, we check whether the critical discount factor that we can calculate based on Expression (1.2) is below the currently considered discount factor. Among the tariffs that satisfy this necessary condition, we then pick the tariff that yields the highest collusive profit.

We run our simulation for three different parameter configurations that refer to different degrees of product differentiation. Note that both the basic valuation v and the transport cost parameter τ can be used to model product differentiation. For example, to investigate an increase in product differentiation, we can either decrease the basic valuation or increase the transport cost parameter (ceteris paribus). We therefore fix the basic valuation at $v = 1$ and only vary the transport costs. More specifically, Assumption 1 requires that the highest value for the transport cost parameter is given by 0.8 and we use 10%, 50%, and 90% of this upper bound to discuss the cases of relatively small, intermediate, and large product differentiation. The figures presented in the text refer to the intermediate set-up with $\tau = 0.4$. The other figures can be found in 1.C.

The main objective of the simulation is to understand how the outcomes with two-part tariffs compare to the outcomes with linear and fixed prices. For the comparison of linear prices and two-part tariffs, we can make an initial hypothesis. We know from Section 1.4 that full collusion (that is, on profit-maximizing prices) is feasible for a larger range of discount factors with linear prices. Because the collusive profits with profit-maximizing prices are the same for both pricing regimes, this means that there is a range of discount factors for which firms using two-part tariffs must deviate from the profit-maximizing two-part tariffs, whereas firms using linear prices can sustain the profit-maximizing prices. In this range, profits are larger with linear prices than with two-part tariffs. On the other hand, we know from other research (for example, Chang, 1991) that collusive profits approach competitive profits if the discount factor tends to zero. Because competitive profits are larger with two-part tariffs than with linear prices, we can expect the profit curves for linear prices and two-part tariffs to intersect at least once. Profits should be larger with two-part tariffs for very small discount factors and larger with linear prices for discount factors sufficiently close to the critical discount factors from Section 1.4.

Our simulation lends support to this hypothesis. Figure 1.2 plots profits and customer surpluses against discount factors for the different pricing schemes. It illustrates that collusive profits are largest and customer surplus is lowest in the case of linear prices if the discount factor is not extremely small. If the discount factor tends to zero, collusive profits approach competitive profits and, hence, collusive profits are largest with two-part tariffs (i.e., they approach $\pi_L^* = \pi_F^* = 0.2$ and $\pi_T^* \approx 0.207$).

The figure further reveals that collusive profits are always lowest with fixed fees. Customer surplus is largest when collusion on profit-maximizing fixed fees is sustainable. When firms have to collude partially on fixed fees, customer surplus is similar to that in the case of two-part tariffs. Consequently, customer surplus is also smaller with fixed fees than with linear prices if discount factors are sufficiently small. The reason is that collusive profits tend to competitive profits that are equal under both pricing schemes. In other words, firms extract roughly the same part of customers' utility if discount factors tend to zero. At the same time, customers do not mix and, hence, incur transport costs when firms set fixed fees instead of linear prices.

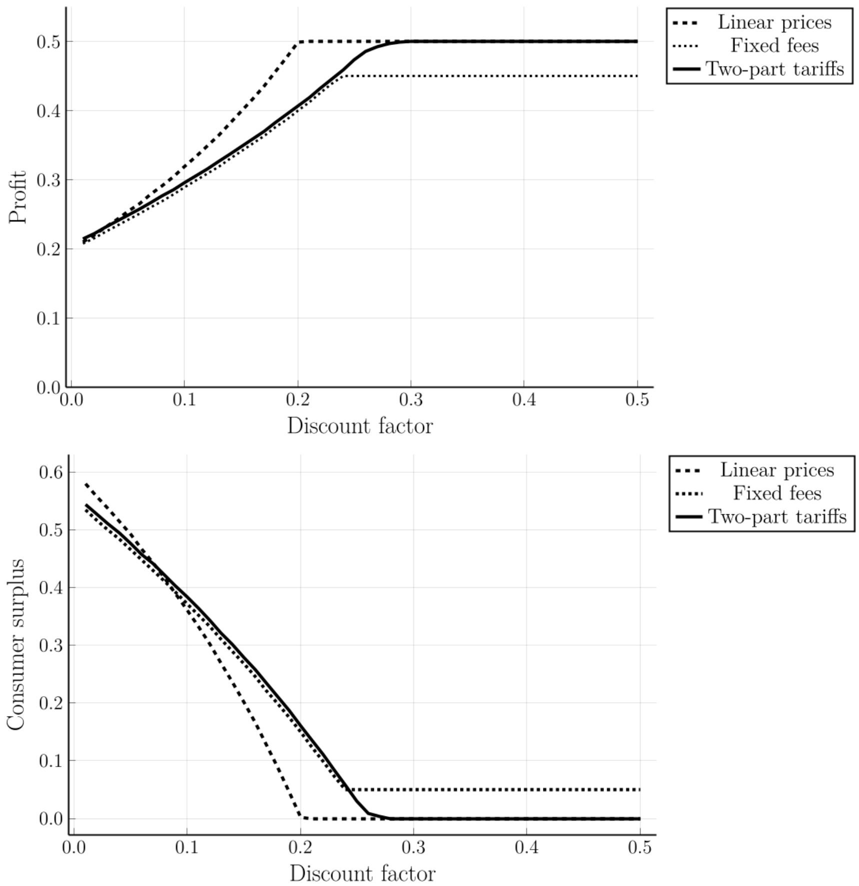


Figure 1.2: Profits and customer surpluses under partial collusion ($v = 1$, $\tau = 0.4$).

The corresponding figures for the cases of low and high product differentiation are Figures 1.3 and 1.4. Both figures qualitatively support our findings, although the quantitative measures, such as thresholds and distances, vary.

To put these results in perspective, remember that Proposition 1 shows that simple (that is, single-part) pricing schedules make it easier for firms to coordinate on profit-maximizing prices. The main insight of this subsection is that firms are likely to benefit from a ban of the fixed price component of a two-part tariff – even if they can only sustain collusion below the critical discount factor. At the same time, the ban is likely to harm customers most. Again, this finding is especially important because the ban is most beneficial for customers in absence of collusion.[22] We also obtain new insights with regard to the ban of the linear price component. When we compare fixed prices to two-part tariffs, we find that the ban

[22]Note that due to the robustness of the result, it does not matter for the policy maker whether the actual industry discount factor is known. This proves to be quite convenient because generally speaking, little is known about real-life discount factors. A notable exception is Igami and Sugaya (2021) who analyze the vitamin cartels.

	Competition	Collusion
Critical discount factor	–	$\overline{\delta}_L < \overline{\delta}_F < \overline{\delta}_T$ if $\tau < \tau^{(1)}$ $\overline{\delta}_F < \overline{\delta}_L < \overline{\delta}_T$ if $\tau > \tau^{(1)}$
Producer surplus	$\pi_F^* = \pi_L^* < \pi_T^*$	$\pi_F^c < \pi_T^c = \pi_L^c$
Customer surplus	$CS_F^* < CS_T^* < CS_L^*$	$CS_T^c = CS_L^c < CS_F^c$
Social welfare	$W_F^* < W_T^* < W_L^*$	$W_F^c < W_T^c = W_L^c$

Table 1.2: Comparison of critical discount factors, profits, customer surpluses, and welfare.

could facilitate collusion on the one hand, but on the other hand, it can be beneficial for customers if firms partially collude anyway.

1.6 Summary

This paper investigates firms' incentives to collude in a framework in which customers have the ability to combine products from different firms to achieve a better fit of their preferences. Motivated by various examples from the banking and insurance industry, we consider two policy interventions (banning linear or fixed prices) and investigate the case of partial collusion in Section 1.5.3. Table 1.2 summarizes our findings. First, firms can be restricted to use linear prices only, which leads to an increase in customer surplus in a static environment. However, it is shown that such a restriction makes it easier for firms to collude and harms customers. Additionally, our investigation of partial collusion shows that in the presence of collusion, linear prices are again most likely to lead to the highest profits and the lowest customer surplus among all pricing schemes. In summary, we conclude that the possibility to have higher customer surplus in absence of collusion comes along with an increasing scope for collusion and lower customer surplus in the presence of collusion.

Second, we consider a ban of linear prices, so that firms must compete (or collude) with fixed prices. Although collusion on profit-maximizing prices is easier with fixed prices than with two-part tariffs, we find that firms prefer to partially collude with two-part tariffs and fixed fees can harm customers the least among the three collusive pricing regimes. In summary, fixed fees can be less harmful to customers in presence of collusion, whereas they are most harmful to customers in absence of collusion (Hoernig and Valletti, 2007).

In summary, the present analysis has important implications for competition and consumer protection policy. The previous literature has shown that customers can benefit from policy interventions. Our paper highlights that such interventions can have undesired consequences and the implications are thus ambiguous: The possibility to achieve a higher customer surplus in absence of collusion may come at the expense of an increasing scope for collusion. Even in the worst case in which firms always collude, less price instruments can result in lower customer surplus and, hence, harm customers.

Our results raise the question whether firms can benefit from single-part pricing schemes even if they are not constraint by legal restrictions. The idea is that firms have access to multi-part pricing in general, but can also collude not only on the price level, but also on not using certain price components of a multi-part tariffs. We investigate this question in Appendix 1.A and show that in an environment where firms can use two-part tariffs in general, they can indeed use a self-imposed constraint to fixed prices to stabilize collusive agreements at profit-maximizing prices. However, we also show that this result is not robust once we allow for partial collusion.

Finally, as with every (theoretical) analysis, our analysis is limited by the assumptions of our model and the usual caveats apply. The assumptions of unit demand and the uniform distribution of consumers along the linear city are just two examples. However, in light of our general argument on the comparison of the price-setting incentives under linear prices and two-part tariffs (Section 1.5.1), we are confident that our main result remains intact even if one relaxes these assumptions.

Appendices

1.A Self-Imposed Constraints to Single-Part Pricing Schemes in Absence of Legal Restrictions

Our main analysis builds on the assumption that firms face legal or regulatory restrictions and are thus constrained in their pricing setting. In this section, we explore what happens in absence of such interventions. We allow firms to always set two-part tariffs, but also enable them to use more sophisticated collusive agreements. More precisely, we allow the firms to collude not only on the price level, but also on not using a particular price component of their two-part tariffs. For instance, firms might commit to fixing the linear price to zero and setting the optimal fixed fee under this self-imposed constraint.

The extension is relevant in those cases in which collusion on profit-maximizing two-part tariffs is not possible. If firms' pricing is not constrained, we already know that sufficiently patient firms (that is, $\delta \geq \overline{\delta}_T$) are able to collude. We also know that they gain the highest possible collusive profits because they use only the linear price component and are thus able to extract the full customer surplus, while minimizing customers' transport costs. By contrast, less patient firms (that is, $\delta < \overline{\delta}_T$) cannot sustain collusion at profit-maximizing collusive two-part tariffs and might consider to commit to not using a certain price component.

When firms commit to a particular pricing schedule, this commitment is only binding for firms that participate in collusion. Deviating and competing firms can still decide to use other pricing schemes. The comparison of different critical discount factors boils down to the comparison of the different types of profits. When we compare a commitment to either linear or fixed prices to the unconstrained case of two-part tariffs, the competitive profits are always the same because firms prefer two-part tariffs (Hoernig and Valletti, 2007). Thus, possible differences in the critical discount factors can only arise from differences in collusive or deviation profits.

We discuss linear prices first. This case is trivial because firms set zero fixed fees in the case of two-part tariffs (Lemma 1) and, hence, committing to profit-maximizing linear prices leads to the same prices. Because the linear price is only binding for colluding firms, deviating and competing firms can use two-part tariffs; hence, there is no difference from the analysis of two-part tariffs in the Section 1.4. Let $\widetilde{\delta}_L$ denote the critical discount factor when firms commit to using linear prices only. The following corollary summarizes our finding:

Corollary 3. *It holds that* $\widetilde{\delta}_L = \overline{\delta}_T > \overline{\delta}_L$.

In contrast to linear prices, the effect of a commitment to fixed fees is not clear a priori. As shown in Lemma 1, collusive profits are lower with fixed fees than with two-part tariffs. This makes it harder for firms to sustain collusion. On the other hand, the deviating firm now has access to an additional price component and deviates from a fixed instead of a linear price, so that deviation profits could be lower which would increase the likelihood of collusion. To better understand the impact on the deviator's behavior, the following lemma summarizes optimal deviation prices and profits:

Lemma 6. *The deviating firm sets its fixed and linear prices denoted by $f_F^{d\prime}$ and $p_F^{d\prime}$, such that*

$$f_F^{d\prime} + p_F^{d\prime} = f_F^d.$$

The resulting deviation profit is given by

$$\pi_F^{d\prime} = \begin{cases} \pi_F^d = v - \frac{5\tau}{4} & \text{if } 0 < \tau \leq \frac{4v}{13} \\ \pi_F^d = \frac{(4v+3\tau)^2}{128\tau} & \text{if } \frac{4v}{13} < \tau \leq \tau^{(2)}, \end{cases}$$

where $\tau^{(2)} := 4v/(26\sqrt{5} - 53) \approx 0.7785483v$.

Lemma 6 shows that a deviating firm cannot take advantage of the additional linear price compared to the case of a ban on linear prices. The reason is that it is not profitable for the deviating firm to set both price components so low that customers mix. Customers who buy one unit exclusively at one firm are indifferent between paying a linear price or a fixed fee because they always have to pay the full linear price, and not just a share of it. Thus, the total price customers pay is the same as in the previous case of fixed prices, but the deviator can now split it into two arbitrary parts, namely the linear price and the fixed fee.

Note that firms only collude if $\tau < \tau^{(2)}$. This threshold is lower than the threshold defined in Assumption 1. If $\tau > \tau^{(2)}$, collusive profits with fixed fees are smaller than the competitive profits with two-part tariffs and firms never use fixed fees to facilitate collusion.

The following lemma specifies the critical discount factor denoted by $\widetilde{\delta}_F$:

Lemma 7. *The critical discount factor with fixed fees is given by*

$$\widetilde{\delta}_F := \begin{cases} \frac{-4v+9\tau}{26\tau\sqrt{5}-44\tau-8v} & \text{if } 0 < \tau \leq \frac{4v}{13} \\ \frac{(4v-5\tau)^2}{16v^2+24v\tau+873\tau^2-416\tau^2\sqrt{5}} & \text{if } \frac{4v}{13} < \tau \leq \tau^{(2)}. \end{cases}$$

If the critical discount factor $\widetilde{\delta}_F$ is lower than the critical discount factor $\overline{\delta}_T$, firms may be able commit to using fixed fees to facilitate collusion. The following proposition describes under which conditions this requirement is satisfied for collusion on profit-maximizing prices:

Proposition 4. *A threshold $\tau^{(3)}$ exists ($\approx 0.776497731291448v$), such that $\widetilde{\delta}_F < \overline{\delta}_T$ holds for all $\tau < \tau^{(3)}$.*

As noted above, a commitment to fixed fees has two opposing effects. On the one hand, collusive profits decline, that is, collusion tends to be harder to sustain. On the other hand, a cheating firm must deviate from fixed fees, which is less profitable, and, hence, collusion tends to be easier to sustain. Proposition 4 states that the deviation effect dominates the collusion effect for a large[23] range of transport cost parameters.

There is only a small range for very large degrees of product differentiation (between $\tau^{(2)}$ and $\tau^{(3)}$) where two-part tariffs could be beneficial. In fact, the bound $\tau^{(3)}$ arises from a comparison with the lower bound of the critical discount factor in the case of two-part tariffs. The true critical discount factor can, however, be larger. We can further characterize the relationship between the two critical discount factors by making use of the fact that the critical discount factor in the case of a self-imposed commitment to fixed fees converges to 1 as the degree of product differentiation approaches its maximum considered in our analysis (that is, if $\tau \to \tau^{(2)}$).

Corollary 4. *A threshold $\tau^{(4)} > \tau^{(3)}$ exists, such that $\widetilde{\delta}_F > \overline{\delta}_T$ holds for all $\tau > \tau^{(4)}$.*

Because the size of the remaining interval ($[\tau^{(3)}, \tau^{(2)}]$) is very small, we refrain from providing a more detailed characterization of the cutoff point $\tau^{(4)}$.

Proposition 4 suggests that firms can benefit from a commitment to single-part pricing schemes. However, this result is based on the assumption that colluding firms choose profit-maximizing prices. If we allow firms to collude on prices that are different from the profit-maximizing prices (partial collusion), this result turns out not to be robust. To see why, let us assume that firms could optimally commit to using only a fixed fee f (or only a linear price p). In this case, the optimal solution implies that for a given discount factor δ, the two-part tariffs that maximize profits under the constraint that the critical discount factor is below or equal to the actual discount factor result in a (weakly) smaller profit than the profit with the self-imposed pricing scheme. However, in the counterfactual with the two-part tariffs (f_T, p_L), the firms can simply choose the same fixed fee $f_T = f$ and dismiss the linear components $p_L = 0$, and would gain the same profit (or set $f_T = 0$ and $p_L = p$). This is because two-part tariffs nest the cases of linear prices and fixed fees as special cases. Furthermore, because deviating and competing firms can always use two-part tariffs, partially colluding firms can use the two-part tariffs as if this was the optimal fixed fee (or the optimal linear price) if this is the optimal strategy overall. We can thus conclude:

Corollary 5. *If firms are allowed to collude on prices that differ from the profit-maximizing prices (partial collusion), firms gain the largest profit under two-part tariffs.*

The above line of reasoning can easily be adopted for fully nonlinear tariffs, that is, firms are always better off when they (partially) collude with fully nonlinear tariffs than when they restrict themselves to single-part pricing schemes to stabilize collusive agreements. As with the two-part tariffs, this immediately follows from the fact that fully nonlinear tariffs nest the other pricing schemes as special cases.

In summary, we find that firms can use a commitment to single-part pricing schemes to stabilize collusive agreements, but this strategy turns out not to be profitable in the sense

[23]Indeed, $\tau^{(3)}$ is approximately given by $0.776497731291448v$, which is close to the threshold $\tau^{(2)} \approx 0.7785483v$ defined in Lemma 6 and also close to the upper bound $0.8v$ defined in Assumption 1.

that collusion on prices that are different from the profit-maximizing prices might result in larger profits.

1.B Proofs

Proof of Lemma 1.
Linear prices:
The case of collusion with linear prices corresponds to the monopoly case in Hoernig and Valletti (2011), which they use as a benchmark for their analysis. The authors assume that a monopolist offers both products that are located at the extreme points of the line. This is the same optimization problem as in the case of two colluding firms which are located at the same extreme points. The firms – just like the monopolist – have an incentive to maximize welfare (that is, induce the efficient allocation where $\underline{x} = 0$ and $\overline{x} = 1$) which is then fully extracted by setting an optimal linear price of v.

Fixed fees:
First, we assume that firms set prices in such a way that at least some customers buy from both firms. We thus have:

$$\underline{x} \leq \overline{x} \quad \Leftrightarrow \quad \sqrt{\frac{f_2}{\tau}} \leq 1 - \sqrt{\frac{f_1}{\tau}}$$

Note that if the constraint binds, the indifferent customer who is located at $x = \underline{x} = \overline{x}$ is indifferent between buying from both firms or buying from one firm exclusively.

Both firms maximize their joint profit that is given by

$$\Pi = f_1 \cdot \overline{x} + f_2 \cdot (1 - \underline{x}) \, .$$

We calculate the derivative of the profit function with respect to both fixed prices:

$$\frac{\partial \Pi}{\partial f_1} = 1 - \frac{3}{2}\sqrt{\frac{f_1}{\tau}}$$
$$\frac{\partial \Pi}{\partial f_2} = 1 - \frac{3}{2}\sqrt{\frac{f_2}{\tau}}$$

When we set the derivatives to zero, we find the candidate solution $f_1 = f_2 = 4\tau/9$. Evaluating the aforementioned constraint at these values yields $2/3 \leq 1/3$, which is obviously not true. Because the first derivatives are always larger than zero for $0 \leq f_1, f_2 \leq 4\tau/9$, we conclude that, given that the aforementioned assumption has to be satisfied, firms prefer to set their prices in a way such that the constraint binds and the indifferent customer is indifferent between combining both products or buying from one firm exclusively. Because the decision of the indifferent customer has no influence on the joint profit, it is sufficient to analyze the case in which all customers buy exclusively from one firm.

Now we turn to the case in which customers do not mix. Given Assumption 1, firms set prices, such that all customers have non-negative utilities and all customers buy. Thus, total demand is not affected by the price level and firms set prices, such that the customer who is indifferent between buying from firm 1 and buying from firm 2 is also indifferent between buying and not buying. Let x be the position of this customer. Then, firm 1 sets $v - \tau \cdot x^2$ and realizes a market share of x. Analogously, firm 2 sets $v - \tau \cdot (1-x)^2$ and realizes a market share of $1 - x$. The joint profit is given by Π:

$$\Pi = x \cdot \left[v - \tau \cdot x^2\right] + (1-x) \cdot \left[v - \tau \cdot (1-x)^2\right]$$
$$\frac{\partial \Pi}{\partial x} = -3 \cdot \tau \cdot x^2 + 3 \cdot \tau \cdot (1-x)^2$$
$$\frac{\partial^2 \Pi}{\partial x^2} = -6 \cdot \tau \left(x + (1-x)\right) < 0$$

The first derivative is zero if $x = 1/2$. Therefore, the optimal collusive price equals $v - t/4$.

Two-part tariffs:
Customers do not buy if they have to pay more than the reservation price v. Therefore, the highest possible profit is v. Note that if both firms charge no fixed fees (that is, $f_i = 0$) and optimal linear prices (that is, $p_i = p_L^c = v$), the total profit is equal to v, that is, the profit is maximized.

We show that no other combination of fixed fees and linear prices is a collusive equilibrium. First note that if the fixed fee is zero, symmetric linear prices lower than v lead to a total profit lower than v and, hence, will not be chosen in equilibrium.

Note further that firms set symmetric linear prices. In the case of asymmetric linear prices, customers suffer from strictly positive transport costs and, hence, customers' total utility that firms can extract is lower than v.

Finally, we show that firms have no incentive to set strictly positive fixed fees. If at least one fixed fee was larger than zero, at least one indifferent customer ($\underline{x}$ and/or $\overline{x}$) is not located at the extreme, that is, at least some customers do not mix and suffer from strictly positive transport costs. As a result, customers' total utility that firms can extract is lower than v, and, hence, the prices will not be chosen in equilibrium. □

Proof of Lemma 2.
We assume without loss of generality that firm 1 deviates from the collusive agreement.

Linear prices:
Firm 1 can set its price, such that it either monopolizes the market or firm 2 realizes a strictly positive market share. Note that customers do not mix in the first case (that is, $\underline{x} = \overline{x} = 1$), while there is a share of customers who mix in the latter case (that is, $\underline{x} < \overline{x} = 1$). We derive prices and payoffs for both cases. When firm 1 monopolizes the market, the highest possible price satisfies the constraint $\underline{x} = 1$. Thus, we find $p_L^{d_1} = \pi_L^{d_1} = v - 2\tau$. Otherwise, in the second case, we plug the collusive price, p_L^c, as the price of firm 2 into firm 1's profit

function and maximize the profit with respect to firm 1's own price. This leads to

$$p_L^{d_2} = \frac{2v - 4\tau + A}{3}, \; \pi_L^{d_2} = \frac{(-2v + 4\tau - A)(v^2 - vA - 4v\tau + 2\tau A - 20\tau^2)}{108\tau^2}.$$

Note that $\underline{x}_L^{d_2} := \underline{x}_L \left(p_1 = p_L^{d_2}, p_2 = p_L^c\right) < 1$ holds for all $0 < \tau \leq 4v/5$. Comparing profits $\pi_L^{d_1}$ and $\pi_L^{d_2}$, we find that sharing the market is always more profitable than monopolization. As a result, firm 1 sets $p_L^{d_2}$ and earns $\pi_L^{d_2}$.

Fixed fees:
Firm 1 can set its fixed price, such that it either monopolizes the market or firm 2 realizes a strictly positive market share. In the first case, firm 1 has to compensate the loyal customers of firm 2 for suffering from higher transport costs. The customer at location $x = 1$ suffers from the highest transport costs of τ. As a result, prices and profits are given by

$$f_F^{d_1} = \pi_F^{d_1} = f_F^c - \tau = v - \frac{5\tau}{4}.$$

Next, we analyze the case in which firm 1 does not monopolize the market. First assume that it sets its price f, such that at least some customers mix and, hence, buy from both firms. In this case, $\underline{x} < \overline{x}$ holds. Firm 2 sticks to the collusive price schedule and sets $f_F^c = v - \tau/4$ (see Lemma 1). We can plug the prices into the aforementioned inequality and obtain

$$\sqrt{\frac{v}{\tau} - \frac{1}{4}} < 1 - \sqrt{\frac{f}{\tau}}.$$

The right-hand side of the inequality is always smaller than or equal to one. At the same time, we can rearrange the left-hand side and obtain

$$\sqrt{\frac{v}{\tau} - \frac{1}{4}} \geq 1 \quad \Leftrightarrow \quad \frac{v}{\tau} - \frac{1}{4} \geq 1 \quad \Leftrightarrow \quad \frac{v}{\tau} \geq \frac{5}{4} \quad \Leftrightarrow \quad \frac{4}{5}v \geq \tau.$$

Assumption 1 ensures that the first equivalence sign is correct and the final inequality holds. Because the left-hand side is always larger than or equal to one, but the right-hand side is less than or equal to one, the initial inequality $\underline{x} < \overline{x}$ does not hold. In other words, it is not possible that the deviating firms sets a fixed fee, such that at least some customers buy from both firms.

Turning to the case in which customers do not mix, we plug the collusive price, f_F^c, into firm 1's profit function and maximize the profit with respect to firm 1's own price. The resulting price and profit are

$$f_F^{d_2} = \frac{v}{2} + \frac{3\tau}{8}, \quad \pi_F^{d_2} = \frac{(4v + 3\tau)^2}{128\tau}.$$

The second case is relevant if and only if the profit is larger than the profit in the first case and the customer who is indifferent between buying from firm 1 or firm 2 is located in the interval. Both conditions lead to $\tau > \frac{4v}{13}$.

Two-part tariffs:
We have to distinguish between three cases to calculate the deviation profits: (i) Firm 1 monopolizes the market, that is, $\underline{x} = \overline{x} = 1$, (ii) (at least some) customers buy from both firms and the entire market is served, and (iii) there are some customers who use the outside option and do not buy from any firm. Note that customers will never buy from firm 2 exclusively. This is because they will pay a price of v, which equals their valuation, and, in addition, suffer from the transport costs. This implies $\underline{x} < \overline{x} = 1$ for case (ii).

It is difficult to derive a (closed-form) solution for the third case. Therefore, we calculate the prices and profits only for the first two cases.

Case (i): Firm 1 sets the highest possible price, such that $\underline{x} = 1$. It follows $p_T^{d_1} = v - 2\tau$. In addition, it sets the fixed fee, such that the customer at location $x = 1$ is indifferent between buying and not buying, that is, $U_1\left(x = 1; p_1 = p_T^{d_1}\right) = 0$. Thus, it sets $f_T^{d_1} = \tau$. The profit is $\pi_T^{d_1} = v - \tau$.

Case (ii): Firm 1 maximizes its profit under the constraint $\overline{x} = 1$, which is equivalent to $f_T^{d_2} = (p_1 - v)^2/4\tau$. The maximization yields $p_T^{d_2} = v/3 + 2\tau/3$ and, hence, $f_T^{d_2} = (v-\tau)^2/9\tau$. The profit is $\pi_T^{d_2} = (-v^3 + 12v^2\tau + 6v\tau^2 + 10\tau^3)/54\tau^2$.

Next, we derive a threshold that determines which case would be optimal for firm 1, ignoring the existence of case (iii). The constraint

$$\underline{x}\left(p_1 = p_T^{d_2}, f_1 = f_T^{d_2}\right) \leq 1 \qquad \Leftrightarrow \qquad \tau \geq \frac{v}{4}$$

and the comparison of the profits

$$\pi_T^{d_1} < \pi_T^{d_2} \qquad \Leftrightarrow \qquad \tau > \frac{v}{4}$$

lead to the threshold.

Because we did not calculate prices and profits for the third case, the profits for the other two cases are lower bounds, that is, firm 1 might be able to achieve an even larger profit if it sets its price in a way such that some customers choose the outside option and do not buy. We establish an upper bound for the deviation profit in Lemma 4, where we investigate fully nonlinear tariffs. The reason is that with fully nonlinear tariffs, firm 1 would deviate from the same collusive prices (see Corollary 1), but has access to a tariff structure that is more flexible and nests two-part tariffs as a special case. Because the lower bound equals the upper bound for $\tau < v/4$, we can replace the inequality sign with an equality sign. □

Proof of Lemma 3. The critical discount factors result immediately from inserting the respective profits into Expression (1.2). Competitive profits are derived by Anderson and Neven (1989) and Hoernig and Valletti (2007), whereas collusive and deviation profits are given by Lemma 1 and Lemma 2. Note that the critical discount factor (1.2) is monotonically increasing in the deviation profit, so that in the case of two-part tariffs, the lower bound for the profit results in a lower bound for the critical discount factor. □

Proof of Proposition 1. The proposition results immediately from pairwise comparisons of the critical discount factors and the corresponding bounds. □

Proof of Lemma 4. First, we introduce the notation in line with Hoernig and Valletti (2011, p. 6). Let

$$\begin{aligned}
&u(x,q) = v - v(1-q) - \tau(1-q-x)^2, \\
&\tilde{U}(x,\hat{x}) = u(x,Q(\hat{x})) - P(\hat{x}), \text{ and} \\
&U(x) = \tilde{U}(x,x).
\end{aligned}$$

The function $u(x,q)$ captures the residual utility when a customer at location x buys quantity q from firm 1 and quantity $1-q$ from firm 2 before paying the price of firm 1. It is useful to think about this residual utility as the largest possible amount that firm 1 can charge from this specific customer. $\tilde{U}(x,\hat{x})$ is the utility of a customer at location x when the customer buys the quantity that the customer at location $\hat{x}$ was supposed to buy. Finally, $U(x)$ is the utility of a customer at location x when the customer actually buys the supposed quantity.

The maximization problem is then given by:

$$\begin{aligned}
&\max_{P,Q}\left[\pi_1 = \int_0^{\bar{x}} P(x)\,dx\right] \\
\text{s.t. } &(IC) \quad U(x) \geq \tilde{U}(x,\hat{x}) \quad \forall x,\hat{x} \in [0,\bar{x}], \\
&(PC) \quad U(x) \geq 0 \quad \forall x \in [0,\bar{x}]
\end{aligned}$$

Note that our participation constraint differs from that in Hoernig and Valletti (2011). In their analysis, customers evaluate the possibility of buying from firm 1 against the possibility of buying from firm 2 exclusively. This is not reasonable in our analysis because firm 2 charges the collusive tariff $P(q) = vq$. If a customer buys exclusively from firm 2, the customer (at $x < 1$) would suffer from total costs of more than v, which exceeds the customer's willingness to pay. Therefore, customers who choose not to buy from firm 1 will not buy any product.

The beginning of our proof closely follows the first two steps of the proof of Proposition 1 in Hoernig and Valletti (2011). First note that the participation constraint binds for $x = \bar{x}$, that is,

$$u(\bar{x},Q(\bar{x})) - P(\bar{x}) = 0.$$

The incentive constraint has to bind for all customers $x \in [0,\bar{x}]$ and can be rewritten as follows

$$u(x,Q(x)) - P(x) \overset{def}{=} U(x) \overset{(IC)}{=} \tilde{U}(x,\bar{x}) \overset{def}{=} U(\bar{x}) - \int_x^{\bar{x}} \frac{\partial u(s,Q(s))}{\partial x} ds.$$

Note that the first and third equality follow from the definitions of the respective objects, while the second equality is just the binding incentive constraint. Rewriting this equality yields:

$$P(x) = u(x, Q(x)) + \int_x^{\bar{x}} \frac{\partial u(s, Q(s))}{\partial x} ds - U(\bar{x}).$$

The deviation profit of firm 1 is then given by integrating over all customers in $x \in [0, \bar{x}]$:

$$\begin{aligned}
\pi_1 &= \int_0^{\bar{x}} \left[u(x, Q(x)) + \int_x^{\bar{x}} \frac{\partial u}{\partial x}(s, Q(s))\, ds - U(\bar{x}) \right] dx \\
&= \int_0^{\bar{x}} \left[u(x, Q(x)) + \int_x^{\bar{x}} \frac{\partial u}{\partial x}(s, Q(s))\, ds \right] dx - \bar{x} \cdot U(\bar{x}).
\end{aligned}$$

Hoernig and Valletti (2011) analytically show that one can further simplify this expression:

$$\pi_1 = \int_0^{\bar{x}} \left[u(x, Q(x)) + x \cdot \frac{\partial u}{\partial x}(x, Q(x)) \right] dx - \bar{x} \cdot U(\bar{x}). \tag{1.3}$$

This completes the first step of the proof in Hoernig and Valletti (2011). In their second step, they derive the first-order condition for the optimal $Q(x)$ depending on the rival's tariff. While the rival's tariff is unknown at this stage in their analysis, we know the optimal collusive tariff the rival will set. This leads to the following characterization of $Q(x)$:

$$v + 2\tau(1 - Q(x) - x) = 2\tau x \quad \Leftrightarrow \quad Q(x) = \frac{v}{2\tau} + 1 - 2x.$$

Note that we require $Q(x) \in [0, 1]$, which leads to the following thresholds:

$$\begin{aligned}
Q(x) \geq 0 \quad &\Leftrightarrow \quad x \leq \frac{v}{4\tau} + \frac{1}{2} \equiv \overline{\theta}, \\
Q(x) \leq 1 \quad &\Leftrightarrow \quad x \geq \frac{v}{4\tau} \equiv \underline{\theta}.
\end{aligned}$$

We can further check under which conditions the customers located at $\underline{\theta}$ and $\overline{\theta}$ are located in $[0, 1]$. First, note that $\underline{\theta} > 0$ and $\overline{\theta} > 0$ always hold. Second, we can express $\overline{\theta}$ in terms of $\underline{\theta}$, that is, $\overline{\theta} = 1/2 + \underline{\theta}$. Comparing $\underline{\theta}$ and $\overline{\theta}$ to one leads to two conditions for the product differentiation parameter:

$$\begin{aligned}
\underline{\theta} \leq \frac{1}{2} \quad &\Leftrightarrow \quad \tau \geq \frac{v}{2} \equiv \Theta_2 \\
\underline{\theta} \leq 1 \quad &\Leftrightarrow \quad \tau \geq \frac{v}{4} \equiv \Theta_1
\end{aligned}$$

We further have to ensure that $Q(x)$ is well-defined in the sense that a customer at x buying $Q(x)$ must have a (weakly) positive residual utility:

$$v - v \cdot (1 - Q(x)) - \tau (1 - Q(x) - x)^2 \geq 0 \quad \Leftrightarrow \quad x \in \left[\underbrace{-\frac{v\tau + \sqrt{4\tau^3 v + 2\tau^2 v^2}}{2\tau^2}}_{<0}, \underbrace{\frac{-v\tau + \sqrt{4\tau^3 v + 2\tau^2 v^2}}{2\tau^2}}_{\equiv \tilde{\theta}} \right].$$

It is also straightforward to verify that $\tilde{\theta} \leq \overline{\theta}$ always holds. Further constraints arise from the following comparisons:

$$\begin{aligned} \tilde{\theta} \geq \underline{\theta} \quad &\Leftrightarrow \quad \tau \leq \frac{v}{16}, \\ \tilde{\theta} \leq 1 \quad &\Leftrightarrow \quad \tau \geq \frac{v}{2} = \Theta_2. \end{aligned}$$

The constraints lead to four areas that we have to investigate in detail:

1. Area I ($\tau \leq v/16$): All customers buy exclusively from firm 1, that is, $1 \leq \tilde{\theta} \leq \underline{\theta} < \overline{\theta}$.

2. Area II ($v/16 < \tau \leq \Theta_1$): All customers buy exclusively from firm 1, that is, $1 \leq \underline{\theta} \leq \tilde{\theta} \leq \overline{\theta}$.

3. Area III ($\Theta_1 < \tau \leq \Theta_2$): Customers in $[0, \underline{\theta}]$ buy exclusively from firm 1, and customers in $[\underline{\theta}, 1]$ buy from both firms, that is, $\underline{\theta} \leq 1 \leq \tilde{\theta} \leq \overline{\theta}$.

4. Area IV ($\Theta_2 < \tau$): Customers in $[0, \underline{\theta}]$ buy exclusively from firm 1, and customers in $\left[\underline{\theta}, \tilde{\theta}\right]$ buy from both firms, that is, $\underline{\theta} \leq \tilde{\theta} \leq \overline{\theta} \leq 1$.

Note that the first two areas have the same outcome, so that we can pool them in one case.

Case 1 ($\tau \leq \Theta_1$): If the deviating firm monopolizes the market, it extracts the entire utility from the customer located at $\bar{x} = 1$, that is, $U(\bar{x}) = 0$.

$$P(\bar{x}) = u(\bar{x}, Q(\bar{x})) = v - \tau.$$

The resulting profit is

$$\pi_1 = \int_0^1 \left[\underbrace{v - \tau \cdot x^2}_{=u(x, Q(x))} + \underbrace{(-1) \cdot 2 \cdot \tau \cdot x^2}_{x \cdot \frac{\partial u}{\partial x}(x, Q(x))} \right] dx = \left[v \cdot x - \tau x^3 \right]_0^1 = v - \tau.$$

Case 2 ($\Theta_1 \leq \tau \leq \Theta_2$): Customers in the interval $[0, \underline{\theta}]$ will buy exclusively from firm 1, whereas customers in $[\underline{\theta}, 1]$ will buy from both firms. This allows us to split the profit function of the deviating firm into two components. Further, note that $U(\bar{x}) = 0$ for $\bar{x} = 1$

because the deviating firm could otherwise gain a larger profit by charging a higher price:

$$\begin{aligned}
\pi_1 &= \int\limits_0^1 \left[u\left(x, Q\left(x\right)\right) + x \cdot \frac{\partial u}{\partial x}\left(x, Q\left(x\right)\right) \right] dx \\
&= \int\limits_0^{\underline{\theta}} \left[u\left(x, 1\right) + x \cdot \frac{\partial u}{\partial x}\left(x, 1\right) \right] dx + \int\limits_{\underline{\theta}}^1 \left[u\left(x, Q\left(x\right)\right) + x \cdot \frac{\partial u}{\partial x}\left(x, Q\left(x\right)\right) \right] dx
\end{aligned}$$

We calculate the values of both components separately. For the first component, we get

$$\int\limits_0^{\underline{\theta}} \left[\underbrace{v - \tau \cdot x^2}_{=u(x,Q(x))} + \underbrace{(-1) \cdot 2 \cdot \tau \cdot x^2}_{x \cdot \frac{\partial u}{\partial x}(x,Q(x))} \right] dx = \left[v \cdot x - \tau x^3 \right]_0^{\underline{\theta}} = v \cdot \underline{\theta} - \tau \cdot \underline{\theta}^3 = \frac{v^2}{4\tau} - \frac{v^3}{64\tau^2}.$$

Turning to the second component, we get

$$\int\limits_{\underline{\theta}}^1 \left[\underbrace{v\left(\frac{v}{2\tau} + 1 - 2x\right) - \tau\left(x - \frac{v}{2\tau}\right)^2}_{=u(x,Q(x))} + \underbrace{2\tau x\left(x - \frac{v}{2\tau}\right)}_{x \cdot \frac{\partial u}{\partial x}(x,Q(x))} \right] dx = \left[\frac{v^2 x}{4\tau} + vx - vx^2 + \frac{\tau x^3}{3} \right]_{\underline{\theta}}^1 = \frac{64\tau^3 - v^3}{192\tau^2}.$$

Adding up the values of both components, we get

$$\pi_1 = \frac{16\tau^3 + 12\tau v^2 - v^3}{48\tau^2}.$$

Case 3 ($\Theta_2 \leq \tau$): Finally, we turn to the case in which some customers choose the outside option and do not buy from any firm. By definition, we have $u\left(\bar{x}, Q\left(\bar{x}\right)\right) = 0$ for $\bar{x} = \tilde{\theta}$. This implies $P\left(Q\left(\bar{x}\right)\right) = 0$ and $U\left(\bar{x}\right) = 0$. As in the previous case, we can split the profit function into two components:

$$\begin{aligned}
\pi_1 &= \int\limits_0^{\tilde{\theta}} \left[u\left(x, Q\left(x\right)\right) + x \cdot \frac{\partial u}{\partial x}\left(x, Q\left(x\right)\right) \right] dx \\
&= \int\limits_0^{\underline{\theta}} \left[u\left(x, 1\right) + x \cdot \frac{\partial u}{\partial x}\left(x, 1\right) \right] dx + \int\limits_{\underline{\theta}}^{\tilde{\theta}} \left[u\left(x, Q\left(x\right)\right) + x \cdot \frac{\partial u}{\partial x}\left(x, Q\left(x\right)\right) \right] dx.
\end{aligned}$$

For the first component, we get

$$\int\limits_0^{\underline{\theta}} \left[\underbrace{v - \tau \cdot x^2}_{=u(x,Q(x))} + \underbrace{(-1) \cdot 2 \cdot \tau \cdot x^2}_{x \cdot \frac{\partial u}{\partial x}(x,Q(x))} \right] dx = \left[v \cdot x - \tau x^3 \right]_0^{\underline{\theta}} = \frac{v^2}{4\tau} - \frac{v^3}{64\tau^2}.$$

Turning to the second component, we get

$$\int_{\underline{\theta}}^{\tilde{\theta}} \left[\underbrace{v\left(\frac{v}{2\tau}+1-2x\right)-\tau\left(x-\frac{v}{2\tau}\right)^2}_{=u(x,Q(x))} + \underbrace{2\tau x\left(x-\frac{v}{2\tau}\right)}_{x\cdot\frac{\partial u}{\partial x}(x,Q(x))} \right] dx = \left[\frac{v^2 x}{4\tau}+vx-vx^2+\frac{\tau x^3}{3}\right]_{\underline{\theta}}^{\tilde{\theta}}$$

$$= \frac{5\left(\frac{\sqrt{\tau^2 v(2\tau+v)}\,(5v+4\tau)\sqrt{2}}{5}-\frac{45v\tau\left(v+\frac{48\tau}{25}\right)}{32}\right)v}{6\tau^3}.$$

Adding up the values of both components, we get

$$\pi_1 = \frac{5\left(\frac{\sqrt{\tau^2 v(2\tau+v)}\,(5v+4\tau)\sqrt{2}}{5}-\frac{57v\tau\left(v+\frac{32\tau}{19}\right)}{40}\right)v}{6\tau^3}.$$

Critical discount factors: The critical discount factor results immediately from inserting the different types of profit into Expression (1.2). □

Proof of Proposition 2. Note that Proposition 1 shows that the critical discount factor is always larger under two-part tariffs than under linear prices or fixed fees. Thus, the proposition immediately results from the comparison of $\overline{\delta}_T$ and $\overline{\delta}_N$.

A comparison of the competitive profits reveals that competitive profits are larger under fully nonlinear prices than under two-part tariffs. In addition, collusive prices and profits are the same and deviation profits are larger under fully nonlinear prices. The latter follows from the fact that the collusive prices are the same and that fully nonlinear tariffs nest two-part tariffs as a special case. Since the critical discount factor (1.2) increases in both the competitive and deviation profits, it follows $\overline{\delta}_N > \overline{\delta}_T$. □

Proof of Corollary 2. Our analysis is similar to that in Chang (1991), except that customers in our model have the possibility to combine products. In the following, we show that this possibility does not change the result of Chang (1991).

In the first step, we assume that firms set collusive prices above the competitive prices (as in Chang, 1991). This means that all customers buy exclusively from one firm and mixing does not occur. To adopt the results of Chang (1991), we have to show that a deviating firm has no incentive to set prices that induce customers to mix. To see this, assume that the price is such that at least some customers combine products of both firms, that is,

$$\underline{x} < \overline{x} \quad \Leftrightarrow \quad \sqrt{\frac{f_2}{\tau}} < 1-\sqrt{\frac{f_1}{\tau}} \quad \Leftrightarrow \quad \sqrt{\frac{f_1}{\tau}} < 1-\sqrt{\frac{f_2}{\tau}}.$$

Consider the case in which firm 2 sets the collusive price $f_2 \geq f_F^* = \tau$ and firm 1 deviates. Then, $\sqrt{f_2/\tau} \geq 1$ and the right-hand side is not positive. Because we are only interested in the positive values of the square roots, the inequality cannot be satisfied irrespective of f_1. This implies that the deviating firm sets its price, such that all customers buy exclusively from one firm.

Second, we have to show that colluding firms have no incentive to set prices below the competitive prices to induce customer to mix. We have to distinguish between two cases. In the first case, firms set prices below competitive prices, but the prices are so large that customer do not mix. In this case, they would earn a profit that is smaller than the competitive profit (that is, $f_F^c/2$ instead of $f_F^*/2$), and because the competitive prices constitute a Nash equilibrium, they could always set these prices without the risk that the other firm could profitably deviate.

The second case covers prices below competitive prices that induce (at least some) customers to combine products from both firms. A necessary conditions for this customer behavior is $C \equiv \overline{x} - \underline{x} \geq 0$. The joint profit function is given by $\pi = f_1 \cdot \overline{x} + f_2 \cdot \underline{x}$. The derivative of the profit with respect to each fixed fee is independent of the other fixed fee and is strictly positive as long as $f_1 < 4\tau/9$ and $f_2 < 4\tau/9$. Because $f_1 = 4\tau/9$ and $f_2 = 4\tau/9$ violates the condition $C \geq 0$, we face a corner solution, where the condition on C binds with equality (that is, $C = 0$). We can rewrite C and express f_1 in terms of f_2, insert $f1\,(f_2)$ into the profit π, and use the first-order conditions on π with respect to f_2 to determine the optimal fixed fees. The result is $f_1 = f_2 = \tau/4$. This leads to an upper bound for the profit in the case of mixing customers of $\tau/8$. Comparing this upper bound to the collusive profits derived by Chang (1991) reveals that firms prefer to collude on prices above the competitive prices. $\square$

Proof of Lemma 5. Consider the case in which the firms want to sustain collusion for a given discount factor δ. They have to set a collusive price p_L^c, such that collusion is sustainable. We start with the analysis of the optimal behavior of a deviating firm. Similar to the case of collusion on profit-maximizing prices, there are two possible scenarios. First, the deviating firm sets a price, such that it captures the entire market. Then, the deviating firm would set a price of $p_L^c - 2\tau$, which follows from the fact that the monopolization of the market requires $\underline{x} \leq 1$ and that the deviating firm wants to set the highest possible price given this constraint such that $\underline{x} = 1$. The resulting profit would be $p_L^c - 2\tau$. Alternatively, the deviating firm could leave the other firm with a strictly positive market share ($\underline{x} < 1$). The first-order condition implies that the price would be

$$p_L^{d,p} = \frac{2p_L^c - 4\tau + C}{3},$$

and the profit would be

$$\pi_L^{d,p} = \frac{(2p_L^c - 4\tau + C) \cdot \left(2\tau C - p_L^c C - 20\tau^2 - 4\tau p_L^c + (p_L^c)^2\right)}{108\tau^2},$$

with $C := \sqrt{28\tau^2 - 4\tau p_L^c + \left(p_L^c\right)^2}$. A comparison of the profits in both cases reveals that the profit in the second case is always strictly larger than the profit in the first case.

For a given price level p_L^c, we can calculate a critical discount factor using Expression

(1.2). The optimal collusive price is then given by

$$p_L^c(\delta) = \arg\max_{p_L^c} \pi_L^{c,p} \quad \text{s.t.} \quad \delta \geq \frac{\pi_L^{d,p} - \pi_L^{c,p}}{\pi_L^{d,p} - \pi_L^*},$$

where the collusive profit is given by $\pi_L^{c,p} = p_L^c/2$ and the competitive profit is the same as before.

The expression on the right-hand side of the constraint is strictly decreasing in the price p_L^c for $p_L^c > \tau$. To see this, we can take the derivative of the expression with respect to the collusive price and then compare the derivative to zero. This comparison reveals that the derivative is always negative for $p_L^c > \tau$. Note that τ is the competitive price level and that firms would never charge a price below this level because they would prefer to compete otherwise. Because the collusive profit is strictly increasing in p_L^c, this implies that the colluding firms will choose the largest possible price, such that the constraint binds with equality, that is, p_L^c is defined by

$$\delta = \frac{\pi_L^{d,p} - \pi_L^{c,p}}{\pi_L^{d,p} - \pi_L^*}.$$

Because of the strict monotonicity, p_L^c is also uniquely defined by this equality.

Finally, it remains to show that the optimal collusive price $p_L^c(\delta)$ is decreasing in the discount factor. As noted above, the right-hand side of the above equation is strictly decreasing in the price. We can think of δ as a function of p_L^c. Because of the strict monotonicity, this function is bijective and its inverse function exists, with the inverse being also strictly decreasing. This concludes the proof. □

Proof of Proposition 3. If firms collude on profit-maximizing prices under both pricing regimes (that is, if $\delta > \overline{\delta}_F$ and $\delta > \overline{\delta}_L$), profits are always larger with linear prices ($= v/2$) than with fixed prices ($= v/2 - \tau/8$). This also means that if collusion on profit-maximizing prices is sustainable with linear prices, but not with fixed prices (that is if $\delta < \overline{\delta}_F$ and $\delta > \overline{\delta}_L$), profits are larger with linear prices ($= v/2$) than with fixed prices ($< v/2 - \tau/8$).

Next, we look at the case in which collusion on profit-maximizing prices is sustainable with fixed prices, but not with linear prices (that is, if $\delta > \overline{\delta}_F$ and $\delta < \overline{\delta}_L$). With fixed fees, firms will set a price of $v - \tau/4$ and earn a profit of $v/2 - \tau/8$. Let us assume that with linear prices, firms would set the linear price to the same price level, that is, $p = v/2 - \tau/8$. Then, they would earn the same profit. It is straightforward to show that the critical discount factor that results from inserting this price in the formula stated in Lemma 5 is lower than $\overline{\delta}_F$. Because the critical discount factor stated in Lemma 5 is monotonically decreasing in the price, it follows that firms can increase the linear price until the critical discount factor equals the actual discount factor. This means that the profits will be larger with linear prices than with fixed prices.

Finally, we look at the case in which full collusion is never possible (that is, if $\delta < \overline{\delta}_F$ and $\delta < \overline{\delta}_L$). Let $f_F^c(\delta)$ and $p_L^c(\delta)$ be the optimal collusive prices that allow the firms to sustain collusion with fixed fees and linear prices for a given discount factor of δ. Then, each firm earns $f_F^c(\delta)/2$ with fixed fees and $p_L^c(\delta)/2$ with linear prices. This is because with fixed prices, customers do not mix and thus each firm covers half of the market. By contrast, with linear prices, all customers combine the products optimally, which also leads to a market share of 1/2. Because both profits result from dividing the corresponding prices by 2, it is sufficient to compare the levels of the optimal prices instead of the profits.

Lemma 5 does not provide a closed-form solution for $p_L^c(\delta)$, so that we cannot compare the resulting prices for a given discount factor. Instead, we will show that for a given price level p, the discount factor is lower with linear prices than with fixed fees; that is, if we rewrite $p = f_F^c(\delta)$ and $p = p_L^c(\delta)$ for δ, δ is larger in the first case with fixed fees than in the second case with linear prices. Because of the strictly monotonic relationship between the discount factors and the prices, this means that for a given discount factor δ, the price level is larger with linear prices than with fixed fees.

Lemma 5 provides the discount factor for the case of linear prices (set p_L^c to p). We can rewrite the optimal fixed fee from Corollary 2 and get

$$\delta = \begin{cases} \frac{f-2\tau}{2f-3\tau} & \text{if } \delta < \frac{1}{3} \\ \frac{f-\tau}{f+3\tau} & \text{if } \delta \geq \frac{1}{3}. \end{cases}$$

The comparison of the discount factors is straightforward and leads to the proposition. □

Proof of Corollary 3. The inequality $\overline{\delta}_T > \overline{\delta}_L$ is part of Proposition 1. Therefore, we only prove $\widetilde{\delta}_L = \overline{\delta}_T$. In the case of two-part tariffs, colluding firms set the linear price equal to the basic valuation, v, and the fixed fee to zero. Therefore, firms deviate from the same prices in the cases of linear prices and two-part tariffs. As a result, optimal deviation prices and profits are the same. Additionally, Hoernig and Valletti (2007) show that firms always use two-part tariffs in the case of competition. In summary, linear prices and two-part tariffs lead to the same profits in the cases of collusion, deviation, and punishment (competition). Thus, the critical discount factors are also equal. □

Proof of Lemma 6. First assume that firm 1 is the deviating firm and sets its price schedule (p, f) such that at least some customers mix and, hence, buy from both firms. In this case, $\underline{x} < \overline{x}$ holds. Firm 2 sticks to the collusive price schedule and sets $f_F^c = v - \tau/4$ (see Lemma 1). We can plug the prices into the aforementioned inequality and obtain

$$\begin{aligned} & \sqrt{\frac{v-\frac{\tau}{4}}{\tau}} - \frac{p}{2\tau} < 1 - \sqrt{\frac{f}{\tau}} - \frac{p}{2\tau} \\ \Leftrightarrow \quad & \sqrt{\frac{v}{\tau} - \frac{1}{4}} < 1 - \sqrt{\frac{f}{\tau}} \end{aligned}$$

The right hand side of the inequality always is smaller or equal to 1. At the same time, we can rearrange the left hand side and obtain

$$\sqrt{\frac{v}{\tau}-\frac{1}{4}} \geq 1 \quad \Leftrightarrow \quad \frac{v}{\tau}-\frac{1}{4} \geq 1 \quad \Leftrightarrow \quad \frac{v}{\tau} \geq \frac{5}{4} \quad \Leftrightarrow \quad \frac{4}{5}v \geq \tau$$

Assumption 1 ensures that the first equivalence sign is correct and the final inequality holds. Since the left hand side is always larger than or equal to 1, but the right hand side is less than or equal to 1, the initial inequality $\underline{x} < \overline{x}$ does not hold. In other words: it is not possible that the deviating firms sets a pricing schedule such that at least some customers buy from both firms.

Consider now the case where customers do not mix. As each customer buys one unit exclusively from one firm, (s)he has to pay the full linear price and not just a share of it. As a result, customers are indifferent between paying fixed and linear prices. For simplicity, we investigate the case where the linear price is equal to 0 first and find that the optimal fixed price is the same price that we already derived in Lemma 2, f_F^d. Because customers are indifferent between paying linear and fixed prices, we conclude that each combination of a linear price $p \geq 0$ and a fixed price $f \geq 0$ with $p + f = f_F^c$ is optimal and, hence, yields the optimal collusive profit. $\square$

Proof of Lemma 7. The critical discount factor results immediately from inserting the respective profits into Expression (1.2). Competitive profits are derived by Hoernig and Valletti (2007), whereas collusive and deviation profits are given by Lemma 1 and Lemma 6. $\square$

1.C Numerical Simulation

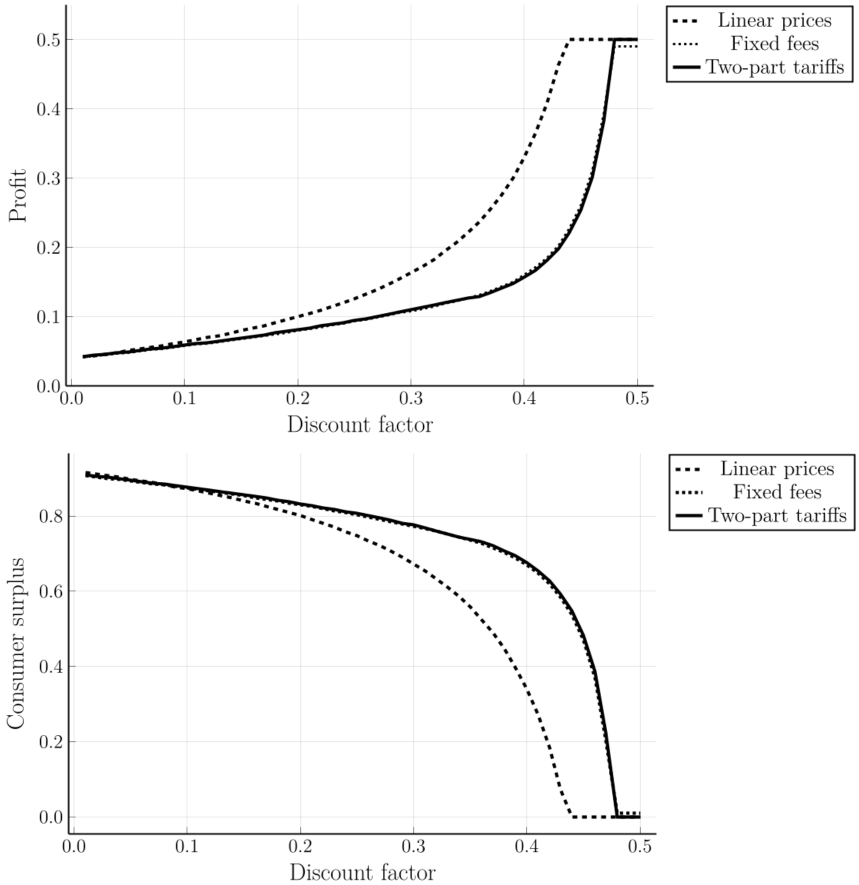


Figure 1.3: Profits and customer surpluses under partial and full collusion ($v = 1$, $t = 0.08$).

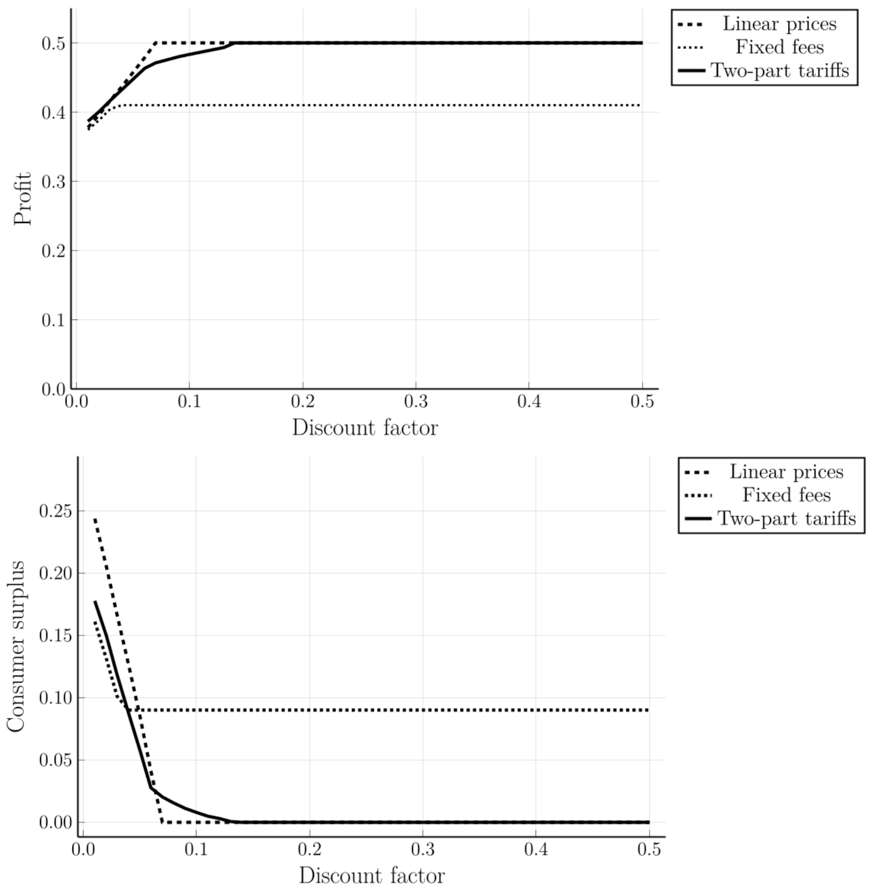


Figure 1.4: Profits and customer surpluses under partial and full collusion ($v = 1$, $t = 0.72$).

Bibliography

Abreu, D. (1986). Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory*, 39(1):191–225.

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica*, 56(2):383–396.

Abreu, D., Pearce, D., and Stacchetti, E. (1986). Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251–269.

Anderson, S. P. and Neven, D. J. (1989). Market efficiency with combinable products. *European Economic Review*, 33(4):707–719.

Bain & Company (2012). *Was Versicherungskunden wirklich wollen.*

Bos, I. and Marini, M. A. (2019). Cartel stability under quality differentiation. *Economics Letters*, 174:70–73.

Bos, I., Marini, M. A., and Saulle, R. D. (2020). Cartel formation with quality differentiation. *Mathematical Social Sciences*, 106:36–50.

Chang, M.-H. (1991). The effects of product differentiation on collusive pricing. *International Journal of Industrial Organization*, 9(3):453–469.

Chang, M.-H. (1992). Intertemporal product choice and its effects on collusive firm behavior. *International Economic Review*, 33(4):773–793.

Chen, H.-C. and Ritter, J. R. (2000). The seven percent solution. *Journal of Finance*, 55(3):1105–1131.

Colombo, S. (2010). Product differentiation, price discrimination and collusion. *Research in Economics*, 64(1):18–27.

Colombo, S. and Pignataro, A. (2022). Information accuracy and collusion. *Journal of Economics & Management Strategy*, 31(3):638–656.

d'Aspremont, C., Gabszewicz, J. J., and Thisse, J.-F. (1979). On Hotelling's "Stability in competition". *Econometrica*, 47(5):1145–1150.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies*, 38(1):1–12.

Gabszewicz, J. J., Laussel, D., and Sonnac, N. (2004). Programming and advertising competition in the broadcasting industry. *Journal of Economics & Management Strategy*, 13(4):657–669.

Gal-Or, E. and Dukes, A. (2003). Minimum differentiation in commercial media markets. *Journal of Economics & Management Strategy*, 12(3):291–325.

Gössl, F. and Rasch, A. (2020). Collusion under different pricing schemes. *Journal of Economics & Management Strategy*, 29(4):910–931.

Häckner, J. (1994). Collusive pricing in markets for vertically differentiated products. *International Journal of Industrial Organization*, 12(2):155–177.

Häckner, J. (1995). Endogenous product design in an infinitely repeated game. *International Journal of Industrial Organization*, 13(2):277–299.

Häckner, J. (1996). Optimal symmetric punishments in a bertrand differentiated products duopoly. *International Journal of Industrial Organization*, 14(5):611–630.

Hansen, R. S. (2001). Do investment banks compete in ipos?: the advent of the "7% plus contract". *Journal of Financial Economics*, 59(3):313–346.

Helfrich, M. and Herweg, F. (2016). Fighting collusion by permitting price discrimination. *Economics Letters*, 145:148–151.

Hoernig, S. H. and Valletti, T. M. (2007). Mixing goods with two-part tariffs. *European Economic Review*, 51(7):1733–1750.

Hoernig, S. H. and Valletti, T. M. (2011). When two-part tariffs are not enough: Mixing with nonlinear pricing. *B.E. Journal of Theoretical Economics*, 11(1).

Horstmann, N. and Krämer, J. (2013). Price discrimination or uniform pricing: Which colludes more? *Economics Letters*, 120(3):379–383.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, 39(153):41–57.

Igami, M. and Sugaya, T. (2021). Measuring the uncentive to collude: The vitamin cartels, 1990–99. *Review of Economic Studies*, 89(3):1460–1494.

Jorge, S. F. and Pires, C. P. (2008). Delivered versus mill nonlinear pricing with endogenous market structure. *International Journal of Industrial Organization*, 26(3):829–845.

Liu, Q. and Serfes, K. (2007). Market segmentation and collusive behavior. *International Journal of Industrial Organization*, 25(2):355–378.

Matutes, C. and Regibeau, P. (1992). Compatibility and bundling of complementary goods in a duopoly. *Journal of Industrial Economics*, 40(1):37–54.

Miklós-Thal, J. (2008). Delivered pricing and the impact of spatial differentiation on cartel stability. *International Journal of Industrial Organization*, 26(6):1365–1380.

OECD (1998). *Competition and Related Regulation Issues in the Insurance Industry.*

Peiseler, F., Rasch, A., and Shekhar, S. (2022). Imperfect information, algorithmic price discrimination, and collusion. *Scandinavian Journal of Economics*, 124(2):516–549.

Ross, T. W. (1992). Cartel stability and product differentiation. *International Journal of Industrial Organization*, 10(1):1–13.

Symeonidis, G. (1999). Cartel stability in advertising-intensive and R&D-intensive industries. *Economics Letters*, 62(1):121–129.

Chapter 2

A Bargaining Perspective on Vertical Integration

Coauthors: Geza Sapi and Christian Wey

Abstract:
We analyze vertical integration incentives in a bilaterally duopolistic industry with bargaining in the input market. Vertical integration incentives are a combination of horizontal integration incentives up- and downstream and depend on the strength of substitutability/complementarity and the shape of the unit cost function. Under particular circumstances, vertical integration can convey more bargaining power to the merged entity than a horizontal merger to monopoly. In a bidding game for an exogenously determined target firm, a vertical merger can dominate a horizontal one, while pre-emption does not occur.

Acknowledge:
This paper benefited from helpful comments from the Editor of the Canadian Journal of Economics, Zhiqi Chen, two anonymous referees, and seminar participants at Berlin Humboldt University, Düsseldorf Institute for Competition Economics, German Institute for Economic, Research (DIW Berlin) and the 2018 CISS in Montenegro.
The views expressed in this chapter are solely those of the authors and may not, under any circumstances, be regarded as representing an official position of the European Commission. This is personal research that is not related to any activity of the European Commission.
Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (project 235577387/GRK1974) is gratefully acknowledged.

2.1 Introduction

Competition policy traditionally looks at vertical and horizontal mergers from different perspectives. While horizontal mergers are often regarded as being motivated by the intent to reduce competition, it is more frequently argued that vertical integration is driven by efficiencies, for example, by eliminating double markups, reducing transaction costs or solving some variant of the holdup problem. This is explicitly stated in paragraph 11 of the EC non-horizontal merger guidelines, recognizing that *"[n]on-horizontal mergers are generally less likely to significantly impede effective competition than horizontal mergers,"* (European Union, 2008). A similar view emerges in the Vertical Merger Guidelines of the US Department of Justice and the Federal Trade Commission, noting that *"[v]ertical mergers [...] also raise distinct considerations [than horizontal mergers][...]. For example, vertical mergers often benefit consumers through the elimination of double marginalization, which tends to lessen the risks of competitive harm."*[1]

At least some of this sharp distinction between horizontal and vertical mergers may lie in the tradition of economic analysis to ignore the ability of downstream firms to influence upstream markets. Yet in perhaps most vertically-related industries, supply conditions are determined through bilateral bargaining, where downstream firms may have the ability to actively negotiate contracts with suppliers. Much research has been devoted to how horizontal integration can tip bargaining in favor of the merging parties. This research also gave rise to the recent heated debate on buyer-power in the antitrust arena. The question of how vertical integration can affect bargaining outcomes has, however, remained significantly less studied.

This article intends to take a step toward closing this gap. We investigate the driving forces behind vertical integration, its effects, and social desirability while taking into account that transactions between businesses in input markets arise as a result of bilateral bargaining. To focus on the shift in bargaining power from vertical integration, we apply a model that abstracts away from product market effects such as changes in prices.[2]

We provide conditions for vertical mergers to take place regarding the strength of substitutability or complementarity between final goods and the shape of the unit cost function. We then compare vertical to horizontal integration incentives, and find that vertical merger incentives are a combination of horizontal merger incentives up- and downstream, so that both types of mergers are closely related from a pure bargaining perspective.

We also analyze the strategic incentives of firms to merge in order to pre-empt a potentially harmful merger by a competitor. To investigate this question, we propose a bidding game in which an exogenously determined target firm is up for sale to the highest bidder, either to a horizontally or a vertically-related firm. We show that vertical merger incentives can be stronger than horizontal ones. Consequently, a horizontal merger to monopoly may

[1] See Section 1 of the Vertical Merger Guidelines (U.S. Department of Justice and The Federal Trade Commission, 2020).

[2] Isolating bargaining motivation of mergers from price effects is useful for reasons of tractability. We can also show that adding downstream competition leaves our qualitative results intact. Details are provided in Appendix 2.D.

convey less bargaining power to the merged entity than a vertical integration. In addition, we find that vertical mergers are never motivated by pre-emptive bargaining power considerations.

Our framework is relevant for the analysis of vertical mergers in concentrated input markets, where *both* suppliers and buyers have considerable bargaining power. One example is the US market for pay-TV, which is characterized by an oligopolistic structure on both market sides. Upstream firms produce video content that is licensed as TV channels to downstream firms, which in turn bundle and sell TV programs to households through various channels such as cable or satellite (see Rogerson, 2020; Shapiro, 2021, for details). It is well documented that the terms of supply (e.g., licensing fees) are determined by bargaining between up- and downstream firms (see Salop, 2018; Rogerson, 2020; Shapiro, 2021). Consistent with this, the media has documented temporary bargaining breakdowns.[3] The market faced multiple vertical mergers in the two past decades, including the mergers of News Corp/DirecTV, Comcast/NBCU, and AT&T/Time Warner. For example, in the AT&T/Time Warner case, several Time Warner subsidiaries (e.g., Warner Bros.) produced TV content that they sold to the AT&T subsidiary DirecTV and DirecTV's competitors like Comcast or Charter. Thus, the vertical merger led to a structure where the integrated upstream firms provide TV content not only to their integrated partner DirecTV, but also negotiate with downstream rivals about the provision of their content.

As discussed by Rogerson (2020, pp. 411), since *"bargaining power is so clearly present on both sides of the market, it is not surprising that government authorities have begun to focus their analysis on competitive theories of harm that take the effects of bargaining into account."* While the Federal Communications Commission based its arguments largely on the traditional raising rivals' costs (RRC) theory and input foreclosure paradigms in the News Corp/DirecTV case in 2003, these theories have been successively challenged in subsequent merger cases by a bargaining-based theory, called bargaining leverage over rivals (BLR) theory (Rogerson, 2020). The BLR effect arises because a vertical merger increases the disagreement payoff of the upstream firm, which induces higher retail prices to the detriment of consumers. If bargaining between the supplier and a non-integrated retailer breaks down, the downstream affiliate will earn some extra profit if the input is completely withheld from the rival retailer. Because of vertical integration, the integrated supplier internalizes this positive effect. This improves its threat point and thus bargaining power vis-a-vis the non-integrated retailer.

Two important features distinguish the BLR theory from the RRC theory. First, while the RRC effect arises in setups where bargaining power only exists on the upstream market side, the BLR theory arises from a bargaining framework in which both sides can have (some) bargaining power. Second, the BLR effect does not hinge on the assumption that upstream prices are set before downstream prices, which is required by the RRC theory. Rogerson (2020) also derives a simple statistic, called vertical GUPPI, that can be estimated based

[3]See, for example, the newsflash "TV tussle: DirecTV, Tegna dispute turns TV channels dark in 51 markets including Houston, Seattle," which can be found at `https://eu.usatoday.com/story/tech/2020/12/02/tv-directv-tegna-dispute-results-channel-outages-51-markets/3794473001/`.

on widely available data and that can be used to assess the potential harm due to the BLR effect pre-merger. In the context of our paper, it is worth noting that our effects can also be thought of as originating from a change in the bargaining leverage over rivals. However, while in Rogerson (2020) this effect is related to competitive externalities in the downstream market, our effects do not originate from downstream competition.

In our model of two upstream and two downstream (monopoly) retailers, we get two additional effects of vertical integration which also create bargaining leverage over rivals. First, the upstream supplier's disagreement point when bargaining with the non-integrated retailer is improved when unit costs are increasing. A breakdown in the bargaining then allows the integrated firm to realize an extra profit from increasing sales at the affiliated retailer. Second, the integrated retailer benefits from an improved disagreement point when bargaining with the non-integrated supplier, whenever the goods are substitutes. Here, a breakdown in the bargaining creates an extra profit resulting from the demand increase when the rival supplier's good is not available. Vertical integration leads to a better internalization of this extra profit, while it remains incomplete under separation.

Rogerson (2020) and our paper are part of a large strand of literature analyzing vertical merger incentives. Given that up- and downstream firms have (at least some) market power, vertical integration can be privately and socially desirable because of its potential to reduce the double mark-up problem which arises under linear wholesale prices. However, Luco and Marshall (2020) provide a recent empirical analysis showing that the elimination of double marginalization through vertical integration can also raise anti-competitive concerns in multi-product industries. Other economic theories focus on the anti-competitive effects of vertical mergers by referring to foreclosure and raising rivals' costs effects (see, for example, Salinger, 1988; Ordover et al., 1990; Inderst and Valletti, 2011) or on the use of vertical integration to solve commitment problems (see, for example, Hart et al., 1990).[4] Finally, other contributions discuss firms' incentives to stay vertically separated (see, for example, Bonanno and Vickers, 1988).

An important representative of this strand of literature for our work is De Fontenay and Gans (2005b). They focus on vertical merger incentives in a bargaining framework similar to ours and compare outcomes under upstream competition and monopoly. We extend their analysis to complementary final goods and decreasing unit costs. Doing so yields markedly different results for vertical merger incentives, two of which stand out. First, in their baseline model (with no competitive externalities downstream) and given upstream competition, vertical integration is always preferred to non-integration. Our analysis confirms this result for the particular case of substitute goods and increasing unit costs, but we obtain different results for complementary goods and/or decreasing unit costs. Second, they show that vertical integration incentives are larger under upstream competition than under upstream monopoly, while we show that the impact of upstream competition on vertical integration

[4] The approach of Hart et al. (1990) has been extended to analyze the effects of vertical integration on investment incentives (see, for example, Bolton and Whinston, 1993; Stole and Zwiebel, 1996; Baake et al., 2004; Choi and Yi, 2000). See also Chen (2019) for a recent analysis of how changes in bargaining power affect the incentives of an upstream firm to invest in quality and product variety.

incentives can go either way.

Finally, two other papers are worth mentioning. First, our analysis builds on the model of Inderst and Wey (2003) who analyze horizontal merger incentives up- and downstream, as well as the choice of a manufacturer between two technologies influencing production costs. One of their main findings is that upstream merger incentives depend on the substitutability/complementarity, while downstream merger incentives depend on the shape of the cost function. Our analysis reveals that the same incentives are present in vertical merger considerations, so that vertical merger incentives can be regarded as a combination of merger incentives up- and downstream.

Second, our paper provides a novel perspective on Segal (2003), who discusses various contracts among substitute and complementary firms in the context of cooperative games with random-order values. Our definition of mergers corresponds to what Segal (2003) refers to as collusion. Segal shows that a merger between substitutes likely hurts non-indispensable outsiders, while a merger between complements benefits them. Our model generates additional insights by assigning control of different resources to different firms. While in Segal (2003) firms only differ in terms of the value they generate to the industry as a whole, in our model these differences are systematic for upstream and downstream firms, i.e., suppliers control production and retailers are gatekeepers to consumers, and hence control demand. This gives rise to different incentives for horizontal and vertical mergers depending on the shape of average costs and demand.

The remaining article proceeds as follows: Section 2.2 introduces the model. In Section 2.3, we apply the framework to analyze vertical merger incentives. Section 2.4 compares horizontal and vertical merger incentives in more detail and derives conditions determining which of these incentives is strongest. Finally, Section 2.5 concludes. All proofs are in Appendix 2.A.

2.2 Model

Our setup follows Inderst and Wey (2003), and extends their analysis to vertical mergers. Consider an industry with two upstream suppliers $s \in S^0 = \{A, B\}$ and two downstream retailers $r \in R^0 = \{a, b\}$. We denote the set of all firms by $\Omega = S^0 \cup R^0$ and subsets by Ψ.

Each supplier controls the production of one input, with inputs being differentiated. Supplier s incurs costs of production which are given by $C_s(q_{sr} + q_{sr'})$ where q_{sr} is the quantity exchanged between s and r. We use primes (s' and r') to refer to the alternative supplier and retailer, respectively. We allow the average unit costs, given by $\overline{C}_s(q) = C_s(q)/q$, to be either strictly increasing or decreasing in q[5].

Downstream retailers procure inputs from the suppliers and turn them into final goods that they sell to final consumers. For simplicity, we assume that one unit of an input is turned into one unit of a final good. Since the inputs are differentiated, the final goods are

[5]Our analysis is also relevant for the case where average costs are U-shaped. We discuss this issue in footnote 12.

also differentiated.

Demand at the retailers is independent, hence, there are no competitive externalities downstream.[6] This means that changes in the industry structure affect only the distribution of rents, but not product market outcomes such as the (input and output) quantities, prices or the total surplus generated. This is an important simplification: while it abstracts away from short run price effects, which are typically a key concern in antitrust analysis, doing so also allows us to isolate the pure bargaining effects of various vertical and horizontal mergers. We relax the assumption in Appendix 2.D and show that our main finding remains intact if we allow downstream externalities.

Retailer r faces the indirect demand function $p_{sr}(q_{sr}, q_{s'r})$ when selling the final good produced from the input of supplier s. We consider cases where the two final goods are either substitutes or complements at each outlet.

The degree of substitutability/complementarity and the degree of strictly increasing/decreasing unit costs will be the important determinants in our analysis. To simplify the presentation of our results, let $\Delta_p(q) := p_{sr}(q,q) - p_{sr}(q,0)$ and $\Delta_C(q) := \overline{C}_s(2q) - \overline{C}_s(q)$ for $q > 0$. If final goods are strict complements (substitutes), then $\Delta_p(q) > 0$ ($\Delta_p(q) < 0$) for all $q > 0$ and we simply write $\Delta_p > 0$ ($\Delta_p < 0$). Similarly, if unit costs are strictly increasing (decreasing), then $\Delta_C(q) > 0$ ($\Delta_C(q) < 0$) for all $q > 0$ and we write $\Delta_C > 0$ ($\Delta_C < 0$).

Some of our results rely on a comparison between $\Delta_p(q)$ and $\Delta_C(q')$ for particular $q, q' > 0$ that result from the corresponding proofs. To simplify the notation, we omit the arguments in the main text and use the notation $\Delta_p \gtreqless \Delta_C$.

The retailers incur no other costs than the costs of buying the goods. Supply contracts between upstream and downstream firms are determined by bargaining, and involve lump sum transfers that do not impact product market outcomes. This means that firms use efficient contracts and double marginalization does not occur. We follow other authors[7] studying the effects of integration in a bargaining framework and adopt the Shapley value as a solution concept of the bargaining game.

The Shapley value allocates to each independently negotiating party its expected marginal contribution to coalitions, where the expectation is taken over all coalitions in which the party may belong, with all coalitions assumed to occur with equal probability. Formally, let Ψ denote the set of independently negotiating parties and $|\Psi|$ the cardinality of this set. The payoff of firm $\psi \in \Psi$ is given by

$$U_\psi^\Psi = \sum_{\widetilde{\Psi} \subseteq \Psi \,|\, \psi \in \widetilde{\Psi}} \frac{\left(\left|\widetilde{\Psi}\right| - 1\right)! \left(|\Psi| - \left|\widetilde{\Psi}\right|\right)!}{|\Psi|!} \left[W_{\widetilde{\Psi}} - W_{\widetilde{\Psi} \setminus \psi}\right], \tag{2.1}$$

[6] For example, we can think of retailers operating in different geographic markets.

[7] Examples include Hart and Moore (1990); Stole and Zwiebel (1996); Rajan and Zingales (1998); Inderst and Wey (2003); Segal (2003); De Fontenay and Gans (2005b); Montez (2007) and Kranton and Minehart (2000). While the Shapley value is an axiomatic solution concept, there are numerous justifications for the Shapley value as an outcome of a non-cooperative bargaining processes (see, e.g., Gul 1989; Stole and Zwiebel 1996; Inderst and Wey 2003; De Fontenay and Gans 2005a,b, and Winter (2002) for a survey).

where $\widetilde{\Psi} \subseteq \Psi \,|\, \psi \in \widetilde{\Psi}$ represents a set $\widetilde{\Psi} \subseteq \Psi$, such that ψ is a member of coalition $\widetilde{\Psi}$, and $W_{\widetilde{\Psi}}$ denotes the maximum surplus achieved by the firms in coalition $\widetilde{\Psi}$. For simplicity, we write $\widetilde{\Psi}\backslash\psi$ for $\widetilde{\Psi}\backslash\{\psi\}$. The maximum industry profit is given by

$$W_\Omega(\{q_{sr}\}_{sr\in S^0\times R^0}) = \sum_{r\in R^0} [p_{Ar}(q_{Ar}, q_{Br})q_{Ar} + p_{Br}(q_{Br}, q_{Ar})q_{Br}] - \sum_{s\in S^0} C_s(q_{sa} + q_{sb}).$$

The maximum surplus of a coalition follows from the maximum industry profit by only considering the links between members of the coalition. For example, the coalition $\Omega\backslash A$ does not include supplier A and, hence, the links between A and the two retailers are missing. This means that supplier A cannot provide the retailers with inputs ($q_{Aa} = q_{Ab} = 0$). Analogously, the coalition $\Omega\backslash a$ has no links with retailer a and, hence, this retailer has no access to inputs, i.e., $q_{Aa} = q_{Ba} = 0$.

In the terminology of cooperative game theory $W(\cdot)$ is often referred to as the *characteristic function.* W_Ψ is assumed to be continuous, strictly quasi-concave for all $\Psi \subseteq \Omega$ and superadditive,[8] i.e., $W_\Psi \geq W_{\widetilde{\Psi}}$ for every Ψ and $\widetilde{\Psi}$ with $\widetilde{\Psi} \subset \Psi \subseteq \Omega$. Importantly, since at least one supplier and retailer is necessary for production, $W_\Psi = 0$ if Ψ does not contain at least one firm from each market side.

The Shapley value corresponds to the idea that in bargaining, a party should reap its marginal contribution to an existing agreement between other parties. However, the marginal contribution of a firm depends on the agreements already in place between other firms. In a well-known interpretation of the Shapley value, players are randomly ordered in a sequence. Since several random orderings are possible, each of them is assumed to be equally likely. Each player gets as a payoff its marginal contribution to the coalition formed by the preceding players in the sequence. The Shapley value is the expected payoff taken over all possible orderings.[9]

To see why this interpretation applies to formula (2.1), we can split (2.1) into three components. The first component is the sum operator which iterates over all possible coalitions to which firm ψ may marginally contribute. The third expression—the expression in brackets—is the marginal contribution of firm ψ, i.e., the difference in industry profits with and without firm ψ. Finally, the second component is the fraction and needs more attention.

The fraction may seem to be complicated at first glance, but it has a relatively simple interpretation. First, it is important to note that, in mathematics, the factorial of a set can be used to denote the number of possible orderings. This means that if a set contains n players, there are $n!$ different ways to order them. In the context of the Shapley value, the different orderings describe which party joins the coalition in which position. Since we focus on the coalitions to which firm ψ contributes marginally, we know that firm ψ comes last.

[8]Superadditivity means that the marginal contribution of an arbitrary firm to an arbitrary coalition is non-negative. To ensure that this assumption is met we assume that downstream markets are independent and contracts are efficient. In principle, one could choose less restrictive assumptions. For instance, if retailers were sufficiently differentiated, this assumption would also hold in the presence of downstream competition.

[9]We provide an example in Appendix 2.B.

The remaining $|\Psi|-1$ firms in the coalition can be ordered in $(|\Psi|-1)!$ ways. Similarly, the second part of the numerator describes the number of orderings of the parties outside of the coalition. Taken together, the numerator describes the number of orderings in which a fixed set of firms enters a coalition, with firm ψ entering last. By dividing this expression by the number of all orderings and assuming that all orderings occur with the same probability, we get the likelihood of such an event.

Before we turn to the analysis, we introduce the symmetry assumption that we use in some parts of our analysis to derive clear-cut results. Note that the assumption is not necessary for all results and will be explicitly invoked at various segments of the text.

Assumption 2 (Symmetry)**.** *Suppliers and retailers are symmetric:* $C_s(\cdot) = C_{s'}(\cdot) = C(\cdot)$, $q_{sr} = q_{s'r'}$ *and* $p_{sr}(\cdot) = p_{s'r'}(\cdot)$ *for any* $s, s' \in S^0$ *and any* $r, r' \in R^0$.

2.3 Vertical Merger Incentives

The first part of our analysis is concerned with the derivation of the vertical merger incentives. For this purpose, it is important to be clear about what we mean by a merger. Throughout this paper, we consider a merger as combining two otherwise independent bargaining units into a single firm. Whereas under non-integration each supplier and retailer bargains separately, under integration the negotiations of the merged entity are controlled by one common agent, which reduces the number of negotiating parties by one. This is a realistic way to think about mergers in which the merged firms are united under a common management, which conducts negotiations with other entities. It would happen, for example, if the key executives of the acquired company were replaced by the new owner.[10]

We can now calculate equilibrium payoffs under different market structures. We use the notation $\{s, s', r, r'\}$ to denote a market structure, where the commas separate non-merged and therefore individually negotiating entities. For example, $\{AB, a, b\}$ stands for the market structure with an upstream monopoly facing a duopoly of retailers. Similarly, $\{Aa, B, b\}$ denotes the market structure consisting of supplier A being vertically integrated with retailer a, and supplier B as well as retailer b negotiating independently. For each market structure, the profits of the negotiating parties are immediately given by the Shapley value. Appendix 2.C provides an overview of all payoffs under the market structures that are relevant for our analysis. By comparing pre- to post-merger payoffs, we can then derive the vertical integration incentives for various pre-merger market structures.

Proposition 5. *Whether a vertical merger between supplier* s *and retailer* r *increases their joint payoff depends on the pre-merger market structure in the following way:*

(i) If suppliers are integrated and retailers are separated $(\Psi = \{AB, a, b\})$*, the joint profit of supplier* AB *and retailer* r *weakly increases by vertically merging if* $W_{\Omega\backslash r} + W_{\Omega\backslash r'} \geq$

[10]Note that this definition differs from the one in De Fontenay and Gans (2005b). They follow the property rights literature (Grossman and Hart, 1986; Hart and Moore, 1990) and distinguish between the owner and the manager of a firm. After a merger, the manager of a purchased entity remains indispensable in further negotiations and acts as an independent negotiating party.

W_Ω, whereas it decreases if the opposite holds.

(ii) If suppliers are separated and retailers are integrated ($\Psi = \{A, B, ab\}$), the joint profit of supplier s and retailer ab weakly increases by vertically merging if $W_{\Omega\setminus s} + W_{\Omega\setminus s'} \geq W_\Omega$, whereas it decreases if the opposite holds.

(iii) If suppliers and retailers are non-integrated ($\Psi = \{A, B, a, b\}$), the joint profit of supplier s and retailer r weakly increases by vertically merging if

$$\left(W_{\Omega\setminus s'r'} - W_{\Omega\setminus sr}\right) + W_{\Omega\setminus s} + W_{\Omega\setminus r} \geq W_\Omega, \tag{2.2}$$

whereas it decreases if the opposite holds.

In order to give an economic interpretation for Proposition 5, the following corollary connects the conditions stated in Proposition 5 with the economic fundamentals.

Corollary 6. *Vertical merger incentives depend on the initial market structure, the degree of substitutability or complementarity between the final goods and the shape of the unit cost function in the following way:*

(i) With suppliers integrated and retailers separated ($\Psi = \{AB, a, b\}$), a vertical merger between supplier AB and retailer r takes place (does not take place) if both suppliers have strictly increasing (decreasing) unit costs.

(ii) With suppliers separated and retailers integrated ($\Psi = \{A, B, ab\}$), a vertical merger between supplier s and retailer ab takes place (does not take place) if the final goods are strict substitutes (complements).

(iii) Invoke Assumption 2 (symmetry) and take the scenario with all firms separated ($\Psi = \{A, B, a, b\}$). Supplier s and retailer r merge (stay separated) if $\Delta_p < \Delta_C$ ($\Delta_p > 0$ and $\Delta_C < 0$).

We now provide some intuition on vertical merger incentives. First take the pre-merger case of a monopolist retailer (downstream) facing separated suppliers upstream. In this situation, vertical integration between the retailer and one supplier is profitable for the merging parties if the final goods are substitutes. Why is this so? It is convenient to focus on the effects of integration on the non-merged supplier: Since only the distribution of payoffs is affected, not overall output, any gains of the merging parties must correspond exactly to the losses of the non-merged supplier.

If final goods are substitutes, each supplier wants to be the first to reach an agreement with the retailer. This is because the bargaining between a supplier and the retailer revolves around the sharing of the marginal rent generated by the negotiating parties: With final goods being substitutes, the additional rent generated by the first supplier to reach an agreement with the retailer is larger than that generated by the second supplier. Therefore, suppliers prefer negotiating over infra-marginal input quantities to bargaining "on the margin." This explains why, with substitutes, the non-merging supplier loses if the other market actors integrate vertically.

With vertical integration between the retailer and the rival upstream firm, the non-merging supplier cannot be the first to reach an agreement with the retailer, because vertical integration guarantees that an agreement between the rival and the retailer is in place. The non-merging supplier is left with having to bargain at the margin over the lower surplus it generates by coming second to the retailer.

The same logic holds if final goods are complements. In that case, each supplier prefers to be the second in reaching an agreement with the retailer: Complementary final goods imply that the additional surplus generated by the second supplier to reach an agreement with the retailer is larger than that generated by the first one, because adding a complement to the market boosts demand for *both* final goods. Vertical integration with complements would ensure that the integrated supplier cannot be the second to reach an agreement with the integrated retailer. This would benefit the non-merging party and therefore harm the firms considering integration.

Take now the situation in which a monopoly supplier negotiates pre-merger with two retailers. Vertical integration between the supplier and a retailer takes place if unit costs are strictly increasing. The reasoning is as follows: If unit costs are strictly increasing, each retailer prefers to be the first to reach an agreement with the supplier, i.e., to negotiate over infra-marginal input quantities. The retailer coming second faces higher unit costs and is therefore left with a smaller surplus over which to negotiate with the supplier. Vertical integration corresponds to a sure agreement between the integrated upstream and downstream firms, leaving the non-merging retailer with being the second as the only option. This erodes the bargaining power of the second retailer and therefore benefits the merging parties.

If unit costs are strictly decreasing, each retailer prefers to be the second to reach an agreement with the supplier and to negotiate for the marginal input quantities. Once a supplier-retailer agreement is in place, the additional rent generated by another retailer is larger since unit costs decrease with the input quantity needed to supply that retailer. In this case, a vertical merger is not attractive since it forces the integrated retailer to be the first.[11] [12]

Finally, we explain the intuition behind vertical integration incentives under pre-merger full separation. We focus on the most instructive case, namely when all firms are symmetric as assumed in Corollary 6, and postpone discussing the role of asymmetry to later. Under such circumstances, vertical merger incentives correspond to a mix of vertical integration

[11]An interesting question is whether an integrated firm could commit to not supplying its own retail entity until an agreement with another retailer is in place. We are not aware of such a practice in the context of bargaining.

[12]The mechanisms of our analysis also apply to the case where average costs are U-shaped. The vertical merger incentive is driven by a comparison of unit costs at a high output level, where all downstream firms are served, and at a low output level, where only one downstream firm is served. When average costs are U-shaped, its functional form needs to be known in order to make such a comparison. If unit costs are smaller at the low output level than at the high output level, a downstream rival is harmed by vertical integration. If the opposite holds, the rival benefits.

incentives under an upstream and downstream monopoly. These incentives can point in different directions. Whether incentives for vertical integration arise therefore depends on the relative strength of these forces.

With all firms initially separated, whether a vertical merger is profitable or not depends on the degree of complementarity or substitutability of the final goods compared to how strong unit costs increase or decrease. This relationship is illustrated in Figure 2.1. The strength of complementarity and substitutability is captured by Δ_p while the extent to which unit costs increase or decrease is measured by Δ_C.

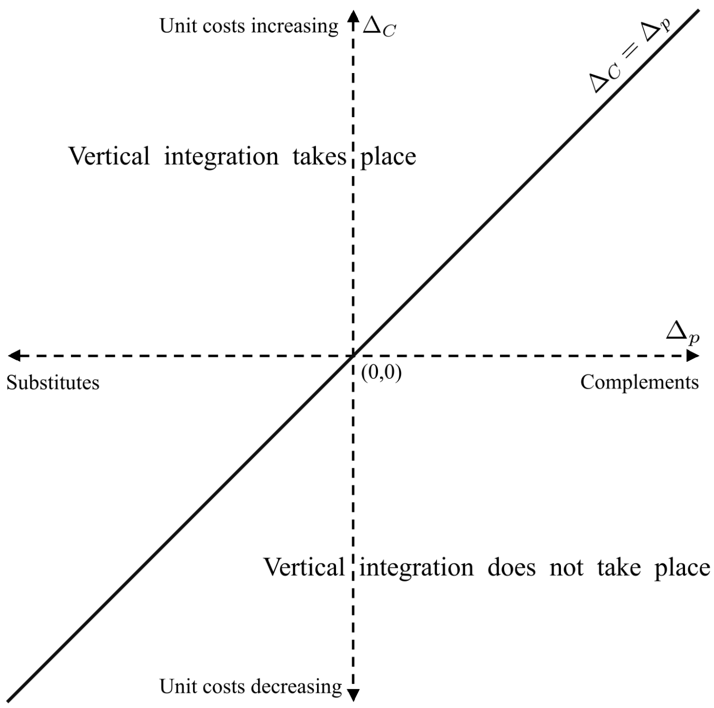


Figure 2.1: Vertical integration incentives

A vertical merger implies that the integrated firms are always the first to reach an agreement with each other. If this is what they would want in the absence of the merger, integration is unambiguously profitable. This is the case when final goods are substitutes ($\Delta_p < 0$) and unit costs are increasing ($\Delta_C > 0$). If unit costs are increasing, retailers want to be the first to reach an agreement with each supplier. Being the second means having to negotiate for a lower surplus because unit costs are higher for the additional input quantity to be supplied. If final goods are substitutes, suppliers prefer to be the first to reach an agreement with each retailer. The supplier that comes second must take into account the negative price externality it is imposing on the other final good and, hence, is left to negotiate for a lower surplus. In summary, with substitutes and strictly increasing unit costs, both retailers and suppliers prefer to be the first to reach an agreement with firms on the other market side. This is exactly what a vertical merger guarantees and is therefore unambiguously profitable.

The logic is the same for why vertical mergers are not preferred if final goods are complements ($\Delta_p > 0$) and unit costs are strictly decreasing ($\Delta_C < 0$). Under such circumstances, both retailers and suppliers prefer to be the second to reach an agreement with firms on the other market side, because that is when their marginal contribution is largest. A vertical merger undermines this opportunity because it guarantees being the first to reach an agreement and is therefore unprofitable with no ambiguity.

Interesting situations arise when final goods are substitutes (complements) and unit costs are strictly decreasing (increasing). In these cases, the interests of the suppliers and retailers are not aligned. For example, with substitutes and strictly decreasing unit costs, suppliers prefer to be the first to reach an agreement with each retailer, whereas retailers want to be the second. Since vertical integration implies that the merging parties are always the first to reach an agreement with each other, it benefits the merging supplier but harms the merging retailer. The profitability of such a merger therefore depends on whether the gains of the former exceed the losses of the latter. This is the case if final goods are sufficiently strong substitutes while unit costs are sufficiently slowly decreasing. The same logic applies in reverse if final goods are complements and unit costs are strictly increasing.

In the discussion about vertical integration incentives under pre-merger full separation, we remained silent on the role of asymmetries between firms. We address this issue now. While all of what has been said so far stays valid, asymmetries between firms have some implications for vertical merger incentives. According to Claim (iii) of Proposition 5, vertical integration between supplier s and retailer r is profitable if

$$\left(W_{\Omega\setminus s'r'} - W_{\Omega\setminus sr}\right) + W_{\Omega\setminus s} + W_{\Omega\setminus r} \geq W_{\Omega}. \tag{2.3}$$

Under symmetry, the term in brackets cancels out, but not under asymmetry. Expression (2.3) implies that vertical integration is more likely to take place if the vertically integrated firm is relatively large compared to the non-merging ones (i.e., if the difference $W_{\Omega\setminus s'r'} - W_{\Omega\setminus sr}$ is large). This is the case if the vertically integrated firms s and r are able to produce a relatively large surplus on their own compared to the surplus produced by the non-merging firms s' and r', which rely solely on each other. This is more likely if final goods are substitutes and unit costs are increasing.[13] While the thresholds for vertical integration to take place depicted in Figure 1 may shift to the North-West, the qualitative result behind Figure 1 remains intact: vertical integration incentives are stronger when unit costs increase fast and final goods are stronger substitutes.

Finally, it remains to note that in our setup, vertical integration incentives are not unambiguously larger under upstream competition than under monopoly. This is especially true in the case of substitutes and strictly increasing unit costs. Therefore, our findings are in contrast to the results derived by De Fontenay and Gans (2005b), who find that vertical integration incentives are always stronger with upstream competition in the aforementioned case. To see this, we can compare the conditions for vertical integration under both market

[13] This is the combination when inframarginal surplus is the largest. The merged firm is guaranteed this inframarginal surplus without negotiation.

structures as given in Claims (i) and (iii) of Proposition 5.

Vertical integration incentives are stronger under upstream monopoly than under competition if

$$W_{\Omega\setminus r} + W_{\Omega\setminus r'} > \left(W_{\Omega\setminus s'r'} - W_{\Omega\setminus sr}\right) + W_{\Omega\setminus s} + W_{\Omega\setminus r}, \tag{2.4}$$

whereas they are weaker if the opposite holds. To demonstrate that arrangements exist in which vertical integration incentives under an upstream monopoly are stronger than under competition, we focus on the case of symmetry. Then, condition (2.4) reduces to $W_{\Omega\setminus r} > W_{\Omega\setminus s}$, which holds if an additional retailer increases total surplus by a relatively small amount, while the marginal contribution of a supplier is rather large. This is likely to be the case, for example, if unit costs are strongly increasing while final goods are relatively weak substitutes (or even complements). Upstream competition can therefore either enhance or reduce the incentives for vertical integration.

2.4 Comparing Vertical and Horizontal Merger Incentives

In this section, we compare vertical and horizontal merger incentives based purely on bargaining power considerations. Throughout this section, we invoke Assumption 2 (symmetry) to obtain clear-cut results.

We proceed in three steps: We first explain horizontal merger incentives. Since in this case our model corresponds to Inderst and Wey (2003), we summarize their results on horizontal integration and then explain why vertical merger incentives are a combination of up- and downstream merger incentives. Second, we compare the gains from horizontal and vertical mergers. Third, we analyze a bidding game where up- and downstream firms bid for an exogenously picked target firm (either a supplier or a retailer).

2.4.1 Horizontal Mergers

Inderst and Wey (2003) derive conditions under which horizontal mergers are profitable from the perspective of bargaining power. Adapting Corollary 1 of Inderst and Wey (2003), retailers merge if

$$W_{\Omega\setminus a} + W_{\Omega\setminus b} > W_{\Omega}, \tag{2.5}$$

whereas they stay separated if the inequality is reversed. Similarly, suppliers merge if

$$W_{\Omega\setminus A} + W_{\Omega\setminus B} > W_{\Omega}, \tag{2.6}$$

and they stay separated if the opposite holds.

This implies that upstream firms merge (stay separated) if final goods are strict substitutes (complements), while downstream firms merge (stay separated) if upstream firms have strictly increasing (decreasing) unit costs (Proposition 2 of Inderst and Wey, 2003).

It should be noted that merger incentives on each market side are independent of whether firms are merged or not on the other side.

Since vertical merger incentives are affected by the same economic determinants, we conclude that they can be regarded as a combination of horizontal integration incentives up- and downstream. The intuition behind this result is as follows: Depending on the substitutability or complementarity as well as the shape of the unit cost function, firms on each market side want to finish their negotiations with firms on the other market side either first or second. A horizontal merger ensures an agreement with both firms on the other market side, because the merged entity becomes a monopolist and is therefore indispensable. A vertical merger ensures an agreement only with one firm on the other side of the market. However, contrary to a horizontal merger, a vertical merger involves firms from both market sides, so that there is a coexistence of integration incentives up- and downstream.

2.4.2 Comparison of Horizontal and Vertical Merger Gains

We turn to the comparison of horizontal and vertical merger gains and define the gain of a merger Δ_x, $x \in \{U, D, V\}$ as the difference in the joint pre- and post-merger profits of the merging firms. The subscript U refers to an upstream merger, D to a downstream merger and V to a vertical merger.

$$\begin{aligned}
\Delta_U &= U_{AB}^{\{AB,a,b\}} - U_A^{\{A,B,a,b\}} - U_B^{\{A,B,a,b\}} \\
\Delta_D &= U_{ab}^{\{A,B,ab\}} - U_a^{\{A,B,a,b\}} - U_b^{\{A,B,a,b\}} \\
\Delta_V &= U_{Aa}^{\{Aa,B,b\}} - U_A^{\{A,B,a,b\}} - U_a^{\{A,B,a,b\}}
\end{aligned}$$

Note that we added a superscript to the payoff U_i in order to distinguish between the different market structures under which payoffs are computed. Moreover, we focus on a vertical merger between supplier A and retailer a since firms on both market sides are symmetric.

The result that vertical merger incentives are a combination of up- and downstream incentives leads directly to a conclusion about the ordering of merger gains. Vertical merger incentives consist equally of upstream and downstream horizontal merger incentives. However, each horizontal merger incentive enters at only half strength because only one firm is directly affected. As long as horizontal merger incentives upstream and downstream are not equally strong, vertical integration incentives must be strictly between the upstream and downstream merger incentives.[14] Proposition 6 summarizes this conclusion.

Proposition 6. *The gains from horizontal upstream, horizontal downstream, and vertical mergers are ordered as follows:*

$$\Delta_U \gtreqless \Delta_V \gtreqless \Delta_D \quad \Leftrightarrow \quad W_{\Omega\setminus s} \gtreqless W_{\Omega\setminus r}.$$

[14]In the special case of equally strong horizontal merger incentives, vertical incentives will be equal as well, and firms are indifferent between all types of mergers.

The following corollary links the condition $W_{\Omega\setminus s} \gtrless W_{\Omega\setminus r}$ to the primitives of our model.

Corollary 7. *The following implications hold for all $s \in S^0$ and $r \in R^0$.*

$$\begin{aligned} -\Delta_p < \Delta_C \quad &\Rightarrow \quad W_{\Omega\setminus s} < W_{\Omega\setminus r} \\ -\Delta_p > \Delta_C \quad &\Rightarrow \quad W_{\Omega\setminus s} > W_{\Omega\setminus r} \end{aligned}$$

If final goods are substitutes ($\Delta_p < 0$) and unit costs are decreasing ($\Delta_C < 0$), suppliers want to merge, while retailers want to stay separated. In other words, the gain of a horizontal upstream merger is positive, while the gain of a downstream merger is negative. Thus, the incentive for the suppliers to merge is the strongest and the incentive for the retailers is the weakest. Analogously, the order is reversed if final goods are complements ($\Delta_p > 0$) and unit costs are increasing ($\Delta_C > 0$).

In the case of substitutes ($\Delta_p < 0$) and strictly increasing unit costs ($\Delta_C > 0$), the gains of both horizontal up- and downstream mergers are positive, such that the ratio of the strengths of both integration incentives determines the ordering. This is similar to the case of complements ($\Delta_p > 0$) and strictly decreasing unit costs ($\Delta_C < 0$) in which both merger gains are negative.

2.4.3 Bidding Game

Can bargaining incentives drive horizontal and vertical mergers to prevent a takeover by others? And which firm can be expected to prevail in a takeover auction? We investigate these questions in a bidding game, where a single firm is up for sale to the highest bidder in the industry. Bidders evaluate their gain from winning the auction against the possible outcomes when not winning the auction. In the latter case, the counterfactual becomes another firm potentially taking over the target. This has implications for bidding incentives.

We assume that one firm, either up- or downstream, is up for sale. This firm will be referred to as the target firm. The other firms in the market bid to acquire the target, which is sold to the highest bidder. We also consider the existence of an outside option, i.e., the target firm will only be sold if the highest bid exceeds its profit under full separation. We will refer to this minimum bid level as the reservation price. Each possible buyer has a maximum willingness-to-pay (hereafter referred to as WTP), which consists of two parts. The first part is the gain that a buyer realizes due to the merger, whereas the second part is given by the loss if a competitor merges instead.

Horizontal integration incentives are said to be stronger (weaker) than vertical integration incentives if the bidder on the same market side as the target has a higher (lower) WTP to merge with the target than all bidders from the other market side.

The auction is modeled as a two-stage game, with firms submitting sealed bids for the target in the first stage. At the end of the stage, the firm with the highest bid merges with the target if the bid exceeds the reservation price. In the second stage, the acquirer pays out its bid and supply contracts are negotiated. We solve the game using backward induction, and can start immediately with the first stage since second stage profits are determined by the Shapley value.

We first turn to the case where a supplier is the target and assume w.l.o.g. that firm A is up for sale. Firms B and a submit bids β_B and β_a, respectively.

$$\beta_B = \begin{cases} U_{AB}^{\{AB,a,b\}} - U_B^{\{A,B,a,b\}} & \text{if } \beta_a < U_A^{\{A,B,a,b\}} \\ U_{AB}^{\{AB,a,b\}} - U_B^{\{A,B,a,b\}} + U_B^{\{A,B,a,b\}} - U_B^{\{Aa,B,b\}} & \text{if } \beta_a > U_A^{\{A,B,a,b\}} \end{cases} \tag{2.7}$$

$$\beta_a = \begin{cases} U_{Aa}^{\{Aa,B,b\}} - U_a^{\{A,B,a,b\}} & \text{if } \beta_B < U_A^{\{A,B,a,b\}} \\ U_{Aa}^{\{Aa,B,b\}} - U_a^{\{A,B,a,b\}} + U_a^{\{A,B,a,b\}} - U_a^{\{AB,a,b\}} & \text{if } \beta_B > U_A^{\{A,B,a,b\}} \end{cases} \tag{2.8}$$

The case distinction accounts for the fact that a merger with a competitor is not necessarily a credible threat if the bidder itself refuses to merge. A takeover by a rival constitutes a credible threat only if the WTP of the competitor exceeds the reservation price of the target. Otherwise, the target is not sold and the distribution of bargaining rents remains unaffected. Consequently, under such circumstances, the bidder's bargaining position remains unaffected in case of non-merging and its WTP equals its bargaining gain in case of merging.

Equations (2.7) and (2.8) can be rewritten in terms of the industry profit as derived in Appendix 2.C. On this basis, we have to check for each ordering of bids β_B, β_a and $U_A^{\{A,B,a,b\}}$ whether the bids actually exceed the target firm's reservation price, so that the sale of the target actually takes place. For example, consider the case

$$\beta_B < \beta_a < U_A^{\{A,B,a,b\}}.$$

In this case the target firm has a higher reservation price than the bids, and consequently remains unsold. The WTP of firms B and a reduce to

$$\beta_B = \frac{1}{12}\left[2\, W_{\Omega\backslash a} + 2\, W_{\Omega\backslash A} + W_\Omega\right] \quad \text{and} \quad \beta_a = \frac{1}{12}\left[4\, W_{\Omega\backslash a} + W_\Omega\right].$$

We can then derive the conditions under which various ordering of bids occur, as follows:

$$\begin{aligned} &\beta_B < \beta_a \quad \Leftrightarrow \quad W_{\Omega\backslash A} < W_{\Omega\backslash a} \\ &\beta_B < U_A^{\{A,B,a,b\}} \quad \Leftrightarrow \quad W_\Omega < 2W_{\Omega\backslash A} \\ &\beta_a < U_A^{\{A,B,a,b\}} \quad \Leftrightarrow \quad W_\Omega < W_{\Omega\backslash A} + W_{\Omega\backslash a} \end{aligned}$$

The computations in all other cases are straightforward and can be found in the proof of Proposition 7, which summarizes the results.

We turn to the case where a retailer is up for sale and assume w.l.o.g. that firm a is the target. Supplier A and retailer b submit bids β_A and β_b, respectively.

$$\beta_A = \begin{cases} U_{Aa}^{\{Aa,B,b\}} - U_A^{\{A,B,a,b\}} & \text{if } \beta_b < U_a^{\{A,B,a,b\}} \\ U_{Aa}^{\{Aa,B,b\}} - U_A^{\{A,B,a,b\}} + U_A^{\{A,B,a,b\}} - U_A^{\{A,B,ab\}} & \text{if } \beta_b > U_a^{\{A,B,a,b\}} \end{cases} \tag{2.9}$$

$$\beta_b = \begin{cases} U_{ab}^{\{A,B,ab\}} - U_b^{\{A,B,a,b\}} & \text{if } \beta_A < U_a^{\{A,B,a,b\}} \\ U_{ab}^{\{A,B,ab\}} - U_b^{\{A,B,a,b\}} + U_b^{\{A,B,a,b\}} - U_b^{\{Aa,B,b\}} & \text{if } \beta_A > U_a^{\{A,B,a,b\}} \end{cases} \tag{2.10}$$

As before, these relationships can be rewritten using Appendix 2.C and the conditions under which each firm prevails can be derived for all possible orderings of bids β_A, β_b and the target's reservation price $U_a^{\{A,B,a,b\}}$. The following proposition sums up our main results from this analysis:

Proposition 7. *The outcome of the auction is independent of whether a supplier or a retailer is up for sale. If, in all possible merger constellations, the joint profit of the merging firms is lower than their joint profit under full separation, no merger takes place. Otherwise a retailer completes the takeover if $W_{\Omega\backslash a} > W_{\Omega\backslash A}$, whereas a supplier acquires the target firm if $W_{\Omega\backslash a} < W_{\Omega\backslash A}$.*

The result that the target firm is not sold if neither vertical nor horizontal mergers with the target are profitable can be explained as follows. In our model changes in bargaining power only affect the distribution of rents. If a merger is unprofitable, the merged firms face a loss compared to their joint profit under full separation and, hence, the non-integrated firms benefit. Consequently, firms never have an incentive to prevent an unprofitable merger of their competitors and will bid less than the reservation price in order to stay separated.

In the remaining cases, a firm from the market side on which the competitive pressure is largest completes the takeover. To see this, note that the case of strictly decreasing unit costs and complements is excluded because no merger occurs in this case. Thus, on at least one market side firms have an incentive to finish negotiations first so that they are not affected by a negative externality due to substitutability or strictly increasing unit costs. An increase in the strength of the externality has two effects. On the one hand, the contribution of the firm signing a contract second decreases, i.e., $W_{\Omega\backslash A}$ or $W_{\Omega\backslash a}$ increases. On the other hand, there is an increase in the incentive to conclude negotiations first as this can be considered as an indicator of the competitive pressure. Therefore, the question of which firm completes the takeover can be translated into a comparison of the competitive pressures on both market sides.

Further insights can be derived by comparing Proposition 6 and Proposition 7.

Corollary 8. *Consider the cases in which a merger takes place. The firm with the largest gain in profits due to the merger with the target firm completes the takeover.*

As shown in Proposition 6, the gain of a vertical merger is always in between the gains of both types of horizontal mergers. If, as Corollary 8 states, the bidder with the highest gain acquires the target firm, how can a vertical merger occur? The striking difference between our auction model and the simple comparison of merger gains is that the auction does not allow for horizontal mergers on the other market side than that of the target firm. Therefore, a vertical merger is the best way to realize the merger incentives of the other market side. To put it simply, vertical mergers are driven by the merger incentives of the other market side.

The second striking difference is that firms take into account the other market participants as bidders. The idea is that firms might acquire the target firm in order to pre-empt a merger with another bidder. If, like in our model, the firm with the highest gain in profits completes the acquisition, pre-emption is never the determining factor for the decision.[15] We conclude:

Corollary 9. *Firms never acquire the target firm in order to pre-empt the merger of another market participant.*

Corollary 9 shows that bargaining power considerations neither strengthen nor weaken pre-emption decisions. The intuition is the following: Changes in bargaining power, *ceteris paribus*, lead to a change in the distribution of rents, but the total surplus generated remains unaffected. Thus, if a merger is profitable, the loss of the non-integrated firms is equal to the gain of the merged firms. Consequently, the loss of a single non-integrated competitor is (weakly) smaller than the gain of the integrated firm and, hence, the incentive to prevent the merger is (weakly) smaller than the incentive of the other firms to carry out the merger.

Finally, we briefly address counter-mergers. The idea is that the remaining non-integrated parties may want to merge to counteract possible negative effects caused by the merger of their competitors. Take the case of a horizontal merger with the target firm. As shown by Inderst and Wey (2003) (Proposition 2), horizontal merger incentives on each market side are not affected by whether firms are merged or not on the other market level. Thus, a horizontal counter-merger takes place if the target firm is a supplier and unit costs are strictly increasing or if the target is a retailer and final goods are substitutes.

Now turn to the case of a vertical merger with the target firm and keep in mind that, in our model, mergers only affect the distribution of rents. As shown in Corollary 8, a vertical merger only takes place if it is profitable, i.e., the joint profit of the merged firms increases compared to the case of full separation. This means, in turn, that the joint profit of the non-merging firms decreases. A vertical counter-merger leads to two symmetric vertically integrated firms, so that the surplus is shared equally. A counter-merger, therefore, always takes place because this leads to an increase in the joint profit of the non-integrated firms to pre-auction level.

[15] Our result differs from Colangelo (1995) which shows that vertical mergers can be driven by the incentive to prevent a horizontal merger. However, his model is not tailored to the analysis of bargaining power, but merger decisions are affected by the monopolization of the downstream market, the elimination of double markups, and price discrimination against non-integrated firms.

2.5 Conclusion

We propose a model of a bilaterally duopolistic industry where upstream producers bargain with downstream retailers over supply conditions. In the applied framework, integration does not affect the total output produced, but it does affect the distribution of rents among players. We make four contributions in this article.

First, we identify conditions for vertical mergers to occur and show that vertical integration incentives can be regarded as a combination of horizontal merger incentives up- and downstream. Second, we directly compare the strength of horizontal and vertical merger incentives and find that vertical merger incentives always fall between up- and downstream horizontal merger incentives. Third, we show that a horizontal merger to monopoly may convey less bargaining power to the merged entity than vertical integration. Fourth, we find that a vertical merger is never motivated by pre-emptive bargaining power considerations.

While many of our results are general, this article has some limitations. Our analysis focuses on the pure bargaining effects of mergers, taking product market outcomes as constant. In particular, we assume that there are no competitive externalities downstream and that contracts in the input market are efficient. This allows us to identify the main forces behind bargaining, in isolation of price and efficiency considerations. These considerations outside our model, however, must remain an integral part of merger analysis. As for the absence of downstream competition, we offer an alternative approach in Appendix 2.D that allows us to relax this assumption. However, both our model and the model used in Appendix 2.D are not able to include other contractual relationships that could, for example, lead to a double markup problem, thereby ignoring potential efficiency incentives for vertical integration.

While the aforementioned assumptions allow the application of the Shapley value, it is worth noting that empirical applications of bargaining models often use an alternative concept, the so-called Nash-in-Nash bargaining (see Collard-Wexler et al. 2019 for a microfoundation and the references given there for examples). We can show that the outcomes of our analysis differ if we use Nash-in-Nash bargaining in combination with passive beliefs[16]. More precisely, if we assume that production decisions and bargaining are made simultaneously, vertical integration does not have any effect on firm profits. If, however, the integrated firm is allowed to adjust its production decision with respect to its own integrated retailer, vertical integration is always profitable. This is because the integrated firm can better respond to a bargaining breakdown which shifts its threat point in negotiations with non-integrated firms.

This means that under Nash-in-Nash bargaining and passive beliefs, the ability of firms to condition decisions on changes in the market structure due to bargaining breakdowns benefits firms. Here, the benefits of the Shapley value come into play. In the model of Inderst and Wey (2003), the Shapley value is derived from, among others, the assumption that firms use contingent contracts, i.e., firms can specify contracts contingent on the market

[16]Details are provided in Appendix 2.E.

structure. This is a simplified mechanism that allows firms to respond to off-equilibrium events like bargaining breakdowns. In contrast to Nash-in-Nash bargaining with passive beliefs, it allows all firms, not just the integrated firm, to adjust their decisions, which can be interpreted as the possibility for firms to re-negotiate contracts in the case of long-lasting blackouts.

The question of which model is preferable remains, and this is generally difficult to answer. On the one hand, it is likely that the possibility to re-negotiate contracts in the case of long-lasting blackouts will depend on the industry structure and the institutional environment. On the other hand, even if the researcher has a particular industry in mind, it might be difficult to determine the appropriate model since long-lasting blackouts are off-equilibrium outcomes and therefore rarely observed in practice (see Salop, 2018, for a related discussion in the context of the US pay-TV example).

Another restriction in our analysis is the assumption of symmetry for some results. Imposing this assumption helps obtain clear and simple results, at the cost of omitting potential effects from asymmetry between firms. We expect that asymmetry may qualify the strength of various effects identified in our model, but would not turn these around. Uncovering the role of asymmetries in more detail would be an interesting avenue for further research.

Finally, while this article confines itself to the analysis of vertical merger incentives also in comparison to horizontal ones, many possible extensions arise naturally. Extending the bilateral duopoly setup to more firms as well as taking into account investment incentives could be fruitful topics for further research.

2.A Proofs

The proofs require the application of the Shapley value. Appendix 2.C gives an overview of the profits under various market structures.

Proof of Proposition 5. The proof follows immediately by comparing the change in payoffs of the merging parties as summarized in Table 2.1.

Change in market structure	Change in payoffs of vertically merging parties (ΔU)
$\{AB, a, b\}$	$[U_{AB} + U_a]_{\{AB,a,b\}} = \frac{1}{6}\left[4W_\Omega - W_{\Omega\backslash a} + 2W_{\Omega\backslash b}\right]$
$\downarrow$	$[U_{ABa}]_{\{ABa,b\}} = \frac{1}{2}\left[W_{\Omega\backslash b} + W_\Omega\right]$
$\{ABa, b\}$	$\Delta U_{ABa} = \frac{1}{6}\left[W_{\Omega\backslash a} + W_{\Omega\backslash b} - W_\Omega\right]$
$\{A, B, ab\}$	$[U_A + U_{ab}]_{\{A,B,ab\}} = \frac{1}{6}\left[4W_\Omega - W_{\Omega\backslash A} + 2W_{\Omega\backslash B}\right]$
$\downarrow$	$[U_{Aab}]_{\{Aab,B\}} = \frac{1}{2}\left[W_{\Omega\backslash B} + W_\Omega\right]$
$\{Aab, B\}$	$\Delta U_{Aab} = \frac{1}{6}\left[W_{\Omega\backslash A} + W_{\Omega\backslash B} - W_\Omega\right]$
$\{A, B, a, b\}$	$[U_A + U_a]_{\{A,B,a,b\}} = \frac{1}{6}\left[3W_\Omega - W_{\Omega\backslash Aa} + W_{\Omega\backslash Bb} - W_{\Omega\backslash A} + W_{\Omega\backslash B} - W_{\Omega\backslash a} + W_{\Omega\backslash b}\right]$
$\downarrow$	$[U_{Aa}]_{\{Aa,B,b\}} = \frac{1}{6}\left[2W_{\Omega\backslash Bb} + W_{\Omega\backslash b} + W_{\Omega\backslash B} - 2W_{\Omega\backslash Aa} + 2W_\Omega\right]$
$\{Aa, B, b\}$	$\Delta U_{Aa} = \frac{1}{6}\left[\left(W_{\Omega\backslash Bb} - W_{\Omega\backslash Aa}\right) + W_{\Omega\backslash A} + W_{\Omega\backslash a} - W_\Omega\right]$

Table 2.1: Change in payoffs by vertical integration

□

Proof of Corollary 6. We proceed by proving each claim separately.
Claim (i). With suppliers integrated and retailers separated ($\Psi = \{AB, a, b\}$), the condition under which a vertical merger between supplier AB and retailer r takes place is given by Claim (i) in Proposition 5. This is identical to the condition under which a horizontal merger between retailers takes place in Inderst and Wey (2003). The proof of Claim (i) follows immediately from Corollary 1(ii) and Proposition 2 of Inderst and Wey (2003).
Claim (ii). With suppliers separated and retailers integrated ($\Psi = \{A, B, ab\}$), the condition for a vertical merger between supplier s and retailer ab to take place is given by Claim (ii) of Proposition 5. This is identical to the condition for a horizontal merger between suppliers to take place in Inderst and Wey (2003). The proof of Claim (ii) follows immediately from Corollary 1(i) and Proposition 2 of Inderst and Wey (2003).
Claim (iii). Under Assumption 2 (symmetry), the condition for a vertical merger to take place in Claim (iii) of Proposition 5 reduces to

$$W_{\Omega\backslash s} + W_{\Omega\backslash r} > W_\Omega. \tag{2.11}$$

We focus w.l.o.g. on a merger of supplier A with retailer a. The proof for any other supplier-retailer combination would proceed analogously. We first show that a vertical merger takes place if final goods are substitutes and unit costs are strictly increasing. Let q_{sr}^{Ψ} denote the quantity of input s used by retailer r if the subset $\Psi \subseteq \Omega$ of firms participate. Condition (2.11) can be written as

$$\left[\sum_{r\in R^0} p_{Br}(q_{Br}^{\Omega\backslash A},0)q_{Br}^{\Omega\backslash A} - C_B(q_{Br}^{\Omega\backslash A} + q_{Br'}^{\Omega\backslash A})\right] + \left[\sum_{s\in S^0} p_{sb}(q_{sb}^{\Omega\backslash a}, q_{s'b}^{\Omega\backslash a})q_{sb}^{\Omega\backslash a} - \sum_{s\in S^0} C_s(q_{sb}^{\Omega\backslash a})\right]$$
$$> \left[\sum_{s\in S^0}\sum_{r\in R^0} p_{sr}(q_{sr}^{\Omega}, q_{s'r}^{\Omega})q_{sr}^{\Omega} - \sum_{s\in S^0} C_s(q_{sr}^{\Omega} + q_{sr'}^{\Omega})\right]. \tag{2.12}$$

Note that the sum of payoffs on the LHS in (2.11) does not increase if the optimal quantities $q_{rs}^{\Omega\backslash A}$ and $q_{rs}^{\Omega\backslash a}$ are replaced by q_{rs}^{Ω}. It follows that (2.11) holds if

$$\left[\sum_{r\in R^0} p_{Br}(q_{Br}^{\Omega},0)q_{Br}^{\Omega} - C_B(q_{Br}^{\Omega} + q_{Br'}^{\Omega})\right] + \left[\sum_{s\in S^0} p_{sb}(q_{sb}^{\Omega}, q_{s'b}^{\Omega})q_{sb}^{\Omega} - \sum_{s\in S^0} C_s(q_{sb}^{\Omega})\right]$$
$$> \left[\sum_{s\in S^0}\sum_{r\in R^0} p_{sr}(q_{sr}^{\Omega}, q_{s'r}^{\Omega})q_{sr}^{\Omega} - \sum_{s\in S^0} C_s(q_{sr}^{\Omega} + q_{sr'}^{\Omega})\right].$$

Under Assumption 2 (symmetry), this inequality can be written as

$$4p(q^{\Omega}, q^{\Omega})q^{\Omega} - 2C(2q^{\Omega}) < 2p(q^{\Omega},0)q^{\Omega} - C(2q^{\Omega}) + 2p(q^{\Omega}, q^{\Omega})q^{\Omega} - 2C(q^{\Omega}).$$

Dividing by $2q^{\Omega}$ and rearranging yields

$$p(q^{\Omega}, q^{\Omega}) - p(q^{\Omega}, 0) < \overline{C}(2q^{\Omega}) - \overline{C}(q^{\Omega}),$$

or identically, $\Delta_p\left(q^{\Omega}\right) < \Delta_C\left(q^{\Omega}\right)$. The RHS is positive if unit costs are strictly increasing while the LHS is negative if final goods are substitutes. Consequently, if final goods are substitutes and unit costs are strictly increasing, Condition (2.11) holds.

Next, we show that if final goods are complements and unit costs are strictly decreasing, no vertical merger takes place. A vertical merger does not occur if inequality (2.12) is reversed, such that

$$\left[\sum_{r\in R^0} p_{Br}(q_{Br}^{\Omega\backslash A},0)q_{Br}^{\Omega\backslash A} - C_B(q_{Br}^{\Omega\backslash A} + q_{Br'}^{\Omega\backslash A})\right] + \left[\sum_{s\in S^0} p_{sb}(q_{sb}^{\Omega\backslash a}, q_{s'b}^{\Omega\backslash a})q_{sb}^{\Omega\backslash a} - \sum_{s\in S^0} C_s(q_{sb}^{\Omega\backslash a})\right]$$
$$< \left[\sum_{s\in S^0}\sum_{r\in R^0} p_{sr}(q_{sr}^{\Omega}, q_{s'r}^{\Omega})q_{sr}^{\Omega} - \sum_{s\in S^0} C_s(q_{sr}^{\Omega} + q_{sr'}^{\Omega})\right].$$

Under Assumption 2 (symmetry), this can be written as

$$\left[2p(q^{\Omega\backslash A},0)q^{\Omega\backslash A} - C(2q^{\Omega\backslash A})\right] + \left[2p(q^{\Omega\backslash a}, q^{\Omega\backslash a})q^{\Omega\backslash a} - 2C(q^{\Omega\backslash a})\right]$$
$$< \left[2p(q^{\Omega}, q^{\Omega})q^{\Omega} - C(2q^{\Omega})\right] + \left[2p(q^{\Omega}, q^{\Omega})q^{\Omega} - C(2q^{\Omega})\right].$$

Each bracket on the RHS corresponds to half of the industry surplus if all firms participate. We can replace the optimal quantities on the RHS by other quantities and find that if the new inequality holds, the above inequality with optimal quantities would also hold. In the first bracket, we replace q^{Ω} by $q^{\Omega\backslash A}$ and in the second bracket by $q^{\Omega\backslash a}$. Doing so yields

$$2p(q^{\Omega\backslash A},0)q^{\Omega\backslash A} - 2C(q^{\Omega\backslash a}) < 2p(q^{\Omega\backslash A},q^{\Omega\backslash A})q^{\Omega\backslash A} - C(2q^{\Omega\backslash a}).$$

By rearranging and dividing both sides by $2q^{\Omega\backslash a}$, we get

$$\left[p(q^{\Omega\backslash A},q^{\Omega\backslash A}) - p(q^{\Omega\backslash A},0)\right]\frac{q^{\Omega\backslash A}}{q^{\Omega\backslash a}} > \overline{C}(2q^{\Omega\backslash a}) - \overline{C}(q^{\Omega\backslash a}),$$

which is equivalent to $\Delta_C\left(q^{\Omega\backslash a}\right) < \Delta_p\left(q^{\Omega\backslash A}\right)\frac{q^{\Omega\backslash A}}{q^{\Omega\backslash a}}$. The LHS of this inequality is negative if unit costs are strictly decreasing while the RHS is positive when final goods are complements. We can conclude that if final goods are complements and unit costs are strictly decreasing, no vertical merger between a supplier and a retailer takes place.

□

Proof of Proposition 6. We use Assumption 2 (symmetry) and consider w.l.o.g. $s = A$ and $r = a$. Using Appendix 2.C, we rewrite the first inequality as follows.

$$\begin{aligned} &U_{AB}^{\{AB,a,b\}} - U_A^{\{A,B,a,b\}} - U_B^{\{A,B,a,b\}} \gtreqless U_{Aa}^{\{Aa,B,b\}} - U_A^{\{A,B,a,b\}} - U_a^{\{A,B,a,b\}} \\ &\quad\Leftrightarrow \quad W_{\Omega\backslash B} \gtreqless W_{\Omega\backslash Bb} + W_{\Omega\backslash a} - W_{\Omega\backslash Aa} \end{aligned}$$

We apply the symmetry assumption.

$$W_{\Omega\backslash B} \gtreqless W_{\Omega\backslash Bb} + W_{\Omega\backslash a} - W_{\Omega\backslash Aa} \quad \Leftrightarrow \quad W_{\Omega\backslash B} \gtreqless W_{\Omega\backslash a} \quad \Leftrightarrow \quad W_{\Omega\backslash s} \gtreqless W_{\Omega\backslash r}$$

The second inequality can be rewritten in a similar way.

$$\begin{aligned} &U_{ab}^{\{A,B,ab\}} - U_a^{\{A,B,a,b\}} - U_b^{\{A,B,a,b\}} \gtreqless U_{Aa}^{\{Aa,B,b\}} - U_A^{\{A,B,a,b\}} - U_a^{\{A,B,a,b\}} \\ &\Leftrightarrow \quad W_{\Omega\backslash b} \gtreqless W_{\Omega\backslash Bb} - W_{\Omega\backslash Aa} + W_{\Omega\backslash A} \\ &\Leftrightarrow \quad W_{\Omega\backslash b} \gtreqless W_{\Omega\backslash A} \\ &\Leftrightarrow \quad W_{\Omega\backslash r} \gtreqless W_{\Omega\backslash s} \end{aligned}$$

□

Proof of Corollary 7. In the following, $\alpha(f)$ denotes the competitor of firm f on the same market side. The inequality $W_{\Omega\backslash r} > W_{\Omega\backslash s}$ can be written as

$$\begin{aligned} &\sum_{s'\in S^0} p_{s'\alpha(r)}\left(q_{s'\alpha(r)}^{\Omega\backslash r}, q_{\alpha(s')\alpha(r)}^{\Omega\backslash r}\right) q_{s'\alpha(r)}^{\Omega\backslash r} - \sum_{s'\in S^0} C_{s'}\left(q_{s'\alpha(r)}^{\Omega\backslash r}\right) \\ &\quad > \sum_{r'\in R^0} p_{\alpha(s)r'}\left(q_{\alpha(s)r'}^{\Omega\backslash s}, 0\right) q_{\alpha(s)r'}^{\Omega\backslash s} - C_{\alpha(s)}\left(q_{\alpha(s)r'}^{\Omega\backslash s} + q_{\alpha(s)\alpha(r')}^{\Omega\backslash s}\right). \end{aligned} \tag{2.13}$$

Under Assumption 2 (symmetry), the RHS remains unchanged if we replace the quantity $q^{\Omega\backslash s}_{\alpha(s)\alpha(r')}$ by $q^{\Omega\backslash s}_{\alpha(s)r'}$. Furthermore, the LHS does not increase if we replace the quantities by $q^{\Omega\backslash s}_{\alpha(s)r'}$ because the original quantities maximize the expression. We define $q^s := q^{\Omega\backslash s}_{\alpha(s)r'}$. Therefore, inequality (2.13) holds if the following inequality is fulfilled.

$$2q^{\Omega\backslash s} \cdot p(q^{\Omega\backslash s}, q^{\Omega\backslash s}) - 2C(q^{\Omega\backslash s}) > 2q^{\Omega\backslash s} \cdot p(q^{\Omega\backslash s}, 0) - C(2q^{\Omega\backslash s})$$

Dividing both sides by $2q^{\Omega\backslash s}$ yields

$$p(q^{\Omega\backslash s}, q^{\Omega\backslash s}) - \overline{C}(q^{\Omega\backslash s}) > p(q^{\Omega\backslash s}, 0) - \overline{C}(2q^{\Omega\backslash s}),$$

which can be rearranged to $-\Delta_p \left(q^{\Omega\backslash s}\right) < \Delta_C \left(q^{\Omega\backslash s}\right)$. As a result, we find that $W_{\Omega\backslash r} > W_{\Omega\backslash s}$ is fulfilled if inequality $-\Delta_p \left(q^{\Omega\backslash s}\right) < \Delta_C \left(q^{\Omega\backslash s}\right)$ holds.

The argument for $W_{\Omega\backslash r} < W_{\Omega\backslash s}$ is analogous. This inequality can be written as

$$\begin{aligned} &\sum_{s' \in S^0} p_{s'\alpha(r)} \left(q^{\Omega\backslash r}_{s'\alpha(r)}, q^{\Omega\backslash r}_{\alpha(s')\alpha(r)}\right) q^{\Omega\backslash r}_{s'\alpha(r)} - \sum_{s' \in S^0} C_{s'} \left(q^{\Omega\backslash r}_{s'\alpha(r)}\right) \\ &\quad < \sum_{r' \in R^0} p_{\alpha(s)r'} \left(q^{\Omega\backslash s}_{\alpha(s)r'}, 0\right) q^{\Omega\backslash s}_{\alpha(s)r'} - C_{\alpha(s)} \left(q^{\Omega\backslash s}_{\alpha(s)r'} + q^{\Omega\backslash s}_{\alpha(s)\alpha(r')}\right). \end{aligned} \tag{2.14}$$

Under Assumption 2 (symmetry), the LHS remains unchanged if we replace the quantity $q^{\Omega\backslash r}_{\alpha(s')\alpha(r)}$ by $q^{\Omega\backslash r}_{s'\alpha(r)}$. Furthermore, the RHS does not increase if we replace the quantities by $q^{\Omega\backslash r}_{s'\alpha(r)}$ because the original quantities maximize the expression. We define $q^r := q^{\Omega\backslash r}_{s'\alpha(r)}$. Therefore, inequality (2.14) holds if the following inequality is fulfilled.

$$2q^{\Omega\backslash r} \cdot p(q^{\Omega\backslash r}, q^{\Omega\backslash r}) - 2C(q^{\Omega\backslash r}) < 2q^{\Omega\backslash r} \cdot p(q^{\Omega\backslash r}, 0) - C(2q^{\Omega\backslash r})$$

Dividing both sides by $2q^{\Omega\backslash r}$ yields

$$p(q^{\Omega\backslash r}, q^{\Omega\backslash r}) - \overline{C}(q^{\Omega\backslash r}) < p(q^{\Omega\backslash r}, 0) - \overline{C}(2q^{\Omega\backslash r}),$$

which can be rearranged to $-\Delta_p \left(q^{\Omega\backslash r}\right) > \Delta_C \left(q^{\Omega\backslash r}\right)$. As a result, we find that $W_{\Omega\backslash r} < W_{\Omega\backslash s}$ is fulfilled if inequality $-\Delta_p \left(q^{\Omega\backslash r}\right) > \Delta_C \left(q^{\Omega\backslash r}\right)$ holds.

□

Proof of Proposition 7. We start with the case where firm A is up for sale and compare the WTP for all possible orderings of (2.7), (2.8) and $U_A^{\{A,B,a,b\}}$.
Outcome 1: No acquisition takes place, i.e., $U_A^{\{A,B,a,b\}} > \beta_B$ and $U_A^{\{A,B,a,b\}} > \beta_a$. Rewriting these inequalities yields:

$$\begin{aligned} U_A^{\{A,B,a,b\}} > \beta_B \quad &\Leftrightarrow \quad W_\Omega > 2\, W_{\Omega\backslash B} \\ U_A^{\{A,B,a,b\}} > \beta_a \quad &\Leftrightarrow \quad W_\Omega > W_{\Omega\backslash a} + W_{\Omega\backslash B} \end{aligned}$$

Outcome 2: Supplier B wins the auction, i.e., $\beta_B > U_A^{\{A,B,a,b\}}$ and $\beta_B > \beta_a$. Note that the value of β_B depends on $U_A^{\{A,B,a,b\}} \lesseqgtr \beta_a$ which can be rewritten as $3W_\Omega \lesseqgtr 2W_{\Omega\backslash a} + 4W_{\Omega\backslash B}$.

$$\beta_B > U_A^{\{A,B,a,b\}} \quad \Leftrightarrow \quad 2W_{\Omega\backslash B} > W_\Omega$$

$$\beta_B > \beta_a \quad \Leftrightarrow \quad \begin{cases} W_\Omega > 2W_{\Omega\backslash a} & \text{if } 3W_\Omega > 2W_{\Omega\backslash a} + 4W_{\Omega\backslash B} \\ W_{\Omega\backslash B} > W_{\Omega\backslash a} & \text{if } 3W_\Omega < 2W_{\Omega\backslash a} + 4W_{\Omega\backslash B} \end{cases}$$

Outcome 3: A retailer wins the auction, i.e., $\beta_a > U_A^{\{A,B,a,b\}}$ and $\beta_a > \beta_B$. Note that the value of β_a depends on $U_A^{\{A,B,a,b\}} \lesseqgtr \beta_B$ which can be rewritten as $W_\Omega \lesseqgtr 2W_{\Omega\backslash B}$.

$$\beta_a > U_A^{\{A,B,a,b\}} \quad \Leftrightarrow \quad \begin{cases} W_\Omega < W_{\Omega\backslash a} + W_{\Omega\backslash B} & \text{if } W_\Omega > 2W_{\Omega\backslash B} \\ 3W_\Omega < 2W_{\Omega\backslash a} + 4W_{\Omega\backslash B} & \text{if } W_\Omega < 2W_{\Omega\backslash B} \end{cases}$$

$$\beta_a > \beta_B \quad \Leftrightarrow \quad \begin{cases} 4W_{\Omega\backslash B} < W_\Omega + 2W_{\Omega\backslash a} & \text{if } W_\Omega > 2W_{\Omega\backslash B} \\ W_{\Omega\backslash B} < W_{\Omega\backslash a} & \text{if } W_\Omega < 2W_{\Omega\backslash B} \end{cases}$$

If $W_\Omega > 2W_{\Omega\backslash B}$ and $W_\Omega > W_{\Omega\backslash B} + W_{\Omega\backslash a}$ hold, outcome 1 is the only possible solution. Otherwise, if $W_\Omega < 2W_{\Omega\backslash B}$ or $W_\Omega < W_{\Omega\backslash B} + W_{\Omega\backslash a}$, it follows from the above conditions that outcome 2 occurs under the condition $W_{\Omega\backslash B} > W_{\Omega\backslash a}$ and outcome 3 under the condition $W_{\Omega\backslash B} < W_{\Omega\backslash a}$.

We turn to the case where retailer a is up for sale and compare all orderings of (2.9), (2.10) and $U_a^{\{A,B,a,b\}}$.

Outcome 1: No firm acquires the target, i.e., $U_a^{\{A,B,a,b\}} > \beta_A$ and $U_a^{\{A,B,a,b\}} > \beta_b$.

$$U_a^{\{A,B,a,b\}} > \beta_b \quad \Leftrightarrow \quad W_\Omega > 2\,W_{\Omega\backslash b}$$

$$U_a^{\{A,B,a,b\}} > \beta_A \quad \Leftrightarrow \quad W_\Omega > W_{\Omega\backslash A} + W_{\Omega\backslash b}$$

Outcome 2: A supplier wins the auction, i.e., $\beta_A > U_a^{\{A,B,a,b\}}$ and $\beta_A > \beta_b$. Note that the value of β_A depends on $U_a^{\{A,B,a,b\}} \lesseqgtr \beta_b$ which can be rewritten as $W_\Omega \lesseqgtr 2W_{\Omega\backslash b}$.

$$\beta_A > U_a^{\{A,B,a,b\}} \quad \Leftrightarrow \quad \begin{cases} W_\Omega < W_{\Omega\backslash A} + W_{\Omega\backslash b} & \text{if } W_\Omega > 2W_{\Omega\backslash b} \\ 3W_\Omega < 2W_{\Omega\backslash A} + 4W_{\Omega\backslash b} & \text{if } W_\Omega < 2W_{\Omega\backslash b} \end{cases}$$

$$\beta_A > \beta_b \quad \Leftrightarrow \quad \begin{cases} 4W_{\Omega\backslash b} < W_\Omega + 2W_{\Omega\backslash A} & \text{if } W_\Omega > 2W_{\Omega\backslash b} \\ W_{\Omega\backslash b} < W_{\Omega\backslash A} & \text{if } W_\Omega < 2W_{\Omega\backslash b} \end{cases}$$

Outcome 3: Retailer b wins the auction, i.e., $\beta_b > U_a^{\{A,B,a,b\}}$ and $\beta_b > \beta_A$. Note that the value of β_b depends on $U_a^{\{A,B,a,b\}} \lesseqgtr \beta_A$ which can be rewritten as $3W_\Omega \lesseqgtr 2W_{\Omega\backslash A} + 4W_{\Omega\backslash b}$.

$$\beta_b > U_a^{\{A,B,a,b\}} \quad \Leftrightarrow \quad W_\Omega < 2W_{\Omega\backslash b}$$

$$\beta_b > \beta_A \quad \Leftrightarrow \quad \begin{cases} W_\Omega > 2W_{\Omega\backslash A} & \text{if } 3W_\Omega > 2W_{\Omega\backslash A} + 4W_{\Omega\backslash b} \\ W_{\Omega\backslash b} > W_{\Omega\backslash A} & \text{if } 3W_\Omega < 2W_{\Omega\backslash A} + 4W_{\Omega\backslash b} \end{cases}$$

If $W_\Omega > 2W_{\Omega\setminus b}$ and $W_\Omega > W_{\Omega\setminus A} + W_{\Omega\setminus b}$ hold, outcome 1 is the only possible solution. Otherwise, if $W_\Omega < 2W_{\Omega\setminus b}$ or $W_\Omega < W_{\Omega\setminus A} + W_{\Omega\setminus b}$, it follows from the above conditions that outcome 2 occurs under the condition $W_{\Omega\setminus A} > W_{\Omega\setminus b}$ and outcome 3 under the condition $W_{\Omega\setminus A} < W_{\Omega\setminus b}$.

□

Proof of Corollary 8. Note that the outcomes of both Propositions 6 and 7 depend on the condition $W_{\Omega\setminus a} \lesseqgtr W_{\Omega\setminus A}$. Corollary 8 results immediately from comparing the outcomes.

□

2.B Example for the Application of the Shapley Value

We focus on an industry structure with an upstream monopoly and non-integrated retailers (i.e., $\Psi = \{AB, a, b\}$) to demonstrate the use of the Shapley value. In this case six orderings are possible, those displayed in Table 2.2. We focus on the payoff of supplier AB.

		Marginal contribution		
	Ordering	AB	a	b
1	AB, a, b	0	$W_{\Omega\backslash b}$	$W_\Omega - W_{\Omega\backslash b}$
2	AB, b, a	0	$W_\Omega - W_{\Omega\backslash a}$	$W_{\Omega\backslash a}$
3	a, AB, b	$W_{\Omega\backslash b}$	0	$W_\Omega - W_{\Omega\backslash b}$
4	b, AB, a	$W_{\Omega\backslash a}$	$W_\Omega - W_{\Omega\backslash a}$	0
5	a, b, AB	W_Ω	0	0
6	b, a, AB	W_Ω	0	0

Table 2.2: Marginal contributions in various orderings

In orderings 1 and 2, supplier AB comes first. Its marginal contribution is zero because without a retailer preceding it, the supplier cannot bring its input to the market. Supplier AB comes second in orderings 3 and 4. In ordering 3, supplier AB's contribution is to enable production with retailer a, together creating $W_{\Omega\backslash b}$ of surplus. This is the surplus that can be created without retailer b. Similarly, in ordering 4, supplier AB enables production with retailer b and therefore generates $W_{\Omega\backslash a}$ of surplus.

In orderings 5 and 6, supplier AB comes last. Since the retailers preceding have no final goods to sell absent a supplier, firm AB's marginal contribution corresponds to the full industry surplus W_Ω in these orderings. Finally, taking expectations of the orderings with equal probabilities, the Shapley value yields as a payoff for the supplier

$$U_{AB} = \frac{1}{6}\left[0 + 0 + W_{\Omega\backslash b} + W_{\Omega\backslash a} + W_\Omega + W_\Omega\right] = \frac{1}{6}\left[W_{\Omega\backslash b} + W_{\Omega\backslash a} + 2W_\Omega\right].$$

2.C Payoffs Under Various Market Structures

Market structure	Payoffs
Full separation $\{A, B, a, b\}$	$U_A = \frac{1}{12}\left[W_{\Omega\backslash Bb} + W_{\Omega\backslash Ba} + W_{\Omega\backslash b} - W_{\Omega\backslash Ab} + W_{\Omega\backslash a} - W_{\Omega\backslash Aa} + W_{\Omega\backslash B} - 3W_{\Omega\backslash A} + 3W_\Omega\right]$ $U_B = \frac{1}{12}\left[-W_{\Omega\backslash Bb} - W_{\Omega\backslash Ba} + W_{\Omega\backslash b} + W_{\Omega\backslash Ab} + W_{\Omega\backslash a} + W_{\Omega\backslash Aa} - 3W_{\Omega\backslash B} + W_{\Omega\backslash A} + 3W_\Omega\right]$ $U_a = \frac{1}{12}\left[W_{\Omega\backslash Bb} - W_{\Omega\backslash Ba} + W_{\Omega\backslash b} + W_{\Omega\backslash Ab} - 3W_{\Omega\backslash a} - W_{\Omega\backslash Aa} + W_{\Omega\backslash B} + W_{\Omega\backslash A} + 3W_\Omega\right]$ $U_b = \frac{1}{12}\left[-W_{\Omega\backslash Bb} + W_{\Omega\backslash Ba} - 3W_{\Omega\backslash b} - W_{\Omega\backslash Ab} + W_{\Omega\backslash a} + W_{\Omega\backslash Aa} + W_{\Omega\backslash B} + W_{\Omega\backslash A} + 3W_\Omega\right]$
Upstream monopoly $\{AB, a, b\}$	$U_{AB} = \frac{1}{6}\left[W_{\Omega\backslash b} + W_{\Omega\backslash a} + 2W_\Omega\right]$ $U_a = \frac{1}{6}\left[W_{\Omega\backslash b} - 2W_{\Omega\backslash a} + 2W_\Omega\right]$ $U_b = \frac{1}{6}\left[-2W_{\Omega\backslash b} + W_{\Omega\backslash a} + 2W_\Omega\right]$
Vertically integrated upstream monopoly $\{ABa, b\}$	$U_{ABa} = \frac{1}{2}\left[W_{\Omega\backslash b} + W_\Omega\right]$ $U_b = \frac{1}{2}\left[-W_{\Omega\backslash b} + W_\Omega\right]$
Downstream monopoly $\{A, B, ab\}$	$U_A = \frac{1}{6}\left[W_{\Omega\backslash B} - 2W_{\Omega\backslash A} + 2W_\Omega\right]$ $U_B = \frac{1}{6}\left[-2W_{\Omega\backslash B} + W_{\Omega\backslash A} + 2W_\Omega\right]$ $U_{ab} = \frac{1}{6}\left[W_{\Omega\backslash B} + W_{\Omega\backslash A} + 2W_\Omega\right]$
Vertically integrated downstream monopoly $\{Aab, B\}$	$U_{Aab} = \frac{1}{2}\left[W_{\Omega\backslash B} + W_\Omega\right]$ $U_B = \frac{1}{2}\left[-W_{\Omega\backslash B} + W_\Omega\right]$
Full integration $\{ABab\}$	$U_{ABab} = W_\Omega$
Single vertical integration $\{Aa, B, b\}$	$U_{Aa} = \frac{1}{6}\left[2W_{\Omega\backslash Bb} + W_{\Omega\backslash b} + W_{\Omega\backslash B} - 2W_{\Omega\backslash Aa} + 2W_\Omega\right]$ $U_B = \frac{1}{6}\left[-W_{\Omega\backslash Bb} + W_{\Omega\backslash b} - 2W_{\Omega\backslash B} + W_{\Omega\backslash Aa} + 2W_\Omega\right]$ $U_b = \frac{1}{6}\left[-W_{\Omega\backslash Bb} - 2W_{\Omega\backslash b} + W_{\Omega\backslash B} + W_{\Omega\backslash Aa} + 2W_\Omega\right]$
Double vertical integration $\{Aa, Bb\}$	$U_{Aa} = \frac{1}{2}\left[W_{\Omega\backslash Bb} - W_{\Omega\backslash Aa} + W_\Omega\right]$ $U_{Bb} = \frac{1}{2}\left[-W_{\Omega\backslash Bb} + W_{\Omega\backslash Aa} + W_\Omega\right]$

Table 2.3: Payoffs under various market structures

2.D Incentives in the Presence of Downstream Externalities

2.D.1 Introduction

In our main analysis, we adopt the model of Inderst and Wey (2003) that omits other driving forces behind (vertical) integration than changes in bargaining power. In particular, this means that we do not allow for competition between the two retailers. The purpose of this appendix is to demonstrate that relaxing this assumption does not affect our main result. More specifically, the introduction of other forces driving mergers may lead to different market outcomes, but the impact of bargaining power considerations remains visible and still has its share in the decision of whether to merge or not.

To demonstrate this, we build on a framework proposed by De Fontenay and Gans (2014) that incorporates the framework of Inderst and Wey (2003) as a special case. They use a bargaining game similar to Inderst and Wey (2003) and introduce the possibility of externalities. This enables us to model competition in downstream markets since both retailers are now allowed to exert an externality on each other. The authors show that under a set of fairly reasonable assumptions,[17] firms' equilibrium profits are given by a generalized version of the Myerson-Shapley value.

The structure of this appendix file is as follows: Section 2.D.2 introduces some additional notations and briefly describes the result of De Fontenay and Gans (2014). In Section 2.D.3, we present Proposition 8 which summarizes integration incentives for different types of mergers in the presence of downstream competition and discuss how it relates to the results in our main article. Finally, we prove Proposition 8 in Section 2.D.4.

2.D.2 Notation and Model

Before we apply the framework of De Fontenay and Gans (2014), we first need to introduce some additional notations.[18] Let P denote a partition of Ω and P^N the set of all partitions. Broadly speaking, a partition divides Ω into mutually exclusive non-empty subsets.[19] If we consider a given partition P, we only consider links between firms in the same set. In addition, and as before, we only allow contractual relationships between suppliers and retailers, but not between firms on the same market side. For example, if we consider the partition $P = \{\{A, a\}, \{B, b\}\}$, supplier A (B) and retailer a (b) share a link. However,

[17] It is beyond the scope of this appendix to provide a complete overview of the framework proposed by De Fontenay and Gans (2014). However, we would like to emphasize that most of the assumptions are either similar to those of Inderst and Wey (2003) or standard in the literature on vertically-related industries. For example, following Inderst and Wey (2003), they assume that firms assign different representatives to each of the other negotiation parties and that bilateral bargaining takes place simultaneously. Furthermore, firms use binding and contingent contracts. The latter means that they can condition on the success of other negotiations. An example of assumptions that are standard in the literature is the adoption of passive beliefs.

[18] To be consistent with the notation of our main article, we alter the original notation of De Fontenay and Gans (2014). The most important changes are that we replace Φ_i by U_i, u_i by π_i, and x by q. We also avoid the terminology of link structures.

[19] More formally, a partition P is a set of sets $P = \{P_1, \dots, P_M\}$ with $P_i \neq \emptyset$ for all i, $\bigcup_{i=1}^{M} P_i = \Omega$, and $P_i \cap P_j = \emptyset$ for all $i \neq j$.

there is no link between supplier A (B) and retailer b (a) because these firms are in different sets.

Each firm i is endowed with a (dis-)utility function π_i.[20] Depending on whether the firm is a supplier or retailer, this is either the profit in the downstream market or the costs.

$$\pi_i = \sum_{s \in S^0} p_{si}\left(q_{si}, q_{s'i}, q_{si'}, q_{s'i'}\right) q_{si} \quad \text{for} \quad i \in R^0$$
$$\pi_i = -C_i\left(q_{ir} + q_{ir'}\right) \quad \text{for} \quad i \in S^0$$

Note that contrary to the analysis in our main article, the inverse demand p_{sr} now depends on four instead of two quantities of final goods because we allow for downstream competition. If we consider a merger between two firms i and j, its post-merger utility function is simply the sum of the two pre-merger utility functions, i.e., $\pi_{ij} = \pi_i + \pi_j$.

For a given market structure, firms' bargaining results in bilateral efficient input quantities, which means that for a given set of quantities exchanged by the other parties, each pair of negotiating firms chooses its quantity to maximize its joint utility. This clearly features a Nash equilibrium notion, but focuses on bargaining pairs rather than single firms. As before, the input quantity is zero if two firms do not share a link. To simplify the notation, we denote the utility of firm i by π_i^P when we face a partition P and firms exchange bilateral efficient input quantities. De Fontenay and Gans (2014) prove the existence of a perfect Bayesian equilibrium in which each agent i receives

$$U_i(K) = \sum_{P \in P^N} \sum_{S \in P} \underbrace{(-1)^{|P|-1}(|P|-1)! \left[\frac{1}{|N|} - \sum_{\substack{i \notin S' \in P \\ S' \neq S}} \frac{1}{(|P|-1)(|N|-|S'|)}\right]}_{\text{multiplier}} \underbrace{\sum_{j \in S} \pi_j^P}_{\text{coalition value}} \tag{2.15}$$

The formula may seem complicated at first glance, but as with the formula for the Shapley value, it is straightforward to apply. The first sum symbol refers to the sum over all partitions P in the set of partitions P^N. For each partition, we calculate the sum over all sets S in that partition. Finally, for each set S, we calculate a multiplier and the coalition value of S, which is the sum of the profits of the firms in S.

2.D.3 Result

With this result of De Fontenay and Gans (2014) in hand, it is straightforward to calculate firms' profits under various market structures. By comparing pre- and post-merger profits, we derive conditions under which mergers are profitable.

Since the introduction of downstream competition adds an additional layer of complexity

[20] De Fontenay and Gans (2014) use a set of assumptions about the utility functions to ensure tractability. Any sum of two utility functions $\pi_i + \pi_j$ with $i \neq j$ has to be bounded, continuous, and differentiable in $q = (q_{Aa}, q_{Ab}, q_{Ba}, q_{Bb})$, concave, and continuously differentiable in q_{ij}.

to our analysis, we adopt the assumption of symmetry to simplify the presentation for the reader. Our main finding is:

Proposition 8. *(i) The upstream firms merge if*

$$2 \cdot \sum_{i \in \Upsilon = \{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}} > \sum_{i \in \Upsilon = \{A,B,a,b\}} \pi_i^{\{\Upsilon\}} \tag{2.16}$$

and they stay separated if the inequality is reversed.

(ii) The downstream firms merge if

$$\begin{aligned} &2 \cdot \sum_{i \in \Upsilon = \{A,ab\}} \pi_i^{\{\Upsilon,B\}} - 2 \cdot \sum_{i \in \Upsilon = \{A,a,b\}} \pi_i^{\{\Upsilon,B\}} + 2 \cdot \sum_{i \in \Upsilon = \{A,B,a\}} \pi_i^{\{\Upsilon,b\}} > \\ &3 \cdot \sum_{i \in \Upsilon = \{A,B,a,b\}} \pi_i^{\{\Upsilon\}} - 2 \cdot \sum_{i \in \Upsilon = \{A,B,ab\}} \pi_i^{\{\Upsilon\}} \end{aligned} \tag{2.17}$$

and they stay separated if the inequality is reversed.

(iii) An upstream firm and a downstream firm merge if

$$\begin{aligned} &\sum_{i \in \Upsilon = \{Aa,b\}} \pi_i^{\{\Upsilon,B\}} + \sum_{i \in \Upsilon = \{Aa,B\}} \pi_i^{\{\Upsilon,b\}} + \\ &2 \cdot \pi_{Aa}^{\{\{Aa\},\{B,b\}\}} - 2\pi_{Aa}^{\{\{Aa\},\{B\},\{b\}\}} > \\ &3 \cdot \sum_{i \in \Upsilon\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} - 2 \cdot \sum_{i \in \{Aa,B,b\}} \pi_i^{\{\Upsilon\}} \end{aligned} \tag{2.18}$$

and they stay separated if the inequality is reversed.

Proposition 8 highlights that even in the presence of downstream competition, some parts of the vertical merger incentives share a strong link to horizontal merger incentives up- and downstream. Starting with the merger incentives of the suppliers, we find that inequality (2.16) looks similar to its Shapley value equivalent (2.6). The right-hand side of the inequality represents the total industry profit in the absence of the merger, while the left-hand side is the industry profit in the scenario where supplier B is isolated and does not share any link with the retailers. It is worth noting that although (2.16) and (2.6) look similar, the downstream competition still has an effect. When supplier B is able to maintain a relationship with both retailers, the input quantity sold to one of the retailers also affects the other retailer through the downstream competition and, hence, we observe the downstream externality.

While the inequality that quantifies the horizontal upstream merger incentives looks very similar to its Shapley value equivalent, this is different for the horizontal downstream merger incentives (2.17) and its equivalent (2.5). The right-hand side contains two terms which, apart from different scaling factors, quantify the difference in the total industry profit before and after the merger. In the absence of downstream externalities, the total industry profit remains the same after the merger, leaving the one-time industry profit after applying the different scaling factors. In this case, the right-hand side simplifies and equals the right-hand side of (2.16) which, in turn, looks similar to the right-hand side of (2.5).

The first two terms of the left-hand side of (2.17) introduce a component which is not part of the Shapley value equivalent. It is the difference in the total industry profit before and after a retail merger if only one supplier shares links with the retailers. This difference is zero in the absence of downstream externalities, but it can take different values in general.

The last term of the left-hand side of (2.17) is the total industry profit if only one retailer is active in the market and shares links with the suppliers. This expression is also present in the Shapley value equivalent (2.5).

Turning to the vertical merger incentives and its relationship to horizontal merger incentives up- and downstream, the right-hand side of inequality (2.18) looks similar to the right-hand side of (2.17). It also measures, apart from different scaling factors, the difference in the total industry profit before and after the merger. As in the case of horizontal downstream merger incentives, this expression simplifies in the absence of downstream competition and then equals the right-hand side of (2.16) which, in turn, looks similar to the Shapley value equivalent (2.2).

The most important part for our analysis is the first line of (2.18) which represents the horizontal merger incentives up- and downstream. This is clear when comparing the line to the left-hand side of (2.16) and to the last term of the left-hand side of (2.17), respectively. The expressions capture the effect of cutting a supplier's or a retailer's links, so that this firm cannot maintain relationships with firms on the other market side. We also observe these components in the Shapley value equivalent (2.2).

Finally, the second line of (2.18) introduces an additional component which is not part of the Shapley value equivalent (2.2). It captures the magnitude of the externality when the non-integrated firms form a coalition without sharing any links with the integrated firm. In the absence of downstream competition, it is not important whether supplier B provides inputs to retailer b and, hence, this expression would be zero. However, in the presence of downstream competition, the relationship between supplier B and retailer b exerts an externality on the integrated firm, so that the expression can take different values.

In summary, we find that even after the introduction of downstream competition, components driving horizontal merger incentives up- and downstream have an impact on the vertical merger incentives. The question of to which extent these horizontal integration incentives dominate the decision of whether to merge vertically or not depends on the type of downstream competition and, hence, is model-dependent. With a particular model of competition in mind, researchers may be able to derive additional insights for industry-specific applications.

2.D.4 Proof

To finalize this Online Appendix, we prove Proposition 8. We have already mentioned what steps we need to go through. The idea is to use the generalized version of the Myerson-

Shapley value to calculate firms' profits and then compare the respective profits to quantify integration incentives. To make it easier for the reader to follow our analysis, we proceed in three steps. First, we derive the multipliers that enter the formula for the profits. Then, we apply these multipliers to calculate the profits. Finally, we compare the respective profits.

Partitions, Sets, and Multipliers

Recall formula (2.15). Loosely speaking, to calculate the profit, we have to run a nested loop. The outer loop contains all partitions and for each partition, we then loop over all sets in that partition. In this subsection, we provide an overview of all partitions as well as the corresponding sets and the related multipliers. Since the set of all partitions depends on the initial market structure, we provide tables for all four cases (full separation and the three types of mergers).

The first table presents the multipliers in the case of full separation. The first column shows the different partitions (labeled P) and the corresponding sets within these partitions. The remaining columns show the different multipliers. Since the multipliers depend on the firm for which we compute the profit, there are four columns related to the four different firms.

	A	B	a	b
$P = \{\{A, B, a, b\}\}$				
$\{A, B, a, b\}$	$1/4$	$1/4$	$1/4$	$1/4$
$P = \{\{A, B, b\}, \{a\}\}$				
$\{A, B, b\}$	$1/12$	$1/12$	$-1/4$	$1/12$
$\{a\}$	$-1/4$	$-1/4$	$3/4$	$-1/4$
$P = \{\{A, B, a\}, \{b\}\}$				
$\{A, B, a\}$	$1/12$	$1/12$	$1/12$	$-1/4$
$\{b\}$	$-1/4$	$-1/4$	$-1/4$	$3/4$
$P = \{\{A, a, b\}, \{B\}\}$				
$\{A, a, b\}$	$1/12$	$-1/4$	$1/12$	$1/12$
$\{B\}$	$-1/4$	$3/4$	$-1/4$	$-1/4$
$P = \{\{B, a, b\}, \{A\}\}$				
$\{B, a, b\}$	$-1/4$	$1/12$	$1/12$	$1/12$

$\{A\}$	$3/4$	$-1/4$	$-1/4$	$-1/4$
$P = \{\{A, b\}, \{B, a\}\}$				
$\{A, b\}$	$1/4$	$-1/4$	$-1/4$	$1/4$
$\{B, a\}$	$-1/4$	$1/4$	$1/4$	$-1/4$
$P = \{\{A, B\}, \{a, b\}\}$				
$\{A, B\}$	$1/4$	$1/4$	$-1/4$	$-1/4$
$\{a, b\}$	$-1/4$	$-1/4$	$1/4$	$1/4$
$P = \{\{A, a\}, \{B, b\}\}$				
$\{A, a\}$	$1/4$	$-1/4$	$1/4$	$-1/4$
$\{B, b\}$	$-1/4$	$1/4$	$-1/4$	$1/4$
$P = \{\{A, b\}, \{B\}, \{a\}\}$				
$\{A, b\}$	$-1/6$	$1/6$	$1/6$	$-1/6$
$\{B\}$	$1/6$	$-1/3$	0	$1/6$
$\{a\}$	$1/6$	0	$-1/3$	$1/6$
$P = \{\{A, B\}, \{a\}, \{b\}\}$				
$\{A, B\}$	$-1/6$	$-1/6$	$1/6$	$1/6$
$\{a\}$	$1/6$	$1/6$	$-1/3$	0
$\{b\}$	$1/6$	$1/6$	0	$-1/3$
$P = \{\{A, a\}, \{B\}, \{b\}\}$				
$\{A, a\}$	$-1/6$	$1/6$	$-1/6$	$1/6$
$\{B\}$	$1/6$	$-1/3$	$1/6$	0
$\{b\}$	$1/6$	0	$1/6$	$-1/3$
$P = \{\{B, b\}, \{A\}, \{a\}\}$				
$\{B, b\}$	$1/6$	$-1/6$	$1/6$	$-1/6$
$\{A\}$	$-1/3$	$1/6$	0	$1/6$
$\{a\}$	0	$1/6$	$-1/3$	$1/6$
$P = \{\{B, a\}, \{A\}, \{b\}\}$				

$\{B, a\}$	$1/6$	$-1/6$	$-1/6$	$1/6$
$\{A\}$	$-1/3$	$1/6$	$1/6$	0
$\{b\}$	0	$1/6$	$1/6$	$-1/3$
$P = \{\{a, b\}, \{A\}, \{B\}\}$				
$\{a, b\}$	$1/6$	$1/6$	$-1/6$	$-1/6$
$\{A\}$	$-1/3$	0	$1/6$	$1/6$
$\{B\}$	0	$-1/3$	$1/6$	$1/6$
$P = \{\{A\}, \{B\}, \{a\}, \{b\}\}$				
$\{A\}$	$1/2$	$-1/6$	$-1/6$	$-1/6$
$\{B\}$	$-1/6$	$1/2$	$-1/6$	$-1/6$
$\{a\}$	$-1/6$	$-1/6$	$1/2$	$-1/6$
$\{b\}$	$-1/6$	$-1/6$	$-1/6$	$1/2$

Table 2.4: Partitions, sets, and multipliers in the case of full separation

The second table shows the partitions, sets, and multipliers in the case of a horizontal upstream merger.

	AB	a	b
$P = \{\{AB, a, b\}\}$			
$\{AB, a, b\}$	$1/3$	$1/3$	$1/3$
$P = \{\{AB, b\}, \{a\}\}$			
$\{AB, b\}$	$1/6$	$-1/3$	$1/6$
$\{a\}$	$-1/3$	$2/3$	$-1/3$
$P = \{\{AB, a\}, \{b\}\}$			
$\{AB, a\}$	$1/6$	$1/6$	$-1/3$
$\{b\}$	$-1/3$	$-1/3$	$2/3$
$P = \{\{a, b\}, \{AB\}\}$			

$\{a, b\}$	$-1/3$	$1/6$	$1/6$
$\{AB\}$	$2/3$	$-1/3$	$-1/3$
$P = \{\{AB\}, \{a\}, \{b\}\}$			
$\{AB\}$	$-1/3$	$1/6$	$1/6$
$\{a\}$	$1/6$	$-1/3$	$1/6$
$\{b\}$	$1/6$	$1/6$	$-1/3$

Table 2.5: Partitions, sets, and multipliers in the case of a horizontal upstream merger

The third table shows the partitions, sets, and multipliers in the case of a horizontal downstream merger.

	A	B	ab
$P = \{\{A, B, ab\}\}$			
$\{A, B, ab\}$	$1/3$	$1/3$	$1/3$
$P = \{\{A, ab\}, \{B\}\}$			
$\{A, ab\}$	$1/6$	$-1/3$	$1/6$
$\{B\}$	$-1/3$	$2/3$	$-1/3$
$P = \{\{A, B\}, \{ab\}\}$			
$\{A, B\}$	$1/6$	$1/6$	$-1/3$
$\{ab\}$	$-1/3$	$-1/3$	$2/3$
$P = \{\{B, ab\}, \{A\}\}$			
$\{B, ab\}$	$-1/3$	$1/6$	$1/6$
$\{A\}$	$2/3$	$-1/3$	$-1/3$
$P = \{\{A\}, \{B\}, \{ab\}\}$			
$\{A\}$	$-1/3$	$1/6$	$1/6$
$\{B\}$	$1/6$	$-1/3$	$1/6$
$\{ab\}$	$1/6$	$1/6$	$-1/3$

Table 2.6: Partitions, sets, and multipliers in the case of a horizontal downstream merger

The fourth table shows the partitions, sets, and multipliers in the case of a vertical merger.

	Aa	B	b
$P=\{\{Aa, B, b\}\}$			
$\{Aa, B, b\}$	1/3	1/3	1/3
$P=\{\{Aa, b\}, \{B\}\}$			
$\{Aa, b\}$	1/6	−1/3	1/6
$\{B\}$	−1/3	2/3	−1/3
$P=\{\{Aa, B\}, \{b\}\}$			
$\{Aa, B\}$	1/6	1/6	−1/3
$\{b\}$	−1/3	−1/3	2/3
$P=\{\{B, b\}, \{Aa\}\}$			
$\{B, b\}$	−1/3	1/6	1/6
$\{Aa\}$	2/3	−1/3	−1/3
$P=\{\{Aa\}, \{B\}, \{b\}\}$			
$\{Aa\}$	−1/3	1/6	1/6
$\{B\}$	1/6	−1/3	1/6
$\{b\}$	1/6	1/6	−1/3

Table 2.7: Partitions, sets, and multipliers in the case of a vertical merger

Profits

We are now able to calculate the profits according to (2.15). To demonstrate how we proceed, suppose we consider a particular market structure (e.g., no integration) and want to compute the profit U_i of firm i. The first step is to select the table that refers to the market structure of interest. Then, we iterate over all partitions and sets in this table. In doing so, we take the multipliers from the column that belongs to player i and multiply each of them by the corresponding coalition value.

Next, we simplify the expression by omitting terms that refer to isolated players. We call a set of players isolated if it contains players from only one side of the market. In other words, these players cannot exchange inputs with the other market side and, hence, their coalition value is zero. In addition, we apply the symmetry assumption, i.e., firms on each side of the market are considered symmetric. This allows us to simplify the expression even further.

We start with the case of full separation and first look at the profit of a supplier.

$$
\begin{aligned}
U_A = & \frac{1}{4} \sum_{i \in \Upsilon = \{A,B,a,b\}} \pi_i^{\{\Upsilon\}} \\
& + \frac{1}{12} \sum_{i \in \Upsilon = \{A,B,b\}} \pi_i^{\{\Upsilon,\{a\}\}} - \underbrace{\frac{1}{4} \pi_a^{\{\{A,B,b\},\{a\}\}}}_{= 0 \text{ (isolated retailer)}} \\
& + \frac{1}{12} \sum_{i \in \Upsilon = \{A,B,a\}} \pi_i^{\{\Upsilon,\{b\}\}} - \underbrace{\frac{1}{4} \pi_b^{\{\{A,B,a\},\{b\}\}}}_{= 0 \text{ (isolated retailer)}} \\
& + \frac{1}{12} \sum_{i \in \Upsilon = \{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}} - \underbrace{\frac{1}{4} \pi_B^{\{\{A,a,b\},\{B\}\}}}_{= 0 \text{ (isolated supplier)}} \\
& - \frac{1}{4} \sum_{i \in \Upsilon = \{B,a,b\}} \pi_i^{\{\Upsilon,\{A\}\}} + \underbrace{\frac{3}{4} \pi_A^{\{\{B,a,b\},\{A\}\}}}_{= 0 \text{ (isolated supplier)}} \\
& + \underbrace{\frac{1}{4} \sum_{i \in \Upsilon = \{A,b\}} \pi_i^{\{\Upsilon,\{B,a\}\}} - \frac{1}{4} \sum_{i \in \Upsilon = \{B,a\}} \pi_i^{\{\Upsilon,\{A,b\}\}}}_{= 0 \text{ (symmetry)}} \\
& + \underbrace{\frac{1}{4} \sum_{i \in \Upsilon = \{A,B\}} \pi_i^{\{\Upsilon,\{a,b\}\}}}_{= 0 \text{ (isolated suppliers)}} - \underbrace{\frac{1}{4} \sum_{i \in \Upsilon = \{a,b\}} \pi_i^{\{\Upsilon,\{A,B\}\}}}_{= 0 \text{ (isolated retailers)}} \\
& + \underbrace{\frac{1}{4} \sum_{i \in \Upsilon = \{A,a\}} \pi_i^{\{\Upsilon,\{B,b\}\}} - \frac{1}{4} \sum_{i \in \Upsilon = \{B,b\}} \pi_i^{\{\Upsilon,\{A,a\}\}}}_{= 0 \text{ (symmetry)}} \\
& - \frac{1}{6} \sum_{i \in \Upsilon = \{A,b\}} \pi_i^{\{\Upsilon,\{B\},\{a\}\}} + \underbrace{\frac{1}{6} \pi_B^{\{\{A,b\},\{B\},\{a\}\}}}_{= 0 \text{ (isolated supplier)}} + \underbrace{\frac{1}{6} \pi_a^{\{\{A,b\},\{B\},\{a\}\}}}_{= 0 \text{ (isolated retailer)}} \\
& - \underbrace{\frac{1}{6} \sum_{i \in \Upsilon = \{A,B\}} \pi_i^{\{\Upsilon,\{a\},\{b\}\}}}_{= 0 \text{ (isolated suppliers)}} + \underbrace{\frac{1}{6} \pi_a^{\{\{A,B\},\{a\},\{b\}\}}}_{= 0 \text{ (isolated retailer)}} + \underbrace{\frac{1}{6} \pi_b^{\{\{A,B\},\{a\},\{b\}\}}}_{= 0 \text{ (isolated retailer)}} \\
& - \frac{1}{6} \sum_{i \in \Upsilon = \{A,a\}} \pi_i^{\{\Upsilon,\{B\},\{b\}\}} + \underbrace{\frac{1}{6} \pi_B^{\{\{A,a\},\{B\},\{b\}\}}}_{= 0 \text{ (isolated supplier)}} + \underbrace{\frac{1}{6} \pi_b^{\{\{A,a\},\{B\},\{b\}\}}}_{= 0 \text{ (isolated retailer)}}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{6}\sum_{i\in\Upsilon=\{B,b\}}\pi_i^{\{\Upsilon,\{A\},\{a\}\}}-\underbrace{\frac{1}{3}\pi_A^{\{\{B,b\},\{A\},\{a\}\}}}_{=0\text{ (isolated supplier)}}\\
&+\frac{1}{6}\sum_{i\in\Upsilon=\{B,a\}}\pi_i^{\{\Upsilon,\{A\},\{b\}\}}-\underbrace{\frac{1}{3}\pi_A^{\{\{B,a\},\{A\},\{b\}\}}}_{=0\text{ (isolated supplier)}}\\
&+\underbrace{\frac{1}{6}\sum_{i\in\Upsilon=\{a,b\}}\pi_i^{\{\Upsilon,\{A\},\{B\}\}}}_{=0\text{ (isolated retailers)}}-\underbrace{\frac{1}{3}\pi_A^{\{\{a,b\},\{A\},\{B\}\}}}_{=0\text{ (isolated supplier)}}\\
&+\underbrace{\frac{1}{2}\pi_A^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated supplier)}}-\underbrace{\frac{1}{6}\pi_B^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated supplier)}}\\
&-\underbrace{\frac{1}{6}\pi_a^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}-\underbrace{\frac{1}{6}\pi_b^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&=\frac{1}{4}\sum_{i\in\Upsilon=\{A,B,a,b\}}\pi_i^{\{\Upsilon\}}\\
&+\frac{1}{12}\sum_{i\in\Upsilon=\{A,B,b\}}\pi_i^{\{\Upsilon,\{a\}\}}+\frac{1}{12}\sum_{i\in\Upsilon=\{A,B,a\}}\pi_i^{\{\Upsilon,\{b\}\}}\\
&+\frac{1}{12}\sum_{i\in\Upsilon=\{A,a,b\}}\pi_i^{\{\Upsilon,\{B\}\}}-\frac{1}{4}\sum_{i\in\Upsilon=\{B,a,b\}}\pi_i^{\{\Upsilon,\{A\}\}}\\
&\underbrace{-\frac{1}{6}\sum_{i\in\Upsilon=\{A,b\}}\pi_i^{\{\Upsilon,\{B\},\{a\}\}}+\frac{1}{6}\sum_{i\in\Upsilon=\{B,b\}}\pi_i^{\{\Upsilon,\{A\},\{a\}\}}}_{=0\text{ (symmetry)}}\\
&\underbrace{-\frac{1}{6}\sum_{i\in\Upsilon=\{A,a\}}\pi_i^{\{\Upsilon,\{B\},\{b\}\}}+\frac{1}{6}\sum_{i\in\Upsilon=\{B,a\}}\pi_i^{\{\Upsilon,\{A\},\{b\}\}}}_{=0\text{ (symmetry)}}\\
&=\frac{1}{4}\sum_{i\in\Upsilon=\{A,B,a,b\}}\pi_i^{\{\Upsilon\}}+\frac{1}{6}\sum_{i\in\Upsilon=\{A,B,b\}}\pi_i^{\{\Upsilon,\{a\}\}}-\frac{1}{6}\sum_{i\in\Upsilon=\{A,a,b\}}\pi_i^{\{\Upsilon,\{B\}\}}
\end{aligned}
$$

We stick to the case of full separation and turn to the profit of a retailer.

$$
\begin{aligned}
U_a=&\frac{1}{4}\sum_{i\in\Upsilon=\{A,B,a,b\}}\pi_i^{\{\Upsilon\}}\\
&-\frac{1}{4}\sum_{i\in\Upsilon=\{A,B,b\}}\pi_i^{\{\Upsilon,\{a\}\}}+\underbrace{\frac{3}{4}\pi_a^{\{\{A,B,b\},\{a\}\}}}_{=0\text{ (isolated retailer)}}\\
&+\frac{1}{12}\sum_{i\in\Upsilon=\{A,B,a\}}\pi_i^{\{\Upsilon,\{b\}\}}-\underbrace{\frac{1}{4}\pi_b^{\{\{A,B,a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&+\frac{1}{12}\sum_{i\in\Upsilon=\{A,a,b\}}\pi_i^{\{\Upsilon,\{B\}\}}-\underbrace{\frac{1}{4}\pi_B^{\{\{A,a,b\},\{B\}\}}}_{=0\text{ (isolated supplier)}}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{12}\sum_{i\in\Upsilon=\{B,a,b\}}\pi_i^{\{\Upsilon,\{A\}\}}-\underbrace{\frac{1}{4}\pi_A^{\{\{B,a,b\},\{A\}\}}}_{=0\text{ (isolated supplier)}}\\
&\underbrace{-\frac{1}{4}\sum_{i\in\Upsilon=\{A,b\}}\pi_i^{\{\Upsilon,\{B,a\}\}}+\frac{1}{4}\sum_{i\in\Upsilon=\{B,a\}}\pi_i^{\{\Upsilon,\{A,b\}\}}}_{=0\text{ (symmetry)}}\\
&\underbrace{-\frac{1}{4}\sum_{i\in\Upsilon=\{A,B\}}\pi_i^{\{\Upsilon,\{a,b\}\}}}_{=0\text{ (isolated suppliers)}}+\underbrace{\frac{1}{4}\sum_{i\in\Upsilon=\{a,b\}}\pi_i^{\{\Upsilon,\{A,B\}\}}}_{=0\text{ (isolated retailers)}}\\
&\underbrace{+\frac{1}{4}\sum_{i\in\Upsilon=\{A,a\}}\pi_i^{\{\Upsilon,\{B,b\}\}}-\frac{1}{4}\sum_{i\in\Upsilon=\{B,b\}}\pi_i^{\{\Upsilon,\{A,a\}\}}}_{=0\text{ (symmetry)}}\\
&+\frac{1}{6}\sum_{i\in\Upsilon=\{A,b\}}\pi_i^{\{\Upsilon,\{B\},\{a\}\}}-\underbrace{\frac{1}{3}\pi_a^{\{\{A,b\},\{B\},\{a\}\}}}_{=0\text{ (isolated retailer)}}\\
&\underbrace{+\frac{1}{6}\sum_{i\in\Upsilon=\{A,B\}}\pi_i^{\{\Upsilon,\{a\},\{b\}\}}}_{=0\text{ (isolated suppliers)}}-\underbrace{\frac{1}{3}\pi_a^{\{\{A,B\},\{a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&-\frac{1}{6}\sum_{i\in\Upsilon=\{A,a\}}\pi_i^{\{\Upsilon,\{B\},\{b\}\}}+\underbrace{\frac{1}{6}\pi_B^{\{\{A,a\},\{B\},\{b\}\}}}_{=0\text{ (isolated supplier)}}+\underbrace{\frac{1}{6}\pi_b^{\{\{A,a\},\{B\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&+\frac{1}{6}\sum_{i\in\Upsilon=\{B,b\}}\pi_i^{\{\Upsilon,\{A\},\{a\}\}}-\underbrace{\frac{1}{3}\pi_a^{\{\{B,b\},\{A\},\{a\}\}}}_{=0\text{ (isolated retailer)}}\\
&-\frac{1}{6}\sum_{i\in\Upsilon=\{B,a\}}\pi_i^{\{\Upsilon,\{A\},\{b\}\}}+\underbrace{\frac{1}{6}\pi_A^{\{\{B,a\},\{A\},\{b\}\}}}_{=0\text{ (isolated supplier)}}+\underbrace{\frac{1}{6}\pi_b^{\{\{B,a\},\{A\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&\underbrace{-\frac{1}{6}\sum_{i\in\Upsilon=\{a,b\}}\pi_i^{\{\Upsilon,\{A\},\{B\}\}}}_{=0\text{ (isolated retailers)}}+\underbrace{\frac{1}{6}\pi_A^{\{\{a,b\},\{A\},\{B\}\}}}_{=0\text{ (isolated supplier)}}+\underbrace{\frac{1}{6}\pi_B^{\{\{a,b\},\{A\},\{B\}\}}}_{=0\text{ (isolated supplier)}}\\
&-\underbrace{\frac{1}{6}\pi_A^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated supplier)}}-\underbrace{\frac{1}{6}\pi_B^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated supplier)}}\\
&+\underbrace{\frac{1}{2}\pi_a^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}-\underbrace{\frac{1}{6}\pi_b^{\{\{A\},\{B\},\{a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&=\frac{1}{4}\sum_{i\in\Upsilon=\{A,B,a,b\}}\pi_i^{\{\Upsilon\}}\\
&-\frac{1}{4}\sum_{i\in\Upsilon=\{A,B,b\}}\pi_i^{\{\Upsilon,\{a\}\}}+\frac{1}{12}\sum_{i\in\Upsilon=\{A,B,a\}}\pi_i^{\{\Upsilon,\{b\}\}}\\
&+\frac{1}{12}\sum_{i\in\Upsilon=\{A,a,b\}}\pi_i^{\{\Upsilon,\{B\}\}}+\frac{1}{12}\sum_{i\in\Upsilon=\{B,a,b\}}\pi_i^{\{\Upsilon,\{A\}\}}
\end{aligned}
$$

$$
\begin{aligned}
&\underbrace{+\frac{1}{6}\sum_{i\in\Upsilon=\{A,b\}}\pi_i^{\{\Upsilon,\{B\},\{a\}\}}-\frac{1}{6}\sum_{i\in\Upsilon=\{A,a\}}\pi_i^{\{\Upsilon,\{B\},\{b\}\}}}_{=0\text{ (symmetry)}}\\
&\underbrace{+\frac{1}{6}\sum_{i\in\Upsilon=\{B,b\}}\pi_i^{\{\Upsilon,\{A\},\{a\}\}}-\frac{1}{6}\sum_{i\in\Upsilon=\{B,a\}}\pi_i^{\{\Upsilon,\{A\},\{b\}\}}}_{=0\text{ (symmetry)}}\\
&=\frac{1}{4}\sum_{i\in\Upsilon=\{A,B,a,b\}}\pi_i^{\{\Upsilon\}}-\frac{1}{6}\sum_{i\in\Upsilon=\{A,B,b\}}\pi_i^{\{\Upsilon,\{a\}\}}+\frac{1}{6}\sum_{i\in\Upsilon=\{A,a,b\}}\pi_i^{\{\Upsilon,\{B\}\}}
\end{aligned}
$$

Next, we focus on a horizontal upstream merger and calculate the profit of the integrated firm.

$$
\begin{aligned}
U_{AB}=&\frac{1}{3}\sum_{i\in\Upsilon=\{AB,a,b\}}\pi_i^{\{\Upsilon\}}\\
&+\frac{1}{6}\sum_{i\in\Upsilon=\{AB,b\}}\pi_i^{\{\Upsilon,\{a\}\}}-\underbrace{\frac{1}{3}\pi_a^{\{\{AB,b\},\{a\}\}}}_{=0\text{ (isolated retailer)}}\\
&+\frac{1}{6}\sum_{i\in\Upsilon=\{AB,a\}}\pi_i^{\{\Upsilon,\{b\}\}}-\underbrace{\frac{1}{3}\pi_b^{\{\{AB,a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&\underbrace{-\frac{1}{3}\sum_{i\in\Upsilon=\{a,b\}}\pi_i^{\{\Upsilon,\{AB\}\}}}_{=0\text{ (isolated retailers)}}+\underbrace{\frac{2}{3}\pi_{AB}^{\{\{a,b\},\{AB\}\}}}_{=0\text{ (isolated suppliers)}}\\
&-\underbrace{\frac{1}{3}\pi_{AB}^{\{\{AB\},\{a\},\{b\}\}}}_{=0\text{ (isolated suppliers)}}+\underbrace{\frac{1}{6}\pi_a^{\{\{AB\},\{a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}+\underbrace{\frac{1}{6}\pi_b^{\{\{AB\},\{a\},\{b\}\}}}_{=0\text{ (isolated retailer)}}\\
&=\frac{1}{3}\sum_{i\in\Upsilon=\{AB,a,b\}}\pi_i^{\{\Upsilon\}}+\frac{1}{3}\sum_{i\in\Upsilon=\{AB,b\}}\pi_i^{\{\Upsilon,\{a\}\}}
\end{aligned}
$$

Then, we repeat the exercise for a horizontal downstream merger.

$$
\begin{aligned}
U_{ab}=&\frac{1}{3}\sum_{i\in\Upsilon=\{A,B,ab\}}\pi_i^{\{\Upsilon\}}\\
&+\frac{1}{6}\sum_{i\in\Upsilon=\{A,ab\}}\pi_i^{\{\Upsilon,\{B\}\}}-\underbrace{\frac{1}{3}\pi_B^{\{\{A,ab\},\{B\}\}}}_{=0\text{ (isolated supplier)}}\\
&\underbrace{-\frac{1}{3}\sum_{i\in\Upsilon=\{A,B\}}\pi_i^{\{\Upsilon,\{ab\}\}}}_{=0\text{ (isolated suppliers)}}+\underbrace{\frac{2}{3}\pi_{ab}^{\{\{A,B\},\{ab\}\}}}_{=0\text{ (isolated retailers)}}\\
&+\frac{1}{6}\sum_{i\in\Upsilon=\{B,ab\}}\pi_i^{\{\Upsilon,\{A\}\}}-\underbrace{\frac{1}{3}\pi_A^{\{\{B,ab\},\{A\}\}}}_{=0\text{ (isolated supplier)}}
\end{aligned}
$$

$$
\begin{aligned}
&+ \underbrace{\frac{1}{6}\pi_{A}^{\{\{A\},\{B\},\{ab\}\}}}_{=0\text{ (isolated supplier)}} + \underbrace{\frac{1}{6}\pi_{B}^{\{\{A\},\{B\},\{ab\}\}}}_{=0\text{ (isolated supplier)}} - \underbrace{\frac{1}{3}\pi_{ab}^{\{\{A\},\{B\},\{ab\}\}}}_{=0\text{ (isolated retailers)}} \\
&= \frac{1}{3}\sum_{i\in\Upsilon=\{A,B,ab\}}\pi_i^{\{\Upsilon\}} + \frac{1}{3}\sum_{i\in\Upsilon=\{A,ab\}}\pi_i^{\{\Upsilon,\{B\}\}}
\end{aligned}
$$

Finally, we turn to the vertical merger and, again, calculate the profit of the integrated firm.

$$
\begin{aligned}
U_{Aa} = &\frac{1}{3}\sum_{i\in\{Aa,B,b\}}\pi_i^{\{\Upsilon\}} \\
&+ \frac{1}{6}\sum_{i\in\Upsilon=\{Aa,b\}}\pi_i^{\{\Upsilon,\{B\}\}} - \underbrace{\frac{1}{3}\pi_{B}^{\{\{Aa,b\},\{B\}\}}}_{=0\text{ (isolated supplier)}} \\
&+ \frac{1}{6}\sum_{i\in\Upsilon=\{Aa,B\}}\pi_i^{\{\Upsilon,\{b\}\}} - \underbrace{\frac{1}{3}\pi_{b}^{\{\{Aa,B\},\{b\}\}}}_{=0\text{ (isolated retailer)}} \\
&- \frac{1}{3}\sum_{i\in\Upsilon=\{B,b\}}\pi_i^{\{\Upsilon,\{Aa\}\}} + \frac{2}{3}\pi_{Aa}^{\{\{B,b\},\{Aa\}\}} \\
&- \frac{1}{3}\pi_{Aa}^{\{\{Aa\},\{B\},\{b\}\}} + \underbrace{\frac{1}{6}\pi_{B}^{\{\{Aa\},\{B\},\{b\}\}}}_{=0\text{ (isolated supplier)}} + \underbrace{\frac{1}{6}\pi_{b}^{\{\{Aa\},\{B\},\{b\}\}}}_{=0\text{ (isolated retailer)}} \\
= &\frac{1}{3}\sum_{i\in\{Aa,B,b\}}\pi_i^{\{\Upsilon\}} \\
&+ \frac{1}{6}\sum_{i\in\Upsilon=\{Aa,b\}}\pi_i^{\{\Upsilon,\{B\}\}} + \frac{1}{6}\sum_{i\in\Upsilon=\{Aa,B\}}\pi_i^{\{\Upsilon,\{b\}\}} \\
&+ \frac{1}{3}\pi_{Aa}^{\{\{B,b\},\{Aa\}\}} - \frac{1}{3}\pi_{Aa}^{\{\{Aa\},\{B\},\{b\}\}}
\end{aligned}
$$

Horizontal and Vertical Integration Incentives

Finally, we compare pre- and post-merger profits calculated in the previous section. We start with the *horizontal upstream merger incentives.*

$$U_{AB}\left(\{AB,a,b\}\right) > U_A\left(\Omega\right) + U_B\left(\Omega\right)$$

We insert the profits calculated in the previous section and find

$$
\begin{aligned}
\frac{1}{3}\sum_{i\in\Upsilon=\{AB,a,b\}}\pi_i^{\{\Upsilon\}} + \frac{1}{3}\sum_{i\in\Upsilon=\{AB,b\}}\pi_i^{\{\Upsilon,\{a\}\}} > &\frac{1}{2}\sum_{i\in\Upsilon=\{A,B,a,b\}}\pi_i^{\{\Upsilon\}} \\
&+\frac{1}{3}\sum_{i\in\Upsilon=\{A,B,b\}}\pi_i^{\{\Upsilon,\{a\}\}} \\
&-\frac{1}{3}\sum_{i\in\Upsilon=\{A,a,b\}}\pi_i^{\{\Upsilon,\{B\}\}}
\end{aligned}
$$

Note that

$$\sum_{i\in\Upsilon=\{AB,a,b\}} \pi_i^{\{\Upsilon\}} = \sum_{i\in\Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}}$$

and

$$\sum_{i\in\Upsilon=\{AB,b\}} \pi_i^{\{\Upsilon,\{a\}\}} = \frac{1}{3} \sum_{i\in\Upsilon=\{A,B,b\}} \pi_i^{\{\Upsilon,\{a\}\}}$$

This is because the cost function of each supplier depends only on its own input quantity and not on the input quantity of the other supplier. When the suppliers merge, they bargain as a single unit and, hence, both cost functions enter the objective function of bargaining pairs that include the integrated supplier. However, as we take the partial derivative of the objective function with respect to each input quantity separately, the derivative of one of the two cost functions is always zero. Therefore, the derivatives that determine the input quantities are the same regardless of whether the suppliers are integrated or not.

We use these identities to simplify the inequality even further.

$$\frac{1}{3} \sum_{i\in\Upsilon=\{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}} > \frac{1}{6} \sum_{i\in\Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}}$$

We multiply both sides by 6 and get the final inequality stated in the proposition.

$$2 \cdot \sum_{i\in\Upsilon=\{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}} > \sum_{i\in\Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}}$$

Next, we turn to the *horizontal downstream merger incentives.*

$$U_{ab}\left(\{A,B,ab\}\right) > U_a\left(\Omega\right) + U_b\left(\Omega\right)$$

We insert the profits calculated in the previous section.

$$\begin{aligned}\frac{1}{3} \sum_{i\in\Upsilon=\{A,B,ab\}} \pi_i^{\{\Upsilon\}} + \frac{1}{3} \sum_{i\in\Upsilon=\{A,ab\}} \pi_i^{\{\Upsilon,\{B\}\}} > &\frac{1}{2} \sum_{i\in\Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} \\ &-\frac{1}{3} \sum_{i\in\Upsilon=\{A,B,b\}} \pi_i^{\{\Upsilon,\{a\}\}} \\ &+\frac{1}{3} \sum_{i\in\Upsilon=\{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}}\end{aligned}$$

We rearrange the terms.

$$\begin{aligned}\frac{1}{3} \sum_{i\in\Upsilon=\{A,ab\}} \pi_i^{\{\Upsilon,\{B\}\}} - \frac{1}{3} \sum_{i\in\Upsilon=\{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}} + \frac{1}{3} \sum_{i\in\Upsilon=\{A,B,b\}} \pi_i^{\{\Upsilon,\{a\}\}} > \\ \frac{1}{2} \sum_{i\in\Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} - \frac{1}{3} \sum_{i\in\Upsilon=\{A,B,ab\}} \pi_i^{\{\Upsilon\}}\end{aligned}$$

We multiply both sides by 6 and get the final inequality.

$$2 \cdot \sum_{i \in \Upsilon=\{A,ab\}} \pi_i^{\{\Upsilon,\{B\}\}} - 2 \cdot \sum_{i \in \Upsilon=\{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}} + 2 \cdot \sum_{i \in \Upsilon=\{A,B,b\}} \pi_i^{\{\Upsilon,\{a\}\}} > 3 \cdot \sum_{i \in \Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} - 2 \cdot \sum_{i \in \Upsilon=\{A,B,ab\}} \pi_i^{\{\Upsilon\}}$$

Finally, we turn to the *vertical merger incentives.*

$$U_{Aa}\left(\{Aa, B, b\}\right) > U_A\left(\Omega\right) + U_a\left(\Omega\right)$$

We insert the profits calculated in the previous section and find

$$\frac{1}{3} \sum_{i \in \{Aa,B,b\}} \pi_i^{\{\Upsilon\}} + \frac{1}{6} \sum_{i \in \Upsilon=\{Aa,b\}} \pi_i^{\{\Upsilon,\{B\}\}} + \frac{1}{6} \sum_{i \in \Upsilon=\{Aa,B\}} \pi_i^{\{\Upsilon,\{b\}\}} + \frac{1}{3} \pi_{Aa}^{\{\{B,b\},\{Aa\}\}} - \frac{1}{3} \pi_{Aa}^{\{\{Aa\},\{B\},\{b\}\}} >$$
$$\frac{1}{4} \sum_{i \in \Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} + \frac{1}{6} \sum_{i \in \Upsilon=\{A,B,b\}} \pi_i^{\{\Upsilon,\{a\}\}} - \frac{1}{6} \sum_{i \in \Upsilon=\{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}}$$
$$+ \frac{1}{4} \sum_{i \in \Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} - \frac{1}{6} \sum_{i \in \Upsilon=\{A,B,b\}} \pi_i^{\{\Upsilon,\{a\}\}} + \frac{1}{6} \sum_{i \in \Upsilon=\{A,a,b\}} \pi_i^{\{\Upsilon,\{B\}\}}$$

Note that some terms cancel out.

$$\frac{1}{3} \sum_{i \in \{Aa,B,b\}} \pi_i^{\{\Upsilon\}} + \frac{1}{6} \sum_{i \in \Upsilon=\{Aa,b\}} \pi_i^{\{\Upsilon,\{B\}\}} + \frac{1}{6} \sum_{i \in \Upsilon=\{Aa,B\}} \pi_i^{\{\Upsilon,\{b\}\}} + \frac{1}{3} \pi_{Aa}^{\{\{B,b\},\{Aa\}\}} - \frac{1}{3} \pi_{Aa}^{\{\{Aa\},\{B\},\{b\}\}} > \frac{1}{2} \sum_{i \in \Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}}$$

We rearrange the terms.

$$\frac{1}{2} \sum_{i \in \Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} - \frac{1}{3} \sum_{i \in \{Aa,B,b\}} \pi_i^{\{\Upsilon\}} < \frac{1}{6} \sum_{i \in \Upsilon=\{Aa,b\}} \pi_i^{\{\Upsilon,\{B\}\}} + \frac{1}{6} \sum_{i \in \Upsilon=\{Aa,B\}} \pi_i^{\{\Upsilon,\{b\}\}} + \frac{1}{3} \pi_{Aa}^{\{\{B,b\},\{Aa\}\}} - \frac{1}{3} \pi_{Aa}^{\{\{Aa\},\{B\},\{b\}\}}$$

We multiply both sides by 6 and get the final inequality.

$$3 \cdot \sum_{i \in \Upsilon=\{A,B,a,b\}} \pi_i^{\{\Upsilon\}} - 2 \cdot \sum_{i \in \{Aa,B,b\}} \pi_i^{\{\Upsilon\}} < \sum_{i \in \Upsilon=\{Aa,b\}} \pi_i^{\{\Upsilon,\{B\}\}} + \sum_{i \in \Upsilon=\{Aa,B\}} \pi_i^{\{\Upsilon,\{b\}\}}$$

$$+2 \cdot \pi_{Aa}^{\{\{B,b\},\{Aa\}\}} - 2 \cdot \pi_{Aa}^{\{\{Aa\},\{B\},\{b\}\}}$$

2.E Incentives with Nash-in-Nash Bargaining

2.E.1 Brief Overview of the Assumptions

In this appendix, we use the model of Inderst and Wey (2000) to investigate vertical integration incentives in a Nash-in-Nash bargaining setup. The model is roughly the same as the one in Inderst and Wey (2003) and our paper, except for the bargaining part. That is, two suppliers produce differentiated inputs which are turned into final goods by two retailers. Suppliers can have increasing or decreasing average costs, there is no downstream competition, and final goods are either substitutes or complements at both retail outlets. For simplicity, we will focus on the case where both suppliers and both retailers are symmetric. Furthermore, we have adopted the notation of our paper.

As for the bargaining part of their model, each supplier s negotiates with each retailer r over a contract that involves a quantity q_{sr} and a transfer v_{sr}. All negotiations take place simultaneously and each supplier-retailer pair negotiates separately. They impose two key assumption. First, they assume passive beliefs, i.e., both parties believe that all other bargaining pairs will settle on the equilibrium quantities. Second, bargaining is efficient, i.e., firms choose the input quantity q_{sr} such that they maximize their joint profit given that all other firms set the equilibrium input quantities.

2.E.2 Results of Inderst and Wey (2000)

Let us first briefly recap the results of Inderst and Wey (2000) for the case of full separation, which we will denote with a superscript S. The transfer paid by a retailer r to a supplier s is given by

$$v_{rs}^S = \frac{1}{2}\left[q^* p\left(q^*, q^*\right) - q^*\left[p\left(q^*, 0\right) - p\left(q^*, q^*\right)\right]\right] + \frac{1}{2}\left[C\left(2q^*\right) - C\left(q^*\right)\right], \tag{2.19}$$

where q^* is the equilibrium input quantity.

The profit of a retailer r is

$$U_r^S = q^* p\left(q^*, q^*\right) + q^*\left[p\left(q^*, 0\right) - p\left(q^*, q^*\right)\right] - \left[C\left(2q^*\right) - C\left(q^*\right)\right].$$

The profit of a supplier s is

$$U_s^S = q^* p\left(q^*, q^*\right) - q^*\left[p\left(q^*, 0\right) - p\left(q^*, q^*\right)\right] - C\left(q^*\right).$$

The joint pre-merger profit of a retailer r and a supplier s is

$$U_{r+s}^S = U_r^S + U_s^S = 2q^* p\left(q^*, q^*\right) - C\left(2q^\star\right). \tag{2.20}$$

2.E.3 Results with Nash-in-Nash Bargaining

Next, we derive the optimal transfers and profits if supplier A and retailer a merge vertically.

Bargaining with a retailer

Let us focus on the negotiations of the integrated firm Aa with retailer b. The following equation determines how the joint surplus is split between Aa and b.

$$\begin{aligned}[2q^*p(q^*,q^*) - C(2q^*) + v_{Ab} - v_{Ba}] - [q_{Aa}p(q_{Aa},q^*) + q^*p(q^*,q_{Aa}) - C(q_{Aa}) - v_{Ba}] \\ = [2q^*p(q^*,q^*) - v_{Ab} - v_{Bb}] - [q^*p(q^*,0) - v_{Bb}]\end{aligned}$$

$$\begin{aligned}\Leftrightarrow\ [2q^*p(q^*,q^*) - C(2q^*) + v_{Ab}] - [q_{Aa}p(q_{Aa},q^*) + q^*p(q^*,q_{Aa}) - C(q_{Aa})] \\ = [2q^*p(q^*,q^*) - v_{Ab}] - q^*p(q^*,0)\end{aligned}$$

The left side of the (first) equation is the gains from trade of Aa and the right side is the gains from trade of b.

If production decisions and negotiations are made simultaneously, $q_{Aa} = q^*$ holds. This follows from the assumption of passive beliefs. (That is, $q_{Aa} = q^*$ is optimal given that the integrated firm Aa expects all other input quantities to be q^*.) Solving for v_{Ab} yields

$$v_{Ab} = \frac{1}{2}[2q^*p(q^*,q^*) - C(q^*) + C(2q^*) - q^*p(q^*,0)],$$

which is equal to the transfer under full separation (2.19).

If, however, the integrated firm is allowed to adjust q_{Aa} in a second stage, it will set

$$q^*_{Aa} = \arg\max_{q_{Aa}}\ q_{Aa}p(q_{Aa},q^*) + q^*p(q^*,q_{Aa}) - C(q_{Aa}).$$

Because the vertically integrated firm responds optimally in the second stage, it can improve its threat point compared to the case of fully separated firms, i.e.,

$$q^*_{Aa}p(q^*_{Aa},q^*) + q^*p(q^*,q^*_{Aa}) - C(q^*_{Aa}) > 2q^*p(q^*,q^*) - C(q^*).$$

Solving for v_{Ab} yields

$$\begin{aligned}v_{Ab} &= \frac{1}{2}[q^*_{Aa}p(q^*_{Aa},q^*) + q^*p(q^*,q^*_{Aa}) - C(q^*_{Aa}) + C(2q^*) - q^*p(q^*,0)] \\ &> \frac{1}{2}[2q^*p(q^*,q^*) - C(q^*) + C(2q^*) - q^*p(q^*,0)].\end{aligned}$$

This means, a merger between supplier A and retailer a results in a higher price that retailer b must pay for input A.

Bargaining with a supplier

Now we turn to the case where the vertically integrated firm Aa negotiates with supplier B. The following equation determines how the joint surplus is split between Aa and B.

$$[2q^*p(q^*,q^*) - C(2q^*) + v_{Ab} - v_{Ba}] - [q_{Aa}p(q_{Aa},0) - C(q + q_{Aa}) + v_{Ab}]$$

$$= [v_{Ba} + v_{Bb} - C(2q^*)] - [v_{Bb} - C(q^*)]$$

$$\Leftrightarrow \ [2q^* p(q^*, q^*) - C(2q^*) - v_{Ba}] - [q_{Aa} p(q_{Aa}, 0) - C(q^* + q_{Aa})]$$
$$= v_{Ba} - C(2q^*) + C(q^*)$$

The left side of the (first) equation is the gains from trade of Aa and the right side is the gains from trade of B.

Analogous to the above case, $q_{Aa} = q^*$ holds if production decisions and negotiations are made simultaneously. Solving for v_{Ba} yields

$$v_{Ba} = \frac{1}{2}[2q^* p(q^*, q^*) - C(q^*) + C(2q^*) - q^* p(q^*, 0)],$$

which is equal to the transfer under full separation (2.19).

If, however, the integrated firm is allowed to adjust q_{Aa} in a second stage, it will set

$$q^*_{Aa} = \arg\max_{q_{Aa}} \ q_{Aa} p(q_{Aa}, 0) - C(q^* + q_{Aa}).$$

Because the vertically integrated firm responds optimally in the second stage, it can improve its threat point compared to the case of fully separated firms, i.e.,

$$q^*_{Aa} p(q^*_{Aa}, 0) - C(q^* + q^*_{Aa}) > q^* p(q^*, 0) - C(2q^*).$$

Solving for v_{Ba} yields

$$\begin{aligned} v_{Ba} &= \frac{1}{2}[2q^* p(q^*, q^*) - [q^*_{Aa} p(q^*_{Aa}, 0) - C(q + q^*_{Aa})] - C(q^*)] \\ &< \frac{1}{2}[2q^* p(q^*, q^*) - [q^* p(q^*, 0) - C(2q^*)] - C(q^*)]. \end{aligned}$$

This means, a merger between supplier A and retailer a results in a lower price that the vertically integrated retailer a must pay for input B.

Profits

Finally, the profit of Aa is

$$2q^* p(q^*, q^*) - C(2q^*) + v_{Ab} - v_{Ba}. \tag{2.21}$$

The above analysis shows that $v_{Ab} = v_{Ba}$ if production decisions and negotiations are made simultaneously. This leads to the result that the profit under vertical separation (2.20) equals the profit under vertical integration (2.21).

If, however, the vertically integrated firm is allowed to adjust its production decision in a second stage, the transfer from retailer b to the integrated firm Aa (v_{Ab}) is larger than under full separation and the transfer from the integrated firm Aa to supplier B is lower. This leads to the results that the profit under vertical integration (2.21) is always larger than the profit under vertical separation (2.20).

Bibliography

Baake, P., Kamecke, U., and Normann, H.-T. (2004). Vertical foreclosure versus downstream competition with capital precommitment. *International Journal of Industrial Organization*, 22(2):185–192.

Bolton, P. and Whinston, M. D. (1993). Incomplete contracts, vertical integration, and supply assurance. *Review of Economic Studies*, 60(1):121–148.

Bonanno, G. and Vickers, J. (1988). Vertical separation. *Journal of Industrial Economics*, 36(3):257–265.

Chen, Z. (2019). Supplier innovation in the presence of buyer power. *International Economic Review*, 60(1):329–353.

Choi, J. P. and Yi, S.-S. (2000). Vertical foreclosure with the choice of input specifications. *RAND Journal of Economics*, 31(4):717–743.

Colangelo, G. (1995). Vertical vs. horizontal integration: pre-emptive merging. *Journal of Industrial Economics*, 43(3):323–337.

Collard-Wexler, A., Gowrisankaran, G., and Lee, R. S. (2019). "Nash-in-Nash" bargaining: A microfoundation for applied work. *Journal of Political Economy*, 127(1):163–195.

De Fontenay, C. C. and Gans, J. S. (2005a). Vertical integration and competition between networks. *Review of Network Economics*, 4(1):4–18.

De Fontenay, C. C. and Gans, J. S. (2005b). Vertical integration in the presence of upstream competition. *RAND Journal of Economics*, 36(3):544–572.

De Fontenay, C. C. and Gans, J. S. (2014). Bilateral bargaining with externalities. *Journal of Industrial Economics*, 62(4):756–788.

European Union (2008). Guidelines on the assessment of non-horizontal mergers under the council regulation on the control of concentrations between undertakings. *Official Journal of the European Union*, (C 265):6–25. `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52008XC1018(03)`.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719.

Gul, F. (1989). Bargaining foundations of Shapley value. *Econometrica*, 57(1):81–95.

Hart, O. and Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98(6):1119–1158.

Hart, O., Tirole, J., Carlton, D. W., and Williamson, O. E. (1990). Vertical integration and market foreclosure. *Brookings papers on economic activity. Microeconomics*, 1990:205–286.

Inderst, R. and Valletti, T. (2011). Incentives for input foreclosure. *European Economic Review*, 55(6):820–831.

Inderst, R. and Wey, C. (2000). Market structure, bargaining, and technology choice. *WZB Discussion Paper No. FS IV 00-12*, pages 1–33.

Inderst, R. and Wey, C. (2003). Bargaining, mergers, and technology choice in bilaterally oligopolistic industries. *RAND Journal of Economics*, 34(1):1–19.

Kranton, R. E. and Minehart, D. F. (2000). Networks versus vertical integration. *RAND Journal of Economics*, 31(3):570–601.

Luco, F. and Marshall, G. (2020). The competitive impact of vertical integration by multi-product firms. *American Economic Review*, 110(7):2041–2064.

Montez, J. V. (2007). Downstream mergers and producer's capacity choice: why bake a larger pie when getting a smaller slice? *RAND Journal of Economics*, 38(4):948–966.

Ordover, J. A., Saloner, G., and Salop, S. C. (1990). Equilibrium vertical foreclosure. *American Economic Review*, 80(1):127–142.

Rajan, R. G. and Zingales, L. (1998). Power in a theory of the firm. *Quarterly Journal of Economics*, 113(2):387–432.

Rogerson, W. P. (2020). Modelling and predicting the competitive effects of vertical mergers: The bargaining leverage over rivals effect. *Canadian Journal of Economics/Revue canadienne d'économique*, 53(2):407–436.

Salinger, M. A. (1988). Vertical mergers and market foreclosure. *Quarterly Journal of Economics*, 103(2):345–356.

Salop, S. C. (2018). The AT&T/Time Warner merger: How Judge Leon garbled Professor Nash. *Journal of Antitrust Enforcement*, 6(3):459–477.

Segal, I. (2003). Collusion, exclusion, and inclusion in random-order bargaining. *Review of Economic Studies*, 70(2):439–460.

Shapiro, C. (2021). Vertical mergers and input foreclosure lessons from the AT&T/Time Warner case. *Review of Industrial Organization*, 59(2):303–341.

Stole, L. A. and Zwiebel, J. (1996). Organizational design and technology choice under intrafirm bargaining. *American Economic Review*, 86(1):195–222.

U.S. Department of Justice and The Federal Trade Commission (2020). Vertical merger guidelines.
https://www.ftc.gov/system/files/documents/reports/us-department-justice-federal-trade-commission-vertical-merger-guidelines/vertical_merger_guidelines_6-30-20.pdf.

Winter, E. (2002). The shapley value. *Handbook of game theory with economic applications*, 3:2025–2054.

Chapter 3

Rising Markups and the Role of Consumer Preferences

Coauthors: Alexander MacKay, Nathan Miller and Joel Stiebale

Abstract: We characterize the evolution of markups for consumer products in the United States from 2006 to 2019. We use detailed data on prices and quantities for products in more than 100 distinct product categories to estimate demand systems with flexible consumer preferences. We recover markups under an assumption that firms set prices to maximize profit. Within each product category, we recover separate yearly estimates for consumer preferences and marginal costs. We find that markups increase by about 30 percent on average over the sample period. The change is attributable to decreases in marginal costs that are not passed through to consumers in the form of lower prices. Our estimates indicate that consumers have become less price sensitive over time.

Acknowledge: We thank Chris Conlon, Charlie Murry, and Ariel Pakes for helpful comments. We are grateful for the comments and suggestions from many seminar and conference participants. Researchers own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Computational support and infrastructure was provided by the "Centre for Information and Media Technology" (ZIM) at Heinrich Heine University Düsseldorf. Hendrik Döpper gratefully acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG) (project 235577387/ GRK1974).

3.1 Introduction

Firms with market power set prices that reflect marginal costs, consumer preferences, and the prices of related products. Economic theory indicates that differences between prices and marginal costs—the markups—have wide-ranging implications for market outcomes. All else equal, an increase in markups transfers wealth from consumers to producers and can cause consumers to change their purchase decisions. These effects lead to less efficient resource allocation and, through reduced production, affect the markets for inputs, such as labor. Changes in markups may also affect the long-run dynamics in an industry by distorting investment and innovation incentives (Aghion et al., 2005). Thus, the growing empirical evidence that markups are rising in the United States and abroad (e.g., De Loecker et al., 2020; Ganapati, 2021a; De Loecker and Eeckhout, 2021) raises important questions for economic policy.

In this paper, we study the markups that arise in the U.S. economy for a vast number of firms and products. Our objective is to understand the supply and demand conditions that influence firms' pricing decisions. Through an analysis of economic mechanisms, we are able to connect markups to other economic outcomes, such as consumer surplus and deadweight loss, and provide context for various policy considerations. For example, with no changes in demand, rising markups may arise from reduced competition (e.g., due to anticompetitive mergers) or from cost-reducing technological progress.[1] Alternatively, rising markups could reflect shifts in consumer preferences, rather than such supply-side changes.

Although measures of prices are often available, marginal costs are typically unobserved to the researcher. Hence, one must interpret the available data through the lens of economic theory to recover markups. Our approach is to estimate differentiated-products demand systems for more than 100 consumer product categories—such as cereals, shampoo, and over-the-counter cold medications—using prices, quantities, and consumer demographics. With demand estimates in hand, we impute the marginal costs and markups that rationalize prices under the assumption of profit maximization. We repeat this procedure separately for each year over 2006–2019. Our approach is standard in industrial organization (e.g., Berry et al., 1995), although most previous applications focus on a single product category, such as ready-to-eat cereal (Nevo, 2001; Backus et al., 2021), beer (Miller and Weinberg, 2017), or yogurt (Villas-Boas, 2007; Hristakeva, 2020). We implement the methodology at scale to obtain markups for thousands of products, across categories, geographic regions, and over time.

We estimate that average markups increase by about 30 percent between 2006 and 2019, with the average Lerner index increasing from approximately 0.45 to 0.60.[2] We find that the aggregate trend is driven by changes within products over time, rather than consumer substitution toward higher markup products. Larger absolute increases obtain for products

[1] In environments with incomplete pass-through, cost reductions do not yield corresponding declines in price.

[2] The Lerner index is calculated as $\frac{p-c}{p}$, where p and c are price and marginal cost, respectively (Lerner, 1934). As long as marginal cost does not exceed price, it can take values from zero to one.

with higher initial markups; however, in percentage terms, the changes that we estimate are similar for high- and low-markup products. Thus, we interpret our results as indicating that the full distribution of product-level markups may be shifting upward over time. Our findings of increasing average markups is consistent with the findings of De Loecker et al. (2020), despite using a different methodology (supply and demand) and data (prices and quantities).

Our paper makes at least three distinct contributions. First, we use models of supply and demand to evaluate changes to markups over time and potential causes, including changes in costs, concentration, demographics, and consumer preferences. Second, we identify a secular decline in price sensitivity for consumer products, which is a key driver of the increasing markups we observe. Using auxiliary data, we document that this trend corresponds to a decline in coupon use and time spent shopping. Third, our flexible demand modeling approach allows us to evaluate the implications for consumer welfare across the income distribution.

Rising markups must be due to either price increases or marginal cost reductions. We observe that real prices increase during the early years of the sample period and then fall during the later years. Specifically, from 2006 to 2012, average real prices increase by seven percent. After 2012, average real prices decline and, by 2019, are only two percent higher than in 2006. Although price increases partially account for rising markups initially, by the latter years of the sample, cost reductions account for most of the aggregate markup trend.

In many models with imperfect competition, including the one that we estimate, cost changes are not completely passed through to prices. In such settings, falling marginal costs would typically lead lower prices but higher markups. However, incomplete pass-through cannot, on its own, explain the combination of lower marginal costs and slightly higher prices that emerges from the data and our estimates. Our estimates indicate that demand-side changes help to account for these trends. We find that demand for consumer products has become less elastic over time. In particular, consumer price sensitivity declines by about 30 percent from 2006 to 2019. Consumer price sensitivity can reflect both the strength of brand-specific preferences and the perceived value of lower prices; in the model, less price sensitive consumers require a greater difference in prices to switch to a less-preferred brand.

We exploit the unique panel structure of our data to explore factors that predict markup trends. In regressions with product and time fixed effects, we find that products with larger increases in markups tend to have greater reductions in both marginal cost and price sensitivity. Indeed, these two factors explain a substantial majority of the differential trends in product-level markups. Changes in consumer demographics and market concentration also are correlated with markups but have much less explanatory power. We then use counterfactual simulations to examine how equilibrium markups would had evolved in response to our estimated changes in price sensitivity and marginal costs if demographics, product assortments, product ownership, and other demand parameters were constant over time. The results confirm that these two factors can account for almost all of the time-series variation in markups.

In many markets, including the consumer products markets we examine, one might expect costs to decline over time as firms improve their production and distribution technologies. Thus, perhaps more surprising is the decline in consumers' price sensitivity. To explore potential mechanisms, we analyze whether changes in price sensitivity are associated with changing retail patterns, such as the growth of online retail and warehouse clubs, or firm-level investments in R&D or marketing. However, we find that these factors account for only a small fraction of the differential category-level trends in price sensitivity. This suggests that lower price sensitivity might instead arise from exogenous shifts in consumer behavior, such as increase in opportunity cost of time. Consistent with this hypothesis, we find that the use of coupons, which involve some small efforts by consumers, has been falling in the U.S. in aggregate since the early 1990s. Over our sample period, total coupons redeemed and coupon redemption rates have fallen by 50 percent and 30 percent, respectively. In addition, according to time use data, time spent shopping on consumer products fell by approximately 20 percent during our sample period.

In our final analyses, we explore consumer surplus and welfare. Our findings indicate that consumer surplus per capita has increased during our sample period despite rising markups. We attribute this to changing preferences, particularly lower price sensitivity. The changes in consumer surplus vary across the income distribution. While consumers with incomes above the median had substantial gains in surplus during the second half of our sample period, the lowest income quartile experienced substantial losses in some time periods and had approximately the same level of consumer surplus at the end of our sample period as they had in 2006.

Changes in markups have been costly for consumers despite the increase in consumer surplus. In a counterfactual simulation, we find that consumer surplus would have been 14 percent higher in 2019 if markups were scaled down to 2006 levels. Furthermore, under the counterfactual of marginal cost pricing, consumer surplus in 2019 increases by 50 percent and total welfare increases by 9 percent. Taken together, these analyses suggest an important impact market of power on resource allocation, aggregate welfare, and the distribution of income—subjects of longstanding interest (e.g., Harberger, 1954).[3]

Our analysis uses detailed product-level sales from the Kilts NielsenIQ Retail Scanner Data, which consists of a large sample of retail stores. The sales data primarily come from mass merchandisers, grocery stores, and drug stores. Out of a wider set of broad-basket retailers (i.e., also including warehouse clubs and dollar stores), consumer spending on these three retail channels comprised 83 percent of revenues in 2007 and 82 percent in 2019. Thus, our focal channels represent a substantial share of spending on consumer products throughout our sample period. Within these channels, our data consists of a sample of product categories and retailers. We complement the sales data with the Kilts

[3]The types of consumer products we focus on (e.g., food, personal care, etc.) represented 10-15 percent of consumer expenditures in 2015. In magnitudes, this is an significant segment, as it is larger than spending on utilities and public transportation (10.0 percent), medical care (8.4 percent), and new and used vehicles (6.6 percent), but smaller than spending on shelter (32.8 percent). Spending shares are obtained from the 2015 calculation of CPI-U importance weights: https://www.bls.gov/cpi/tables/relative-importance/home.htm

NielsenIQ Consumer Panel Data, which contain household-level purchases and demographic information. These data allow us to control for potential selection across retail channels by consumers with different demographics, as well as allowing for differences in product preferences across households.

A significant contribution of this paper is the application of flexible demand models across categories and over time. We employ the random coefficients logit demand model of Berry et al. (1995) and allow consumer preferences to vary with observable and unobservable demographic characteristics. Typical empirical applications of this model return one set of preference parameters. By contrast, we apply the model across 133 categories, and, critically for our analysis of changing preferences, separately in each of year of our sample. In order to estimate a large number of models, we employ micro-moments of consumer purchases to identify heterogeneity parameters and use covariance restrictions to resolve price endogeneity (MacKay and Miller, 2023). Our approach yields a panel of preference parameters from 1,862 estimated models.

Though we primarily focus on aggregate trends across a broad set of product categories, our empirical approach yields estimates that are consistent with more narrow studies that focus on individual product markets. These comparisons prove useful for assessing the potential simplifications of our model and our identification strategy for the price parameter. For example, for coffee, our estimates of marginal costs move one-for-one with the world commodity price index, and, like Nakamura and Zerom (2010), we estimate the commodity price is roughly half of total marginal costs. For ready-to-eat cereals, we estimate costs and margins in line with those of Backus et al. (2021), who employ additional product characteristics and use an instrumental variables strategy.[4] More broadly, for categories that we can find random coefficients logit estimates, we find that our model yields similar elasticities/markups.

Our research contributes to a growing empirical literature on the evolution of markups. Our finding of increasing markups across a number of categories is broadly consistent with De Loecker et al. (2020); given our distinct modeling approach, we are able to provide insights into specific supply and demand mechanisms. A number of studies recover markups from estimates of demand elasticities, as we do, focusing on specific industries over time. Ganapati (2021b) finds that the markups of wholesalers increased over 1992-2012 due to greater scale economies and the expansion of distribution networks, and with consumers benefiting from lower prices and access to higher quality goods. Grieco et al. (2022) find that the markups of automobile manufacturers decreased over 1980-2018 due to greater competition, despite dramatic increases in product quality and reductions in marginal costs. Miller et al. (2022) show that technology adoption in the cement industry over 1974-2019 increased markups and reduced marginal costs, with price levels changing only modestly. Consistent with our results, these studies highlight the role of technological change as a determinant of long run economic outcomes.[5]

[4]In Appendix 3.F, we summarize results that we obtain using the Backus et al. (2021) approach to construct additional product characteristics for ready-to-eat cereals. These are similar to our baseline estimates.

[5]Also related is Peltzman (2020), which analyzes accounting data on manufacturing firms over 1982-2012

Two other articles explore the relationship between changing consumer preferences and markups. Berry and Jia (2010) find that an increase in consumer price sensitivity helps explain a modest decline in the markups of airline carriers over 1999–2006. This result suggests the caveat that the decreases in price sensitivity that we find for consumer products may not extend throughout the economy. As price sensitivity reflects the strength of brand preferences, it may increase in some sectors even as it decreases in others. Finally, Brand (2021) considers the hypothesis that increases in product variety lead to lower price sensitivity. He estimates demand in nine of the consumer product categories that we consider, both in 2006 and 2017, and finds less elastic demand and higher markups in the later year. Key distinguishing factors in our analysis include both the scope of our analysis—we consider a much broader set of product categories in every year—and our use of individual consumer data to link substitution patterns to variation in demographics in the cross section and over time. In addition, we deal with the issue of price endogeneity.

The paper proceeds as follows: In Section 2, we discuss our approach for recovering markups and specify the model of demand and supply. We discuss the data in Section 3. In Section 4, we describe the estimator and our identification strategy, and we validate the results of our empirical approach for selected industries. Section 5 describes the evolution of markups over time and discusses possible determinants of market power. In Section 6, we investigate the role of changes in price sensitivity and its determinants. In Section 7, we calculate consumer surplus and welfare over time for different scenarios. Section 8 concludes.

3.2 Methods

3.2.1 The Demand Approach to Recovering Markups

We follow the demand approach to recover markups. This approach is often used when data on prices and quantity are available, and it is a staple of the industrial organization literature. The approach invokes the assumption that firms maximize profits and then recovers an estimate for marginal costs that rationalizes observed prices. Take the case of a single-product firm that sets a price, P, given a residual demand schedule, $Q(P)$, and total costs, $C(Q)$. Differentiating its profit function with respect to price and rearranging yields a first order condition for profit maximization of the form:

$$\frac{P - C'}{P} = -\frac{1}{\varepsilon} \tag{3.1}$$

where $\varepsilon \equiv \frac{\partial Q(P)}{\partial P}\frac{P}{Q(P)}$ is the price elasticity of demand. The left-hand-side of the equation is the Lerner index, a widely-used measure of markups (Lerner, 1934; Elzinga and Mills, 2011). Knowledge of the demand elasticity identifies the Lerner index. With data on price, one also can recover marginal cost, the additive markup (i.e., $P - C'$), and the price-over-cost markup (i.e., P/C').

and finds support for rising markups and increasing total factor productivity.

The demand approach gained prominence in industrial organization after various methodological advances made it possible to estimate demand systems for markets that contain many differentiated products (e.g., Berry, 1994; Berry et al., 1995). With a demand system in hand, welfare statistics such as consumer surplus can be calculated, and it also becomes possible to conduct counterfactual simulations for policy evaluation or an exploration of causal mechanisms. However, in part due to the computation burden of demand estimation, most applications focus on a single industry or consumer product category. An advance of our paper is that it employs a flexible demand model across many product categories simultaneously.

The main alternative is the so-called *production approach* that was pioneered in Hall (1988) and De Loecker and Warzynski (2012), and is applied to the evolution of markups in De Loecker et al. (2020) and De Loecker and Eeckhout (2021). Under an assumption of cost minimization, the multiplicative markup (i.e., P/C') equals the product of (i) the elasticity of output with respect to a variable input and (ii) the ratio of revenue to expenditures on the variable input. Thus, firm-level markups can be recovered by estimating output elasticities and then scaling with accounting data on revenues and expenditures. As with many research designs, challenges arise in implementation. For example, Raval (2020) finds that using different variable inputs can yield different markups, and Bond et al. (2021) demonstrates that markups may not be identified if revenue is used as a proxy for output.[6] Due to these and other concerns, some scholars have argued that the existing evidence of rising markups is rather suggestive than definitive (e.g., Basu, 2019; Berry et al., 2019; Syverson, 2019).

Importantly, the demand approach we pursue is distinguished from the production approach in that we construct markups at the (much more narrow) level of a product in a specific market. Our estimates are based on observed prices and quantities at this level, instead of firm-level revenue information that aggregates across many products and markets. Thus, we view large-scale evidence on the evolution of markups obtained with the demand approach as a useful complement to the evidence that has been obtained with the production approach (e.g., De Loecker et al., 2020; De Loecker and Eeckhout, 2021).[7] Implementation of the demand approach comes with its own challenges. As suggested by equation (3.1), inferences about markups are inextricably linked to the demand elasticities, so an identification strategy is needed to obtain consistent estimates of the demand-side parameters in the presence of price endogeneity. Perhaps more fundamentally, the demand-side approach requires the researcher to specify the structure of the demand system and the nature of competition between firms.

We maintain the assumptions of differentiated-products Bertrand competition and random coefficients logit demand, which have been widely used in the literature to study consumer products. There may be some product categories for which our assumptions may be inappropriate. Our strategy to mitigate any such misspecification bias is to aggregate results

[6] See also Doraszelski and Jaumandreu (2019) and De Ridder et al. (2022).

[7] One working paper implements both approaches in the context of the U.S. brewing industry, and finds that they deliver similar results (De Loecker and Scott, 2022).

across product categories. Implemented at scale, this allows us to explore how product-level markups have evolved, the reasons for any such changes, and the consequences for consumers and firms.

3.2.2 Demand Model

For each product category and each year, we apply the random coefficients logit model of Berry et al. (1995). We work with scanner data that are aggregated to the level of a retail chain, quarter, and geographic region. As in Backus et al. (2021), we assume that each consumer is affiliated with a single retail chain and geographic region, in the sense that they select among the products sold by one chain in their region. Let there be $j = 0, \ldots, J_{crt}$ products available for purchase in chain c, region r, and quarter t, including an outside good ($j = 0$). Each affiliated consumer chooses among these products. The indirect utility that consumer i receives from a purchase of product $j > 0$ is

$$u_{ijcrt} = \beta_i^* + \alpha_i^* p_{jcrt} + \xi_{jr} + \xi_{cr} + \xi_t + \Delta\xi_{jcrt} + \epsilon_{ijcrt} \tag{3.2}$$

where p_{jcrt} is the retail price, the terms $(\xi_{jr}, \xi_{cr}, \xi_t)$ are product$\times$region, chain$\times$region, and quarter fixed effects, respectively, $\Delta\xi_{jcrt}$ is a structural error term, and ϵ_{ijcrt} is a consumer-specific logit error term. A consumer that selects the outside good receives $u_{i0crt} = \epsilon_{i0crt}$.

We assume that the consumer-specific coefficients, β_i^* and α_i^*, depend on a set of observed and unobserved demographic variables according to

$$\alpha_i^* = \alpha + \Pi_1 D_i \tag{3.3}$$
$$\beta_i^* = \beta + \Pi_2 D_i + \sigma v_i \tag{3.4}$$

where D_i contains the observed demographics and $v_i \sim N(0,1)$ contains an unobserved consumer demographic. We restrict the unobserved demographics to affect only the constant, rather than also prices, because we find that separately identifying both effects is difficult in practice. Allowing β to be absorbed by the product fixed effects, the structural parameters to be estimated are $\theta = (\alpha, \Pi_1, \Pi_2, \sigma)$.

Note that we have omitted subscripts for year and product category. However, as we estimate demand separately for each category-year, all structural parameters and fixed effects are allowed to vary freely by product category and year.

Quantity demanded is given by $q_{jcrt}(p_{crt}; \theta) = s_{jcrt}(p_{crt}; \theta) M_{crt}$, where $s(\cdot)$ is the market share, p_{crt} is a vector of prices, and M_{crt} is the "market size" of the chain-region-period, a measure of potential demand. We refer readers to Nevo (2000b) for equations that characterize market shares and the demand elasticities. We use a market size that is proportional to the population and the number of retail stores operated by the chain within each region. We provide details on the calculation in Appendix 3.B, and we show that our main trends are robust to alternative measures in Appendix 3.E.5.

Our specification accommodates vertical differentiation among the inside goods because higher quality (more expensive) products may attract relatively price-insensitive consumers. This can be an important modeling feature in the context of markup trends, especially to the extent that prices or consumer incomes change over time. Our specification also incorporates heterogeneity in the utility that consumers receive from the inside goods, which allows the data to determine the extent of substitution between the inside and outside goods.[8] In principle, product characteristics other than price could be incorporated into the demand model. We do not pursue this across our categories because it would require matching to auxiliary datasets on characteristics, which would be difficult to implement at scale.[9]

Data on non-price characteristics would allow for a more flexible treatment of horizontal differentiation in the model. It is generally recognized in industrial organization that this can have benefits for counterfactuals involving specific cross-product substitution patterns, such as merger simulation (e.g., Nevo, 2000a) or studies of entry and exit (e.g., Ciliberto et al., 2021). Whether it has first-order implications for markup trends depends on the prevalence of changes in product ownership, such as those that would be introduced by mergers, entry, or exit. Averaging over the product categories, we do not observe meaningful changes in concentration over the sample period (Figure 3.21 in the Appendix).[10] Furthermore, our analysis includes a screen for within-category product differentiation to account for potentially substantive unobserved product characteristics. We obtain similar results with and without this screen (Figure 3.12 in the Appendix). Finally, we test the robustness of our results to including product characteristics for ready-to-eat cereals using a specification that is similar to Backus et al. (2021). We document these results in Appendix 3.F. Putting all of the above results together, we conclude that our treatment of non-price characteristics is unlikely to drive our results.

On the other hand, we find that the consumer heterogeneity parameters we do include meaningfully affect the estimated elasticities and markups. To test this, we also estimate our model using a standard logit demand specification, where we set ($\Pi_1 = 0$, $\Pi_2 = 0$, $\sigma = 0$) for all categories and years. Relative to this specification, we find that our baseline estimates yield more elastic demand and smaller markups. We report these results in Appendix 3.E.7.

3.2.3 Supply Model

Consumer products are produced by manufacturers and sold through retail chains. We assume that each manufacturer sets prices to maximize its profit, taking as given the prices of its competitors and passive cost-plus pricing on the part of retailers. Thus, the retail

[8] An alternative approach that allows data to influence substitution between the inside and outside goods involves specifying a random coefficients nested logit (RCNL) model with the outside good in its own nest (e.g., Grigolon and Verboven, 2014). With the RCNL model, the speed of estimation slows dramatically for higher values of the nesting parameter, making the model inappropriate for our application.

[9] Consider the approach that Backus et al. (2021) take to estimate demand for ready-to-eat cereals. They obtain auxiliary data from Nutritionix about the nutritional content of the products, such as the grain (e.g., wheat or corn) and the sugar content. These data then are consolidated into a handful of principal components that serve as product characteristics in the demand model. For many of the product categories we consider, and all of the non-food categories, nutritional content is unavailable or unlikely to drive consumer substitution.

[10] Bhattacharya et al. (2022) provide a detailed examination of the mergers in the same retail scanner data.

markup becomes part of the marginal cost that the manufacturer must pay to sell their products (Gandhi and Nevo, 2021). This assumption is maintained elsewhere (e.g., Miller and Weinberg, 2017; Backus et al., 2021) and is supported by evidence from the empirical literature.[11]

The first order conditions for profit maximization can be expressed in terms of the additive markup:

$$p_{crt} - c_{crt} = -\left(\Omega_{crt} \circ \left[\frac{\partial s_{crt}(p_{crt})}{\partial p_{crt}}\right]'\right)^{-1} s_{crt}(p_{crt}) \tag{3.5}$$

where the vectors p_{crt}, s_{crt}, and c_{crt} collect the prices, market shares, and marginal costs of products $j = 1, \ldots, J_{crt}$, and Ω_{crt} is an "ownership matrix" in which each j^{th}, k^{th} element equals one if products j and k are produced by the same manufacturer, and zero otherwise. We assume that marginal costs are constant in output. For consumer products, we view this as a reasonable approximation, and the assumption is often maintained in the literature (Villas-Boas, 2007; Chevalier et al., 2003; Hendel and Nevo, 2013; Miller and Weinberg, 2017; Backus et al., 2021).

An implication of optimal price-setting behavior is that firms find it profitable to adjust their markups with demand conditions, which enter equation (3.5) through market shares and demand derivatives. Therefore, our model explicitly allows for price endogeneity, which we address in estimation. We decompose marginal cost according to:

$$c_{jcrt} = \eta_{jr} + \eta_{cr} + \eta_t + \Delta\eta_{jcrt} \tag{3.6}$$

where $(\eta_{jr}, \eta_{cr}, \eta_t)$ are product×region, chain×region, and quarter fixed effects, and $\Delta\eta_{jcrt}$ is a supply-side structural error term. As in our demand specification, all fixed effects can vary freely by product category and year because we estimate separate models for each category-year combination. Thus, our model allows for changes in brand-specific technologies over time, and, on an annual frequency, these changes may be correlated with changes in demand (e.g., a plant closure). The supply-side structural error term incorporates "cost shifters" that have been used in the literature to estimate demand, including changes in materials costs and distribution costs that affect products and chains differentially.

3.3 Data

3.3.1 Data Sources and Estimation Samples

Our primary sources of data are the Retail Scanner Data and Consumer Panel Data of Kilts NielsenIQ, which span the years 2006–2019. The scanner data contain unit sales and revenue

[11] For instance, De Loecker and Scott (2022) find evidence for perfect wholesale-retail pass-through indicating competitive retail markets. There is also evidence that retail prices respond to cost shocks (Butters et al., 2022) but not shocks to retailer demand (Arcidiacono et al., 2020). Finally, evidence suggests that retail markups have been relatively stable over the period 1980-2014, despite large changes in demand (Anderson et al., 2018). Our modeling assumptions are also consistent with nonlinear contracts that specify slotting fees or other fixed transfers.

at the level of the universal product code (UPC), store, and week. The consumer panel data contain the purchases of a sample of panelists by UPC code, retailer, and day, along with demographic information on the panelists. We employ aggregation and a number of screens to construct samples that are suitable for the model laid out in the previous section.

We take as given the consumer product categories ("modules") that are specified by NielsenIQ. Within each category are UPCs that consumers are likely to view as substitutes. Our baseline sample comprises 133 product categories that cover 55 percent of revenues in the Retail Scanner Data. We obtain these categories by first identifying the top 200 categories by revenue, and then applying a screen based on observed price dispersion to avoid categories with highly dissimilar products. We discuss our category selection procedure in more detail in Section 3.3.2.

Within these categories, we define products at the brand level, which consolidates thousands of UPC codes into a more manageable set. Brands are defined by NielsenIQ and are fairly narrow. For example, in ready-to-eat cereals, "Cheerios," "Honey Nut Cheerios," and "Multigrain Cheerios" are three distinct brands.[12] Within a brand, we aggregate sales across UPCs by unit of measurement, which characterizes volume (e.g., liters), mass (e.g., ounces), or count (e.g., six-pack), depending on the category.[13] We measure price using the ratio of revenue to equivalent unit sales, following the standard practice to adjust for differences in package size (e.g., Nevo, 2001; Miller and Weinberg, 2017; Backus et al., 2021). Within each category, we treat the top 20 brands by revenue as distinct products, and we collapse the remaining brands into a single composite "fringe" product that we assume is priced by an independent firm. The top 20 brands within each category account for approximately 85 percent of category revenues and typically include a private label product.[14]

We focus our analysis on the stores that NielsenIQ classifies as mass merchandisers, grocery stores, or drug stores. Our data on prices and quantities comes from a sample of retailers within these channels.[15] More broadly, these retail channels comprise a substantial part of overall spending on consumer products. Based on auxiliary data on the revenues of large U.S. retailers, we estimate that, in 2019, they accounted for 82 percent of revenues among broad-basket retailers (i.e., mass merchandisers, grocery, drug stores, dollar stores, and warehouse clubs). This share of revenue appears to be stable in our sample period, as the estimated share in 2007 is 83 percent. Among all channels, we estimate that mass merchandisers, grocery stores, and drug stores account for over 50 percent of consumer product spending, where the broader sample includes specialty retailers (e.g., electronics,

[12] Other examples include "Oreo," "Oreo Double Stuf", and "Mini Oreo" (cookies) and "Yoplait," "Yoplait Go-gurt," "Yoplait Whips!," "Yoplait Thick & Creamy," and "Yoplait Light Thick & Creamy" (yogurt).

[13] In a handful of categories, UPC codes differ in terms of whether units are reported in terms of volume, mass, or count. For those categories, we use only those UPC codes associated with the highest-revenue metric.

[14] To explore the sensitivity of the analysis to the cap of 20 branded products per category, we perform robustness checks with a sample that includes only 15 branded products per category. We obtain very similar results. More brands could be added to the model with additional effort to connect brands to their owner, following the same process that we use for the brands currently in the sample (as discussed later in this section).

[15] Our analysis in Appendix 3.D suggests that our findings are not sensitive to compositional changes in the data or due to shifts in shopping behavior across or within retail channels.

Table 3.1: Sample of Product Categories

Rank	Product Category	Observations	Revenue ($ Millions)	Retailer-DMA Combinations	Brands Per Market	Share Top 20 Brands	Share Private Label
1	Cereal - Ready To Eat	231,178	22,557	333	19.3	0.58	0.08
2	Candy - Chocolate	229,065	16,162	335	18.9	0.54	0.03
3	Candy - Non-Chocolate	225,336	9,420	334	18.6	0.61	0.14
4	Deodorants - Personal	221,618	7,186	333	18.3	0.79	0.00
5	Soap - Specialty	214,153	5,563	355	17.5	0.68	0.05
6	Tooth Cleaners	212,056	7,343	333	17.6	0.71	0.00
7	Shampoo - Liquid/Powder	202,923	7,490	332	16.8	0.65	0.04
8	Cookies	202,880	17,191	334	16.8	0.64	0.18
9	Sanitary Napkins	201,864	5,128	333	16.7	0.79	0.18
10	Cold Remedies - Adult	201,134	9,111	332	16.6	0.85	0.40
20	Bottled Water	160,454	23,333	335	13.2	0.90	0.38
40	Baby Formula	133,082	10,616	323	12.1	0.76	0.05
60	Nuts - Bags	107,314	6,500	334	8.9	0.79	0.24
80	Fresh Muffins	85,228	3,899	332	7.6	0.85	0.17
100	Tuna - Shelf Stable	68,711	4,099	332	5.7	0.98	0.13
120	Cream - Refrigerated	52,297	3,402	330	4.6	0.70	0.30
130	Frozen Poultry	33,428	2,145	300	3.9	0.86	0.27
133	Fresh Mushrooms	25,510	2,772	246	3.4	0.95	0.28
	Mean Values	108,442	6,766	319	9.8	0.84	0.16

Notes: This table shows summary statistics for a selection of product categories. The chosen categories are sorted by the number observations in the estimation sample and are indexed by *rank*. *Revenue* provides total sales in millions of nominal US $ from 2006 to 2019. The two groups are separated by a horizontal rule. Statistics are calculated after the data cleaning steps described in the text. The last three columns report raw means across retailer-DMA-year-quarter markets. Shares in this table reflect inside shares (i.e., excluding the outside good).

beauty, apparel). Appendix 3.B.3 provides these summary statistics and describes the auxiliary data.[16]

We use the designated market areas (DMAs) in the NielsenIQ data as the geographic regions. We restrict attention to the 22 DMAs for which there are at least 500 panelists in every year in the consumer panel data. These DMAs account for about half of the total revenue observed in the scanner data. Within each DMA, we aggregate the store-level data up to the level of the retail chain, as many retail chains set common prices among nearby stores (DellaVigna and Gentzkow, 2019). Finally, we aggregate the week-level data up to the level of quarters, following Miller and Weinberg (2017). The average number of retail chains per region is 9.3, and the average number of products per category, retail chain, and region is 10.3. Table 3.1 provides summary statistics for a selection of product categories in

[16]The largest broad-basket channel that we omit is warehouse club, which accounts for 9.0 percent of consumer product spending in 2007 and 9.4 percent in 2019. We observe that the revenue share of dollar stores nearly doubles between 2007 and 2019, consistent with the trend documented in Caoui et al. (2022). Nonetheless, dollars stores account for only 1.5 percent of consumer product spending in 2007 and 2.6 percent in 2019. The share of revenues accounted for by retailers that we do not identify as broad-basket declines slightly over time. This reflects a growth of online retailers that is offset by relative declines in other store formats (e.g., department stores, apparel).

the estimation sample sorted by number of observations.

We employ household demographic data to account for differences in the composition of consumers across markets and changes within markets over time. Specifically, we generate consumer-specific demographic draws by sampling 2,000 consumers from the Consumer Panel Data for each region and year.[17] We sample with replacement and using the projection weights provided by NielsenIQ. Among the available demographics, we select two that we expect should influence demand for many of the consumer products in the data: household income and an indicator for the presence of children in the household. We assume that log of income is what enters demand through equations (3.3) and (3.4). We demean the demographics prior to estimation, and also divide the income measure by its standard deviation. The unobserved demographic is drawn from a standard normal distribution that is independent from the observed demographics.

In estimation, we match the empirical purchasing patterns of households across different demographic types, which allows us to control for heterogeneous preferences and for selection by households into different retailers. Specifically, we use the data to construct "micro-moments" that are the average values of observed demographics for consumers that purchase each product in a given region and year, again using the projection weights. Our model attempts to ensure that, for example, the average income of households that purchased Honey Nut Cheerios in Chicago in 2015 matches the data. When constructing these values, we use purchasing data only at a subset of retailers to match the distribution of retailers that appear in the scanner data (e.g., mass merchandisers, grocery stores, and drug stores). Since our sample of households is not restricted in this way, the model provides some adjustment for selection of consumers into the retailers we observe.

We account for multi-product ownership using auxiliary data, as ownership information is not provided in the NielsenIQ databases. We start with a manual search in which we identify the company that owns each product. Because multiple company names could be associated with the same manufacturer when a conglomerate has multiple subsidiary companies, we use data from Capital IQ to obtain the ultimate parent company for each product. This process provides a snapshot of product ownership at the end of our sample period. We backcast ownership for the preceding years using information on mergers and acquisitions (M&A) from the Zephyr database, compiled by Bureau van Dijk. Compared with most other M&A databases, Zephyr has the advantage that there is no minimum deal value for a transaction to be included. We assume that prices are chosen to maximize the profit of the ultimate parent company. Finally, we match our sample with firm-level financial data from Compustat to obtain information on marketing expenditures and R&D. We use these variables to explain variation in price sensitivities across brands and time. This information is available for about half of the observations in our sample because Compustat covers publicly traded firms.

[17]By sampling at the region-year level, we implicitly assume that the consumers of retail chains within the same region have the same demographics. We take this approach to because we view the consumer panel data as too sparse to reliably sample at the level of a retail chain, region, and year. For a study of consumer demographics and prices as they vary spatially across a city, see Eizenberg et al. (2021).

We deflate prices and incomes using the Consumer Price Index such that they are in real dollars as of the first quarter of 2010.[18]

3.3.2 Selection of Product Categories

Some challenges arise in recovering markups over time using the estimation samples described above. In treating the NielsenIQ categories as well-defined product markets, we create the potential for model misspecification, due to at least two (related) reasons. The first is that products in different categories might be substitutes. For instance, one might suspect some amount of consumer substitution between products in the "Light Beer" and "Beer" categories. In principle, these categories could be combined, possibly with richer demand specification that allows for weaker substitution between light beer and beer. However, looking holistically across the NielsenIQ categories, we are skeptical that cross-category substitution is meaningful for most products. Thus, for our research question, it seems more appropriate to use the NielsenIQ categories rather than making ad hoc adjustments, and that is the approach we take.

The second reason for concern about NielsenIQ product categories—which we view as more important for our application—is that some categories include products that might be very weak substitutes (or possibly not substitutes at all). The "Batteries" category, for example, has some products that are probably close substitutes, such as various brands of AAA batteries, along with other products that are functionally quite different, such as D batteries. We use a relatively tractable specification of the random coefficients logit model in order to scale estimation across categories, and do not consider the model to be sufficiently flexible to handle such rich patterns of product differentiation. This can be problematic if the same demand parameters—and especially the price parameter—are inappropriate for different classes of products within the same category.

To address this potential concern, we use the within-category distribution of prices as a proxy for within-category product heterogeneity, and remove categories in which the 99th percentile of prices is greater than five times the median price. This screen leaves 133 of the top 200 product categories (by revenue) in the baseline sample. The top 200 categories account for 74 percent of revenues in the Retail Scanner Data; the 133 categories in the baseline sample account for 55 percent. Although our screen for within-category heterogeneity focuses attention on categories for which the model is a likely to be a better fit, it does not drive results; we obtain similar markup trends with screens that are more or less strict. In Appendix 3.E.1, we report the product-level markup trends using all 200 of the product categories.

Also worthy of discussion are the compositional changes that occur in the NielsenIQ data as retail stores enter and exit the sample. Such churn appears to be inconsequential over 2006–2017, but significant changes do occur over 2018–2019. Because we estimate

[18]We deflate using the Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average. This CPI measure is predominantly constructed from products and services outside of the categories in our sample. The inflation data are monthly and seasonally adjusted.

independent models separately in each year, compositional changes do not affect the trends we observe from 2006–2017. We control for some aspects of compositional changes in 2018–2019 by including (yearly) chain×region fixed effects in the demand and marginal cost equations and allowing market sizes to scale separately for each retail chain. Moreover, we show in Appendix 3.E.3 that we find nearly identical trends with a balanced panel that includes only brand×chain×region combinations that occur in every year of our sample. In Appendix 3.E.4, we also perform robustness checks where we supplement our baseline data with large retailers present only in the consumer panel data, and we obtain similar results.

3.4 Empirical Strategy

3.4.1 Estimation and Identification

We estimate the equilibrium model developed in Section 3.2 using the generalized method of moments (GMM). We estimate separate models for each category and year, allowing the parameters for estimation, $\theta = (\alpha, \Pi_1, \Pi_2, \sigma)$, to vary arbitrarily across models. The GMM estimator for θ is:

$$\hat{\theta} = \arg\min_{\theta} g(\theta)' W g(\theta), \qquad g(\theta) = \begin{bmatrix} g^{MM}(\theta) \\ g^{CR}(\theta) \end{bmatrix} \tag{3.7}$$

where W is a weighting matrix, $g^{MM}(\theta)$ collects a set of micro-moments that summarizes how well the model matches the correlations between demographics and product purchases that we observe in the NielsenIQ Panelist dataset, and $g^{CR}(\theta)$ implements a covariance restriction between demand-side and cost-side structural error terms. We take a two-step approach to estimation in which we first estimate $\theta_2 = (\Pi_1, \Pi_2, \sigma)$ then estimate the price parameter, α. This reflects that micro-moments identify θ_2 but not α (Berry et al., 2004; Berry and Haile, 2022), and that the covariance restriction exactly identifies α conditional on θ_2 (MacKay and Miller, 2023). In Appendix 3.A, we explain why this segmentation has computation advantages in our setting and provide additional details on the estimation procedure.

For micro-moments, we use variation in purchase patterns across products and regions to capture heterogeneity in preferences. Each element corresponding to product j and demographic k is given by

$$g_{jk}^{MM}(\theta) = \frac{1}{T_j} \sum_{c,r,t} \left(\frac{\sum_i \omega_i s_{ijcrt}(\theta) D_{ik}}{\sum_i \omega_i s_{jcrt}(\theta)} - \mathcal{M}_{jrk} \right) \tag{3.8}$$

where T_j is the number of chain-region-quarter combinations in which product j is sold, ω_i is the weight that we place on consumer i, $s_{ijcrt}(\theta)$ is the consumer-specific choice probability implied by the candidate parameter vector, and $\mathcal{M}_{jrk}$ is the mean demographic observed in the data for product and region. That is, we match the implied average demographic of consumers for each product-chain-region-quarter to the average demographic observed in the data for the corresponding product-region (allowing for differences across years and categories).[19] In our baseline specification, we use two observed demographic variables and at most 21 products, so there can be up to 42 micro-moments. Estimation of θ_2 is standard and identification strategies for these parameters are reasonably well understood.[20] However, micro-moments that can be used to pin down heterogeneity in preferences cannot recover the mean price parameter and resolve price endogeneity.

We address this with covariance restrictions, which are appealing in our setting because they can be implemented at scale for different product categories and years. Specifically, in the second step, we identify the price parameter under the assumption that the demand-side and supply-side structural error terms are uncorrelated in expectation: $\mathbb{E}[\Delta\xi_{jcrt}\Delta\eta_{jcrt}] = 0$. We construct the empirical analog of the moment condition:

$$g^{CR}(\theta) = \frac{1}{T}\sum_{c,r,t} \Delta\xi_{crt}(\theta)'\Delta\eta_{crt}(\theta) \tag{3.9}$$

where the $\Delta\xi_{crt}(\theta)$ and $\Delta\eta_{crt}(\theta)$ terms are recovered for each candidate θ using standard techniques, and T is the number of chain-region-quarter-product combinations for a given year.

Alternative approaches to identify the price parameter typically rely on auxiliary data on cost-shifters or product characteristics, which can be difficult and costly to obtain. An additional benefit of the covariance restrictions approach is that—in contrast to an instruments-based approach—there is no "first-stage" relevance condition that must be satisfied (MacKay and Miller, 2023). Even if product characteristics were available for every category and year, there is no guarantee that, for example, markup-shifter instruments that rely on such characteristics (e.g., Berry et al., 1995; Gandhi and Houde, 2020) would meet the relevance condition for every category of interest.

Moreover, as we have specified the model, the supply-side structural error term ($\Delta\eta_{jcrt}$) incorporates the variation of some of the cost-shifter instruments that have been used to estimate demand in the recent literature, including product-specific shipping costs (Miller and Weinberg, 2017) and the prices of product-specific ingredients (Backus et al., 2021). These and other plausibly exogenous cost-shifters may be highly correlated with the variation that we exploit in estimation.[21]

The marginal cost function and the demand function include fixed effects at the product$\times$region, chain$\times$region, and quarter levels, absorbing some potential sources of endogeneity. For instance, product$\times$region fixed effects capture variation in quality that may be

[19] We allow the average observed demographics to vary by year and category. An alternative approach to the micro-moments would match the implied chain-region demographics to chain-region data, rather than to region-level data. The tradeoff is between the measurement error in the observed component versus the specificity of the moments. However, parameters that fit one set of moments well should also fit the other well.

[20] Berry and Haile (2022) show that micro-moments can identify the non-linear parameters of both observable and unobservable demographics (Π and σ) with variation across and within markets.

[21] See MacKay and Miller (2023) for a more detailed discussion and additional examples.

associated with production/distribution costs and tastes that may vary geographically. Chain×region fixed effects capture consumer heterogeneity across retailers and regions as well as differences in retailer markups and costs. Quarter fixed effects control for seasonal changes in demand and production costs. The residual variation in costs might reflect, for example, that the shipping costs of heavy products rises disproportionately in regions with idiosyncratic increases in gas prices in a particular quarter. All fixed effects shift arbitrarily year-over-year, allowing for longer-run changes in production technology that are correlated with demand.

The covariance restrictions approach to estimation differs in some ways from an instrument-based approach. In particular, the covariance restrictions approach uses all of the endogenous price and quantity data in estimation, rather than only the portions that are attributable to excluded instruments. Although this eliminates the first-stage relevance requirement, it does require the joint estimation of parametric models of supply and demand. Thus, a misspecification of the marginal cost function could affect demand estimates. However, because a fully specified supply-side model is required to recover markups, we view it as sensible to also employ the supply model to estimate structural parameters.[22]

As shown by MacKay and Miller (2023), reduced price sensitivity would suggest that the ratio of the variation in quantities to the variation in prices is falling over time. Intuitively, this reflects demand that is less sensitive to price variation. A change in the price coefficient corresponds to a rotation of the inverse demand curve; more inelastic demand will result in a more "vertical" inverse demand and inverse supply curves on a price-quantity plot and a lower relative variance. Indeed, in our data, within-market price dispersion is increasing while within-market share dispersion is falling, and the changes to the relative variance are highly correlated with the changes in the price sensitivity we estimate.

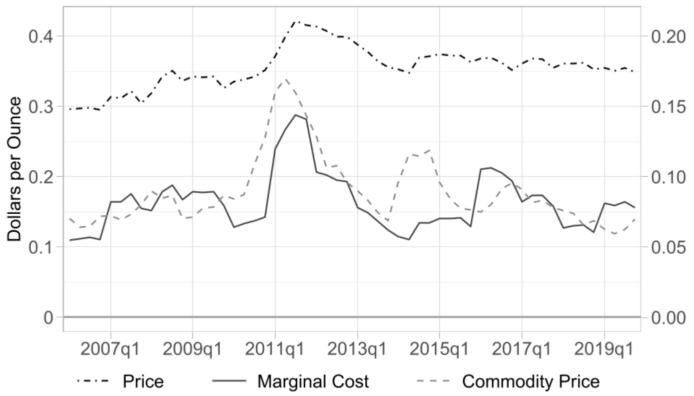
3.4.2 Assessment

We conduct three validation checks to assess the reasonableness of our approach. First, we examine one product category—ground/whole bean coffee—to assess the ability of our method to capture marginal costs. Coffee is somewhat unique among our product categories in that a single ingredient (coffee beans) accounts for a substantial portion of marginal costs and commodity prices for this ingredient are well-established. Second, we compare the own-price elasticities of demand that we obtain to those obtained in the literature. Third, we plot the distribution of elasticities that we obtain with our baseline estimates, and also compare this distribution to two alternative approaches that have been used in the literature.

Marginal Cost Estimates Figure 3.1 plots the time series of quantity-average weighted prices (dot-dash line) and marginal costs (solid line) for coffee. Prices are observed, and

[22]Simulations in MacKay and Miller (2023) suggest that the covariance restriction approach can be robust to modest supply-side misspecification. As an empirical robustness check, we explore an alternative approach that does not require our supply-side model in estimation. In Appendix 3.E.6, we calculate trends in demand (elasticity and price sensitivity) under the assumption that prices are exogenous. Though this often provides biased elasticities (see Section 3.4.2), we interpret the robustness check as being consistent with rotations in the demand curve, i.e., with demand becoming less elastic.

Figure 3.1: Prices and Marginal Costs of Coffee Over Time



Notes: This figure plots the time series of quantity-weighted prices and marginal costs (solid line) for ground/whole bean coffee. Prices are observed and marginal costs are recovered from the profit-maximization conditions. Also shown is the commodity price index for coffee (dashed gray line), which is scaled following the right axis.

marginal cost are recovered according to equation (3.5). The gray dashed line plots the commodity price index for coffee, which is scaled separately on the right axis.[23] Overall, our recovered estimates of marginal costs are strongly correlated with the commodity price index. A regression of average marginal costs on the commodity price yields a coefficient of 0.990 ($p < 0.001$), and the correlation between the two time series is 0.61.[24] Our method is able to capture the large spike in commodity prices in 2011, which is reflected in the spike in marginal costs. We find that, on average, the commodity price is equal to 56 percent of estimated marginal costs. This is consistent with the literature, as Nakamura and Zerom (2010) find that coffee beans account for 45 percent of marginal costs based on data spanning 2001-2004. These results indicate the potential of our empirical approach to recover reasonable marginal cost estimates.

Elasticity Estimates in the Literature Next, we compare our product-level own-price elasticities of demand to those obtained in the literature using similar data and models. In Table 3.2, we report estimates for beer, ready-to-eat cereal, and yogurt, for which comparisons are possible. As shown, we obtain elasticities for beer, ready-to-eat cereal, and yogurt of -4.06, -2.29, and -3.12, respectively. To provide more comparable estimates, we report the median product-level own price elasticities for beer and ready-to-eat cereal, and the mean own-price elasticity from 2006–2010 for yogurt.[25] For beer, we combine beer and light beer categories to match Miller and Weinberg (2017), who do not distinguish between these categories. Miller and Weinberg (2017) report a median elasticity for beer of -4.74,

[23]Data on coffee commodity prices were obtained from Macrotrends.net. Available here: `https://www.macrotrends.net/charts/commodities`, last accessed March 1, 2022

[24]Regressing average marginal costs on the one-period lagged commodity price yields a coefficient of 1.046 and a correlation of 0.66. This slightly stronger relationship may reflect the use of contracts. The relationship is weaker with longer lags.

[25]Every paper differs in the exact data sample used. For example, Hristakeva (2020) uses data from 2001–2010. Because we find rising markups over time for yogurt, restricting it to the earlier years of our sample provides a closer comparison. None of these papers allow preference parameters to vary over time.

Table 3.2: Average Product-Level Own-Price Elasticities of Demand

Category	Our Estimate	Literature Estimate	Citation
Beer	−4.06	−4.74	Miller and Weinberg (2017)
Ready-to-Eat Cereal	−2.29	−2.42	Backus et al. (2021)
Yogurt	−3.12	−4.05	Hristakeva (2020)

Notes: The Miller and Weinberg (2017) estimate is the median product-level elasticity obtained with the RCNL-1 specification. Our corresponding estimate is the median own-price elasticity across all years, combining "Beer" and "Light Beer," which are not distinguished in Miller and Weinberg (2017). The Backus et al. (2021) estimate is the median product-level elasticity obtained with the "prices only" specification; our corresponding estimate is the median own-price elasticity across all years. Hristakeva (2020) reports a mean product-level elasticity from 2001–2010; to make things more comparable, we report our estimated mean own-price elasticity from 2006–2010.

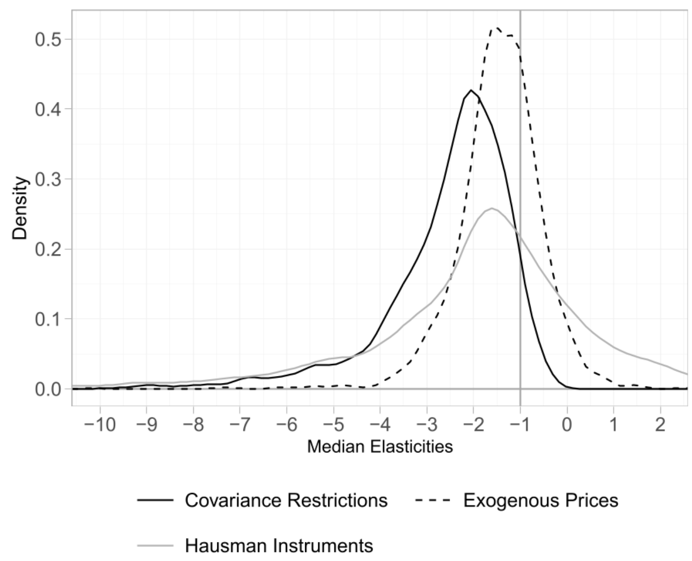
Backus et al. (2021) reports a median elasticity for ready-to-eat cereal of -2.42, and Hristakeva (2020) reports a mean elasticity for yogurt of -4.05. Thus, we conclude that our methodology can obtain reasonable results that are consistent with analyses that make use of specific institutional details to a greater degree.

To provide a more detailed comparison, consider the empirical approach of Backus et al. (2021), which was developed concurrently. In their analysis of ready-to-eat cereals, Backus et al. (2021) use the Kilts NielsenIQ data over a similar time period (2007-2016) with a smaller sample of DMAs, retailers, and weeks. The supply model is quite similar, and the random coefficients logit demand model includes the same consumer demographics that we include in our analysis. One key distinction is that Backus et al. (2021) also collect product characteristics that are included in the demand model. A second key distinction is that, instead of covariance restrictions, Backus et al. (2021) employ two sets of instruments that are constructed from input costs and the characteristics of other products (Berry et al., 1995; Gandhi and Houde, 2020). Despite these differences, we obtain similar elasticities and margins.[26] Furthermore, we run an additional specification for ready-to-eat cereals using product characteristics, and show that this does not materially affect our estimates (Appendix 3.F).

Alternative Identification Strategies For the third validation check, we examine the distribution of median own-price elasticities across all of the 1,862 category-year combinations in our baseline sample. We compare the results to those obtained under two alternative assumptions that can identify the price parameter and be applied at scale. The first alternative assumption is that prices are exogenous. For a given model of supply and demand, price exogeneity holds if both (a) firms do not adjust markups in response to demand shocks and (b) demand shocks are uncorrelated with marginal cost shocks. If the latter condition fails, then prices are endogenous (i.e., correlated with demand shocks) even if firms do not

[26]For cereals, our average unit price is 0.20 and our average estimated marginal cost is 0.10. We find that average markups for this category are relatively stable over time, which is consistent with the De Loecker et al. (2020) estimates for cereals over our sample period.

Figure 3.2: Implied Elasticities Under Alternative Identification Restrictions



Notes: This figure plots the density of the median own-price elasticity by category and year under different identification assumptions. The solid black line shows the density of implied elasticities using covariance restrictions. The dashed line shows the density of implied elasticities assuming exogenous prices. The solid gray line shows the density of implied elasticities using Hausman instruments. The vertical line indicates an elasticity of -1.

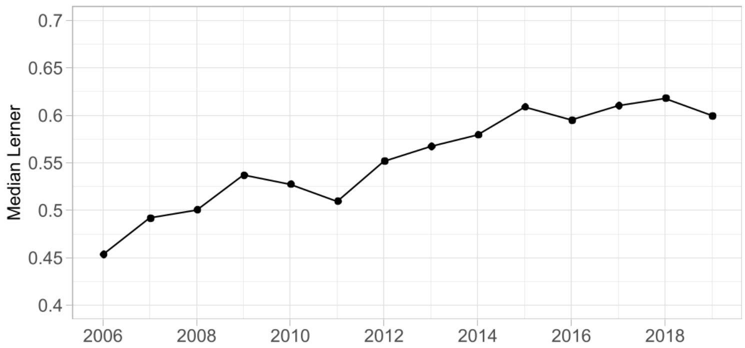
respond directly to demand. Thus, a covariance restriction is necessary for consistent estimation under an assumption of exogenous prices. However, profit maximization generally implies that prices are endogenous, and our covariance restrictions approach to estimation corrects for price endogeneity.

The second alternative approach to estimation uses instruments based on the average price of the same product in other regions (Hausman, 1996). This approach is valid if cost shocks are correlated across regions due to shared manufacturing or distribution facilities, for example, but demand shocks are uncorrelated across regions. These conditions may not be satisfied in many empirical settings. For example, validity can be threatened if firms employ region-wide or national advertising campaigns. Thus, Hausman instruments are at best subject to scrutiny when employed (Berry and Haile, 2021; Gandhi and Nevo, 2021).

Figure 3.2 plots the densities of median own-price elasticities. The solid black line summarizes the results that we obtain with covariance restrictions (our baseline assumption). As shown, the peak of the distribution with covariance restrictions occurs at an elasticity slightly more negative than -2. Relative to our estimates, the distributions of elasticities with exogenous prices (the dashed line) and Hausman instruments (the solid gray line) are shifted to the right, consistent with price endogeneity arising from firms adjusting prices in response to demand shocks. Though covariance restrictions systematically correct for price endogeneity, Hausman instruments do not, and instead yield more elastic demand than exogenous prices in some cases and more inelastic demand in others.

Using covariance restrictions, demand is never upward-sloping, and only 5 percent of the category-year combinations have inelastic demand (i.e., a median elasticity greater than

Figure 3.3: Markups Over Time Across Product Categories



Notes: This figure plots the mean of within-category median markups over time. Markups are defined by the Lerner index, $(p - mc)/p$, and are estimated separately by product category and year. When calculating the mean, we winsorize the upper and lower 2.5 percent of observations across all categories and years.

-1). By contrast, 29 percent of the category-year estimates exhibit inelastic demand with exogenous prices; with Hausman instruments, it is 34 percent. Furthermore, both of those approaches yield several estimates with upward-sloping demand. These results suggest the covariance restrictions approach generates reasonable demand elasticities, and that it is a distinctly good way to approach estimation in our context.

Of course, our ultimate interest is in the evolution of markups across the many different categories in our estimation sample, and we turn to that exercise next.

3.5 The Evolution of Markups in Consumer Products

In this section, we document the evolution of markups across consumer products over time. We start by reporting median markups at the product category level before we discuss how the distribution of markups has shifted. We then move the analysis to the product level which allows us to distinguish within-product variation from variation across products and to decompose the evolution of markups into changes in prices and marginal costs.

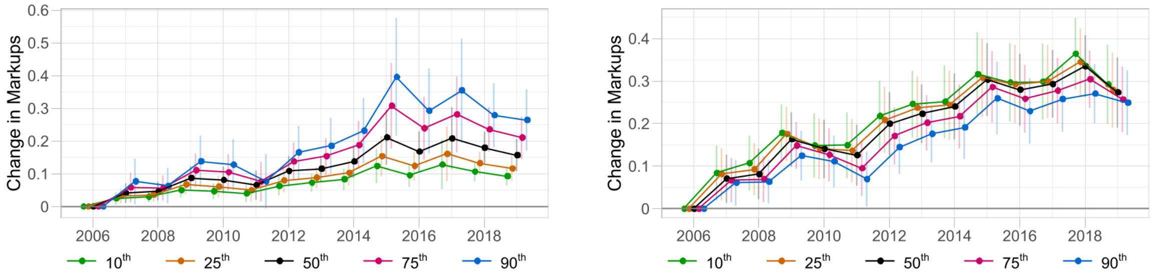
3.5.1 Aggregate Markup Trends

Our estimation procedure yields a panel of 14.4 million product-level observations across 133 categories and 14 years. To evaluate aggregate trends, we first consider changes in the category-level markups in the 1,862 category-year combinations in our data. We take the median markup within each category-year, and we then calculate the mean across categories in each year. Figure 3.3 plots this statistic over time. Averaging across categories, we find an increase in the median Lerner index from approximately 0.45 in 2006 to over 0.60 towards the end of our sample period. This corresponds to an average annual growth rate in markups of 2.3 percent.

Next, we analyze how the distribution of markups within product categories has shifted over time. For this purpose, we regress different percentiles of the markup distribution on year dummies and document the coefficients and confidence intervals in panel (a) of

Figure 3.4: Changes in the Distribution of Markups

(a) Absolute Change (b) Relative Change



Notes: This figure shows coefficients and 95 percent confidence intervals of regressions of percentiles of the markup distribution at the product category level on year dummies using the year 2006 as the base category. In panel (a), outcomes are percentiles of the level of the Lerner index, $(p-c)/p$, in panel (b), outcomes are measured in logarithms.

Figure 3.4. We use the year 2006, the first year of our estimation sample, as the base category. Hence, the estimated coefficients can be interpreted as the change in markups in each year relative to 2006. The results indicate that, while all quartiles of the distribution have increased over time, the upper part of the markup distribution has changed by a higher amount, especially during the second half of our sample period. In panel (b), we repeat the exercise by using the log of the Lerner index, $\ln(\frac{p-c}{p})$. The results show that the *relative* increase in markups is in fact quite similar across the distribution and even slightly more pronounced for lower quartiles. Overall, our estimates indicate that the full distribution of markups is shifting upward over time.[27]

3.5.2 Within-Product Changes in Markups, Prices, and Marginal Costs

The aggregate trends in markups that we document could be explained by firms charging higher markups on existing products or by market entry and exit of brands with different levels of markup. Further, to the extent that within-product changes explain rising markups, this could be due to higher prices, reductions in marginal costs, or both. To evaluate these mechanisms, we analyze the change in markups, prices, and marginal costs at the product level, where our unit of observation is a unique product-chain-region-quarter-year combination.

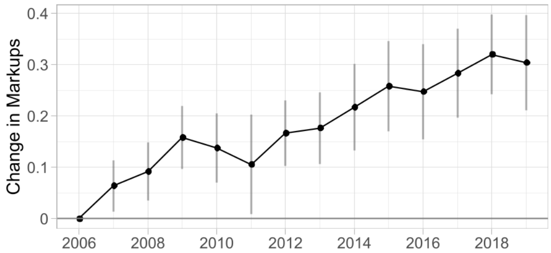
For markups, we regress the log of the Lerner index on quarter, year, and product-chain-region fixed effects, using revenues as weights.[28] The results of this regression are documented in panel (a) of Figure 3.5. The figure displays point estimates and 95 percent confidence intervals for the year fixed effects. The estimates indicate an increase in product-level markups of about 30 percent between 2006 and 2019. The estimated annual growth rate in product-level markups is 2.2 percent per year. With the inclusion of product-chain-

[27] We find similar changes in the distribution of firm-level markups which we calculate as quantity-weighted averages over brands owned by each parent company.

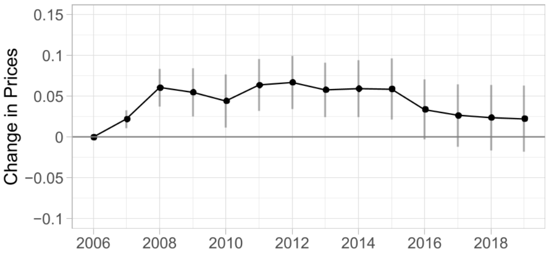
[28] We weight by revenues instead of quantities to assign higher weights to products with higher initial prices. Revenue-weighted relative changes, which we measure by changes in log markups, are consistent with quantity-weighted absolute changes in a consumption basket.

Figure 3.5: Product-Level Changes in Markups, Prices, and Marginal Costs

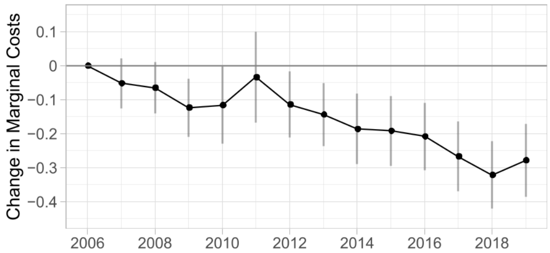
(a) Markups



(b) Prices



(c) Marginal Costs



Notes: This figure shows coefficients and 95 percent confidence intervals of a regressions of the log of the Lerner index, real prices, and real marginal costs at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category.

DMA fixed effects, the nonparametric time trend only captures variation within products. Thus, the estimated change over time is not affected by entry and exit or a reallocation of market shares across products. This indicates that the aggregate markup trends are mainly driven by changes within products over time. We find similar results if we instead use price-over-cost (p/c) markups, as studied by De Loecker et al. (2020). See Appendix 3.E.2.

Table 3.10 in the Appendix provides the full regression results that corresponds to panel (a) of Figure 3.5, alongside alternative specifications in which we replace year fixed effects with a linear time trend, drop product-chain-DMA fixed effects, or use category fixed effects. We obtain qualitatively similar results across these specifications, and estimate average yearly increases in average markups between 1.7 and 2.2 percent. We estimate larger changes when controlling for product-level fixed effects, indicating that within-product changes are greater than the aggregate (revenue-weighted) changes in markups. Though these differences are not substantial, they suggest that some of the product-level increase in markups may be offset by the introduction of lower-markup products over time.[29]

Using our detailed data on prices and our demand estimates, we can decompose the increase in markups into changes in prices and marginal costs (equation (3.5)). For this purpose, we regress log prices and log marginal costs on product-DMA-retailer, quarter, and year fixed effects. Prices and marginal costs are deflated by core CPI and indexed to Q1 of 2010. The yearly coefficients are documented in panels (b) and (c) of Figure 3.5.

[29]Table 3.11 in the Appendix shows results using unweighted regressions. The results are similar. As Table 3.12 shows, we also obtain similar results if we focus on a balanced panel of products, indicating that the overall trends are not primarily driven by the entry and exit of products with different markup growth rates.

Panel (b) shows that real prices increase at the beginning of our sample period, but decline in later years. The average real price for products in our sample increases by 7 percent over 2006 to 2012, but real prices are only 2 percent higher in 2019 than in 2006. Panel (c) of the figure reports the yearly coefficients for log marginal costs. We estimate that marginal costs decline by 2.1 percent per year on average.[30]

In 2017–2019, marginal costs are roughly 25 log points lower than in 2006.[31] Thus, although higher real prices account for part of the increase in markups during the first half of our sample, the higher markups we observe at the end of our sample arise from lower real marginal costs, not higher real prices. Overall, our estimates suggest that declines in real marginal costs have not been passed on to consumers.

3.5.3 Changes in Demand

Why might lower marginal costs not lead to lower prices? Incomplete pass-through arises in many models of imperfect competition, including the one that we estimate. However, on its own, incomplete pass-through cannot explain the combination of lower marginal costs with slightly higher prices. Economic theory suggests other possibilities that could contribute to this phenomenon, including changes in demand. More inelastic demand would put upward pressure on markups, as evident in the first-order conditions in equations (3.5) and (3.1). Another possible explanation for increasing markups is the consolidation of brand ownership, which might occur due to mergers and acquisitions. We do not observe meaningful increases in concentration in our data (see Figure 3.21 in the Appendix).

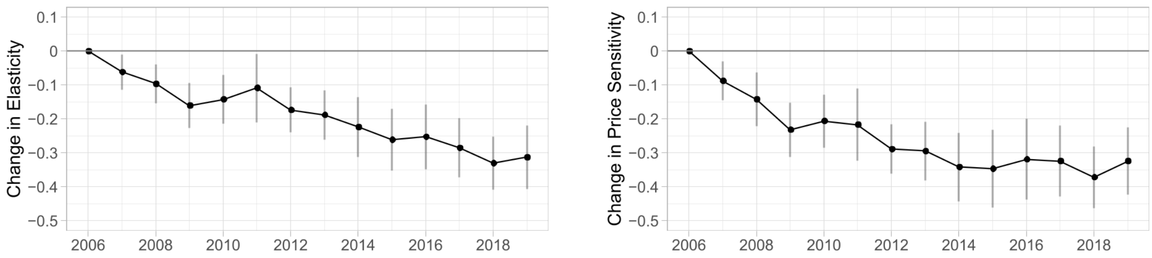
To investigate the possibility of demand-side changes, we first regress the logarithm of the absolute value of own-price elasticities at the product level on the same set of fixed effects used above. We present the results in panel (a) of Figure 3.6. The displayed coefficients show that price elasticities have declined in magnitude, indicating that demand indeed becomes less responsive to prices over time. Price elasticities capture several underlying aspects of consumer preferences and may also reflect supply-side factors such as quality and competition. However, in our sample the main driver appears to changes in the mean price coefficients that we estimate for each category and year in the data. These parameters implicitly adjust for changing consumer demographics and selection by consumers into retailers and products. We repeat the regression exercise using price sensitivity, which we define as the log absolute value of the mean price coefficient (i.e., $\log(-\alpha)$), as the dependent variable. Panel (b) shows that the declines in price sensitivity were large through 2012, corresponding with the increase in real prices we observe over the same period. In econometric and simulation-based exercises that we present shortly, this reduced price sensitivity

[30] An interesting feature of our results is that marginal costs increase between 2009 and 2011, as the Producer Price Index (PPI) for farm products was increasing. The coincident declining markups indicate that these costs were not fully passed through to consumer prices. A similar but more modest increase in the PPI for farm products over 2006-2007 is not evident in our marginal cost estimates. Another explanation for declining markups over 2009-2011 in these years is trading-down behavior of consumers during the recession (Jaimovich et al., 2019).

[31] Figure 3.19 in the Appendix uses nominal (i.e., non-deflated) prices and marginal costs, and shows that nominal marginal costs are relatively constant over time.

Figure 3.6: Changes in Demand

(a) Log Absolute Elasticity (b) Price Sensitivity



Notes: This figure shows coefficients and 95 percent confidence intervals of a regression of log absolute elasticity and price sensitivity at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category.

emerges as an important determinant of rising markups.

Our estimates allow us to examine other changes in demand as well. For instance, the fixed effects allow us to characterize changes in perceived product quality over time, relative to that provided by the outside good. We measure product quality as the value that an average consumer obtains from the product (relative to the outside good); to improve comparability across categories we standardize values using the category-level means and standard deviations. Figure 3.20 in the Appendix shows that perceived product quality declines over the sample period. Improvements in the outside good—which includes shopping through online retailers for example—could contribute to this trend. The same appendix figure plots changes in the coefficients that characterize how observed consumer demographics affect the consumer-specific price coefficient and category-level constant (Π_1, Π_2). As we discuss below, these changes have relatively little impact on markup trends.

To summarize, our decomposition of effects indicates that the increase in markups was driven by lower real marginal costs, without commensurate reductions in real prices. Firms were able to charge higher markups because consumers became less price sensitive over time.

3.5.4 Panel Data Analysis

To evaluate the relative importance of demand and supply channels in driving changes in product-specific markups, we use a regression analysis that exploits the unique panel structure of our estimates across products and over time. Specifically, we regress product-level log markups on consumer preference parameters, marginal costs, consumer demographics, and market concentration. We use category and year fixed effects, such that the regression coefficients capture deviations from aggregate trends. We focus on the ability of the regressors to explain changes in product-level markups, as reflected by their contribution to the R^2.

For the consumer preference parameters, we include price sensitivity and perceived product quality, as defined in the previous section. We standardize the product qualities and marginal costs, separately by for each category, so that they have a variance of one.[32] For

[32] Standardization improves comparability across categories and also eases interpretation of the coefficients.

Table 3.3: Factors Predicting Cross-Category Variation in Markup Trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Marginal Cost (Standardized)	-0.564^{***}					-0.450^{***}	-0.449^{***}
	(0.024)					(0.023)	(0.023)
Price Sensitivity		-0.721^{***}				-0.392^{***}	-0.393^{***}
		(0.030)				(0.022)	(0.022)
Quality (Standardized)			-0.142^{***}			0.006	0.007
			(0.022)			(0.006)	(0.006)
Income (Log)				0.052^{**}		0.059^{***}	0.058^{***}
				(0.025)		(0.013)	(0.013)
Children at Home				-0.175^{***}		-0.076^{***}	-0.083^{***}
				(0.064)		(0.026)	(0.027)
Parent HHI					0.236		0.236^{***}
					(0.186)		(0.046)
Brand HHI					0.091		-0.097^{**}
					(0.178)		(0.048)
Retailer HHI					0.203^{***}		0.074^{***}
					(0.077)		(0.025)
Brand-Category-DMA-Retailer FEs	X	X	X	X	X	X	X
Time Period FEs	X	X	X	X	X	X	X
Observations	14,407,410	14,407,410	14,407,410	14,407,410	14,407,353	14,407,410	14,407,353
R^2 (Within)	0.719	0.468	0.047	0.000	0.003	0.826	0.827

Notes: This table reports regression results where the dependent variable is log markups. Observations are at the brand-category-DMA-retailer-year-quarter level, and brand-category-DMA-retailer and year-quarter fixed effects are included in each specification. Standard errors are clustered at the category level and are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

consumer demographics, we use log income and the presence of young children at home. Finally, for market concentration, we examine three constructions of the Herfindahl-Hirschman Index (HHI). Parent HHI is calculated for the upstream parent companies of the products (i.e. for the brand manufacturers). Brand HHI is calculated under the counterfactual of single-product firms, and serves to isolate changes in market concentration that are unrelated to product ownership. Finally, Retailer HHI is calculated for the retailers, separately for each category and region. We measure the HHIs on a zero-to-one scale.[33]

Table 3.3 summarizes the results. Each regression includes fixed effects for each product-market (i.e., brand $\times$ category $\times$ retailer $\times$ region) and time period (year $\times$ quarter). Thus, the coefficients reflect the correlations of within-product changes over time. Standard errors are clustered at the product category level. The R^2 (within) statistic shows how much of the residual variation in markups—i.e., the portion not absorbed by fixed effects—is accounted for by the explanatory variables.

The results indicate that changes in marginal costs and price sensitivity are highly

We choose this approach to standardization, rather than logs, so as to include observations with negative values.

[33]We use the consumer panel data to construct HHI measures. Our results are qualitatively similar if we instead use the retail scanner data.

correlated with rising markups, and can explain the bulk of the variation in within-product markup changes. Column (1) indicates that marginal cost reductions alone can explain 72 percent of the within-product variation in markups (within $R^2 = 0.719$). The coefficient implies that a one standard deviation reduction in marginal costs is associated with a 56 percent increase in markups. Similarly, column (2) indicates that declines in price sensitivity alone can explain 47 percent of the within-product variation in markups; the coefficient indicates that a 10 percent decrease in price sensitivity is associated with a 7.2 percent increase in markups.

Note that price sensitivity is measured at the category-year level, whereas markups and marginal costs may vary across brands, DMAs, and retailers within each category-year. If we run regressions at the product category level, we find similar coefficients and a higher within R^2 for price sensitivity. We report these results in Table 3.13 in the Appendix.

Columns (3), (4), and (5) examine perceived quality, consumer demographics, and concentration. Although some of the coefficients are statistically significant, each of these measures explains little of the variation in log markups, with within R^2 values less than 0.05.

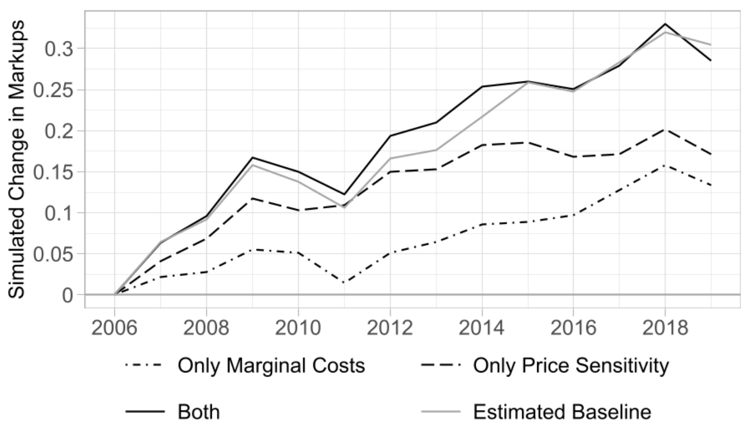
In column (6), we combine price sensitivity, quality, and marginal costs with demographic characteristics. The coefficients on price sensitivity and marginal costs decline modestly, but remain large in magnitude and statistically significant. The coefficient on quality becomes effectively zero. Thus, though declines in relative perceived quality are correlated with increasing markups in the time series, products with greater increases in quality do not realize differential changes in markups. Increases in income remain positively associated with greater markups. Changes in price sensitivity, marginal costs, and demographics explain most of the variation in markups over time. The within R^2 is 0.83.

In column (7), we add our measures of concentration to the specification. We find that changes in retailer concentration remain positively correlated with changes in markups, and the coefficient for parent-manufacturer concentration increases and becomes statistically significant. Yet, these coefficients remain modest. The parent-retailer coefficient of 0.236 in column (7) indicates that a 0.02 change in parent company HHI—i.e., a 200-point change on a 0 to 10,000 scale—is associated with a 0.5 percent increase in markups. The relationship between markups and changes in concentration at the retailer level and brand level (which ignores multi-product ownership) is weaker. Overall, the inclusion of concentration measures does little to change the explanatory power of the regression, as the R^2 barely changes.

3.5.5 Impacts of Marginal Costs and Price Sensitivity on Markups

The previous subsection shows that reductions in marginal costs and price sensitivity are highly correlated with the variation in markup growth across products. Here, we use counterfactual simulations to show the hypothetical causal impact of these two factors on markup trends, holding everything else fixed. Specifically, taking 2006 data as a starting point, we change price sensitivity and/or marginal costs, holding fixed product assortments, consumer demographics, and demand parameters. Given these hypothetical changes, we use equation

Figure 3.7: Simulated Markup Changes



Notes: This figure plots counterfactual log changes in markups from simulations that scale marginal costs (dash-dotted line), price sensitivities (dashed line), or both (solid line) according to the average realized changes that are reported in Figures 3.5 and 3.6. Markups are defined by the Lerner index, $(p - mc)/p$, and changes are reported relative to 2006. Product assortments, consumer demographics, and other demand parameters are held fixed at 2006 values in each simulated year. The solid gray line plots the estimated change in log markups in the realized data for comparison.

(3.5) to solve for equilibrium prices and compute markups. In each simulated year, we apply a uniform relative change to scale product-specific values by the estimated aggregate changes documented in Figures 3.5 and 3.6. Thus, we ask to what extent aggregate changes in marginal costs and price sensitivity can explain aggregate trends in markups.

The results of the counterfactual simulations are depicted in Figure 3.7. The dash-dotted line shows that, relative to 2006, estimated changes in marginal costs would have increased markups by about 13 percent in 2019 if preferences, demographics, product assortments, and ownership had not changed. Changes in price sensitivity (holding marginal costs and other factors fixed) would have increased markups by more than 15 percent towards the end of the sample period, as indicated by the dashed line. The solid black line shows that simulated markups increase by about 28 percent from 2006 to 2019 if we adjust both price sensitivity and marginal costs at the same time. The trajectory of simulated markup changes tracks overall markup trends, depicted by the gray line, closely. Hence, changes in price sensitivity and marginal costs account for nearly all of the time-series variation in markups. Consistent with trends documented in Figures 3.5 and 3.6, changes in markups can be mainly attributed to changes in price sensitivity in the first half of our sample period, while marginal costs are the main driver of rising markups in the second half of our sample.

Economic theory provides a tight theoretical connection between changes in marginal costs and markups. In typical models of imperfect competition, a decline in marginal costs will not be fully passed on to consumers (i.e., cost pass-through is less than one). If costs fall faster than prices, then markups increase. Thus, the relationship that we find between markups and marginal costs is partly a result of imperfectly competitive product markets and declining costs. This logic applies to more general settings: in otherwise stable economic environments, declining costs will yield higher markups due to imperfect competition.

In many markets, we expect costs to decline over time due to innovations in production/distribution technology and operational efficiencies. Our empirical setting is no exception, as many manufacturers sought ways to reduce costs over this time period. For example, Procter & Gamble, one of the largest companies in our data, began a "productivity and cost savings plan" in 2012 that was estimated to reduce annual costs by \$3.6 billion in 2019.[34] Overall, our finding of modest declines in marginal costs is consistent with secular increases in productivity across the economy.

There is also a tight theoretical connection between price sensitivity and markups. All else equal, firms will charge higher prices to less price sensitive consumers. However, in contrast to our finding of declining marginal costs, it is perhaps more surprising that we find that consumer price sensitivity has fallen over time. In the following section, we examine the role of price sensitivity in more detail and discuss potential explanatory factors for the time trend.

3.6 Price Sensitivity and Markups

In this section, we explore the role that price sensitivity plays in explaining changing markups. First, we provide an econometric decomposition that isolates the price sensitivity parameter from observable features of the market that also determine markups in equilibrium. We use this decomposition to provide further evidence for the special role of consumer price sensitivity. We then explore potential mechanisms that could be driving changes in price sensitivity.

We apply an econometric decomposition developed to examine the role that the mean price parameter plays in our analysis. As shown by MacKay and Miller (2023), we can write the product-level additive markups as a function of the mean price parameter (α) and an inverse supply ($\lambda(\cdot)$) for a broad class of oligopoly models. In our model of random coefficients logit demand and Bertrand pricing, this takes the form:

$$p_{jcrt} - c_{jcrt} = -\frac{1}{\alpha}\lambda_{jcrt}(s_{crt}, p_{crt}, \Gamma_{crt}; \Pi_1, \Pi_2, \sigma), \tag{3.10}$$

where s_{crt} and p_{crt} vectors of market shares and prices at the chain-region-quarter level, and Γ_{crt} denotes the matrix of partial demand derivatives (with respect to prices). From an econometric standpoint, $\lambda_{jcrt}(\cdot)$ is a function of market shares, prices, and consumer-specific choice probabilities; it does not depend on the mean price parameter. In Appendix 3.C, we provide the specific functional form of $\lambda(\cdot)$.

Taking the quantity-weighted average within each category and year and dividing by average price, we obtain an expression for the aggregate Lerner index,

$$\overline{L} = \frac{\overline{p} - \overline{c}}{\overline{p}} = -\frac{1}{\alpha}\frac{\overline{\lambda}}{\overline{p}}, \tag{3.11}$$

[34]The Procter & Gamble Company 2019 Annual Report. Available here:
https://www.pg.com/annualreport2019/download/PG-2019-Annual-Report.pdf

In logs, we obtain:

$$\ln \overline{L} = \underbrace{-\ln(-\alpha)}_{-1\times\text{Price Sensitivity}} + \underbrace{\ln\left(\frac{\overline{\lambda}}{\overline{p}}\right)}_{\text{Structural Factors}}, \tag{3.12}$$

where we can decompose the (log) category markups into price sensitivity (i.e., $\ln(-\alpha)$) and a term that captures the net effect of other structural factors: the qualities and marginal costs of products, the ownership of products (i.e., market concentration), the parametric assumptions, and the nonlinear preference parameters. This term can be obtained from directly observable data on product ownership, market shares, prices, and consumer purchasing patterns such as the micro-moments that we use in the first stage of estimation.

The decomposition suggests a regression-based approach to explore the degree to which price sensitivity explains variation in markups across both product categories and over time. We start with cross-sectional regressions—separately for 2006, 2017, and 2019—in which the dependent variable is the category-level aggregate Lerner Index (in logs) and the independent variable is price sensitivity. We present statistics for 2006 and 2019 because they are the first and last years of the sample, and we include 2017 due to the 2018 change in the NielsenIQ data (see the discussion in Section 3.3.2). We also consider a panel regression with observations at the category×year level in which the dependent variable is the year-over-year change in the (log) aggregate Lerner Index and the independent variable is the year-over-year change in price sensitivity.

Table 3.4 summarize the results. The R^2 in columns (1)-(3) indicates that variation in price sensitivity explains a modest fraction of the cross-sectional variation in markups: 16 percent in 2006, 27 percent in 2017, and 7 percent in 2019. This suggests that other structural factors, such as product qualities and multi-product ownership, are relatively more important in explaining variation in markups across categories. Further, this highlights that our demand specification is sufficiently rich to attribute much of the variation in markups across categories to structural factors that are uncorrelated with consumer price sensitivity.[35] As our prior results indicate that decreasing price sensitivity is correlated with higher markups, one might suspect its explanatory power also to increase over time. Consistent with that, the R^2 in 2017 is higher than that of 2006; the lower R^2 is 2019 may be attributable to the compositional shift in the scanner data (which we control for in the analysis of the previous section.)

Column (4) summarizes the results of the panel regression. We find that changes in price sensitivity over time explain 58 percent of the variation in markups over time. Thus, to understand rising markups among the consumer products that we examine, it appears necessary to have an understanding of consumer price sensitivity and how it has changed over time. That is, an econometrician with data on product ownership, market shares, prices, and consumer purchasing patterns—which are sufficient to recover $\lambda(\cdot)$ within a specific modeling context—could make incorrect inferences about markup trends *unless* the model also allows for changes in price sensitivity. This points to a strength of our modeling

[35] This need not be the case with less flexible demand systems. For example, with constant elasticity demand, the Lerner index only varies due to differences in price sensitivity (i.e., $\lambda_t = p_t$ and $\ln(\lambda_t/p_t) = 0$).

Table 3.4: Price Sensitivity and Markups Across Product Categories

	(1) 2006 Log $\overline{L}$	(2) 2017 Log $\overline{L}$	(3) 2019 Log $\overline{L}$	(4) Δ Log $\overline{L}$
Price Sensitivity	-0.134^{***}	-0.200^{***}	-0.090^{***}	
	(0.027)	(0.029)	(0.029)	
Δ Price Sensitivity				-0.575^{***}
				(0.012)
Observations	133	133	133	1,729
R^2	0.162	0.268	0.070	0.571

Notes: This table reports regression results that examine the cross-sectional and time series relationships of price sensitivity and markups, as measured by the log aggregate Lerner index at the category-year level. All regressions include a constant. Columns (1), (2), and (3) capture cross-sectional variation using the years 2006, 2017, and 2019 for the 133 product categories in our baseline sample. Column (4) captures the time series variation by estimating the model in first differences from 2007 through 2019. The regressions are motivated by the decomposition in equation (3.12). Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

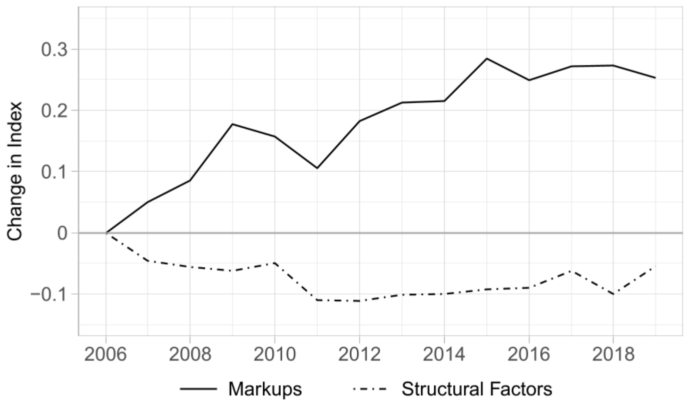
approach: as we estimate demand separately for each category and each year, our estimates of price sensitivity can adjust flexibility over time with the shifts in the empirical variation in the data.

In fact, our analysis implies that a decline in consumer price sensitivity is necessary to generate higher markups in our sample. From 2006 to 2019, the average structural component of equation (3.12) decreased by 0.05. In other words, if price sensitivity had not changed over this period, then the observed changes in other structural features of the market would have implied a five percent decrease in the log Lerner index.[36] Figure 3.8 presents the time series of aggregate log Lerner index as well as the structural components, which are reported relative to 2006 values. The average log Lerner index increased by 0.25 from 2006 to 2019, as shown by the solid black line, while the structural factors decreased from 2006 to 2011 and remained below 2006 levels thereafter. Consistent with our earlier results, this decomposition illustrates that the overall change in markups is tied to changes in consumer price sensitivity.

Why does consumer price sensitivity decline? One possibility is that price-sensitive consumers increasingly select out of mass merchandisers, grocery stores, and drug stores and into other channels that offer lower prices, such as warehouse clubs or dollars stores. However, such an explanation seems to be at odds with aggregate consumer spending patterns. As documented in Table 3.6 in the Appendix, the focal channels in our data comprise the vast majority of broad-basket retail spending in 2007 (83 percent) and 2019 (82 percent).

[36]In our empirical model, the structural component can be obtained from the first step in our estimation routine, where we pin down heterogeneity in demand using micro-moments (Berry and Haile, 2022). Thus, our finding of decrease in the structural component is not sensitive to price endogeneity and does not rely on the moments used to pin down the mean price parameters. See Appendix 3.A for details.

Figure 3.8: Decomposition of Markup Trends



Notes: This figure shows the changes to the aggregate log Lerner Index (black line) and the structural factors (dash-dotted line) specified by equation (3.12). The structural factors incorporate observable changes in prices and the distribution of market shares. The difference between the two lines is captured by changes to price sensitivity.

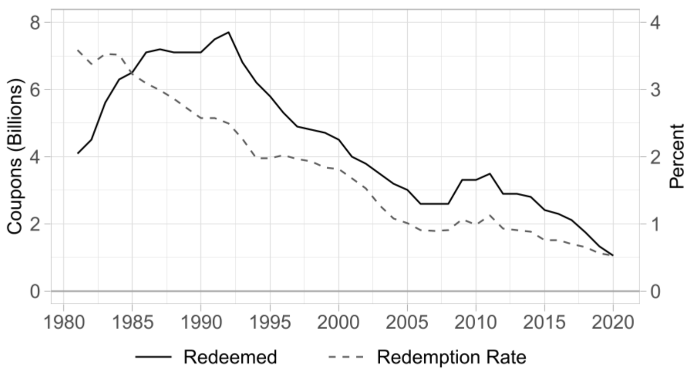
Additional analysis that leverages the consumer panel also suggests that compositional shifts across or within channels do not explain changes in price sensitivity, as we discuss in Appendix 3.D. An alternative possibility is that firms make investment decisions that serve to lower consumer price sensitivity. Such decisions might be reflected in marketing expenditures, R&D expenditures, or the variety of products that they offer for a particular brand. In Appendix 3.D, we also show that changes in these variables do not explain changes in price sensitivity. Therefore, we do not find support for the hypotheses that declining price sensitivity is due to consumer selection across retail channels or firm-level investment decisions.

Changes in price sensitivity may reflect exogenous shifts in preferences that are not the result of changes to supply. To explore this possibility, we examine other information about consumer shopping patterns. In particular, we look at the use of coupons and estimates of time spent shopping for consumer products. Coupon redemptions are a plausible proxy for price sensitivity because they typically involve a small amount of effort in order to obtain a discount on price. To evaluate coupon use, we collect statistics on the number of coupons distributed and redeemed for consumer packaged goods from 1981 through 2020. These statistics reflect industry estimates of coupon use across all channels, including free standing inserts and electronic coupons.[37]

Figure 3.9 plots the aggregate coupon usage over time. The black line reports the number of coupons redeemed each year (left axis). From 1981 to 1992, the number of coupons redeemed roughly doubled, from 4.1 billion to 7.7 billion. Since that year, there has been a steady decline in the number of coupons redeemed, with the exception of a brief bump due to the recession starting in 2009. Over our sample period, the number of coupons redeemed has fallen in half, from 2.6 billion in 2006 to 1.3 billion in 2019.

[37]Industry estimates were obtained from reports by two companies, NCH Marketing from 1981 through 2002, and Inmar Intelligence from 2003 through 2020.

Figure 3.9: Coupon Use Over Time



Notes: This figure shows the annual number of coupons redeemed (left axis) and the redemption rate out of all issued coupons (right axis). From 2006 to 2019, coupon redemptions fell from 2.6 billion to 1.3 billion, and the redemption rate fell from 0.90 percent to 0.56 percent. Annual estimates reflect total coupon usage for consumer products in the United States across all channels, including free standing inserts and electronic coupons.

This trend reflects a decreasing propensity of consumers to use coupons, rather than coupon availability. To highlight this, the dashed line plots the percent of coupons that are redeemed out of all the coupons that were distributed (right axis). Redemption rates are declining over the entire sample period. From 1981 to 1992, the decline reflects the fact that the growth in the distribution of coupons outpaced the growth in coupon redemption rates. From 1992 to 2015, the annual number of coupons issued remained high while redemption rates fell. In 2015, 316 billion coupons were distributed, compared to 309 billion in 1992. From 2016 to 2020, fewer coupons were distributed each year, but redemption fell even faster. The redemption rate fell from 0.90 in 2006 to 0.56 in 2019.

Concurrently, adults in the U.S. spent less time shopping for consumer products. Data from the American Time Use Survey indicate that both the frequency and duration of shopping trips declined over our sample period. For adults between the ages of 25 and 54, time spent on consumer goods purchases fell by 21 percent, from 3.01 to 2.38 hours per week.[38] We also find that, in the consumer panel data, households visit approximately 10 percent fewer unique retailers each week on average in 2019 compared to 2006.

Overall, the declining use of coupons and the reduced time spent purchasing consumer goods suggest a fundamental shift in consumer shopping behavior that is consistent with lower price sensitivity arising from exogenous factors. Both trends indicate that consumers are less willing to exert effort to obtain lower prices. Notably, the decline in coupon use began in the early 1990s, before the rise of online retail. We view this as additional evidence that declining price sensitivity reflects a longer-run secular trend. A potential explanation for this trend is an increase in the opportunity costs of time spent shopping, possibly due to changes in preferences for leisure, or changes to labor supply and the within-household distribution of wages. Consistent with the latter, Griffith et al. (2022) provide evidence that

[38]The American Time Use Survey reports both the frequency of adults participating in an activity in a given day, which declined by 5 percent, and the daily time spent conditional on participation, which declined by 16 percent.

the opportunity cost of time for households in the United Kingdom has increased since the 1980s, and that this change is correlated with an increase in labor force participation and earnings among secondary earners.[39]

3.7 Markups, Welfare, and Consumer Surplus

In this section, we analyze how consumer surplus, producer surplus, and total welfare for consumer products have changed over time. We also examine various counterfactual scenarios in order to estimate the deadweight loss from (changes in) market power and to explore the consequences of rising markups for consumers and firms.

Following Small and Rosen (1981), we calculate consumer surplus as the total expected value that consumers receive from a set of products, given the distribution of the consumer-specific logit error terms (but not their realizations). With the observed set of products, consumer surplus is given by:

$$CS = -\frac{1}{N}\sum_{i}\frac{1}{\alpha_i}\ln\left(\sum_{j}\exp\left(w_{ij}\right)\right) \tag{3.13}$$

where $w_{ij} = \beta_i^* + \alpha_i^* p_{jcrt} + \xi_{jr} + \xi_{cr} + \xi_t + \Delta\xi_{jcrt}$ for the inside products ($j > 0$), $w_{0j} = 0$ for the outside good ($j = 0$), and N denotes the number of consumers.[40] This represents the additional consumer surplus provided by the inside goods, relative to a counterfactual in which only the outside option is available to consumers (as the outside option alone provides zero consumer surplus by assumption). Thus, it can be interpreted as the added value of the focal products under consideration, or, identically, the equivalent variation that would compensate consumers for the loss of these product-market combinations.[41]

Our measure of producer surplus reflects variable profits and is measured as price less marginal costs multiplied with quantities: $PS = \sum_{j>0}(p_j - c_j)q_j$. Our estimation results do not identify fixed costs and, as they are not incorporated into our measure of producer surplus, our results do not inform whether brand manufacturers earn economic profit.[42] We measure welfare (W) as the sum of producer and consumer surplus. The deadweight loss that exists in an observed equilibrium can be calculated by comparing the welfare that obtains with the equilibrium to the welfare that obtains under a counterfactual with prices

[39] An alternative potential explanation, following results in the marketing literature, is that consumers are responding to broad shifts in the pricing behavior of firms. For example, Mela et al. (1997) argues that price-oriented promotions increase consumer price sensitivity in the long run. Therefore, a decline in price sensitivity could potentially be a response to a large-scale decline in price-oriented promotional activity.

[40] In calculating consumer surplus, we use the average price coefficient within each consumer's income decile to avoid dividing by numbers very close to zero. In practice, this matters only for a single category, and we obtain nearly identical results if we use the average price coefficient within income quartiles or across all consumers.

[41] We do not evaluate trends in overall welfare, which would necessitate taking a stance on utility for the outside good. We focus on the relationship between markups and welfare within the products and markets of our sample.

[42] The findings of De Loecker et al. (2020), which look at firm-level accounting statements, indicate that profits have increased along with markups.

Table 3.5: Annual Surplus and Welfare Per Capita

(a) 2006 Preferences and Costs

Specification	CS	PS	W	% change CS	% change W
Baseline	628	261	889	0.0	0.0
Prices Scaled to 2019 Levels	603	263	867	-3.8	-2.4
Markups Scaled to 2019 Levels	551	267	818	-12.2	-8.0
Prices Equal to Marginal Costs	956	0	956	52.4	7.6

(b) 2019 Preferences and Costs

Specification	CS	PS	W	% change CS	% change W
Baseline	974	371	1345	0.0	0.0
Prices Scaled to 2006 Levels	1006	350	1356	3.3	0.8
Markups Scaled to 2006 Levels	1106	280	1386	13.5	3.1
Prices Equal to Marginal Costs	1460	0	1460	49.9	8.6

Notes: This table reports consumer surplus (CS), producer surplus (PS), and welfare (W) per capita based on estimated demand parameters ("Baseline") and for counterfactual scenarios that hold fixed preferences and marginal costs and vary the price levels.

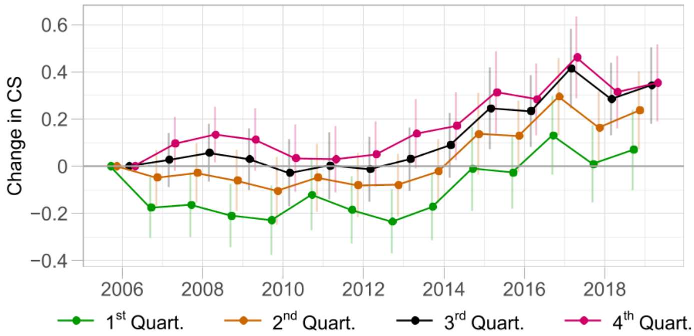
set equal to marginal costs.

Table 3.5 shows per capita consumer, producer surplus, and welfare for 2006 and 2019 using observed prices ("Baseline") and prices under different counterfactual scenarios. To compute counterfactual values, we hold fixed estimated preference parameters and marginal costs, and we simulate consumer choices using different prices. We consider three counterfactual scenarios. First, we scale all prices by the average realized price change for all products in the same category from one year to another (e.g., from 2006 to 2019). Second, we scale all markups by the average realized markup change for all products in that category from one year to another. Because we hold marginal costs fixed, scaling 2006 prices to match 2019 markups results in higher prices than what we observe in the data. Third, we consider a counterfactual where prices equal marginal costs (i.e., no markups). The last two columns in each panel show changes in consumer surplus and welfare relative to the baseline scenario.

Comparing the baseline scenarios, the results indicate that per capita consumer surplus increased by about 50 percent (i.e., about 3 percent annually) between 2006 and 2019, from \$628 to \$974. As average prices did not decline and perceived quality did not increase, the increase in consumer surplus is likely due to lower price sensitivity, i.e., that consumers receive lower disutility from any given price in 2019. Along with higher markups, producer surplus increased over the period, from \$261 to \$371 per capita. Thus, approximately three quarters of the increase in welfare have accrued to consumers.

Markups are costly for consumers. With marginal cost pricing, consumer surplus would be substantially higher in both 2006 and 2019, as shown by the final specification in each panel. Our estimates suggest that markups in 2006 reduced per capita welfare from \$956

Figure 3.10: Consumer Surplus Over Time By Income Group



Notes: This figure reports coefficients and 95 percent confidence intervals of a regression of the log of consumer surplus by purchase on year dummies, controlling for category fixed effects, separately for different quartiles of the income distribution.

to \$889 (about 7 percent). In 2019, markups reduced welfare by about 8 percent.[43]

The changes in markups over this period are economically meaningful. Holding fixed the 2006 preferences, marginal costs, and product assortments, increasing markups to 2019 levels would reduce consumer surplus by 12 percent. However, markups trends do not occur in isolation. Changes in markups are often concurrent with and in response to other factors. For example, declining marginal costs mitigate the impact of rising markups on prices and consumer welfare. When scaling up prices—which are the relevant demand variables—to match 2019 levels, the decrease in consumer surplus is much smaller (3.8 percent). Analogous results obtain if 2019 markups and prices are scaled down to 2006 levels.

Thus, to interpret the impacts of changing markups on welfare, it is necessary to take a stand on what other factors are changing at the same time. Markups are equilibrium objects that are determined by supply and demand. If marginal costs and price sensitivity had not changed, the aggregate trends in markups would have likely looked quite different. This is an important consideration for potential policy responses to markup trends.

In our final analysis, we analyze how the change in consumer surplus varies by income. For this purpose, we calculate the log of consumer surplus per purchasing decision separately by each quartile of the income distribution and for each category-year. We relate these values to category and year fixed effects and document the coefficients across years in Figure 3.10. The results indicate that the increase in per capita consumer surplus between 2006 and 2019 is mainly driven by consumers with relatively high income and takes place during the second half of the sample period. In contrast, the lowest quartile of the income distribution has lower consumer surplus through 2016. The reduction in consumer surplus for the lowest-income households coincides with the increase in real prices in the first half of our sample. After this point, real prices fall and consumer surplus for this quartile increases, recovering to 2006 levels at the end of the sample period. In Figure 3.22 in the Appendix, we repeat the analysis dividing the sample into deciles. The results confirm that changes in consumer surplus are strongly associated with the income distribution. Consumers in the highest income

[43] These estimates of deadweight loss are similar in magnitude to those reported in recent study of publicly-traded firms in the United States (Pellegrino, 2021).

group see increases in consumer surplus over time, while lower income households have, on average, lower consumer surplus over our sample period. These findings suggest that changes in market power and consumer preferences over time have important distributional consequences.

3.8 Conclusion

This paper analyzes the evolution of market power in consumer products in the United States between 2006 and 2019. For this purpose, we combine retail scanner data on quantities and prices with consumer level data across more than 100 product categories. This approach allows us to estimate demand with flexible consumer preferences and recover time-varying markups for individual products under the assumption of profit maximization. Our results indicate that markups increase by about 30 percent during our sample period. In contrast to previous research on the evolution of market power, we estimate similar changes across different quartiles of the markup distribution. In addition, we find similar increases in markups within product categories over time which implies that the results are not driven by a reallocation of market shares towards products with higher markups. We decompose changes in markups into changes in prices and changes in marginal costs. Overall, the nominal prices of products rise at a similar rate as inflation during our sample period. Thus, real prices remain almost constant, and the increase in markups we estimate is primarily due to falling (real) marginal costs. Our results suggest that prices do not decrease along with marginal costs because of changes in consumer preferences. Our estimates suggest that consumers became about 30 percent less price sensitive over the sample period.

The results of a counterfactual simulation exercise indicate that changes in price sensitivity and marginal costs account for nearly all of the time series variation in aggregate markup changes between 2006 and 2019. We also find that these two factors explain most of the cross-category variation in markup trends, while changes in ownership, demographics and perceived quality only play a minor role. Due to decreased price sensitivity, consumer surplus increased during our sample period despite rising markups. The increase in consumer surplus is, however, concentrated among consumers with relatively high income. Nonetheless, changes in markups have been costly for consumers. In a counterfactual simulation, we find that consumer surplus would have been 14 percent higher in 2019 if markups had not changed relative to 2006. If firms would set price equal to marginal costs, consumer surplus in 2019 increases by 50 percent and total welfare increases by 9 percent.

Appendices

3.A Estimation Details

This appendix provides details on the estimation procedure. We estimate the parameters in two steps, which is possible because the mean price parameter and the other ("nonlinear") structural parameters are identified by two independent sets of moments. The parameters for estimation are $\theta = (\alpha, \Pi_1, \Pi_2, \sigma)$. We first estimate $\theta_2 = (\Pi_1, \Pi_2, \sigma)$ and then estimate α, the mean price parameter, in the second step. Our micro-moments identify θ_2 but not α (Berry et al., 2004; Berry and Haile, 2022), and the covariance restriction exactly identifies α given θ_2 (MacKay and Miller, 2023). In principle, a single search could be used to estimate the parameters jointly, as is standard practice for applications that rely on instruments for identification. However, our approach has computational benefits, as we explain below.

3.A.1 First Step

In the first estimation step, we use the micro-moments to pin down the "nonlinear" parameters, i.e., $\theta_2 = (\Pi_1, \Pi_2, \sigma)$. To implement this, we estimate GMM while holding fixed the price parameter at a given value. Because the parameters are identified separately, the specific value chosen for the price parameter has no impact on the micro-moment contributions to the objective function.[44]

For any candidate θ_2, there is a unique vector of the mean product valuations that align the predicted and observed shares (δ). For example, in the special case of $\theta_2 = \overrightarrow{0}$ the mean valuations have a closed-form solution:

$$\delta_{jcrt}\left(\theta_2^{(0)}\right) \equiv \log(s_{jcrt}) - \log(s_{0crt}) \tag{3.14}$$

We proceed to estimate θ_2 based on equation (3.7) while holding fixed the price parameter. For each candidate θ_2, we recover the mean valuations $\{\delta_{jcrt}(\theta_2)\}$ using the contraction mapping of Berry et al. (1995) with a numerical tolerance of 1e-9. We then calculate the micro-moments with $\{\delta_{jcrt}(\theta_2)\}$ and $\bar{\alpha}$. We choose the parameters $\{\delta_{jcrt}(\theta_2)\}$ that minimize the micro-moment contributions to the objective function. We apply equal weights to each micro-moment in estimation.

3.A.2 Second Step

In the second step, we hold fixed the estimated nonlinear parameters and choose the price parameter that minimizes the objective based on the covariance restriction moment. In other words, we estimate α taking as given the estimates of θ_2 obtained in the first step. This is possible because micro-moments do not identify the mean price parameter (Berry

[44] We initialize this step with a price parameter $\bar{\alpha}$ such that the average elasticity when $\theta_2 = \overrightarrow{0}$ is equal to -7, which corresponds to the average starting value that we use in the second step (see below).

and Haile, 2022). To do so, we recover $\Delta\xi_{jcrt}(\theta_2)$ as the residual from the OLS regression of $(\delta_{jcrt}(\theta_2) - \alpha p_{jcrt})$ on the fixed effects for each candidate α. We also obtain marginal costs from equation (3.5), looping over the chain-region-quarter combinations, and then recover $\Delta\eta_{jcrt}(\theta_2)$ as the residual from the OLS regression of marginal costs on the fixed effects. We are then able to calculate the loss function, update the candidate α, and repeat to convergence. We constrain the search to negative values of α. The constraint imposes downward-sloping demand for a consumer with the mean income level.

A complication is that there may be two values for α that satisfy the covariance restriction, with the smaller (more negative) value being the true price parameter under sensible conditions (MacKay and Miller, 2023). Care must then be taken to ensure that the estimator converges to the smaller value. Figure 3.23 illustrates this in the context of ready-to-eat cereals. Each panel traces out the contribution of the covariance restriction to the objective function for different values of α. In 2006, a unique negative α satisfies the covariance restriction, and the constraint we place on the parameter space ($\alpha < 0$) is sufficient to recover the correct estimate. In other years, both possible solutions are negative, and thus could be obtained from estimation, even though the larger (less negative) value is implausibly close to zero.[45]

We proceed by selecting starting values of $\alpha^{(0)} = \phi\tilde{\alpha}$ where $\tilde{\alpha}$ is such that the average elasticity is -1 when $\theta_2 = \overrightarrow{0}$, and $\phi = (2, 4, 6, 8, 10, 12)$. Thus, for each year-category, we estimate with six different starting values. As these starting values are quite negative, the estimator tends to converge on the more negative value of the price parameter that satisfies the covariance restrictions. In the category-years for which the estimator finds both solutions, we select the more negative solution as our estimate of α. This appears to be a robust solution given the θ_2 we estimate.

The two-step approach allows us to more readily evaluate the possibility of multiple solutions for the covariance restriction. In addition, the objective function contribution of the covariance restriction moment can be poorly behaved for unreasonable candidate θ_2 parameters that would be considered if estimation of both θ_2 and α were performed simultaneously. Thus, our two-step approach to estimation yields both speed and numerical stability, both of which are important given the scale of the empirical exercise.

3.A.3 Computation Notes

Our code builds on the BLPestimatoR package for R (Brunner et al., 2020).[46] The package has a slim R skeleton and fast C++ routines for computationally intensive tasks. As micro-moments and covariance restrictions are missing from the package, we added code to cover that part of estimation. All time-critical parts are in C++. In early experiments, we replicated our results for some categories using the PyBLP package for Python (Conlon and

[45] The larger values imply that firms are pricing in the inelastic portion of their residual demand curves. A related complication is that the numerical stability of the moment tends to deteriorate as the candidate α approaches the higher solution, which can lead to convergence issues if the estimator considers parameters near the higher solution.

[46] `https://github.com/cran/BLPestimatoR`, last accessed March 26, 2021

Gortmaker, 2020).[47] We ultimately selected the augmented R package because it allowed us to calculate the micro-moments more quickly; our understanding is that the speed of PyBLP has improved substantially during the course of our research.

In estimation, we use BFGS with a numerical gradient. When searching for θ_2 in the first step of estimation, there are a handful of categories for which BFGS fails to converge, and for those categories we use Nelder-Mead instead. We estimate each category-year combination in parallel using the HILBERT computational cluster at the University of Düsseldorf. There are 2800 estimation routines (200 categories and 14 years). Each routine requires one CPU core and up to 9GB of memory. The longest runs take slightly more than 72 hours and most finish in less than 24 hours. The entire estimation procedure takes around one week.

[47] `https://github.com/jeffgortmaker/pyblp`, last accessed March 26, 2021.

3.B Data Details

3.B.1 Market Size Calculations

Recall from Section 3.2.2 that the quantity demanded in our model is given by $q_{jcrt}(p_{crt};\theta) = s_{jcrt}(p_{crt};\theta)M_{crt}$, where $s(\cdot)$ is the market share, p_{crt} is a vector of prices, and M_{crt} is the market size, a measure of potential demand. As is standard in applications involving random coefficients logit demand, an assumption on market size is needed in order to convert observed quantities into market shares and then estimate the model. Our approach is to use market sizes that scale with the population of the region and the number of stores operated by the retail chain within the region. We apply the following steps separately within each product category:

1. Obtain a time-varying "base" value by multiplying the population (at the region-year level) with the number of stores (at the chain-region-quarter-year level). This obtains $BASE_{crqy} \equiv POP_{ry} \times NS_{crqy}$ where POP_{ry} is the population in region r and year y and NS_{crqy} is the number of stores operated by retail chain c in region r, quarter q, and year y.

2. Obtain the total quantity of the inside products across brands: $Q_{crqy} = \sum_j q_{jcrqy}$.

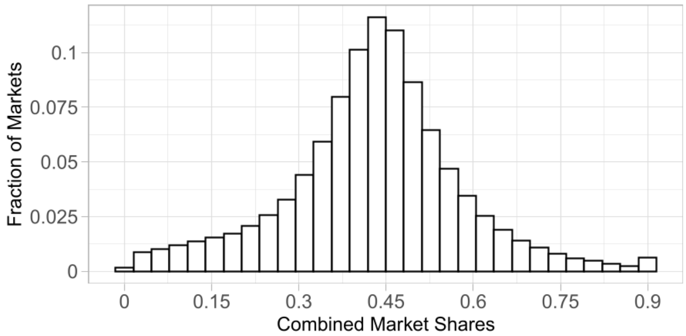
3. Calculate $\gamma_{cr} = E_{q,y}\left[\frac{Q_{crqy}}{BASE_{crqy}}\right]$ as the average quantity-to-base ratio among the periods observed for each retail chain and region. This can be used to convert the base value into units that are meaningful in terms of total quantity-sold. In the calculation of γ_{cr}, we exclude a handful of observations for which the base-adjusted quantity is less than 5 percent of the mean, which helps avoid extraordinary small inside good market shares.

4. We set the market size such that the combined share of the inside goods is around 0.45, on average, and we allow the market size to scale with population and number of stores, as captured by the base value. Specifically, we calculate the market size according to
$$M_{crqy} = \frac{1}{0.45}\gamma_{cr} BASE_{crqy}$$
which generates markets sizes for each retail chain, region, quarter, and year. This yields combined inside shares $\frac{Q_{crqy}}{M_{crqy}} = 0.45\frac{Q_{crqy}}{BASE_{crqy}}\frac{1}{\gamma_{cr}}$.

5. For a small minority of cases (<5 percent of markets), this procedure generates a combined share of the inside goods that exceeds 0.90 in some periods, which is high enough that we encounter numerical problems in estimation. For any category×chain×region combination in which this occurs, we repeat the steps above using the alternative conversion factor $\tilde{\gamma}_{cr} = 0.5 \times \max_{q,y}\left(\frac{Q_{crqy}}{BASE_{crqy}}\right)$, which sets the maximum of the combined shares equal to 0.90.

Figure 3.11 shows the distribution of combined market shares of inside goods. By construction the market shares are centered around 0.45 (step 4), and the small peak around

Figure 3.11: Distribution of Market Shares of Inside Goods



Notes: This figure shows the distribution of market shares of inside goods. Observations are at the chain-region-year-quarter level and reflect the sum of the market shares of all inside goods in a market at a given point in time.

0.9 indicates the imposed maximum that is described in step 5.

We provide robustness checks in Appendix 3.E.5.

3.B.2 Other Notes on Estimation Data

We make a number of adjustments to the NielsenIQ data as we construct the estimation samples. First, we drop two large chains from the Consumer Panel Data that do not appear in the Retail Scanner Data. Second, we impute household income using the midpoint of the bins provided in the Consumer Panel Data data. It is possible to obtain a comparable income measure for the highest-income bin because additional high-income bins are provided from 2006 to 2009; we estimate a midpoint of \$137,500. Third, we observe that many fewer consumers are in the top income bin in 2006 than in 2007 and subsequent years. To produce a more consistent demographic representation of consumers, we rescale the NielsenIQ projection weights in 2006 so that the top bin occurs with the same frequency as it does in 2007. We scale down the projection weights for the other bins in 2006 proportionately. Fourth, to reduce measurement error, we drop products that are extreme outliers in terms of their price—which we implement by dropping observations with a price below the 0.5 percentile or above the 99.5 percentile. We apply this screen before restricting attention to the 22 DMAs. Fifth, we exclude four categories from the ranking that, for some years, exist in the scanner data but not the consumer panel data: prerecorded videos, magazines, cookware, and sunscreens. Finally, product categories belong to the following high-level departments according to NielsenIQ: "Dry Grocery," "Frozen Foods," "Dairy," "Deli," "Packaged Meat," "Fresh Produce," and "Alcoholic Beverages," "Health and Beauty Care," "Non-food Grocery," and "General Merchandise."

3.B.3 Auxiliary Data on Revenues by Retail Channel

As described in the main text, we focus our analysis on retailers that NielsenIQ classifies as mass merchandisers, grocery stores, or drug stores. To provide context about aggregate spending on consumer products and the relative size of these channels, we use auxiliary data

Table 3.6: Share of Revenue by Retail Channel

	2007	2019
Focal Channels		
Mass Merchandisers	0.214	0.218
Grocery Stores	0.219	0.217
Drug Stores	0.088	0.117
Other Broad-Basket Retail Channels		
Warehouse Club	0.090	0.094
Dollar Stores	0.015	0.026
Other Consumer Product Retail Channels		
Convenience Stores, Department Stores, Apparel, etc.	0.374	0.328
Combined Share of Focal Channels		
Among All Consumer Products	0.522	0.552
Among Broad-Basket Retailers	0.833	0.822

Notes: This table displays the share of revenues of broad-basket retailers out of all consumer product spending. We compare broad-basket retailers to "specialized" retailers such as convenience stores, department stores, apparel stores, beauty stores, electronic stores, and online retailers. To construct these estimates, we take the revenues of the largest 100 U.S. retailers. We exclude from this list retailers that do not have consumer products as their primary source of revenue: restaurants, home improvement stores, and auto parts stores. The included retailers represent $1.4 trillion in revenues in 2007 and $2.0 trillion in 2019.

on retailer revenues for large U.S. retailers.

Specifically, we obtain retailer-level revenue data for the largest 100 U.S. retailers. The data are compiled annually by the National Retail Federation, which is the largest retail trade association. The earliest estimates we can find are from 2007, one year after the start of our sample. For 2007 and 2019, we categorize each retailer into one of the following types: mass merchandisers, grocery stores, drug stores, warehouse clubs, dollar stores, and other consumer product stores. Other consumer product stores include convenience stores, department stores, online retailers, and retailers that specialize in a more narrow set of categories (e.g., electronics, beauty, or apparel).[48] We also identify retailers that are restaurants, home improvement stores, and auto parts stores, and we drop these from the analysis because they do not primarily sell consumer products. Because the included retailers also sell products outside of the scope of our analysis (e.g., prescription drugs), the aggregate data may not provide an exact picture of how the retail shares of consumer products evolve over

[48]For Walmart, we adjust the provided estimates to separate Walmart U.S. (mass merchandiser) and Sam's Club (warehouse club) into distinct channels. For Amazon, we adjust the provided estimates in 2019 to include revenues from online sales and third-party seller services in the United States (other), and we separate out Whole Foods (grocery). We use data from Statista for Walmart (https://www.statista.com/statistics/269403/net-sales-of-walmart-worldwide-by-division/), and we obtain 2019 Amazon estimates from Amazon's 2021 10-K filing.

time. Nonetheless, we think the auxiliary data provide useful information. The included retailers represent \$1.4 trillion in revenues in 2007 and \$2.0 trillion in 2019.

Table 3.6 reports the share of consumer product spending in our focal channels (mass merchandisers, grocery stores, and drug stores) and other broad-basket retailers (warehouse clubs and dollar stores) in 2007 and 2019. Our focal channels are the three largest consumer product channels in 2019, and their shares have been fairly stable over our sample period. Combined, the channels represent 83 of spending within broad-basket retailers in 2007 and 82 percent in 2019. Out of all consumer product spending, the focal channels represent 52 percent of spending in 2007 and 55 percent in 2019.

Thus, the focal channels capture the majority of consumer product spending, and their revenue growth has paralleled the average revenue growth among other large U.S. retailers. The largest broad-basket channel that we omit is warehouse club, which accounts for 9.0 percent of revenues in 2007 and 9.4 percent in 2019. The revenue share of dollar stores roughly doubles between 2007 and 2019, consistent with the trend documented in Caoui et al. (2022). Nonetheless, dollars stores account for only 1.5 percent of consumer product spending in 2007 and 2.6 percent in 2019.

The share of revenues allocated to other consumer product channels declined slightly over our sample, from 37 percent in 2007 to 33 percent in 2019. Within this category, online retailers grew substantially, reaching roughly 6 percent of revenues in 2019. However, this increase was offset by relative declines in other store formats, such as department stores and apparel.

3.C Derivation of the Econometric Decomposition

In this appendix, we obtain the structural decomposition used in Section 3.6, following MacKay and Miller (2023). The decomposition is available for a wide class of models, but we focus on random coefficients logit demand with differentiated-products Betrand competition.

First, it is helpful to re-express the indirect utility that consumer i receives from product $j > 0$ (in chain c, region r, and quarter t) as follows:

$$u_{ijcrt} = \delta_{jcrt}(p_{jcrt}; \beta, \alpha) + \mu_{ijcrt}(p_{jcrt}, D_i, v_i; \Pi_1, \Pi_2, \sigma) + \epsilon_{ijcrt} \tag{3.15}$$

where the mean utility of each product, $\delta_{jcrt}(\cdot)$, and contribution of demographics to consumer-specific deviations, $\mu_{ijcrt}(\cdot)$, respectively are given by

$$\begin{aligned} \delta_{jcrt}(p_{jcrt}; \beta, \alpha) &= \beta + \alpha p_{jcrt} + \xi_{jr} + \xi_{cr} + \xi_t + \Delta\xi_{jcrt} \\ \mu_{ijcrt}(p_{jcrt}, D_i, v_i; \Pi_1, \Pi_2, \sigma) &= p_{jcrt}\Pi_1 D_i + \Pi_2 D_i + \sigma v_i \end{aligned}$$

The indirect utility of the outside good remains $u_{ijcrt} = \epsilon_{i0crt}$. The probability with which consumer i selects product j can be expressed

$$s_{ijcrt} \quad (\delta_{crt}, p_{jcrt}, D_i, v_i; \Pi_1, \Pi_2, \sigma) = \tag{3.16}$$

$$\frac{\exp(\delta_{jcrt}(p_{jcrt};\beta,\alpha)+\mu_{ijcrt}(p_{jcrt},D_i,v_i;\Pi_1,\Pi_2,\sigma))}{1+\sum_{k=1}^{J_{crt}} \exp(\delta_{kcrt}(p_{kcrt};\beta,\alpha)+\mu_{ikcrt}(p_{kcrt},D_i,v_i;\Pi_1,\Pi_2,\sigma))} \tag{3.17}$$

where $\delta_{crt} = (\delta_{1crt}, \delta_{2crt}, \dots)$ is the vector of mean utilities. Finally, the market share of product j is obtained by integrating over the joint distribution of consumer demographics:

$$s_{jcrt}(\delta_{crt}, p_{jcrt}; \Pi_1, \Pi_2, \sigma) = \frac{1}{I} \sum_i s_{ijcrt}(\delta_{crt}, p_{jcrt}, D_i, v_i; \Pi_1, \Pi_2, \sigma)$$

For a broad class of oligopoly models, the first order conditions for profit maximization can be expressed in terms of product-level additive markups as follows:

$$p_{jcrt} - c_{jcrt}(\chi_{crt}; \theta) = -\frac{1}{\alpha}\lambda_{jcrt}(q_{crt}, p_{crt}, \Gamma_{crt}; \theta^*), \tag{3.18}$$

where q_{crt} and p_{crt} are vectors of quantities and prices (typically data), Γ_{crt} denotes the matrix of demand derivatives, and θ^* includes all the demand parameters *except* for the mean price parameter (α). Let the set of products sold by the same firm as product j be given by $\mathcal{J}_{f(j)}$. Then, with random coefficients logit demand and Bertrand competition, we have:

$$\lambda_{jcrt} = \frac{s_{jcrt}}{\frac{1}{I}\sum_i s_{ijcrt}(1 - s_{ijcrt})} - \sum_{k \in \mathcal{J}_{f(j)} \setminus j} \frac{s_{kcrt}}{\frac{1}{I}\sum_i s_{ijcrt}s_{ikcrt}} \tag{3.19}$$

where the denominators integrate over the (product of) consumer-specific choice probabilities. From an econometric standpoint, λ_{jcrt} is free from the mean price parameter (α)

because it depends only on market shares and consumer-specific choice probabilities. The market shares are data. From equation (3.17), the consumer-specific choice probabilities depend on $\mu_{crt}(\cdot)$, which obtains immediately from data and $\theta^* = (\Pi_1, \Pi_2, \sigma)$, and on $\delta_{crt}(\cdot)$, which obtains from the contraction mapping of Berry et al. (1995), again given data and θ^*. Related is the observation of Berry and Haile (2022) that micro-moments summarizing how demographics affects consumer choice patterns cannot identify the mean price parameter.

3.D Exploring Alternative Mechanisms

Given the important role of price sensitivity in markups, we next examine potential factors that could explain the change over time. In the main text, we provide evidence that consumers are becoming less price sensitive over time due to exogenous factors (Section 3.6). In this appendix, we consider whether this change could reflect growth in retailers/channels outside of our data or whether this change may be due to firm-level investments that affect consumer behavior, such as increased marketing or product variety.

To assess changes in the composition of retail markets, we construct the share of revenues by retail channel in each product category and each year, including warehouse clubs, dollar stores, and online retail, in addition to mass merchandisers, grocery, and drug stores. We use all available data from the Kilts NielsenIQ consumer panel dataset to construct these measures. Using these data, we obtain similar channel shares to the auxiliary data presented in Appendix 3.B.3. The channels outside of our focal channels realize relatively small growth in shares over this period. The average cross-category share in 2019 was 12.0 percent for warehouse clubs, 2.2 percent for dollar stores, and 1.9 percent for online retailers. In 2006, these values were 11.1 percent, 1.4 percent, and 0.5 percent, respectively. The focal channels capture 86.0 percent share on average in 2006 and 83.9 percent in 2019. Thus, the aggregate compositional shifts in these channels are fairly small for the product categories we study.

Further, we do not find evidence that shifts in consumer spending to retailers outside of our price/quantity data is driving our results. The portion of focal category expenditures in the consumer panel data (which are not limited to a subset of retailers) that are captured by the retail scanner data is flat from 2006 to 2013, while price sensitivity is falling. In part due to changes in the composition of participating retailers, this portion is lower from 2014 to 2017 and higher in 2018 and 2019. The patterns are similar across income groups. To address the potential for the sample composition to impact our findings, we perform a robustness check with a balanced panel of retailers in Appendix 3.E.3. We perform another set of robustness checks in which we supplement our baseline sample from the retail scanner data with large retailers that are in the consumer panel but not in the retail scanner data, which we discuss in Appendix 3.E.4. In both cases, we find very similar trends in markups and price sensitivity.

Finally, our estimated demand parameters provide some evidence that selection over time into different types of retailers may not be driving the trend in price sensitivity we observe. Specifically, we find no trend over time in the coefficients that load onto the interaction of price and household income (Figure 3.20). This indicates that, based on income, there is no disproportionate selection of greater price sensitive consumers to retailers outside of our sample.[49]

Taken together, we think it is unlikely that compositional shifts would account for the 30 percent decline in price sensitivity we estimate over this period. Nonetheless, we explore this further with a regression analysis that exploits panel variation. Some categories are

[49]The random coefficients model endogenizes the consumer's decision to buy from the retailers in our sample, so we are also able to control for some types of selection directly with the model.

Table 3.7: Potential Mechanisms

	(1) Price Sensitivity	(2) Log Abs. Elasticity	(3) Marginal Cost	(4) Perceived Quality
Log Share Warehouse Clubs	−0.014	0.023	0.142	−0.090
	(0.064)	(0.063)	(0.193)	(0.158)
Log Share Dollar Stores	0.064**	0.058**	0.049	0.073
	(0.029)	(0.028)	(0.079)	(0.084)
Log Share Online	−0.090*	−0.074*	−0.121	−0.449***
	(0.047)	(0.041)	(0.136)	(0.144)
Log Marketing Spend	0.012	0.017	0.125**	0.049
	(0.021)	(0.020)	(0.054)	(0.056)
Log R&D	−0.006	−0.006	−0.059	0.016
	(0.023)	(0.020)	(0.057)	(0.072)
Log Num. UPCs	0.100*	0.089*	0.384***	0.455***
	(0.051)	(0.046)	(0.127)	(0.155)
Brand-Category FEs	X	X	X	X
Time Period FEs	X	X	X	X
Observations	1,799	1,799	1,799	1,799
R^2	0.943	0.603	0.122	0.173
R^2 (Within)	0.015	0.013	0.015	0.028

Notes: Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

disproportionately affected by the growth of alternative retail channels. For example, less than one percent of beer was sold online in each year of the sample, whereas the share of online revenues for dry dog food increased from less than 2 percent to over 15 percent during the sample period. If we see a greater decrease in price sensitivity for categories disproportionately affected by the shift to online, that might suggest that consumer selection may be playing some role.

We also investigate whether firm-level investments may yield consumers that are less price sensitive, either through perceived or realized changes to their products. To explore this, we merge our estimates with financial data on marketing and R&D expenses obtained from Compustat. These measures are obtained from annual reports of the parent companies. We also consider whether changes in product variety may account for the changes we observe. We measure product variety as the (log) number of UPCs offered by each brand in each market. We aggregate our data to the category-year level, taking a simple average of each measure. Thus, we seek to evaluate whether categories with disproportional increases in marketing, R&D, or variety also realized greater declines in price sensitivity.

To explore these relationships, we regress price sensitivity ($\ln(-\alpha_t)$) on the logged values of the above measures. We include category fixed effects and year dummies, so that the coefficients reflect time-series variation within each category that departs from the aggregate trend.

Column (1) of Table 3.7 reports the results. We find no significant relationships between share sold in warehouse clubs, marketing expenditures, or R&D expenditures. We find a negative, marginally significant relationship between the share sold online and consumer price sensitivity, and a positive, statistically significant relationship between share sold in dollar stores and price sensitivity. Given the coefficient magnitudes and the absolute size of these channels (shares of less than 2.5 percent in 2019), we think these results most likely reflect other mechanisms, e.g., online retailers entering categories with higher markups and less price sensitive consumers. In support of other mechanisms, a regression with price elasticity as the dependent variable, reported in column (2), returns a coefficient on online sales that is roughly 20 percent smaller. If online sales were skimming off more price sensitive consumers, we would expect elasticities to have a stronger relationship with online sales than the (mean) price sensitivity parameter, as the elasticity also incorporates self-selection based on demographic characteristics (e.g., lower-income consumers). We do not find evidence for this selection.

We find a marginally significant positive relationship between variety and price sensitivity, which indicates that greater variety is weakly correlated with *greater* price sensitivity.[50] Since price sensitivity has decreased over time while variety has increased, we think it is likely that this coefficient reflects other factors. Together, all five measures only explain 1.5 percent of the residual variation in price sensitivity, suggesting that neither retail shopping patterns nor firm-level investments are driving the changes in price sensitivity over time.

Though we focus on explaining price sensitivity, we also run regressions with marginal costs and perceived quality as the dependent variables. We report results in columns (3) and (4). We find a positive and significant relationship with marginal costs and marketing, suggesting that cost decreases were also correlated with less spending on marketing. We also find a large and highly significant relationship between perceived quality and online sales. As perceived quality captures the value to consumers above and beyond outside options (including online sales), this is consistent with the trends we find in Section 3.5. Online retail became an increasingly popular option over the time period, lowering the (relative) utility of in-store purchases. Conversely, we find no effect of warehouse clubs on perceived quality, though the point estimate is negative.

We find that product variety is positively correlated with marginal costs and perceived quality. As both marginal costs and quality are falling over time, while variety is rising, this suggests that greater variety may have helped to mitigate the substitution of consumers to other channels (i.e., online), albeit at higher costs.[51]

[50] Brand (2021) finds the opposite relationship.

[51] This is related to the explanation offered by Brand (2021), who suggests that increased variety may lead to less price sensitivity. However, we do not find that increases in variety are related to lower price sensitivity, and we do not find that changes in quality, which are correlated with variety, drive changes in markups. In the time series, quality declines over time, and we estimate a net relationship with markups very close to zero when controlling for other factors (Table 3.3). Thus, product variety does not appear to

Overall, this analysis suggests that firm-level investments and changes in the composition of retail shopping across channels cannot account for the change in consumer price sensitivity that we document.

be driving the trends we observe.

3.E Alternative Specifications and Robustness Checks

In this section, we present a series of alternative specifications and robustness checks to evaluate the sensitivity of our main findings to particular assumptions. First, we show how the main trend in markups is not sensitive to particular choices of measurement, in terms of which categories are included in our baseline sample and our choice of the Lerner index as our markup measure. We then show that product-level trend in markups looks nearly identical with a balanced panel, confirming that the trend is not due to compositional shifts in products over time. Likewise, we find very similar trends when we extend the sample to large retailers that are present only in the consumer panel data. We also find similar trends in markups and price sensitivity with different approaches to measuring market size.

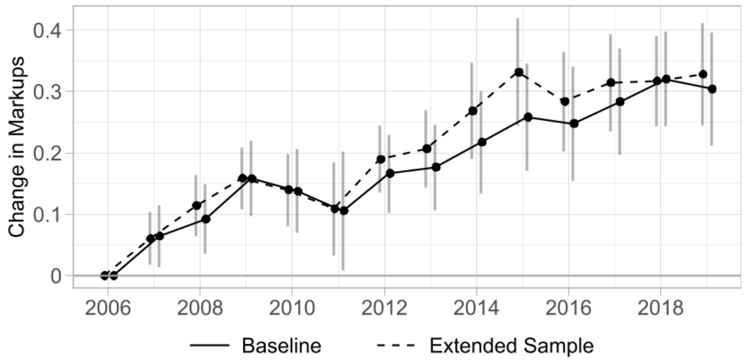
We also examine whether the estimated trends in demand, in terms of more inelastic demand and reduced price sensitivity, are robust to the supply model and the covariance restrictions that we invoke to identify the mean price parameter.[52] We show that a similar trend is obtained when we estimate demand using the assumption that prices are exogenous, which does not invoke the supply model to pin down the demand parameters. Though elasticity estimates under this approach are often unreasonable in terms of levels (see Section 3.4.2), a trend in these parameters would be consistent with a rotation of the demand curve. We find a similar decline in the mean price parameter under this alternative assumption, indicating that our findings of falling price sensitivity are robust to the particular supply-side assumptions we invoke in estimation.

Finally, we examine whether our random coefficient logit demand specification materially affects the estimates relative to a logit specification that does not provide as much flexibility in terms of consumer heterogeneity. Relative to the logit model, the random coefficients specification obtains meaningfully more elastic demand estimates and smaller markups.

[52]As described in the text, the other demand-side parameters are identified by micro-moments.

3.E.1 Category Selection

Figure 3.12: Markups Over Time: Alternative Samples



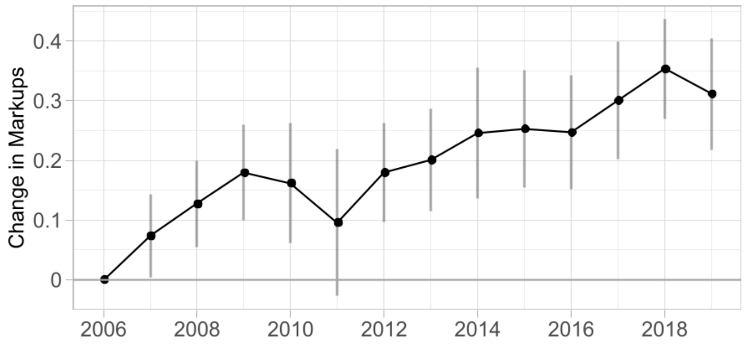
Notes: This figure displays the changes in product-level markups over time for our baseline sample (133 product categories, solid line) and the extended sample (200 product categories, dashed line). The 133 product categories in the baseline sample are selected based on a proxy for within-category product heterogeneity. Point estimates and 95 percent confidence intervals are obtained from regressions of the log of the Lerner index $(p-c)/p$ on year dummies controlling for product-chain-DMA and quarter fixed effects. Observations are at the product-chain-DMA-quarter-year level. The year 2006 is the base category.

In Section 3.3, we describe a category selection procedure in which we first choose the top 200 product categories by revenue, and then screen out categories with large values of within-category price dispersion. All of our baseline results are obtained with the 133 product categories that reflect that screen.

In Figure 3.12, we replicate our product-level markup trends plot using an extended sample of all top 200 categories by revenue. The baseline trend is plotted for comparison. We find similar trends in markups with either selection procedure, with a change of approximately 30 log points from 2006 to 2019.

3.E.2 Markup Measure

Figure 3.13: Markups Over Time: Price-Over-Cost Markups



Notes: This figure shows coefficients and 95 percent confidence intervals from a regression of log markups at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category. Markups are defined as price over marginal cost (p/c) as in De Loecker et al. (2020).

Throughout the paper, we use the Lerner index, $(p-c)/p$, as our measure of markups, which is a typical measure used in the industrial organization literature and in antitrust analysis (Elzinga and Mills, 2011). Other papers studying markups, particularly those in the macroeconomic literature, have used p/c, or price-over-cost markups (e.g., De Loecker et al., 2020). Both measures reflect the same fundamental relationship, but they are measured on different scales. The Lerner index is typically on $[0, 1]$, while price-over-cost markups are typically on $[1, \infty)$.

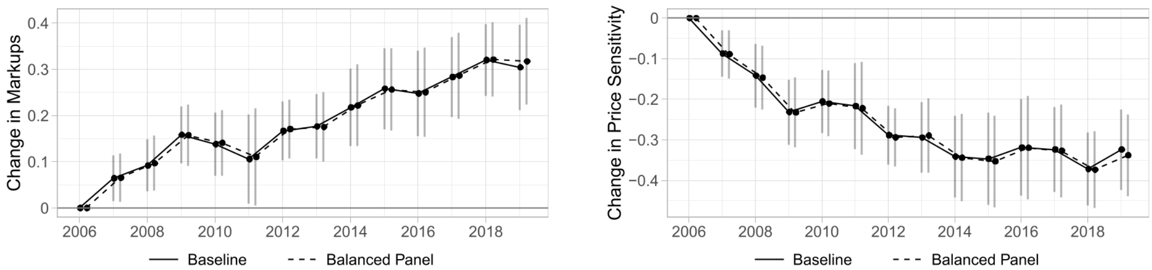
This distinction between the two does not matter for the trends we find in our analysis, which are typically reported in log changes. Figure 3.13 replicates our product-level markup trends, corresponding to panel (a) of Figure 3.5 in the main text, using the price-over-cost markup measure. The trends are nearly identical.

3.E.3 Balanced Panel

Figure 3.14: Balanced Panel

(a) Markup Trend

(b) Price Sensitivity Trend



Notes: This figure shows coefficients and 95 percent confidence intervals from a regression of log markups (panel (a)) and price sensitivity (panel (b)) at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category. The baseline estimates are plotted with a solid line. The dashed line corresponds to an alternative set of estimates from a panel that is balanced by brand×chain×region.

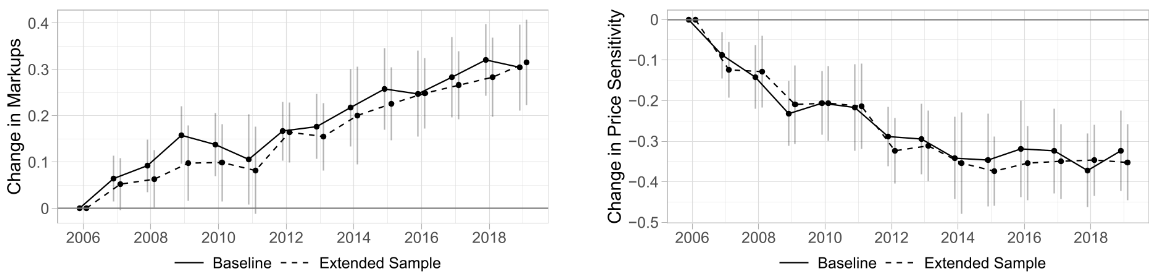
In our main specification, we use an unbalanced panel to maximize sample size and capture changes in aggregate markups due to entry and exit of products. As we discuss in section 3.3, some compositional changes in the NielsenIQ data occur during our sample period due to coverage of certain retail chains. Although our demand estimation controls for chain×region fixed effects, and these fixed effects can change with each year, a possible concern is that retail chains entering the sample may have different growth rates of markups.

In Figure 3.14, we therefore replicate trends of markups and price sensitivities using a balanced panel of brand×chain×region combinations. The trends are similar to those reported in panel (a) of Figure 3.5 and panel (b) of Figure 3.6. The baseline trends are reproduced in the figure for comparison.

3.E.4 Retailer Sample

Figure 3.15: Extended Retailer Sample

(a) Markup Trend (b) Price Sensitivity Trend



Notes: This figure shows coefficients and 95 percent confidence intervals from a regression of log markups (panel (a)) and price sensitivity (panel (b)) at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category. The baseline estimates are plotted with a solid line. The dashed line corresponds to an alternative set of estimates that incorporates large retailers present in the consumer panel data but not in the retail scanner data.

Our baseline data for prices and quantities comes from the retail scanner data, which captures weekly sales by products for a sample of retailers. Though the random coefficients model allows for some forms of selection into the retailers in our sample, one potential concern is there may be a trend in how consumers select outside of our baseline sample in ways that could bias our estimates.

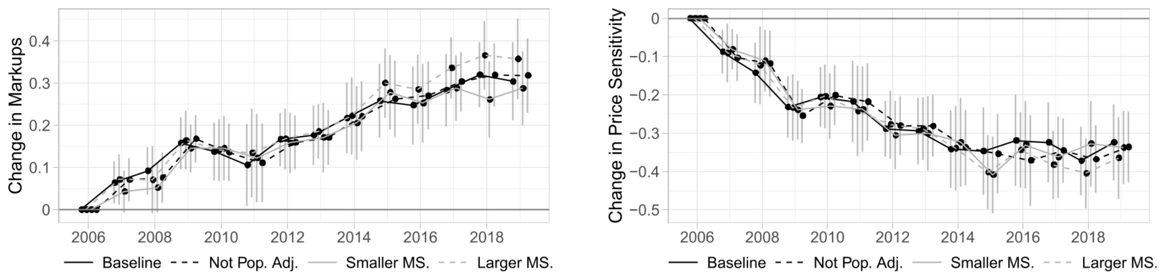
We perform an additional set of robustness checks by supplementing our baseline sample from the retail scanner data with large retailers that are in the consumer panel but not in the retail scanner data. Specifically, we construct product-level price and quantity data for retailers with greater than a 5 percent revenue share in the consumer panel across all of our 133 product categories.[53] We add retailers that are not in the scanner data to our sample, scaling the revenues by DMA-year so that the revenues match for retailers in both samples. We re-run the estimates of our price parameters while holding fixed the estimated nonlinear parameters for this augmented dataset. We find very similar trends in markups and price sensitivity, which are displayed in Figure 3.15. The baseline trends are reproduced in the figure for comparison.

[53] The added retailers have lower product-level prices on average, but there is no differential trend in prices relative to our baseline sample.

3.E.5 Market Size

Figure 3.16: Alternative Market Size Measures

(a) Markup Trend (b) Price Sensitivity Trend



Notes: This figure shows coefficients and 95 percent confidence intervals from a regression of log markups (panel (a)) and price sensitivity (panel (b)) at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category. The baseline estimates are plotted with a solid black line. The gray line corresponds to estimates using an alternative market size calculation that does not vary with population over time, and the dashed line corresponds to estimates that use a alternative values for the average market size. Smaller (larger) market size refers to a specification where we rescale market size such that the average combined market share of inside goods equals 0.6 (0.3).

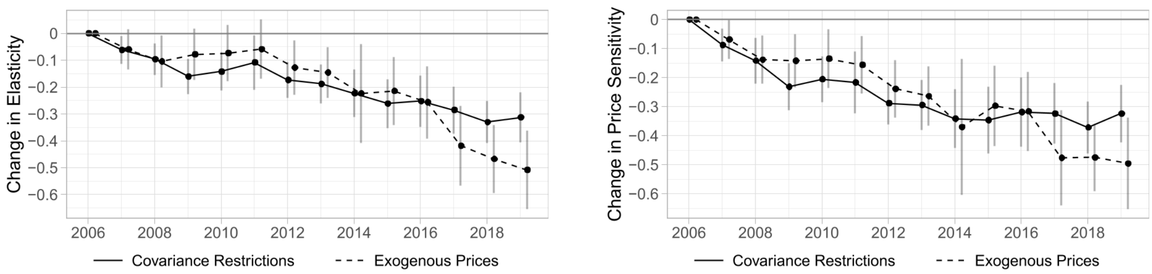
As discussed in Section 3.2.2, we need an assumption about market size to measure market shares of products. In Appendix 3.B.1, we describe how we scale market size to obtain an average market share of inside goods of 0.45 and market growth that varies with the growth of population at the regional level.

To check the robustness of our results towards assumptions about the relevant market, we reran our demand estimation using two alternative definitions of market size. First, we rescale market size to obtain an average combined market share of inside goods of either 0.3 or 0.6. Second, we assume that market size does not vary with population growth. Figure 3.16 shows that these alternative assumptions lead to similar trends in markups and price sensitivity. Thus, the trends we estimate do not hinge on the precise definition of market size.

3.E.6 Changes in Demand Over Time

Figure 3.17: Changes in Demand Over Time

(a) Elasticity Trend (b) Price Sensitivity Trend



Notes: This figure shows coefficients and 95 percent confidence intervals from a regression of the log absolute value of the own-price elasticity (panel (a)) and price sensitivity (panel (b)) at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category. The baseline estimates are plotted with a black line and employ covariance restrictions to estimate mean price parameters. The dashed line corresponds to estimates that instead employ an assumption that prices are exogenous.

We examine whether the estimated trends in demand, in terms of more inelastic demand and reduced price sensitivity, are robust to the supply model and the covariance restrictions that we invoke to identify the mean price parameter. As described in the text, the other demand-side parameters are identified by micro-moments. Thus, here we focus on the mean price parameter, which also has implications for the implied elasticities.

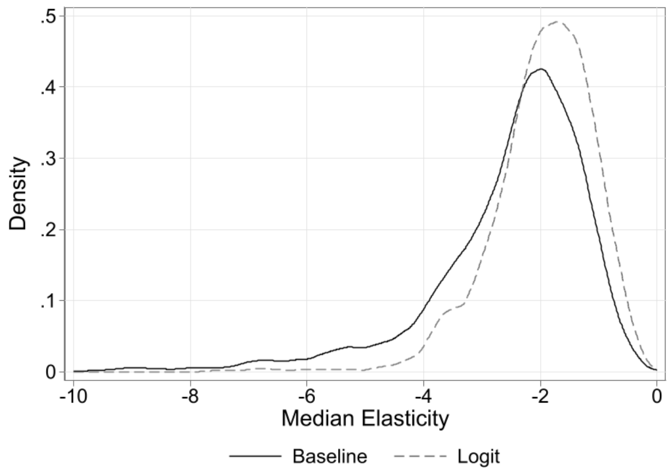
We show that a similar trend is obtained when we estimate demand using the assumption that prices are exogenous, which does not invoke the supply model to pin down the demand parameters. Though elasticity estimates under this approach are often unreasonable in terms of levels (see Section 3.4.2), a change in the estimated parameters would be consistent with a rotation of the demand curve.

Figure 3.17 shows that we find similar trends in elasticities (panel (a)) and the mean price parameter (panel (b)) under the assumption that prices are exogenous. This finding indicates that the reduced-form relationship between prices and quantities is becoming more "vertical" (on a price-quantity graph) over time, consistent with a rotation in the demand curve. The covariance restriction approach finds a similar trend while correcting for price endogeneity. The fact that the trends are similar suggests that our finding of reduced price sensitivity is not sensitive to the particular supply-side assumptions we invoke in estimation.[54]

[54] Of course, as indicated in the main text, a model of firm behavior is required to calculate markups and evaluate whether they are increasing. Regardless of whether firms actually exert market power, a finding of less elastic demand points to a increase in market power *potential*. We thank Chad Syverson for offering this interpretation.

3.E.7 Random Coefficients Logit versus Logit Demand

Figure 3.18: Implied Elasticities for Baseline and Logit Estimates



Notes: This figure plots the density of the median own-price elasticity by category and year. The solid black line shows the density of median elasticities using our baseline specification. The dashed line shows the density of median elasticities from a logit specification without random coefficients. Random coefficients allow for richer consumer heterogeneity.

We examine whether the consumer heterogeneity parameters we include in our baseline specification materially change the estimated elasticities and implied markups. For a comparison, we estimate a standard logit demand model ($\Pi_1 = 0$, $\Pi_2 = 0$, $\sigma = 0$) for all categories and years. Figure 3.18 plots the density of median elasticities in our baseline model (black line) against those in the logit specification (dashed line).

Relative to the logit specification, our baseline estimates obtain more elastic demand estimates and smaller markups. The mean across the category-year median elasticity estimates is -2.57 in our baseline specification and -1.96 in the logit specification. More than twice as many estimates have a median elasticity ¿ -1 (inelastic demand) with the logit specification. Median category-year markups are 0.120 higher in the logit specification (0.686 versus 0.566). These differences are all statistically significant (p-value < 0.001). We obtain an increasing trend in markups with the logit specification, but the trend is steeper, rising from 0.55 to 0.77.

3.F Incorporating Additional Product Characteristics

In this section, we document the point estimates for the ready-to-eat cereals category for our baseline estimates and for an additional test where we include additional product characteristics when estimating demand.

Panel A of Table 3.8 reports the point estimates and standard errors for the mean price parameter and the demographic interactions, including the observed demographics (income and children) and the unobserved $N(0,1)$ draws. Fixed effects are included in estimation but not reported. Panel B of Table 3.8 reports the number of observations, the median own-price elasticity, and the median Lerner index. Each column of the table corresponds to a different year, and each year is estimated independently. We use the standard GMM formula to calculate standard errors while clustering at the DMA level, and we apply a small-sample adjustment that scales up the standard errors to account for the fact that we have a small number of clusters.[55]

Our estimated parameters change some from year to year. For example, from 2016 to 2018, the price parameter changes from -12.93 to -26.44 and back to -13.31. These changes are not due to convergence properties,[56] but instead are due to changes related to demographics and the associated nonlinear parameters. For 2017, the Children$\times$Constant and $N(0,1)\times$Constant coefficient estimates are unusually large, and the price coefficient increases in magnitude in response. To confirm this, we fix the demographic draws and the nonlinear parameters to the 2015 values and re-estimate the price coefficient. When we do this, we obtain a price coefficient of -14.0 and a median elasticity of 0.436, which are closer to the values in the surrounding years. Across all years, holding fixed the demographics and nonlinear parameters at the 2015 values tends to reduce the year-to-year variation in the price coefficient, though the coefficients in most years are only slightly affected, and we still obtain an average markup of approximately 0.50 and no trend in markups for the category.

These blips in parameter estimates can occur in other categories, but they appear to be idiosyncratic and are not frequent. Because we pool our results across more than 100 product categories, the presence of such idiosyncratic blips is not, in our view, a critical issue. We do not see anything systematic across 2017 or in more generally in later years of the sample. Overall, the parameter estimates appear to be fairly stable over time, given the fact that we allow all of our parameters to float independently across years.

We also test for the robustness of our estimates to the inclusion of product characteristics. For this purpose, we follow a similar procedure to Backus et al. (2021). We collect data

[55] An earlier version of this paper did not incorporate the additional small-sample adjustment. The adjustment delivers standard errors of the same order of magnitude as a jackknife estimate of standard errors for the price coefficients. MacKay and Miller (2023) demonstrate how the standard errors from the covariance restriction approach can be substantially smaller than IV standard errors because the estimator exploits observed variation in prices and quantities. We view the reported standard errors as indicating that we have a large number of observations and a good deal of variation in the data; inference for coefficients from specific categories is not central to our project.

[56] Figure 3.23 shows the objective function remains smooth with a single minimum. In fact, we obtain smaller standard errors for this estimate, which suggests that the price coefficient estimate is fairly precise conditional on the nonlinear parameters.

on characteristics at the UPC level, and we merge these characteristics to the UPCs that are associated with each product (brand) in our sample.[57] The characteristics include ingredients, nutritional information, and how the product was marketed. Specifically, we include dummy variables for whether the first ingredient is rice, oat, wheat, corn, protein, almond, or sugar; we include the amount per serving of sugar, fiber, sodium, saturated fat, calories, protein, iron, calcium, and cholesterol; and we include dummy variables for whether the product is marketed as for children, functional/healthy (e.g., heart healthy, antioxidants, etc.), natural, or with low value of "unhealthy" ingredients (e.g., low cholesterol, low fat, etc.). To reduce the dimension of product characteristics, we follow Backus et al. (2021) and project these 20 variables onto the first three principal components ($PC1, PC2, PC3$), which we use in estimation.[58] We interact these variables with our demographics (income and the presence of children) to allow for a product-consumer-specific constant in equation (2). For instance, this can in principle capture that households with children receive higher utility from cereals marketed for children compared to households without children. We do not include the principal components as separate variables without interactions since these are collinear with product fixed effects.

Table 3.9 reports the resulting estimates. Many of the product characteristic interactions are statistically significant, but they do not substantially change our conclusions about markups in the ready-to-eat cereal industry. The price coefficients, elasticities, and implied markups are quite similar to those in our baseline estimates in most years.[59]

[57]Our data on characteristics was obtained from Mintel. On average, we merge characteristics from 53 UPCs to each brand, excluding private label (1,039 merged UPCs) and fringe brands (2,559 merged UPCs). The characteristics are fairly stable within these brands.

[58]The first component is correlated with wheat, protein, fiber, and functional/healthy, the second component is correlated with oats, iron, and calcium, and the third is correlated with rice and low values of unhealthy ingredients

[59]One year where these coefficients do change materially is 2017, which, as we note above, has a bit of instability in our baseline estimates due to the demographic characteristics and associated interactions.

Table 3.8: Estimation Results for RTE Cereals

Panel A: Point Estimates and Standard Errors														
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Price	-18.111	-10.547	-12.987	-10.070	-10.599	-9.128	-10.289	-10.834	-11.999	-11.627	-12.933	-26.440	-13.316	-16.857
	(0.016)	(0.013)	(0.012)	(0.012)	(0.008)	(0.005)	(0.006)	(0.005)	(0.006)	(0.011)	(0.018)	(0.003)	(0.010)	(0.019)
Demographic Interactions														
Income×Price	0.678	1.328	1.157	0.589	0.315	0.729	0.797	1.250	0.852	0.639	0.679	0.898	0.502	0.313
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.010)	(0.001)	(0.001)
Income×Constant	0.150	0.218	0.420	0.215	0.294	-0.006	-0.073	-0.106	-0.050	-0.032	0.026	0.611	0.196	0.314
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.005)	(0.001)	(0.002)
Children×Price	-0.437	-1.432	-0.744	1.141	1.650	2.836	3.321	2.389	2.327	2.405	2.937	2.675	2.454	2.204
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.008)	(0.002)	(0.002)
Children×Constant	7.095	4.727	5.764	2.207	3.579	0.869	0.567	0.681	0.528	0.801	2.288	8.394	4.346	5.172
	(0.021)	(0.022)	(0.016)	(0.013)	(0.014)	(0.003)	(0.000)	(0.001)	(0.001)	(0.008)	(0.024)	(0.014)	(0.010)	(0.023)
Random Coefficient														
N(0,1)×Constant	5.649	3.840	5.226	2.261	4.452	0.689	0.003	0.240	0.243	1.412	4.758	17.462	8.510	10.220
	(0.019)	(0.023)	(0.016)	(0.019)	(0.019)	(0.009)	(0.010)	(0.009)	(0.010)	(0.019)	(0.047)	(0.030)	(0.019)	(0.044)
Panel B: Other Statistics														
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Observations	15,441	16,336	16,604	16,791	17,241	17,329	16,444	16,213	16,443	15,829	15,487	14,365	18,850	17,805
Median Own Elasticity	3.353	1.996	2.573	2.016	2.029	1.744	2.067	2.151	2.349	2.196	2.374	4.732	2.308	2.957
Median Lerner	0.345	0.578	0.454	0.562	0.578	0.627	0.522	0.498	0.455	0.500	0.490	0.253	0.504	0.397

Notes: This table summarizes the results of estimation for the ready-to-eat cereals category for each year in the sample. Panel A provides the parameters and the standard errors, which are clustered at the region level and include a small-sample correction for the number of clusters. Panel B provides the number of product-chain-region-quarter observations, the median own price elasticity of demand, and the median Lerner index.

Table 3.9: Alternative Estimation for RTE Cereals Including Product Characteristics

Panel A: Point Estimates and Standard Errors

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Price	-18.067	-9.733	-10.852	-9.198	-10.118	-9.379	-10.288	-10.884	-11.901	-11.171	-10.962	-13.216	-12.861	-16.471
	(0.019)	(0.007)	(0.006)	(0.004)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)	(0.009)	(0.013)	(0.018)
Demographic Interactions														
Income×Price	1.790	2.527	2.189	1.935	1.181	2.000	1.838	2.357	1.673	1.696	1.483	1.820	0.862	1.042
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
Income×Constant	-0.034	-0.235	-0.243	-0.197	0.035	-0.266	-0.287	-0.325	-0.217	-0.246	-0.222	-0.254	0.185	0.279
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Children×Price	1.367	0.821	-0.335	-0.025	0.891	3.154	2.019	0.889	0.742	0.000	1.150	0.795	1.321	-0.545
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Children×Constant	4.024	1.385	1.314	1.149	1.978	0.579	0.722	0.938	0.748	0.977	0.699	0.657	3.359	4.621
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Product Characteristics														
Income×PC1	0.016	0.021	0.014	0.027	0.023	0.026	0.022	0.023	0.014	0.019	0.017	0.018	0.013	0.015
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)
Children×PC1	-0.124	-0.112	-0.118	-0.109	-0.083	-0.055	-0.078	-0.069	-0.087	-0.104	-0.113	-0.092	-0.077	-0.105
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Income×PC2	-0.018	-0.026	-0.025	-0.019	-0.013	-0.016	-0.012	-0.016	-0.013	-0.007	-0.003	-0.008	0.011	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
Children×PC2	-0.011	-0.025	-0.025	-0.033	-0.027	-0.039	-0.031	-0.001	-0.011	0.021	-0.008	0.011	0.003	0.018
	(0.018)	(0.004)	(0.002)	(0.001)	(0.011)	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)	(0.013)	(0.019)
Income×PC3	-0.027	-0.020	-0.007	-0.006	0.007	-0.005	0.002	0.014	-0.003	-0.018	-0.010	-0.006	-0.011	-0.011
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Children×PC3	-0.217	-0.226	-0.254	-0.223	-0.205	-0.154	-0.190	-0.166	-0.173	-0.179	-0.180	-0.195	-0.170	-0.211
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
Random Coefficient														
N(0,1)×Constant	5.253	1.227	0.341	0.141	2.570	0.355	0.088	0.472	0.174	0.869	0.770	0.110	7.106	9.737
	(0.025)	(0.010)	(0.010)	(0.011)	(0.019)	(0.008)	(0.013)	(0.011)	(0.012)	(0.010)	(0.008)	(0.009)	(0.028)	(0.043)

Panel B: Other Statistics

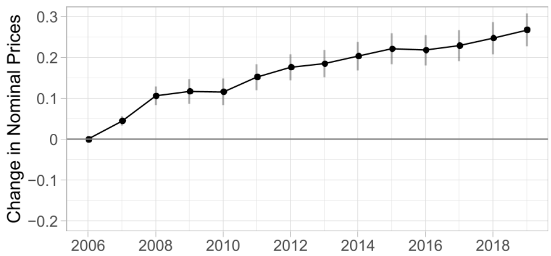
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Observations	15,441	16,336	16,604	16,791	17,241	17,329	16,444	16,213	16,443	15,829	15,487	14,365	18,850	17,805
Median Own Elasticity	3.258	1.732	2.121	1.839	1.932	1.746	2.085	2.191	2.356	2.163	2.057	2.403	2.239	2.930
Median Lerner	0.354	0.640	0.519	0.593	0.594	0.622	0.517	0.491	0.454	0.501	0.522	0.443	0.516	0.400

Notes: This table summarizes the results of estimation for the ready-to-eat cereals category for each year in the sample. Panel A provides the parameters and the standard errors, which are clustered at the region level and include a small-sample correction for the number of clusters. Panel B provides the number of product-chain-region-quarter observations, the median own price elasticity of demand, and the median Lerner index.

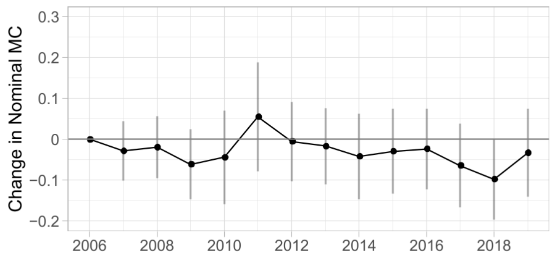
3.G Additional Figures and Tables

Figure 3.19: Product-Level Changes in Nominal Prices and Marginal Costs

(a) Prices

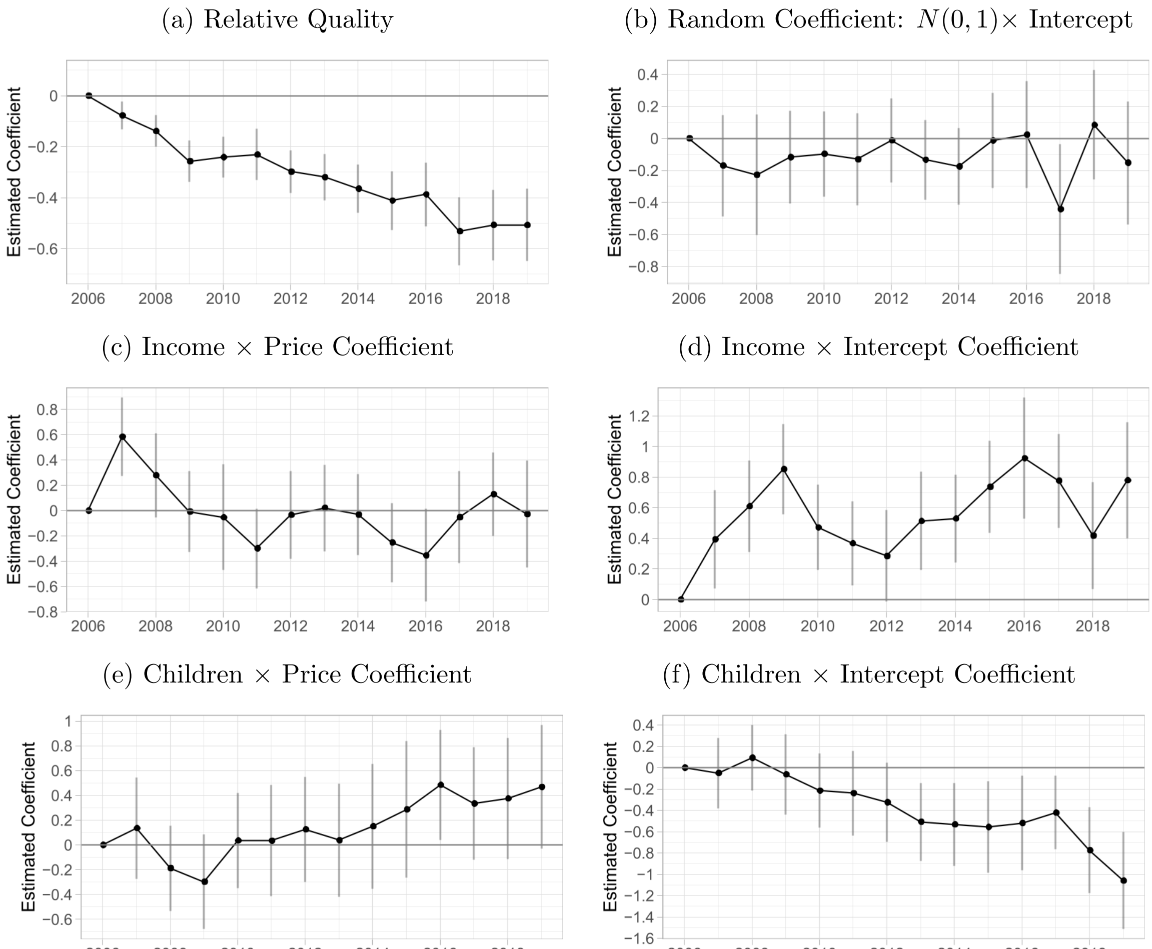


(b) Marginal Costs



Notes: This figure shows coefficients and 95 percent confidence intervals of regressions of the log of nominal prices and marginal costs at the product-chain-DMA-quarter-year level on year dummies controlling for product-chain-DMA and quarter fixed effects. The year 2006 is the base category.

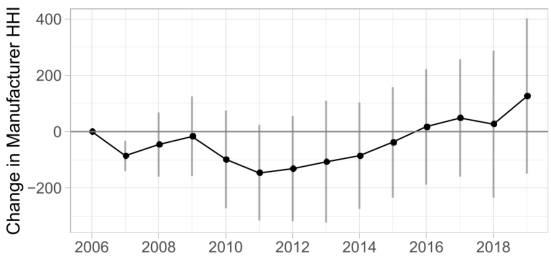
Figure 3.20: Changes in Demand Parameters



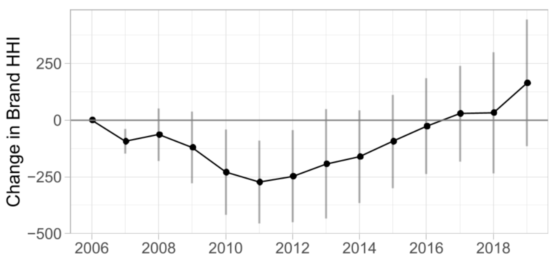
Notes: This figure shows coefficients and 95 percent confidence intervals of a regression of standardized demand parameters on year dummies controlling for product-chain-DMA and quarter fixed effects. Observations are at the product-chain-DMA-quarter-year level. The year 2006 is the base category.

Figure 3.21: Changes in Market Concentration

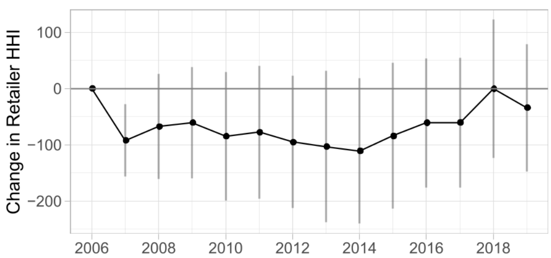
(a) Parent HHI



(b) Brand HHI

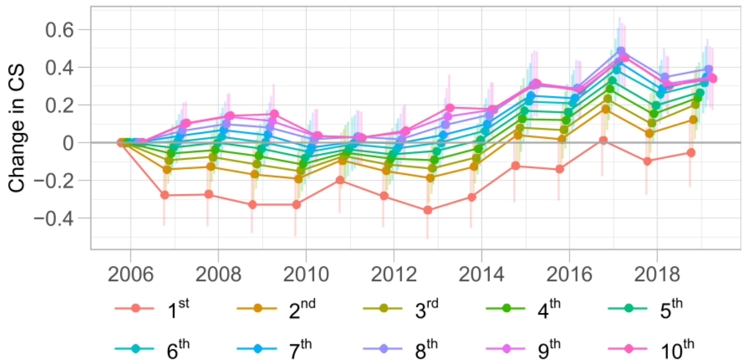


(c) Retailer HHI



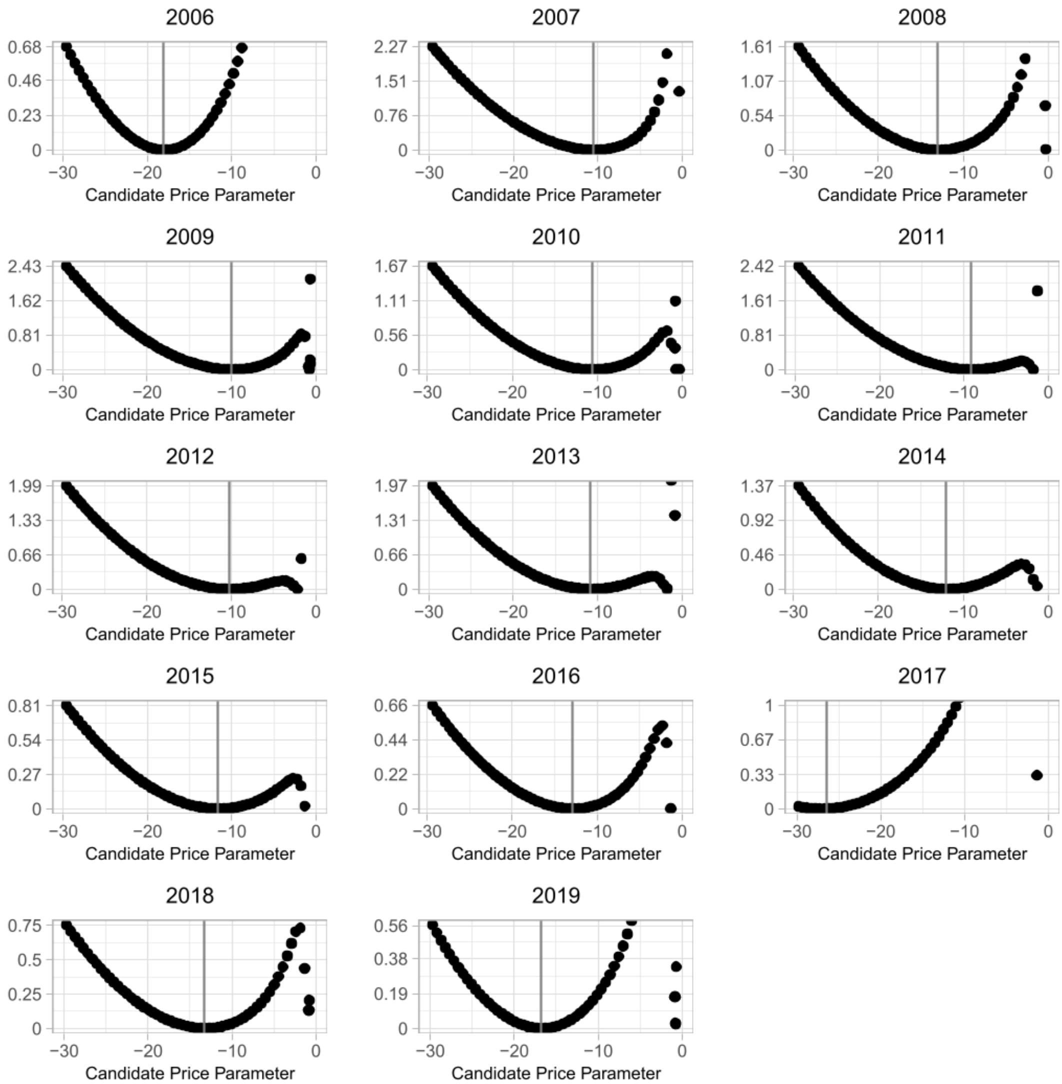
Notes: This figure shows coefficients and 95 percent confidence intervals of a regression of HHI measures on year dummies controlling for product-chain-DMA and quarter fixed effects, with 2006 as the base category. We measure HHI as the sum of squared market shares, where we first adjust market shares so that inside shares sum to one. For this figure, HHI is measured on a 0 to 10,000 scale. Observations are at the product-chain-DMA-quarter-year level.

Figure 3.22: Consumer Surplus Over Time By Income Group, Deciles



Notes: This figure reports coefficients and 95 percent confidence intervals of a regression of the log of consumer surplus by purchase on year dummies, controlling for category fixed effects, separately for different deciles of the income distribution.

Figure 3.23: Contribution of Covariance Restriction to Objective Function: Ready-to-Eat Cereals



Notes: This figure plots the contribution of the covariance restriction to the objective function, scaled by ten thousand, for different candidate price parameters over the range $[-30, 0]$. Other parameters are held fixed at the levels obtained in the first step of estimation.

Table 3.10: Product-Level Markups Over Time, Sales-Weighted Regressions

	(1) Log Markup	(2) Log Markup	(3) Log Markup	(4) Log Markup	(5) Log Markup	(6) Log Markup
Trend	0.018***		0.017***		0.022***	
	(0.003)		(0.003)		(0.003)	
Year=2007		0.061**		0.063**		0.064**
		(0.025)		(0.025)		(0.025)
Year=2008		0.092***		0.093***		0.092***
		(0.028)		(0.028)		(0.029)
Year=2009		0.164***		0.163***		0.158***
		(0.031)		(0.031)		(0.031)
Year=2010		0.144***		0.142***		0.138***
		(0.034)		(0.034)		(0.034)
Year=2011		0.108**		0.105**		0.106**
		(0.048)		(0.048)		(0.049)
Year=2012		0.169***		0.161***		0.167***
		(0.032)		(0.033)		(0.032)
Year=2013		0.180***		0.169***		0.177***
		(0.036)		(0.036)		(0.035)
Year=2014		0.214***		0.203***		0.217***
		(0.042)		(0.043)		(0.042)
Year=2015		0.243***		0.230***		0.258***
		(0.045)		(0.045)		(0.044)
Year=2016		0.222***		0.208***		0.248***
		(0.048)		(0.047)		(0.047)
Year=2017		0.253***		0.238***		0.283***
		(0.044)		(0.043)		(0.044)
Year=2018		0.277***		0.274***		0.320***
		(0.041)		(0.040)		(0.039)
Year=2019		0.255***		0.253***		0.304***
		(0.049)		(0.047)		(0.047)
Quarter FEs	X	X	X	X	X	X
Category, Retailer & DMA FEs			X	X		
Brand-Category-DMA-Retailer FEs					X	X
Observations	14,407,410	14,407,410	14,407,410	14,407,410	14,407,410	14,407,410
R^2	0.013	0.014	0.357	0.359	0.782	0.783

Notes: Dependent variable is the log of the Lerner index. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.11: Product-Level Markups Over Time, Unweighted Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Markup	Log Markup	Log Markup	Log Markup	Log Markup	Log Markup
Trend	0.018***		0.020***		0.023***	
	(0.003)		(0.003)		(0.003)	
Year=2007		0.079***		0.078***		0.086***
		(0.028)		(0.028)		(0.028)
Year=2008		0.098***		0.095***		0.111***
		(0.029)		(0.029)		(0.030)
Year=2009		0.157***		0.154***		0.175***
		(0.034)		(0.034)		(0.035)
Year=2010		0.145***		0.142***		0.165***
		(0.038)		(0.037)		(0.038)
Year=2011		0.103**		0.099**		0.125***
		(0.040)		(0.039)		(0.040)
Year=2012		0.176***		0.173***		0.202***
		(0.037)		(0.037)		(0.037)
Year=2013		0.192***		0.187***		0.215***
		(0.033)		(0.032)		(0.033)
Year=2014		0.226***		0.223***		0.254***
		(0.042)		(0.040)		(0.041)
Year=2015		0.306***		0.310***		0.345***
		(0.052)		(0.050)		(0.052)
Year=2016		0.258***		0.262***		0.298***
		(0.046)		(0.045)		(0.047)
Year=2017		0.278***		0.288***		0.326***
		(0.049)		(0.046)		(0.049)
Year=2018		0.269***		0.286***		0.322***
		(0.039)		(0.038)		(0.040)
Year=2019		0.227***		0.247***		0.282***
		(0.041)		(0.041)		(0.042)
Quarter FEs	X	X	X	X	X	X
Category, Retailer & DMA FEs			X	X		
Brand-Category-DMA-Retailer FEs					X	X
Observations	14,407,410	14,407,410	14,407,410	14,407,410	14,407,410	14,407,410
R^2	0.011	0.014	0.353	0.356	0.760	0.763

Notes: Dependent variable is the log of the Lerner index. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.12: Product-Level Markups Over Time, Balanced Panel, Sales-Weighted Regressions

	(1) Log Markup	(2) Log Markup	(3) Log Markup	(4) Log Markup	(5) Log Markup	(6) Log Markup
Trend	0.021***		0.019***		0.022***	
	(0.003)		(0.003)		(0.003)	
Year=2007		0.064**		0.064**		0.065**
		(0.027)		(0.026)		(0.026)
Year=2008		0.100***		0.100***		0.097***
		(0.030)		(0.030)		(0.030)
Year=2009		0.168***		0.166***		0.157***
		(0.033)		(0.033)		(0.034)
Year=2010		0.155***		0.151***		0.141***
		(0.036)		(0.036)		(0.036)
Year=2011		0.126**		0.120**		0.111**
		(0.053)		(0.053)		(0.053)
Year=2012		0.188***		0.178***		0.171***
		(0.032)		(0.033)		(0.032)
Year=2013		0.192***		0.180***		0.175***
		(0.038)		(0.039)		(0.038)
Year=2014		0.235***		0.222***		0.222***
		(0.043)		(0.046)		(0.045)
Year=2015		0.263***		0.248***		0.256***
		(0.045)		(0.046)		(0.045)
Year=2016		0.246***		0.232***		0.250***
		(0.050)		(0.050)		(0.049)
Year=2017		0.276***		0.259***		0.286***
		(0.047)		(0.047)		(0.047)
Year=2018		0.309***		0.291***		0.321***
		(0.041)		(0.041)		(0.040)
Year=2019		0.301***		0.283***		0.317***
		(0.047)		(0.046)		(0.047)
Quarter FEs	X	X	X	X	X	X
Category, Retailer & DMA FEs			X	X		
Brand-Category-DMA-Retailer FEs					X	X
Observations	4,821,264	4,821,264	4,821,264	4,821,264	4,821,264	4,821,264
R^2	0.018	0.019	0.398	0.399	0.764	0.766

Notes: Dependent variable is the log of the Lerner index. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.13: Factors Predicting Cross-Category Variation in Markup Trends (Category Level)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Marginal Cost (Standardized)	-0.238^{***}					-0.137^{***}	-0.135^{***}
	(0.010)					(0.010)	(0.010)
Price Sensitivity		-0.667^{***}				-0.406^{***}	-0.408^{***}
		(0.027)				(0.037)	(0.037)
Quality (Standardized)			-0.203^{***}			-0.000	0.001
			(0.011)			(0.008)	(0.008)
Income (Log)				-2.373		-0.391	-0.467
				(2.190)		(0.773)	(0.785)
Children at Home				-5.100		-2.296	-2.707
				(6.916)		(2.838)	(2.711)
Parent HHI					1.042^{***}		0.487^{***}
					(0.358)		(0.133)
Brand HHI					-0.374		-0.049
					(0.294)		(0.116)
Retailer HHI					1.705^{**}		0.439
					(0.850)		(0.318)
Category FEs	X	X	X	X	X	X	X
Year FEs	X	X	X	X	X	X	X
Observations	1,862	1,862	1,862	1,862	1,862	1,862	1,862
R^2 (Within)	0.707	0.726	0.496	0.002	0.016	0.848	0.852

Notes: Dependent variable is the log of the mean Lerner index within a category-year. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Bibliography

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *Quarterly Journal of Economics*, 120(5):701–728.

Anderson, E., Rebelo, S., and Wong, A. (2018). Markups across space and time. working paper.

Arcidiacono, P., Ellickson, P. B., Mela, C. F., and Singleton, J. D. (2020). The competitive effects of entry: Evidence from supercenter expansion. *American Economic Journal: Applied Economics*, 12(3):175–206.

Backus, M., Conlon, C., and Sinkinson, M. (2021). Common ownership and competition in the ready-to-eat cereal industry. working paper.

Basu, S. (2019). Are price-cost markups rising in the united states? a discussion of the evidence. *Journal of Economic Perspectives*, 33(3):3–22.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.

Berry, S., Gaynor, M., and Scott Morton, F. (2019). Do increasing markups matter? lessons from empirical industrial organization. *Journal of Economic Perspectives*, 33(3):44–68.

Berry, S. and Jia, P. (2010). An empirical analysis of the airline industry. *American Economic Journal: Microeconomics*, 2:1–43.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, pages 841–890.

Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy*, 112(1):68–105.

Berry, S. T. and Haile, P. A. (2021). Foundations of demand estimation. *Handbook of Industrial Organization, Volume 4*, edited by Kate Ho, Ali Hortacsu, and Alessandro Lizzeri, Elsevier.

Berry, S. T. and Haile, P. A. (2022). Nonparametric identification of differentiated products demand using micro data. working paper.

Bhattacharya, V., Illanes, G., and Stillerman, D. (2022). Have mergers raised prices? evidence from U.S. retail. Working Paper.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, 121:1–14.

Brand, J. (2021). Differences in differentiation: Rising variety and markups in retail food stores. working paper.

Brunner, D., Heiss, F., Romahn, A., and Weiser, C. (2020). Reliable estimation of random coefficient logit demand models. working paper.

Butters, R. A., Sacks, D. W., and Seo, B. (2022). How do national firms respond to local shocks? evidence from excise taxes. *American Economic Review*, 112(5):1737–72.

Caoui, E. H., Hollenbeck, B., and Osborne, M. (2022). The impact of dollar store expansion on local market structure and food access. Working Paper.

Chevalier, J. A., Kashyap, A. K., and Rossi, P. E. (2003). Why don't prices rise during periods of peak demand? Evidence from scanner data. *American Economic Review*, 93(1):15–37.

Ciliberto, F., Murry, C., and Tamer, E. (2021). Market structure and competition in airline markets. *Journal of Political Economy*, 129(11):2995–3038.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with PyBLP. *RAND Journal of Economics*, 51(4):1108–1161.

De Loecker, J. and Eeckhout, J. (2021). Global market power. working paper.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, 135(2):561–644.

De Loecker, J. and Scott, P. (2022). Markup estimation using production and demand data. an application to the us brewing industry. working paper.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–2471.

De Ridder, M., Grassi, B., and Morzenti, G. (2022). The hitchhiker's guide to markup estimation. Working Paper.

DellaVigna, S. and Gentzkow, M. (2019). Uniform pricing in US retail chains. *Quarterly Journal of Economics*, 134(4):2011–2084.

Doraszelski, U. and Jaumandreu, J. (2019). Using cost minimization to estimate markups. working paper.

Eizenberg, A., Lach, S., and Oren-Yiftach, M. (2021). Retail prices in a city. *Americal Economic Journal: Economic Policy*, 13(2):175–206.

Elzinga, K. G. and Mills, D. E. (2011). The Lerner Index of Monopoly Power: Origins and Uses. *American Economic Review*, 101(3):558–564.

Ganapati, S. (2021a). Growing oligopolies, prices, output, and productivity. *Americal Economic Journal: Microeconomics*, 13(3):309–327.

Ganapati, S. (2021b). The modern wholesaler: Global sourcing, domestic distribution, and scale economies. working paper.

Gandhi, A. and Houde, J.-F. (2020). Measuring substitution patterns in differentiated products industries. working paper.

Gandhi, A. and Nevo, A. (2021). Empirical models of demand and supply in differentiated products industries. *Handbook of Industrial Organization, Volume 4*, edited by Kate Ho, Ali Hortacsu, and Alessandro Lizzeri, Elsevier.

Grieco, P. L. E., Murry, C., and Yurukoglu, A. (2022). The evolution of market power in the US automobile industry. working paper.

Griffith, R., Jin, W. M., and Lechene, V. (2022). The decline of home-cooked food. *Fiscal Studies*, 43(2):105–120.

Grigolon, L. and Verboven, F. (2014). Nested logit or random coefficient logit? A comparison of alternative discrete choice models of product differentiation. *Review of Economics and Statistics*, 96(5):916–935.

Hall, R. E. (1988). The relation between price and marginal cost in u.s. industry. *Journal of Political Economy*, 96:921–947.

Harberger, A. C. (1954). Monopoly and resource allocation. *American Economic Review (Papers and Proceedings)*, 44(2):77–87.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In Bresnahan, T. and Gordon, R., editors, *The Economics of New Goods*, pages 209–248. University of Chicago Press, Chicago.

Hendel, I. and Nevo, A. (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review*, 103(7):2722–2751.

Hristakeva, S. (2020). Vertical contracts with endogenous product selection: An empirical analysis of vendor-allowance contracts. working paper.

Jaimovich, N., Rebelo, S., and Wong, A. (2019). Trading down and the business cycle. *Journal of Monetary Economics*, 102:96–121.

Lerner, A. P. (1934). The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies*, 1(3):157–175.

MacKay, A. and Miller, N. H. (2023). Estimating models of supply and demand: Instruments and covariance restrictions. working paper.

Mela, C. F., Gupta, S., and Lehmann, D. R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research*, 34(2):248–261.

Miller, N., Osborne, M., Sheu, G., and Sileo, G. (2022). The evolution of concentration and markups in the unites states cement industry. working paper.

Miller, N. H. and Weinberg, M. C. (2017). Understanding the price effects of the MillerCoors joint venture. *Econometrica*, 85(6):1763–1791.

Nakamura, E. and Zerom, D. (2010). Accounting for incomplete pass-through. *Review of Economic Studies*, 77(3):1192–1230.

Nevo, A. (2000a). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *RAND Journal of Economics*, 31(3):395–421.

Nevo, A. (2000b). A practitioner's guide to estimation of random coefficients logit models of demand. *Journal of Economics and Management Strategy*, 9:513–548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

Pellegrino, B. (2021). Product differentiation and oligopoly: A network approach. working paper.

Peltzman, S. (2020). Productivity, prices and productivity in manufacturing: a Demsetzian perspective. working paper.

Raval, D. (2020). Testing the production approach to markup estimation. working paper.

Small, K. A. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, pages 105–130.

Syverson, C. (2019). Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, 33(3):23–43.

Villas-Boas, S. B. (2007). Vertical Relationships Between Manufacturers and Retailers: Inference with Limited Data. *Review of Economic Studies*, 74(2):625–652.

Chapter 4

The Portfolio Power Theory Revisited: Evidence from Cross-Category Mergers in US Retailing

Abstract: I study 57 cross-category mergers among manufacturers in the US consumer packaged goods retail industry to assess the presence, direction, and size of portfolio effects. In doing so, I exploit differences in the pre-merger bargaining positions of the manufacturers at different retailers. I provide evidence that the manufacturer with the weaker pre-merger bargaining position at a retailer can benefit from increased sales. This increase is driven by changes in quantities, not prices. In addition, I also study the effect on measures of marginal costs and perceived quality. I find that changes in perceived quality help explain these patterns but that marginal costs do not play an important role. Finally, I discuss possible channels that could lead to this result and how these channels are related to the portfolio power theory.

Acknowledge: I thank Chris Conlon, Alexandra Gibbon, Andreas Lichter, Alexander MacKay, Matthias Mertens, Nathan Miller, Felix Montag, Joel Stiebale, and seminar participants at DICE for their helpful comments and suggestions. Part of this research was conducted during a research stay at the Department of Economics of Harvard University. I thank Elie Tamer for the invitation. Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf. I gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) (project 235577387/GRK1974) and the Joachim Herz Foundation.

4.1 Introduction

In her speech to the International Competition Network in May 2022, Lina Khan—Chair of the Federal Trade Commission (FTC)—identified three key areas in which merger enforcement in its current form has no bite and where she seeks adjustments in the future, among them the assessment of non-horizontal mergers. While explicitly referring to "deals that might be described as [..] conglomerate", she said that "[w]e must examine how a range of strategies and effects, including extension strategies and portfolio effects, may warrant enforcement action."[1] Her approach to intensifying merger enforcement is part of a larger policy agenda of US President Joe Biden to reverse trends that led to "less competition" and "more concentration" in the previous decades.[2]

Although the portfolio power theory, sometimes referred to as range or portfolio effects, is not new, the literature lacks a clear definition of what it means by these effects.[3] The idea is usually that if two firms sell their products to the same downstream firms, a merger can benefit them even if their product portfolios do not overlap before the merger. In other words, the increase in the sheer size of a firm's product portfolio can change market outcomes, leaving aside possible substitutability and complementarity considerations within the portfolio.

The channel that is often discussed in this context builds on the idea that up- and downstream firms negotiate with each other over terms of supply. These terms of supply can include financial payments, such as linear wholesale prices or lump sum transfers, but can also include non-financial variables. For instance, the downstream firm could spend more effort on promoting and selling the upstream firm's products. A merger can benefit merging upstream firms by a shift in the so-called gains from trade, that is, the additional gain in profit for a bargaining party due to a collaboration with a firm on the other market side. The idea is that a bargaining breakdown becomes increasingly costly for a downstream firm after the merger because the downstream firm now loses access to the products of both upstream firms and not just to those of a single firm. This increases the incentives for the downstream firm to settle the negotiation with the merged upstream entity and gives the merged entity the possibility to demand larger concessions, such as in the form of a larger

[1] A text version of her speech can be found at `https://www.ftc.gov/system/files/ftc_gov/pdf/Remarks%20of%20Chair%20Lina%20M.%20Khan%20at%20the%20ICN%20Conference%20on%20May%206%2C%202022_final.pdf` (last accessed on October 2, 2023).

[2] At the beginning of her speech, Lina Khan herself referred to Biden's agenda, saying, among others: "As you know, competition law in the United States is currently in the midst of a broad and sweeping reassessment. The significance of this reassessment is perhaps best embodied by President Biden's issuance last summer of an Executive Order on Promoting Competition in the American Economy." In his remarks on the executive order, Biden said: "But what we've seen over the past few decades is less competition and more concentration that holds our economy back. We see it in big agriculture, in big tech, in big pharma. The list goes on" (see `https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/07/09/remarks-by-president-biden-at-signing-of-an-executive-order-promoting-competition-in-the-american-economy/`, last accessed on October 2, 2023).

[3] While the OECD (2001) reports a lack of a clear definition, Watson (2003) cites a definition used by the Office of Fair Trading (UK) in a merger case. However, even Watson (2003) acknowledges that "the scope of the concept is somewhat uncertain."

effort or smaller financial transfers.

The approach of the current US administration and FTC leadership has reminded some antitrust scholars of the “big is bad” doctrine that played an essential role in pre-Chicago merger enforcement in the 1960s and 1970s and has raised concerns that merger enforcement in the future may be less based on economic theory. One reason for this impression is clearly that Lina Khan herself referred to this period in her speech.[4] While she tried to correct this impression later,[5] part of the perception might also be a strong conflict that raged among antitrust scholars in North America and the EU around the turn of the millennium. At this time, the European Commission used the portfolio power theory in its assessment of multiple merger cases, the most famous one being General Electric/Honeywell. While some US scholars deemed the portfolio power theory to be driven by the pre-Chicago “big is bad” thinking and concluded that the decision by the European Commission to block the merger was not based on economic theory (see, for instance, Evans and Salinger, 2002; Patterson and Shapiro, 2001), others pointed to a wrong understanding of the underlying theory and argued that the analysis of economic models does support the decision (see, for instance, Choi, 2001; Reynolds and Ordover, 2002).

Interestingly, while the General Electric/Honeywell decision led to a heated and (to some extent) also fruitful debate about the potentially pro- and anti-competitive effects stemming from the portfolio power theory, surprisingly little empirical research has been conducted since then to assess the presence, direction, and size of portfolio effects in practice (two notable exceptions discussed below). In light of the latest developments in the US, this paper seeks to revisit the portfolio power theory in the context of the US consumer packaged goods retail industry. I study 57 mergers between manufacturers and analyze the impact on the interactions with the retailers and market outcomes.

The mergers in my sample have two characteristics that make them particularly useful for studying possible portfolio effects. First, they are usually cross-category mergers, meaning that the manufacturers have (almost) no overlap in the product categories in which they were active before the merger. Thus, these mergers would likely be classified as conglomerate mergers by antitrust practitioners and are not affected by horizontal merger effects stemming from a reduction in the number of competitors. Second, many mergers are characterized by strong asymmetries in manufacturers’ pre-merger sales to various retailers. I take these pre-merger sales as proxies for the bargaining positions of the manufacturers in negotiations with the retailers. My approach follows the idea that if two manufacturers merge and one manufacturer has a better pre-merger bargaining position with a retailer than the other, the manufacturer with the weaker pre-merger bargaining position may benefit from the merger

[4] In her last sentence, before referring to the portfolio effects, she said: “While the U.S. antitrust agencies energetically grappled with some of these dynamics during the era of industrial-era conglomerates in the 1960s and 70s, we must update that thinking for the current economy.”

[5] The reservations of other scholars resulted not only from her speech but also from a number of other statements and actions on her side. The news article “FTC chair defends track record on antitrust challenges, says big isn’t categorically bad” documents one example of her attempt to reverse this impression. It can be found at `https://www.cnbc.com/2023/07/24/ftc-chair-lina-khan-defends-track-record-on-antitrust-challenges.html` (last accessed on October 2, 2023)

because the joint bargaining position yields an improvement compared to the pre-merger situation.

I provide evidence that manufacturers with weaker pre-merger bargaining positions tend to benefit from mergers, while manufacturers with stronger pre-merger bargaining positions tend to be harmed. These benefits (and losses) come through increases (decreases) in revenues, which are almost entirely driven by increases (decreases) in the quantities sold and not by changes in prices. To dig further into possible mechanisms behind these results, I then use the work of Döpper et al. (2023) to derive measures for marginal costs and non-price characteristics of the products. I provide evidence that changes in marginal costs do not drive portfolio effects but that changes in the non-price attributes play a crucial role. I link these findings to two possible explanations related to the portfolio power theory: My first explanation builds on the argument outlined above that a merger shifts the gains from trade, which increases the incentives for the retailers to settle negotiations with the merging manufacturers and to make larger concessions to the merged entity. In the consumer packaged goods retail industry, negotiations are usually not just about financial payments but also about the effort that a retailer puts into selling and promoting the manufacturers' products. These efforts can take the form of more or better shelf space or increased in-store promotional activities. If a product is more heavily promoted or better placed on the shelf, this could increase consumers' perception of the quality of the manufacturers' products, leading to a larger number of sales. The second channel is that manufacturers can achieve synergy gains by joining forces in the organization of a joint distribution network. This also increases the incentives for retailers to spend more effort on the products of the merging manufacturers because stockouts (or similar problems) are less likely to occur. Finally, I briefly discuss why two alternative explanations—increased (retailer-independent) advertising spending and efficiency gains beyond the distribution network—are less likely to explain the documented patterns.

For future versions of this paper, I intend to provide a structural model that helps clarify the mechanism behind the documented patterns and assess the implications for welfare and profit sharing among manufacturers and retailers in order to discuss the pro- or anti-competitive nature of the portfolio effects in the context of the US consumer packaged goods retail industry.

Concerning the related literature, two other papers studying the portfolio effects of conglomerate mergers are worth mentioning. Park (2009) and Chunga and Jeon (2014) study four and five mergers between South Korean beer and soju manufacturers, respectively. While the South Korean beer market is dominated by a small number of large manufacturers, past and current regulations have led to a market structure with strong regional players in the soju market (one strong regional player per region). Park (2009) uses a structural demand model to investigate the presence of portfolio effects and finds no evidence for such effects. In contrast, Chunga and Jeon (2014) use a reduced-form approach and a slightly different set of mergers. They provide evidence that large beer manufacturers are able to leverage their size in some regions to push the products of the integrated soju manufacturers to wholesalers if the soju manufacturers did not have a strong position in the region prior to the mergers. The fact that this effect is only present if the soju manufacturers were small competitors in the respective regions before the mergers suggests that the portfolio effect helps to increase local competition and thus may be pro-competitive. This paper differs from the contributions of Park (2009) and Chunga and Jeon (2014) in numerous respects. For instance, the soju market is heavily regulated (for instance, ban on wholesale price discrimination and ban on TV and radio advertising), while most of the product categories in my study experience rather little regulation (if any). In addition, I consider a much broader set of mergers as well as a large number of product categories, and while the aforementioned studies only analyze the impact on market shares, I consider various other market outcomes.

My study also contributes to the literature on cross-market mergers, that is, mergers between firms that operate in different (geographic or product) markets and, therefore, would usually not raise concerns by antitrust authorities. Cross-market mergers have recently attracted the attention of scholars in health economics. Lewis and Pflum (2017) and Dafny et al. (2019) provide empirical evidence that cross-market hospital mergers can impact market outcomes and lead to price increases. Both studies are similar to this paper in that they provide an in-depth analysis of the interplay between merging upstream firms and downstream intermediaries that bundle the upstream products. However, this paper differs from these studies in that it focuses on a lack of overlap in product rather than geographic markets, deals with a different industry, and documents the effects of cross-market mergers on revenues that are driven by changes in quantities rather than prices (and thus by changes in the non-price characteristics of the products).

Another strand of literature that has similarities to the one on cross-market mergers deals with cross-border mergers, that is, mergers of firms located in different countries. While some of these mergers may also be affected by a market overlap, others are not, so the literature on cross-border mergers is often concerned with discussing merger effects that arise in the absence of overlapping (geographical) markets. For instance, Guadalupe et al. (2012) study the impact of cross-border acquisitions on Spanish manufacturing firms and find that acquired firms' innovation activities increase post-merger. One channel that they identify is that the acquired firms gain better access to foreign markets through their new parents. This is similar to one of the channels that I discuss in Section 4.5, where the acquired targets benefit in negotiations with the retailers and are able to increase the effort provided by the retailers.

Finally, this paper also contributes to recent discussions about the effectiveness of antitrust enforcement in the EU and the US. Bhattacharya et al. (2023) use the same scanner data as in this study (NielsenIQ) in combination with the SDC Platinum merger database from Thompson Reuters to analyze the effects of mergers that might be potentially relevant for antitrust authorities.[6] They use a structural model that allows them to evaluate coun-

[6] Another study that uses the NielsenIQ data set in combination with SDC Platinum is Majerovitz and Yu (2023). The authors focus on the average horizontal merger, which is characterized by strong asymmetries, typically including a small target and a large acquirer.

terfactual scenarios in which they can vary the intensity of merger enforcement and find that an increase in the intensity would lead to a substantial reduction in Type II errors, however, at the expense of a much larger number of cases to be examined. In another study, Affeldt et al., 2021 focus on potential efficiency gains from mergers, which are often used as a defense against potential merger remedies and prohibitions. They conclude that "[c]ompensating efficiencies appear to be simply too large to be achieved by real world mergers [..]." My paper fits into this strand of literature in that I provide evidence for the existence of merger effects that are often ignored by antitrust authorities. If these effects benefit consumers, they could be used by the merging firms as an additional defense tool. If, in contrast, these effects harm consumers, antitrust authorities might want to block an even larger number of mergers.

The remainder of this paper is structured as follows: Section 4.2 describes the data. Section 4.3 introduces three important definitions (Subsection 4.3.1), provides insights about the cross-category activities of the manufacturers (Subsection 4.3.2), and documents the effects of cross-category mergers on directly observable outcomes like revenues, quantities, and prices (Subsection 4.3.3). Since the scope of directly observable market outcomes is limited, I then use the work of Döpper et al. (2023) to shed some light on other measures like marginal costs in Section 4.4. In doing so, I first describe the model (Subsection 4.4.1) and the empirical strategy (Subsection 4.4.2) before extending my analysis of the effects of cross-category mergers in Subsection 4.4.3. Finally, I discuss possible mechanisms that can drive these results in Section 4.5 before summarizing my main findings in Section 4.6.

4.2 Data

A common problem of empirical studies of vertical chains is that contracts between up- and downstream firms are typically not observed, and data on the vertical relations is missing. Therefore, most of the IO literature combines structural models based on assumptions about firm conduct with data on consumer behavior. My analysis follows this approach and uses two widely used data sets for the US consumer packaged goods retail industry that are provided by NielsenIQ in collaboration with the Kilts Center for Marketing Data Center at the University of Chicago. Both data sets provide information about the consumers' purchasing decisions in a large variety of product categories.[7] The product categories cover both food and non-food products, such as ready-to-eat cereals, shampoo, and bottled water. The difference between the data sets stems from the source from which the data originates.

The first data set—the so-called Retailer Panel—is directly reported by a large set of US retailers. Each retailer provides weekly sales information for its stores. The sales are reported at the level of bar codes where a bar code is defined by the Universal Product Code (UPC). The sales information is complemented with additional information about store, retailer, and product characteristics. The retailers can be categorized into different

[7]NielsenIQ distinguishes between three different product group classifications. I follow Döpper et al. (2023) and use the so-called product modules as an approximation for the product markets. I will refer to these product markets as (product) categories.

retail channels. For this analysis, I restrict my attention to food stores, mass merchandisers, and drug stores[8].

The second data set is the so-called Consumer Panel and contains information about shopping trips of individual households. The households in the sample participate in a program operated by NielsenIQ and report the data themselves. The information about the purchase of a product is complemented by additional information about household, shopping trip, and product characteristics. The households can be reweighted so that they are representative of the US population with respect to a number of observable demographic characteristics.

The different data sources come with different advantages and disadvantages. The Retailer Panel is a useful starting point for my analysis since it covers a large portion of the total household spending in the industry. It does, however, not contain information about the relationship between household characteristics and purchasing decisions because sales are aggregated at the store level. Therefore, it seems reasonable to complement the Retailer Panel with the Consumer Panel if information about individual factors is required.

Döpper et al. (2023) use this strategy to analyze the evolution of market power in the US consumer packaged goods retail industry. To this end, they estimate BLP-style demand systems for 133 product categories between 2006 and 2019. I follow their approach in the sense that I perform the same steps to process the raw NielsenIQ data sets and adopt their estimation strategy to gain insights beyond what can be learned from directly observable measures. More precisely, their approach allows me to recover a measure for marginal costs and a metric to quantify the impact of product characteristics other than the price on consumers' decision-making.

Döpper et al. (2023) use the Retailer Panel to calculate product-level market shares across different regions[9] and retail outlets. Their analysis focuses on 133 product categories, and in each product category, they aggregate the data along three dimensions. First, they choose a different product definition than the UPCs and aggregate sales to the brand level. The reason is that UPCs are often very narrowly defined and do not correspond to what the consumer perceives as a product.[10] For instance, there can be different package sizes of a product and each package size can have a different UPC. This leads to a large number of UPCs, which makes it difficult to infer cross-substitution patterns in practice. Aggregating to the brand level substantially lowers the number of products in a category and allows to circumvent problems related to the large number of alternatives. Döpper et al. (2023) further restrict their attention to the 20 top-selling brands and consolidate the remaining brands into a fringe brand. Second, they aggregate the sales across multiple stores of a retailer in a region. This allows to reduce the likelihood of zero market shares (as discussed

[8]These are the retail channels for which NielsenIQ provides good coverage for all years. The other product channels that I exclude are dollar, club, convenience, and liquor stores.

[9]Döpper et al. (2023) focus on the 22 largest Designated Market Areas, which are coherent areas defined by Nielsen based on media markets. The idea is that consumers in each region are exposed to the same marketing campaigns because they are served by (almost) the same newspapers and TV and radio stations.

[10]Döpper et al. (2023) provide a list of examples of what brand names look like in the dataset (for instance, see their footnote 12).

in, among others, Dubé et al., 2021 and Gandhi et al., 2023) and to keep the data set at a manageable size. Third, they consolidate weekly sales into quarterly sales. This also serves the purpose of a lower likelihood of zero market shares and, in addition, allows to better account for potential concerns arising from the stockpiling behavior of households (as discussed in, among others, Hendel and Nevo, 2006). The three aggregation steps lead to a data set where, for each product category and year, one observation is provided for each brand sold at a retailer in a region in a quarter. For my reduced-form regressions, I further aggregate the data across regions and quarters so that I have brand-retailer-specific metrics for each category and year.

This consolidated data set lacks a link between household characteristics and consumers' purchasing decisions. However, this link is important to account for heterogeneity in consumers' responses to price differences and changes, thereby preventing the estimation of rich substitution patterns. Döpper et al. (2023) use the Consumer Panel in two ways to add this heterogeneity component to the data. First, they calculate the annual distribution of household characteristics at the regional level. In doing so, they restrict their attention to two characteristics, namely the household income and a variable indicating whether a household has children or not. Second, they calculate so-called micro-moments that are used to capture heterogeneity in the target audience of the brands. A micro-moment corresponds to the average characteristic of a consumer buying a certain brand. Döpper et al. (2023) calculate micro-moments for all brands in all product categories and allow them to vary across regions and time.

Finally, three other data sets complement the NielsenIQ data. First, Capital IQ provides a snapshot of ownership information that allows to link brands to manufacturers. Based on this, the Zephyr merger database allows to identify mergers in the sample and keep track of changes in ownership over time.[11] Finally, Döpper et al. (2023) use a Consumer Price Index (CPI) to deflate all monetary measures (like prices). The CPI[12] used in the analysis excludes most of the product categories in the sample so that changes in monetary measures can be interpreted (roughly) as relative to changes in the prices of other goods in the economy.

4.3 Cross-Category Activities and Mergers

4.3.1 Definitions

Cross-category activities of firms are at the core of my analysis. To avoid any confusion about what I mean by cross-category activities or cross-category mergers, I introduce three definitions. I start with a terminology that describes the activities of firms in two or more

[11]The compilation of ownership information in Döpper et al. (2023) is not ideal for my analysis because I do not have information about the owners of the brands that are collapsed into the fringe brand. In addition, information about changes in ownership is available only at the annual level but not at the quarterly level. I am currently working on more detailed ownership information so that this problem is likely to be fixed in future versions of this paper.

[12]The CPI used in the analysis is the "Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average". See `https://fred.stlouisfed.org/series/CPILFESL` (last accessed on October 2, 2023) for details.

product categories.

Definition 1. *A firm is said to be "active cross-category" if its products belong to more than one product category.*

The primary objective of this definition is to describe the activities of manufacturers. The reason is that—as I will demonstrate below—there is large heterogeneity in the firms' product assortment on the manufacturers' side. The definition is, however, also applicable to retailers. All retailers are active in a large variety of product categories and can thus be considered as being active cross-category.

Next, I turn the focus to mergers.

Definition 2. *If two or more firms merge, the merger is said to be a "cross-category merger" if at least one product category exists in which only one of the merging parties was active prior to the merger.*

My analysis solely focuses on mergers of manufacturers. Based on the definition, I can attach a label to each merger indicating whether it is cross-category. This may, however, not be informative about how a merger affects a single product category. Think of two firms, with firm 1 being active in the categories A and B and firm 2 being active in the categories A and C. Both firms are active in more than one product category, therefore they are active cross-category according to Definition 1. In addition, a merger between these two firms would be called a cross-category merger. The reason is that there is only one merging firm active in each of the categories B and C prior to the merger. Although the merger would be a cross-category merger in my terminology, the merger may also generate effects by reducing competition in some product categories where the assortment of the merging parties overlapped before the merger. In my example, this would be category A. To better describe the impact of a merger on a particular product category, I introduce the following definition.

Definition 3. *A product category is affected by a merger if at least one of the merging parties was active in the category prior to the merger. If only one merging party was active in the category before the merger, I say that the category was affected "cross-category." Otherwise, I say that the category is affected "horizontally."*

The term "horizontally" refers to the terminology of a "horizontal merger" and is frequently used in the literature to describe a merger between two or more firms in the same market, which leads to a reduction in the number of competitors and, thus, usually also in competition. As I will discuss later, the set of cross-category mergers used in my analysis contains some cross-category mergers that also affect categories *horizontally*, but the number of categories is rather small. Therefore, I will simply exclude these merger-category combinations in my analysis and focus on the remaining categories without horizontal effects.

Table 4.1: Cross-Category Activities

Panel A: Manufacturers

Number of Firms		Number of Categories per Firm							Sales Share of Firms
Total	Cross-Category	Mean	Median	75% P.	90% P.	Largest 3			Active in 4+ Categories
743	223	2.14	1	2	4	29	31	40	44.15

Panel B: Retailers

Number of Firms		Number of Categories per Firm				
Total	Cross-Category	Mean	10% P.	25% P.	Median	75% P.
101	101	126.68	125	129	132	133

4.3.2 Cross-Category Activities

Although the focus of my analysis is on cross-category mergers, it seems reasonable to establish some facts about cross-category activities first. Panel A of Table 4.1 provides some basic statistics about the activities of the manufacturers. It shows that out of the 743 firms in my sample, around 70% are active in only one product category. In other words, the average (median) firm is not active cross-category.

This is also visible from the distribution of the number of categories per firm, which shows a median of one. The number increases only slightly to 2 and 4 at the 75^{th} and 90^{th} percentile, respectively. Given that I have 133 categories in my sample, these numbers can be considered small. This highlights that the remaining 223 cross-category firms are typically active in a small number of categories.

The large majority of firms that are active in only a few categories is accompanied by a small set of large firms. These firms can be of substantial size. For example, the product assortment of the three largest firms spans 29, 31, and 40 categories. Although these numbers are very large compared to the percentiles listed in Panel A of Table 4.1, it is important to keep in mind that they represent only 22%, 23%, and 30% of the universe of categories in my sample.

Contrary to the manufacturers and as visible from Panel B of Table 4.1, the retailers' assortments typically cover a large portion of the categories in my sample. As mentioned earlier, all 101 retailers are active cross-category. The median retailer covers all categories except one, and even the 10^{th} percentile of the distribution of the number of categories per retailer is 125, which represents almost 94% of the categories in my sample.

From an economic perspective, the two panels of Table 4.1 stress the importance of examining to what extent cross-category effects (and thus portfolio effects) play a role in bargaining and how they shape the relationship between manufacturers and retailers. For instance, if cross-category effects are absent, bargaining outcomes solely depend on the market positions of the firms in a given category (like their market size or their brand valuations). In other words, if a manufacturer is active in a single category and holds a strong market position, it will also have a strong bargaining leverage over the retailers. If, in contrast, cross-category effects are extremely important, the bargaining leverage of such a manufacturer can be expected to be almost negligible. Even the biggest manufacturer in my sample would have a rather weak bargaining position because it is active in "only" around 30% of the categories.

Panel A of Table 4.1 is useful to get a first impression of the cross-category activities of the manufacturers. It is, however, not per se informative about how these activities look like within categories. It could, for instance, be the case that most of the cross-category firms cluster in a small number of categories while other categories are almost unaffected by cross-category firms. The purpose of Table 4.2 is to show that this is indeed not the case and that cross-category activities are a widespread phenomenon.

Table 4.2: Cross-Category Activities of Manufacturers by Product Category

		Number of Firms per Year		Share of Revenues		
Rank	Product Category	Total	Cross-Category	Top 20 Brands	Cross-Category Firms	Private Labels
1	Cereal - Ready to Eat	6	4	0.56	0.48	0.08
2	Candy - Chocolate	7	4	0.52	0.42	0.03
3	Candy - Non-Chocolate	12	5	0.57	0.35	0.09
4	Deodorants - Personal	8	6	0.79	0.78	0.00
5	Soap - Specialty	10	6	0.69	0.61	0.05
6	Tooth Cleaners	5	4	0.74	0.74	0.00
7	Shampoo - Liquid/Powder	9	5	0.60	0.53	0.03
8	Cookies	8	5	0.63	0.46	0.16
9	Sanitary Napkins	5	3	0.75	0.62	0.13
10	Cold Remedies - Adult	10	6	0.88	0.45	0.28
20	Bottled Water	10	7	0.88	0.65	0.22
40	Baby Formula	5	2	0.80	0.37	0.04
60	Nuts - Bags	17	10	0.86	0.42	0.32
80	Fresh Muffins	13	7	0.92	0.71	0.19
100	Tuna - Shelf Stable	14	7	0.99	0.85	0.11
120	Cream - Refrigerated	13	10	0.92	0.46	0.45
130	Frozen Poultry	15	6	0.93	0.34	0.51
133	Fresh Mushrooms	17	2	0.96	0.02	0.44
	Mean Values	12	6	0.85	0.59	0.16

Table 4.2 presents information about the cross-category activities for a subset of categories. The selection of categories is taken from Table 1 in Döpper et al. (2023), and categories are sorted by the number of observations. The value in the first column is the rank resulting from this sorting exercise. The first group of categories (up to the horizontal rule) contains the ten largest categories, while the second part includes a subset of the remaining categories. The last row shows statistics for the average category.

I first focus on the number of firms that are active in the category (column 3) and the corresponding number of cross-category firms (column 4). As indicated in the last row, half of the firms are active cross-category in the average category. Across categories, this ratio varies substantially, but the number of cross-category firms is usually well above zero. Notable exceptions exist in the categories "Baby Formula" and "Fresh mushrooms," where only two cross-category firms are active. In the first case, this is not surprising given that only five firms are active in total, while in the latter case, the category seems indeed to be less affected by cross-category activities.

The number of firms gives a first impression about the activities of cross-category firms within categories, but it may hide important information because this measure treats all firms equally. An alternative would be to look at the share of revenues that is captured by the brands of cross-category firms (column 6). By construction, these brands are a subset of the leading 20 brands in each category; hence, I also report the revenue share of these 20 brands as a benchmark (column 5). The table shows that the 20 brands account for 85 percent of the revenues in the average category. The subset of brands owned by cross-category firms accounts for about 70 percent of this share and almost 60 percent of the total revenues. This means that cross-category firms tend to be large not only because they offer products in multiple categories but also because their market coverage within a given category is large.

It is worth noting that the revenue share of the leading 20 brands may also include the sales of private labels (see column 7 for the corresponding revenue share). Although private label products are typically treated as being produced by the retailers in IO models, and retailers are cross-category firms, I do not treat private label products as products sold by cross-category firms in my analysis. If I included the 16 percent that private labels account for in the average category, around 88 percent of the revenue share of the leading 20 brands would be associated with cross-category firms.

Finally, Table 4.2 shows that across all categories, the revenue share of brands sold by cross-category firms almost never drops below one-third and is often substantially larger. A notable exception is the category "Fresh Mushrooms" where cross-category firms account for only 2 percent of the revenues. This is consistent with the initial inspection based on the number of firms.

4.3.3 Cross-Category Mergers

The previous subsection shows that cross-category activities of both manufacturers and retailers are a widespread phenomenon in the US consumer packaged goods retail industry. In this section, I will explore the effects of cross-category mergers on directly observable market outcomes such as revenues, quantities, and prices. I start my exploration by looking at some statistics that describe the mergers in my data.

Panel A of Table 4.3 shows that out of the 139 mergers in my sample, 95 can be classified as cross-category mergers. Among these cross-category mergers, 57 mergers are suitable for my analysis. I will refer to them as the baseline sample. The difference between the total number of cross-category mergers and the baseline sample is mostly driven by missing data on the acquirer side. In 32 cases, the acquirers are not active in any category in my sample. These mergers still constitute some form of cross-category mergers since the acquirers are not active in the same product categories as the targets, and hence, these mergers do not reduce competition in these categories. However, since my analysis requires information about both merging parties, I restrict my attention to the baseline sample.[13]

[13] There might also be different reasons why firms merge. If an acquirer is not active in any of the 133 categories in my sample, I cannot be sure that this firm is active in the consumer packaged goods retail

Table 4.3: Overview of Cross-Category Mergers

Panel A: Number of Mergers

All Mergers	Cross-Category Mergers	
Total	Total	Baseline
139	95	57

Panel B: Characteristics of Cross-Category Mergers

	All Mergers	Per Merger			
	Unique Values	Mean	25% Q.	50% Q.	75% Q.
Targets	55	-	-	-	-
Acquirers	36	-	-	-	-
Categories (Cross-Category)	115	9.56	3	8	13
Categories (Cross-Category, Target)	68	1.88	1	1	2
Categories (Cross-Category, Acquirer)	105	7.68	2	5	12
Categories (Horizontal)	27	0.47	0	0	1
Brands (Target)	140	2.74	1	2	3
Brands (Acquirer)	643	19.91	5	10	22
Total Sales (Target)	-	34.13	3.24	12.35	31.61
Total Sales (Acquirer)	-	325.66	36.84	97.60	423.22
Avg. Sales Share in Category (Target)	-	0.05	0.01	0.02	0.05
Avg. Sales Share in Category (Acquirer)	-	0.09	0.03	0.06	0.12

Panel B of Table 4.3 provides an overview of the characteristics of the mergers. The second column lists the number of unique values for a variety of characteristics across all baseline mergers. The remaining columns describe the distribution of these characteristics across mergers. Starting with column 2, it shows that the 57 cross-category mergers involve 55 unique targets. This means that almost all targets were bought only once during the 14 years of my sample period. In contrast, the set of acquirers is smaller and consists of only 36 firms, showing that some manufacturers acquire multiple targets. In fact, there are 14 acquirers that conduct more than one acquisition, and the most active acquirer buys four targets over the 14 years of my sample period.

Cross-category mergers can affect market outcomes in categories on both the target and the acquirer side. In total, 115 categories are affected by at least one merger on either side, with 68 categories being affected at least once on the target side and 105 categories at least once on the acquirer side. This suggests that the acquirers are active in more categories than the targets. This is also visible from the distribution of the number of categories per merger (rows 3 to 6). While the average merger affects about 9.5 categories, only about two categories are affected on the target side, and the remaining approximately 7.5 categories are affected on the acquirer side. This pattern also holds true for the three other percentiles reported in the table.

industry at all. For instance, the acquirer could also be a private equity firm.

Row 6 shows that some mergers also affect categories horizontally, that is, both firms are active in these categories before the merger. However, the number of categories is rather small. While there are 27 categories being affected horizontally in total, the average (median) merger shows no overlap in product markets, and even the 75^{th} percentile is only 1.

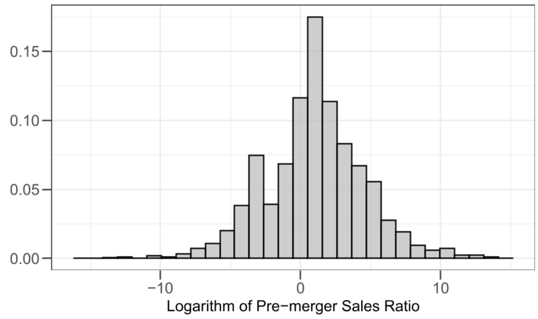
The fact that acquirers operate in many more categories than the targets could also mean that the targets are much smaller than the acquirers. There is, however, a caveat to this idea: the activity of a firm in a category is, per se, not informative about its success within this category. It could, for instance, be the case that an acquirer is active in many more categories but that its brands target niche consumer segments and realize only small market shares, while the target is highly specialized in a single category but is able to capture a large market share. The remaining rows of the table are used to reject this alternative explanation and to show that the acquirers are indeed much larger than the targets. The rows present three different measures to better capture the full extent of what can be described as "being larger": the total number of brands, the total revenues, and the average revenue share in a category. All three measures point to the fact that acquirers are larger. For instance, the average acquirer has more brands (about 20 vs. 3), realizes larger revenues (about 325 vs. 35 million USD), and captures a larger revenue share within a category (about 9 vs. 5 percent). This pattern does not only hold for the average merger but remains valid when looking at an alternative measure for the average (median instead of mean) and different percentiles of the distribution (25^{th} and 75^{th}).

The asymmetry between targets and acquirers provides further guidance for how I can carry out the analysis of merger effects. Recall the idea of the portfolio power theory that a cross-category merger can benefit the merging parties through an improvement in their bargaining position. If the merging firms are highly asymmetric, the shift in the bargaining position is likely larger for the smaller firm because this firm generated only small revenues before the merger and thus was highly dispensable for the retailers. In contrast, the larger merging party generated large revenues already prior to the merger, and its size increased only marginally through the merger. Therefore, the importance of its assortment does not change a lot from a retailer's perspective. In conclusion, this means that my analysis should be primarily concerned with the effects of mergers on the outcomes of the smaller firms, that is the targets. In addition, I will use the fact that firms' activities vary substantially across retailers.

Consider a brand j belonging to a target. I use $f_j(\tau)$ to denote the ownership of brand j at point τ. The time variable τ is measured in event time; that is, it takes the value 0 in the year of the merger. Thus, $f_j(-1)$ and $f_j(0)$ refer to the independent target before the merger and the acquirer after the merger. The key metric of my analysis is the ratio of the revenues of the acquirer relative to those of the target in the year before the merger.

$$\log\left(\frac{total\ sales_{f_j(0),c,-1}}{total\ sales_{f_j(-1),c,-1}}\right) \tag{4.1}$$

Figure 4.1: Histogram of the Pre-Merger Sales Ratio



The indices $f_j(\tau)$ and $f_j(0)$ refer to target and acquirer, while the additional index -1 refers to the last pre-merger period (in event time). The remaining index c denotes the retail chain at which the revenues are generated. Note that the index j belongs only to the ownership variable f_j, but does not enter the total revenues as an additional subscript. This means that the revenues refer to the total revenues of the corresponding firm at retailer c and time -1 and not just those of brand j.

Figure 4.1 shows a histogram of the pre-merger revenue ratios (in logarithm). I keep the observations at the target-acquirer-retailer level, leaving aside that my analysis will take place at the brand level. The figure shows that most of the observations have a positive value. In fact, the 33^{rd} percentile is about -0.01, indicating that about two-thirds of the observations are positive. The mean (1.03) and median (1.07) are both very similar and close to 1, supporting the fact that the distribution looks rather symmetric.

One important observation from Figure 4.1 is that while most of the observations are positive, there is still a substantial fraction that is negative (about one-third). This is one of the two reasons that motivates the use of the logarithm in Expression (4.1) and the subsequent analysis. If the logarithm is negative, the ratio of the revenues must be smaller than one, meaning that the target's revenues at retailer c exceed those of the acquirer. The use of the logarithm allows for opposing effects; that is, the effect is negative if the target's revenues are larger and positive if the target's revenues are smaller. While I impose this relationship by assumption, it is supported by my analysis later (see details below). Another advantage of the logarithm is that it alters the interpretation of the regression coefficients in a meaningful way, allowing me to consider percentage changes in the ratio rather than level changes.

With measure (4.1) in hand, I can now state the main specification.

$$X_{jc\tau} = \alpha_{jc} + \gamma_{year(\tau)} + \sum_{\ell \in [\underline{\tau}, \overline{\tau}]} \left[\beta_{1\ell} + \beta_{2\ell} \cdot \log \left(\frac{total\ sales_{f_j(-1),c,-1}}{total\ sales_{f_j(0),c,-1}} \right) \right] \cdot D(\tau = \ell) + \varepsilon_{jc\tau} \tag{4.2}$$

The variable X_{jct} will be the outcome of interest. For now, this will be the logarithm of either the revenues, quantities, or prices. The indices show the level of aggregation. Observations are at the brand-retailer-event time level, which means that, as noted in Section 4.2, I abstract from regional differences and only exploit the variation across retailers.

As it is common practice in the event study literature, the specification follows a two-way fixed effects design where α_{jc} captures the brand-retailer fixed effects and $\gamma_{year(t)}$ the year fixed effects. The sum operator loops over all event time periods in a time window from 5 years before to 5 years after the merger. Observations before and after this time window are collapsed into two additional bin categories ($\tau < -5$ and $\tau > 5$).[14] The variable $D(\cdot)$ is a dummy variable taking value 1 if the condition in brackets is satisfied and 0 if not. This means the variable takes value 1 if the observation belongs to the event time period ℓ. The coefficients $\beta_{1\ell}$ and $\beta_{2\ell}$ are supposed to capture the effects of the merger. My sample will include only the brands of the targets. Hence, the idea is to compare treated to not-yet-treated brands, assuming that after accounting for the fixed effects and in absence of the treatment, their developments would follow similar trends. The coefficients $\beta_{1\ell}$ capture effects that are common to all brand-retailer combinations, while the coefficients $\beta_{2\ell}$ capture additional effects arising from pre-merger differences in the bargaining positions of the target and the acquirer at retailer c. In practice, the $\beta_{1\ell}$ coefficients turn out to be usually insignificant and close to zero, so I will treat them as an additional set of fixed effects.[15] The $\beta_{2\ell}$ coefficients in the pre-merger periods are supposed to be 0, with the coefficient in the last pre-merger period ($\ell = -1$) being normalized to 0.

The event study literature is currently undergoing significant developments. With this in mind, I will interpret the estimated $\beta_{2\ell}$ coefficients as correlations for now, which may give a first impression of possible cross-category effects. I will discuss a more sophisticated approach based on the recent event study literature and the underlying assumptions later. However, the results will be broadly consistent with the patterns documented based on the initial inspection of the correlations.

Figure 4.2 visualizes the results when I estimate the main specification with Ordinary Least Squares.[16] It shows the $\beta_{2\ell}$ coefficients for the different time periods. The vertical bars show the 95% confidence intervals, with standard errors being clustered at the merger level. The three panels refer to the three different measures of interest. It is clearly visible that the estimated coefficients in the pre-merger time periods are close to zero, independently of the measure under consideration. This provides some support for the idea that treated and not-yet-treated brands undergo similar developments before the merger and after accounting for the fixed effects.

The estimated coefficients for the merger periods are similar for revenues (Panel (a)) and quantities (Panel (b)). This refers to both the direction and the magnitude of the estimated coefficients.[17] In contrast, the coefficients of the prices are very close to zero. Although

[14] I omit the estimates for the bin categories in the following figures. The estimates usually fit the patterns shown in this paper. For my main Figure 4.2, I also report the estimates in the table in Appendix 4.A.

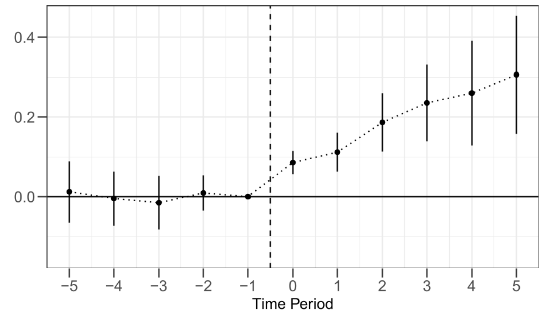
[15] Appendix 4.D compares the results of Figure 4.2 (black) to a version where I omit the $\beta_{1\ell}$ coefficients in the specification (gray). The results are similar.

[16] Appendix 4.A provides the corresponding table with the estimates.

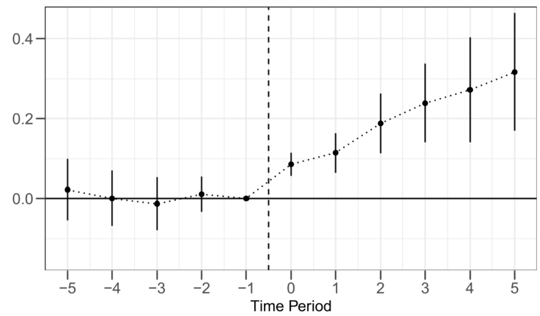
[17] Appendix 4.B shows the result of the exercise when I use revenue and quantity shares of the brands within the product categories as a dependent variable. The results are similar. The result also remains intact if I consider only a balanced panel with a three-year time window. This is visible from Appendix 4.C that compares the outcomes of Figure 4.2 (black) with those of a balanced panel (gray).

Figure 4.2: Changes in Sales, Quantities and Prices

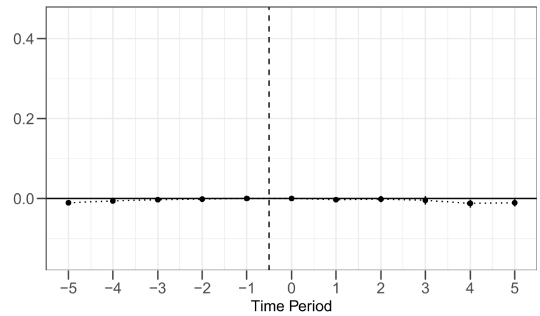
(a) Sales



(b) Quantities



(c) Prices



some of the coefficients are, in fact, statistically significantly different from 0, they can be considered economically negligible.

The post-merger coefficients for revenues and quantities are positive and increasing over time. Since both the dependent variable and the pre-merger revenue ratio are in logarithms, the interpretation of the coefficients relates to changes in percent. If the pre-merger revenue ratio increases by 1%, ceteris paribus, the revenues (and quantities) of the target's brands at the corresponding retailer increase by about 0.1% in the first year after the merger. This effect increases to about 0.3% 5 years after the merger.

To get a sense of the total effect size, consider the 25^{th}, 50^{th}, and 75^{th} percentile of the distribution of the logarithms of the pre-merger revenue ratios as depicted in Figure 4.1, which are -0.92, 1.07, and 3.07, respectively. The benchmark is a target that has the same pre-merger revenues at a retailer as the acquirer so that the logarithm of the pre-merger revenues ratio is 0 and there are no cross-category effects. Compared to this benchmark, the cross-category effects lead to a change in revenues (quantities) by roughly -25%, 39%, and 156% (-25%, 40%, and 165%), ceteris paribus, when evaluated at the three percentiles and at the point estimate.[18] These values indicate that the cross-category effects can reach

[18] The formula to calculate the effect size is $100 \cdot \left(\exp\left(\widehat{\beta}_{25} \cdot \text{ratio}_1\right) - 1\right)$, where $\widehat{\beta}_{25}$ is the point estimate, and ratio_1 is the logarithm of the pre-merger revenue ratio. To derive this formula, let $i = 1$ refer to the case where the logarithm of the pre-merger revenue ratio is given by one of the percentiles listed in the text, and $i = 0$ denotes the benchmark case where the logarithm of the ratio is 0. Let ratio_i denote the logarithm of the pre-merger revenue ratio and y_i the variable of interest. The formula results from the following consideration $\log(y_1/y_0) = \log(y_1) - \log(y_0) = \beta_{25} \cdot \text{ratio}_1 - \beta_{25} \cdot \text{ratio}_0 = \beta_{25} \cdot \text{ratio}_1$ and thus $y_1/y_0 - 1 = \exp(\beta_{25} \cdot \text{ratio}_1) - 1$.

considerable magnitudes.

So far, and as noted above, the analysis documents correlations. The next step is to provide additional evidence that the effects are also causal. To this end, I make use of recent developments in the event study literature. I start my investigation of a potentially causal relationship by discussing the assumptions that would be required for the above analysis to reveal a causal relationship.[19] With the knowledge of which assumptions are unlikely to hold, I can then look for an alternative approach.

I begin with three assumptions that I deem to be unproblematic. The first assumption is the "stable unit treatment value assumption" (SUTVA), which says that the treatment status of one firm does not impact the market outcomes of another firm's brands. In particular, there are no spillovers across firms in my sample. This assumption may seem critical at first glance because mergers clearly impact all firms in the markets in which the merging firms are active. However, my analysis compares treated to not-yet-treated firms. In particular, I do not follow other papers studying mergers (like, for instance, Ashenfelter and Hosken, 2010) and do not use competing brands or private labels for the comparison. This means that the firms in my sample are usually active in different product categories, so spillovers are unlikely to occur.

The second assumption relates to the absence of anticipation effects and says that a merger is not allowed to affect the market outcomes of the merging firms' brands before the merger. In general, the announcement of a merger does not automatically mean that the merger will be carried out in the future. There are various reasons why a proposed merger may be canceled at a later date. For instance, the merger itself requires negotiations between the owners of the target and the acquirer, and they may fail to reach an agreement. Another possibility is that a due diligence conducted after the merger announcement uncovers problems with the target that the acquirer did not anticipate. Because of all these uncertainties in the period between the merger announcement and the final acquisition, it is unlikely that the retailers will start offering better deals to the targets and/or the acquirers before the merger actually takes place. In this context, it is worth noting that in the consumer packaged goods retail industry, firms often negotiate annually, with some smaller negotiations occurring during the year (for instance, to coordinate the joint marketing and sales effort; see, for instance, Anderson and Fox, 2019 on the planning of trade promotions). This means that the retailers make commitments for a relatively long period and may be less willing to respond to rumors in negotiations.

Another line of reasoning is to think of a counterfactual world in which there were anticipation effects. In this case, anticipation effects would probably be relevant for at most one or two years. Figure 4.2 shows coefficients covering up to 5 years before the merger, which means that I would expect to see pre-trends. However, since my analysis does not provide any evidence for pre-trends, it renders anticipation effects unlikely.

[19]The following discussion draws primarily on the survey of Roth et al. (2023). However, due to the fast progress in the event study literature and the importance of this literature for many fields in economics, there are many other good surveys available. Another notable one is de Chaisemartin and D'Haultfœuille (2023).

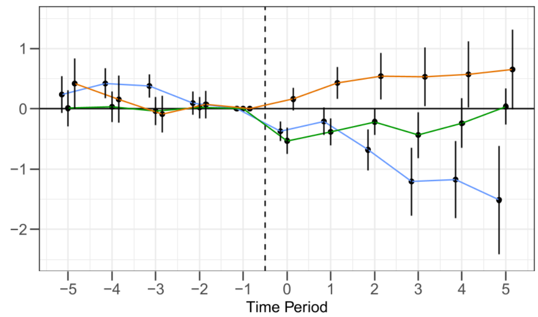
The third assumption is the existence of parallel trends. Depending on the event study approach used, this assumption comes in different forms, but the general idea underlying this assumption is usually similar. Measuring the average treatment effect on the treated requires comparing the outcome of a treated individual to its untreated counterfactual. This poses a problem since it is obviously not possible to observe an individual in both states (that is, being treated and being untreated) at the same time. Therefore, the empirical strategy is to find an appropriate counterfactual scenario. A simple before-after comparison (that is, a change within an individual over time) would not be suitable since other variables than the treatment status may change, and these variables, when not appropriately controlled for, can introduce a bias. Therefore, the literature usually exploits the presence of untreated or not-yet-treated individuals, with the idea that their outcomes follow a similar development except for the impact of the treatment.

While the idea to compare treated to untreated individuals was originally developed for treatments that occur for all treated individuals at the same time, the literature has also applied this approach to so-called staggered adoptions where the individuals are treated at different points in time. The recent event study literature (in particular Goodman-Bacon, 2021) shows that this has previously unexpected consequences in the sense that researchers may compare groups of individuals to each other that they did not intend to compare. More specifically, treated individuals are compared to other individuals who have been treated earlier. These comparisons are often referred to as forbidden comparisons. These comparisons can lead to a bias if the treatment effects are heterogeneous across cohorts, with a cohort being all firms that are treated in a particular year. This bias can even be strong enough to turn around the sign of an estimate for the average treatment effect on the treated and thus causes serious concerns. In the context of my analysis, the assumption that treatment effects are homogeneous across cohorts is difficult to maintain. In particular, my sample period from 2006 to 2019 covers the financial crisis and the subsequent recovery phase, so mergers of different cohorts also experienced different macroeconomic environments.

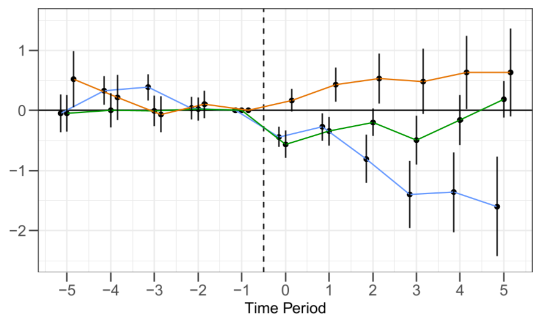
The above considerations are typically discussed in the context of binary treatments. Callaway et al. (2021) highlight that continuous treatments—like in my analysis—further complicate the analysis of causal effects. For instance, they require additional assumptions and stricter versions of some of the previously mentioned assumptions. To simplify my analysis, I convert my continuous measure into a categorical variable. More precisely, I use the logarithms of the pre-merger revenue ratios at the different retailers (as depicted in Figure 4.1) to split my sample into three groups. I use the 33^{rd} and 66^{th} percentiles (-0.01 and 2.22, respectively) as boundaries. I then apply the estimator of Callaway and Sant'Anna (2021) to each category, treating the treatment variable as binary and ignoring potential variation in the treatment intensity in each subsample. The idea of the approach by Callaway and Sant'Anna (2021) is to perform separate estimations of the treatment effects for each cohort. In each year, the firms that are treated in that year are compared only to those who have not yet been treated and will not receive treatment in the time period

Figure 4.3: Changes in Sales, Quantities and Prices

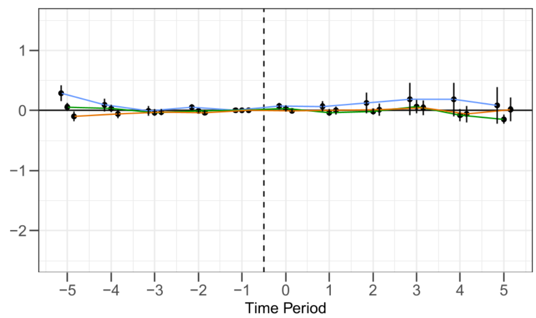
(a) Sales



(b) Quantities



(c) Prices



used for the comparison. The average treatment effect on the treated is then calculated by a weighted average. Apart from the fact that I use a binary instead of a continuous treatment, another caveat is that the number of mergers, which I previously used as clusters for the standard errors, is already quite small in general, and the number in each subsample is even smaller so that I cluster the standard errors only at the brand level.

Figure 4.3 shows the result of this exercise. The three lines refer to the different subsamples, with the blue line referring to ratios below the first tertile, the green line to ratios between the first and second tertile, and the orange line to ratios above the second tertile. Panels (a) and (b) show the results for revenues and quantities. Although some pre-merger coefficients are statistically different from 0 at the 5% level, the figures do not show any meaningful pre-trends in general. In contrast, the post-merger coefficients show clear trends that fit the results of my previous analysis. Since the first tertile is roughly 0, the blue line refers to almost all brands for which the targets have larger revenues than the acquirers at a retailer. For these brands, the effect size is negative, and the decrease in revenues 5 years after the merger is roughly -78%. In contrast, targets whose pre-merger revenues are strongly smaller than those of the acquirers at a retailer experience a strongly positive effect (orange line). While the effect is positive in all years, the effect is statistically different from 0 at the 5% level in only some years. Evaluated at the point estimate, the revenues increase by about 92% 5 years after the merger. Finally, if the pre-merger revenues of target and acquirer do not diverge too strongly (green line), there seems to be a slightly negative effect immediately after the merger, but no effect (neither positive nor negative) is visible 5 years after the merger. With respect to changes in prices (Panel (c)), the coefficients are mostly not statistically different from 0 at the 5% level, and even those that are statistically significant are economically negligible in magnitude. Overall, Figure 4.3 shows similar patterns to those in Figure 4.2, providing evidence that the previously documented patterns are indeed causal.

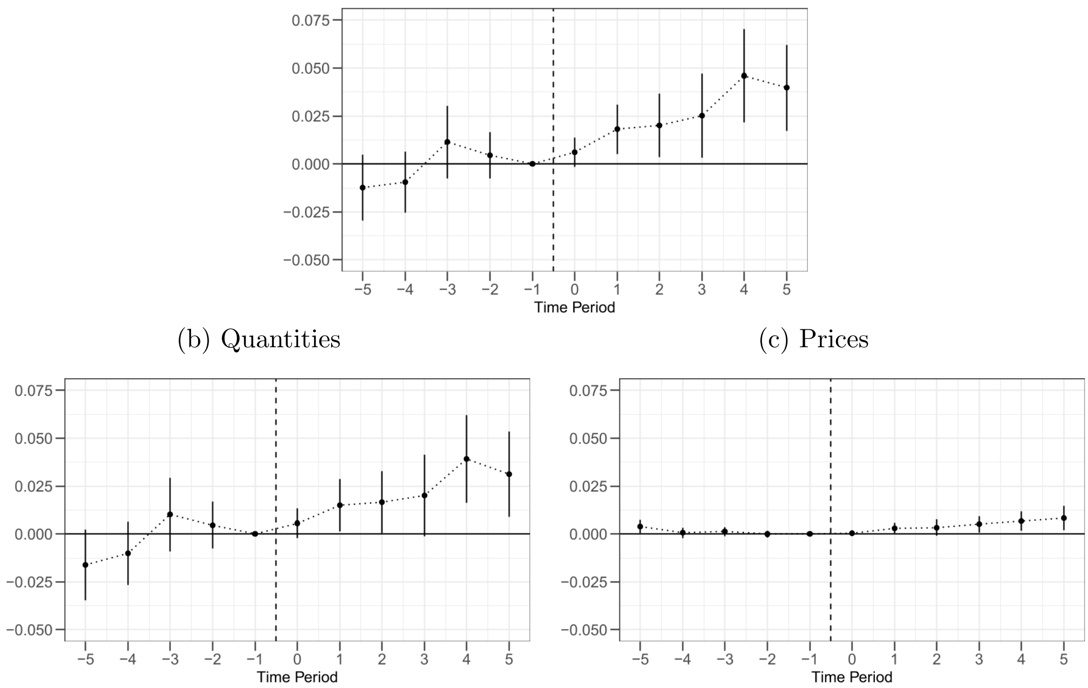
Note that Figure 4.3 also motivates the application of the logarithm to the pre-merger revenue ratios. This is because the direction of the effect takes different signs depending on whether the ratios are above or below 1.

Figure 4.4: Changes in Sales, Quantities and Prices (Acquirers)

(a) Sales

(b) Quantities

(c) Prices



So far, my investigation of the effects of cross-category mergers has focused on the targets. At the end of this section, I will briefly document the effects on the acquirers. To this end, I estimate a modified version of my baseline specification 4.2. Since the underlying data now refers to the acquirers' brands, it seems reasonable to adjust the ratio measure and to use the logarithm of the *inverse* ratio. That is, I consider the ratio of the targets' pre-merger revenues to those of the acquirers.

$$\log\left(\frac{total\ sales_{f_j(-1),c,-1}}{total\ sales_{f_j(0),c,-1}}\right) \tag{4.3}$$

Based on the analysis of the targets, the initial hypothesis is that an acquirer's revenues and quantities increase after the merger if the pre-merger revenues of the acquirer at a retailer are smaller than that of the target, that is if Expression 4.3 is positive.

Figure 4.4 is the analog to Figure 4.2 and shows the results. The bottom line is that the effects go in a similar direction but are of much smaller magnitude. Revenues and quantities increase after the merger, and most of the coefficients are significantly different from 0 at the 5% level. Interestingly, and in contrast to the analysis of the targets, the magnitude of the change in quantities is smaller and thus does not (approximately) match that of the change in revenues (especially in later years). This is because there also seems to be some price effect, although the magnitudes of the corresponding coefficients are still quite small and can be considered economically negligible.

4.4 Merger Effects on Marginal Costs and Perceived Quality

4.4.1 Models of Demand and Supply

So far, my analysis has only been concerned with the effect of cross-category mergers on directly observable market outcomes. In the next step, I use the work of Döpper et al. (2023) to shed some light on other measures that can be inferred from the data based on assumptions about demand and firm conduct. To this end, I will first briefly introduce their models of demand and supply and their empirical strategy. In the following, I adopt the notation and formulas from Döpper et al. (2023).

The demand side builds on the seminal work of Berry et al. (1995) and is a random coefficient Logit model. As before, let c denote the retail chain and r the region. The variable t denotes the quarter. A geographic market is defined as a region-retailer combination, which means that the approach abstracts from retailer competition.[20] Since the model will be estimated separately for each category and year, the combined index crt denotes the market level. In each market, consumers can choose between $0, \dots, J_{crt}$ products, where 0 denotes the outside option of not buying any of the products offered by the manufacturers.

Each consumer i is endowed with certain characteristics (like household demographics) and receives an (indirect) utility of u_{ijcrt} when buying product j. The utility of the outside option ($j = 0$) is normalized to zero. The utility of the other products is given by

$$u_{ijcrt} = \beta_i^* + \alpha_i^* \cdot p_{jcrt} + \xi_{jr} + \xi_{cr} + \xi_t + \Delta\xi_{jcrt} + \varepsilon_{ijcrt}, \tag{4.4}$$

where β_i^* is a consumer-specific constant, p_{jcrt} is the price of product j, α_i^* is a consumer-specific scalar, and ξ_{jr}, ξ_{cr}, and ξ_t are product-region, retailer-region, and quarter fixed effects. $\Delta\xi_{jcrt}$ and ε_{ijcrt} are error terms. The first term is typically called the "structural error term" and captures the reaction of the consumers to unobserved product characteristics, while the second term describes a random consumer-specific taste shock ("Logit" shock). The presence of the structural error term leads to an endogeneity problem when taking the model to the data since unobservable product characteristics might be correlated with observable characteristics like prices.

A key feature of the random coefficient demand model is that it allows for heterogeneity across consumers. Döpper et al. (2023) allow consumers to differ in three characteristics:

[20] Note that this does not mean that this approach completely rules out competition between retailers. Instead, the other retailers in the same region are part of the outside option.

an unobserved demographic which is standard normal distributed, (the logarithm of) the household income, and a variable indicating whether a household has children (= 1) or not (= 0). The consumer's utility depends on these characteristics through the consumer-specific parameters α_i^* and β_i^*. The consumer-specific constant β_i is allowed to vary in all three characteristics, while the consumer-specific reaction to prices α_i^* is only allowed to vary by the two observable characteristics. Formally, this can be expressed by

$$\alpha_i^* = \alpha + \Sigma_1 D_i \qquad \text{and} \qquad \beta_i^* = \beta + \Sigma_2 D_i + \sigma v_i, \tag{4.5}$$

where α and β are the mean parameters that are constant across consumers, D_i describes the observable household demographics, v_i is the unobserved demographic, and Σ_1, Σ_2, and σ describe the impact of these demographics (that is, the size of the consumer-specific deviations from the mean parameters).

Each consumer buys the product which yields the highest utility. Taking the random taste shock into account, the choice probability of consumer i buying product j is given by the typical Logit expression

$$s_{ijcrt} = \frac{\exp\left(u_{ijcrt}\right)}{\sum\limits_{k\in 0,\dots,J_{crt}} \exp\left(u_{ikcrt}\right)}. \tag{4.6}$$

By integrating over the distributions of consumer characteristics, I can derive the market share s_{jcrt} of product j in market crt. Finally, multiplying the market share with the market size M_{crt}[21] leads to the quantity q_{jcrt} sold of product j in market crt.

The demand model is combined with a supply side. Manufacturers are assumed to set prices to maximize (static) profits, i.e., they compete in static Bertrand competition with differentiated goods. Retailers use a cost-plus pricing strategy and place a constant markup on the prices of the manufacturer. Under this assumption, the retail markup becomes part of the manufacturers' marginal costs, and thus, the approach isolates the manufacturers' markups.

The first-order conditions of the manufacturers' maximization problem lead to the following decomposition of the price:

$$p_{crt} = c_{crt} - \left(\Omega_{crt} \circ \left[\frac{\partial s_{crt}\left(p_{crt}\right)}{\partial p_{crt}}\right]'\right)^{-1} s_{crt}\left(p_{crt}\right), \tag{4.7}$$

where p_{crt}, s_{crt}, and c_{crt} are vectors capturing prices, market shares, and marginal costs. Ω_{crt} is the ownership matrix with entries in $\{0,1\}$. If products j and k are owned by the same firm, the entries $[j,k]$ and $[k,j]$ take value 1, otherwise 0. The derivative in square brackets is a matrix that contains the derivative of each market share with respect to each price and thus provides information about the substitution patterns. Finally, $\circ$ denotes the element-wise matrix multiplication (Hadamard product).

[21] Defining the market size across a large number of product categories and years is a non-trivial challenge. See Döpper et al. (2023) for details.

Equation (4.7) decomposes the price into the marginal cost and the markup. The markup depends on observable variables (like market shares and ownership) and the substitution patterns that can be inferred from the demand side. Therefore, it is possible to calculate the markup for a given set of demand side parameters. Since prices are observed as well, Equation (4.7) can also be used to calculate the marginal costs directly.

Finally, as noted earlier, $\Delta\xi_{jcrt}$ can include unobserved product characteristics, which can lead to endogeneity problems. The identification strategy will require splitting the marginal costs into an observed and an unobserved component. To this end, the marginal costs are decomposed using product-region (η_{jr}), retailer-region (η_{cr}), and quarter (η_t) fixed effects. The remaining part ($\Delta\eta_{jcrt}$) denotes the unobserved cost shock.

$$c_{jcrt} = \eta_{jr} + \eta_{cr} + \eta_t + \Delta\eta_{jcrt} \tag{4.8}$$

4.4.2 Estimation and Identification

The aim is to estimate the unknown parameters of the demand model, which can be divided into two sets. The first set, denoted by Θ_1, contains the parameters α and β.[22] These parameters describe the mean values of α_i^* and β_i^* and can be used to calculate the so-called mean utility δ_{jcrt} by setting the remaining parameters that determine the consumer-specific deviations to zero. In contrast, the second set, denoted by Θ_2, contains all parameters that determine the impact of the consumer characteristics, that is, Σ_1, Σ_2, and σ. For each consumer i and product j, the difference between the utility u_{ijcrt} and the mean utility δ_{jcrt} describes the consumer-specific deviation in the utility space.

Döpper et al. (2023) use a modified version of the nested fixed point estimator of Berry et al. (1995) to estimate both sets of unknown parameters[23]. To understand these modifications, it is useful to consider the mechanics of the estimator first. Consider a given set of candidate parameters for Θ_1 and Θ_2. In the first step, the estimator derives the total utility levels for each product in each market and splits them into the mean utilities and the consumer-specific deviations. To do this, it requires only information about the market shares, the consumer characteristics, and the candidate parameters for Θ_2 (but not the candidate parameters for Θ_1). With these estimates in hand, it is then possible to use the candidate parameters from Θ_1 to further split the mean utility into its components and derive an estimate for the structural error term $\Delta\xi_{jcrt}$. This structural error term is then usually interacted with instrument variables. Most importantly, this procedure shows that the estimator uses the candidate parameters for Θ_1 and Θ_2 in two steps, where each step uses only one set (either Θ_1 or Θ_2).

Döpper et al. (2023) use this two-step structure to modify the estimation routine in the following way: In the first step, they use the micro-moments discussed in Section 4.2 to

[22] Technically, ξ_{jr}, ξ_{cr}, and ξ_t are also part of Θ_1. However, the fixed effects are typically not part of the parameters estimated with GMM, but are estimated with Ordinary Least Squares for given candidates for the other parameters.

[23] Apart from the modifications outlined in the following, they adopt some improvements and best practices from Brunner et al. (2017) and Conlon and Gortmaker (2020)

identify the parameters for Θ_2 that determine the consumer-specific deviations from the mean utility. The idea is that when the model is evaluated at the true parameters, it should predict these micro-moments (see Petrin, 2002; see also Berry and Haile, 2020 and Conlon and Gortmaker, 2023 for further details). Recall that a micro-moment describes the characteristics (like the income) of the average consumer buying a certain product. This means that if consumers behave differently because of their characteristics, this shows up in the micro-moments. For instance, if consumers with low incomes are very price sensitive, the average income of a consumer buying an expensive product should be high. If, in contrast, income does not affect price sensitivity, the average income of a consumer buying an expensive brand should be similar to the average income of the entire population.

For each set of candidate parameters for Θ_2, the micro-moments can be calculated for each product and market, and these predicted moments are then compared to the ones observed in the data. Note that the candidate parameters for Θ_1 are irrelevant for this exercise. To see this, consider the choice probability of a consumer i with certain characteristics buying product j. According to Equation (4.6), the choice probability depends on the different utility levels that a consumer can achieve when buying the different products (including the outside option). The estimation routine of Berry et al. (1995) can derive these utility levels and split them into mean utilities and consumer-specific deviations. However, as described above, only the parameters from the second set are required to achieve this. The remaining parameters for Θ_1 can split the mean utility into its components, but the total mean utility remains unaffected and does not change in these parameters. If the mean utility is unaffected, the choice probabilities also remain unaffected.

To summarize, Döpper et al. (2023) can use the first step of the estimation routine of Berry et al. (1995) to get an estimate for Θ_2. In the next step, they are concerned with the estimation of the remaining parameters for Θ_1. Since they already have an estimate for Θ_2, they can fix these parameters in the subsequent estimation routine. In particular, this means that the mean utilities are independent of the candidate parameters for Θ_1 and, thus, remain the same.

With the estimates for Θ_2 in hand, Döpper et al. (2023) can derive two measures of interest for given candidate parameters for Θ_1. First, it is straightforward to derive the structural demand-side error term $\Delta\xi_{jcrt}$. To do this, they simply have to subtract the candidate parameter for the constant β and the price multiplied by the candidate parameter for α from the mean utility and then take the fixed effects. Second, with the choice probabilities and the candidate parameter for α, they can calculate the substitution patterns $\left(\partial s_{crt}\left(p_{crt}\right)/\partial p_{crt}\right)$ required to estimate the marginal costs based on (4.7). By taking fixed effects, they can then derive an estimate for the cost shock $\Delta\eta_{jcrt}$.

Their key identifying assumption is that the covariance between the two error terms is zero for the true parameter in Θ_1:

$$\operatorname{cov}\left(\Delta\xi_{jcrt}, \Delta\eta_{jcrt}\right) = 0. \tag{4.9}$$

MacKay and Miller (2023) and Döpper et al. (2023) discuss (and justify) this assumption in detail. However, it seems worth briefly pointing out two properties of this approach that make it desirable for application across so many product categories. First, and in contrast to the instrumental variables approach used elsewhere in the literature, covariance restrictions do not require to estimate a first stage. This means that the entire potentially endogenous variation is exploited while instrumental variables restrict the variation. Second, the choice of appropriate fixed effects (or other covariates) is important when deriving estimates for the error terms. The aim is to choose them in a way such that the variation that remains in the error term is unique to this error term. In fact, Döpper et al. (2023) point out that with their fixed effects, the variation that remains in the error term is similar to the one used as instruments elsewhere in the literature. They also achieve similar results when using even stricter fixed effects defined at the product-retailer-region level.

4.4.3 Merger Effects on Inferred Measures

Döpper et al. (2023) use the empirical strategy outlined in the previous section to estimate the parameters in Θ_1 and Θ_2 separately for all product categories and years. I use their estimates to infer two new measures that cannot be directly observed from the data. First, I calculate the estimated marginal costs $\widehat{c}_{jcrt}$ based on Expression (4.7). Second, I construct a measure for the perceived quality. To this end, I focus on the mean utility δ_{jcrt} that a consumer gains from buying brand j at retail chain c in region r and quarter t, i.e.,

$$\widehat{\delta}_{jcrt} = \widehat{\beta} + \widehat{\alpha} \cdot p_{jcrt} + \widehat{\xi}_{jr} + \widehat{\xi}_{cr} + \widehat{\xi}_{t} + \Delta\xi_{jcrt}, \tag{4.10}$$

where the hats indicate estimates. The mean utility has the benefit that it is constant for all consumers and is independent of consumer characteristics. In other words, it omits variables related to horizontal product differentiation so that the remaining utility relates to the price and vertical product differentiation. By subtracting the impact of the price ($\widehat{\alpha} \cdot p_{jcrt}$), I can thus infer a measure of the perceived quality. I regress this measure on brand-retailer-region fixed effects and use the estimated fixed effects in the following. This allows me to exclude seasonal effects because I get an average value at the annual level. Finally, with both new measures in hand, I aggregate the observations across regions and quarters so that I end up with one observation for each brand-retailer-year combination.

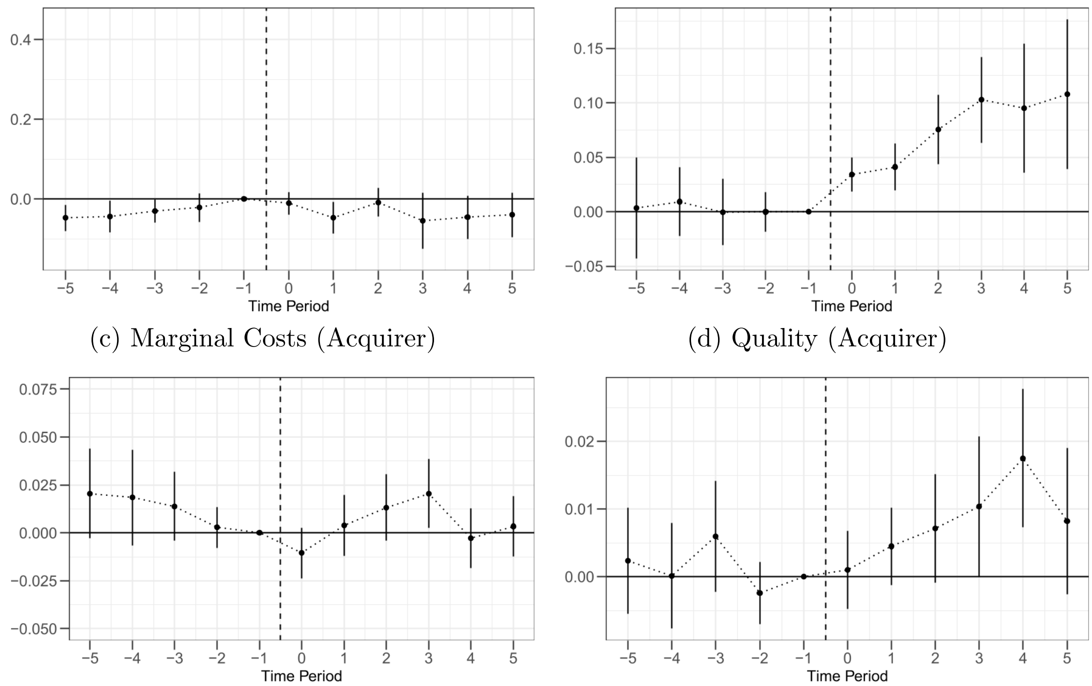
It seems reasonable to briefly provide some intuition for the measure of perceived quality. In particular, it is important to highlight that the perceived quality of a brand may differ from the actual quality. As indicated by the name, it depends on the consumers' opinion of the product. This opinion may change, for instance, if a brand is heavily advertised. Another factor could be the shelf space that a retailer allocates to a brand. If the brand occupies a lot of shelf space, this might be a sign that the retailer "believes" in the high potential of a brand. These (mostly psychological) factors are not explicitly modeled in the demand model, and identifying each of them is potentially challenging on its own. Thus, I consider the fixed effects that enter the mean utility, and hence the perceived quality, as a "reduced-form" approach to capturing the average consumer's opinion of a brand while remaining silent about the psychological channels that lead to such an opinion.

I re-estimate the baseline specification 4.2 using the new measures as dependent variables. Figure 4.5 shows the result of this exercise. The panels in the first row refer to the targets, and those in the second row refer to the acquirers. In line with the previous analysis, I also use the adjusted ratio measure (4.3) for the acquirers.

Figure 4.5: Changes in Marginal Costs and Perceived Quality

(a) Marginal Costs (Target) (b) Quality (Target)

(c) Marginal Costs (Acquirer) (d) Quality (Acquirer)



The first column of Figure 4.2 shows the results for the marginal costs. Note that I use the same scaling for the y-axis as in Figure 4.2 and 4.4, respectively, to simplify the comparison across the graphs for the reader. Panel (a) shows that the marginal costs of the targets are, by and large, unaffected by the merger. Although some coefficients are statistically different from 0 at the 5% level, the magnitudes are so small that they can be considered economically negligible. Interestingly, this is different for the acquirers. The coefficients for changes in the marginal costs of the acquirers show a larger dispersion, but almost all coefficients (except the one in $\tau = 3$) are statistically insignificant at the 5% level. While there seems to be no clear post-merger trend, the pre-merger marginal costs show (if any) a falling trend over time.

The second column shows the results for the perceived quality. Since quality is defined as the sum of the fixed effects, which can take both positive and negative values in general, I use a standardized version of the measure rather than the logarithm as a dependent variable.[24] Therefore, I also do not follow the convention to keep the scaling of the y-axis consistent with those in the other figures.

[24] For the standardized version, I first subtract the mean and then divide the measure by its standard deviation.

The results show that perceived quality stays relatively constant before the merger and increases afterward. While this pattern is rather sharp for the targets, meaning that the pre-merger coefficients are extremely close to 0 and that the post-merger coefficients are all statistically different from 0 at the 5% level, the pattern for the acquirers is more noisy. In particular, most post-merger confidence intervals contain 0, even though it is usually close to the interval boundaries. Overall, however, both panels show trends that match the development of the revenues and quantities.

4.5 Potential Mechanisms and the Portfolio Power Theory

The previous sections paint a rather clear picture of the impact of cross-category mergers on market outcomes. First, cross-category mergers can influence the revenues of both targets and acquirers. The direction of the effect depends on the relative size of the acquirer's revenues to the target's revenues at a given retailer. Second, changes in revenues are almost exclusively driven by changes in quantities and not by changes in prices. Third, by making use of the work of Döpper et al. (2023), I find that marginal costs are almost unaffected and that the changes in quantities can be rationalized within a structural model by changes in the non-price utility part.

The last section is now devoted to a brief discussion of mechanisms that could potentially drive the results. I start with two mechanisms that can be subsumed under the portfolio power theory. The first channel deals with the manufacturers' bargaining power. The idea is that a firm's bargaining power and its ability to influence bargaining outcomes in its own interest depends on its importance for the other firm's profit. This idea is formalized in the Nash-in-Nash bargaining framework, which is frequently used by economists in empirical studies of bargaining (see Draganska et al., 2010; Noton and Elberg, 2018 for examples from the consumer packed goods retail industry and Collard-Wexler et al., 2019 for a micro-foundation). Think of a manufacturer f and a retailer r conducting negotiations over some form of financial payments (for instance, linear prices or fixed fee payments) captured in a vector μ_{fr} and an effort level e_{fr} that describes the effort that a retailer spends on the product of manufacturer f (like shelf space or in-store promotions). Then, according to the Nash-in-Nash bargaining framework, they choose these variables to maximize the following expression:

$$\left(\pi_f\left(\mu, e\right) - \pi_f^{-r}\left(\mu^{-r}, e^{-r}\right)\right)^{\lambda}\left(\pi_r\left(\mu, e\right) - \pi_r^{-f}\left(\mu^{-f}, e^{-f}\right)\right)^{1-\lambda} \tag{4.11}$$

π_f and π_r refer to the profits of the manufacturer and the retailer, and μ and e are vectors capturing the strategic variables of all bargaining pairs. The parameter $\lambda \in (0,1)$ captures the other determinants influencing the abilities of the parties in the negotiations, like, for instance, the negotiation skills of the managers. The superscripts $-f$ and $-r$ refer to a situation where the bargaining between the firms breaks down so that retailer r does not sell products of manufacturer f.

The brackets in Expression 4.11 show the so-called gains from trade, that is, the extra profit that a firm gains through a collaboration with the other firm. A firm gains bargaining leverage over the other firm if the extra profit for the other firm increases. Intuitively, if the products of the other firm are very important for one's own revenues, a bargaining breakdown would be very costly, and the incentive to settle the negotiation increases.

A merger between two manufacturers leads to a change in the gains from trade since the merging firms are now negotiating jointly with the retailers. Before the merger, a bargaining breakdown with one manufacturer still allows a retailer to settle the negotiations with the other manufacturer. However, this is not possible after the merger, and a bargaining breakdown will result in a loss of the products of both firms. This gives the integrated manufacturer a larger bargaining leverage. Dafny et al. (2019) use a similar reasoning in their analysis of cross-market hospital mergers.

There might also be other determinants than the gains from trade that play a role in the negotiations and that could be affected by a merger. In particular, a merger can alter the logistics of the firms and allow them to operate a better distribution network. One key consideration could be the ability of manufacturers to reliably manage deliveries. For instance, if the product of a rather small manufacturer is subject to highly volatile demand and the manufacturer cannot operate a large distribution network due to its size, the retailer might run out of the product and the shelf space remains empty whenever a demand spike occurs. An alternative strategy would be to increase the inventory, leading to increased costs for the retailer. This gives the retailer small incentives to provide the manufacturer with more shelf space.[25] A merger could give the manufacturers the ability to combine their distribution networks. Apart from potential (fixed) cost savings, this might increase both the reliability of regular deliveries (e.g., due to more frequent deliveries and larger truck loads) and the ability to react to irregular delivery requests. These improvements likely depend on manufacturers' past relationships with a retailer since the logistical operations have likely developed to serve retailers that were willing to sell many products from the manufacturers in the past.

Both channels, the bargaining power channel and the improvements in the distribution network, fit the previously documented patterns in that they depend on the manufacturer-retailer-specific relationships. Both also fit the portfolio power theory because they do not depend on the substitutability/complementarity of the products but on the size of the total sales to a retailer. There are two other explanations that are not related to the portfolio power theory and that I deem less likely to explain the patterns. Both explanations have in common that a merger can lead to better access to resources, in particular financial and human resources.

The first explanation is that a merger leads to increased marketing expenditures. As discussed earlier, marketing activities can also be part of the negotiations between manufacturers and retailers (captured by the effort variable in (4.11)). Hence, I relate here

[25] The marketing and operations research literature has devoted an entire subfield to the question of optimal shelf space allocation and, hence, forgone profits due to stock-outs have long been an important topic (see Curhan, 1973; Gilliver and Gordon, 1978; Emmelhainz and Stock, 1991 for examples of early studies on this topic).

to retailer-independent marketing activities. Such activities can change the perception of consumers about the quality of the products of the merging parties (for instance, due to stronger brand preferences). Although this channel might seem to fit the patterns at first glance, there is a good reason to believe that this is actually not the case. This is because my analysis aims to isolate retailer-specific effects and that retailer-independent marketing activities would have affected all retailers in a similar manner.[26]

The same argument applies to potential efficiency gains beyond the previously described improvements in the distribution network. Such efficiency gains can include improvements in the production process due to knowledge spillovers or better access to financial resources that spur investments in new technologies. Efficiency gains are part of a longstanding discussion on the competitive and anti-competitive effects of (horizontal and vertical) mergers, dating back to at least Williamson (1968) (see Affeldt et al., 2021 and the references therein for a recent overview).

In the context of my investigation, there is no evidence for investments in (marginal) cost-reducing production technologies since marginal costs do not fall after the merger. This does, however, not necessarily mean that efficiency gains are absent. Another explanation could be that improvements in production technology lead to quality upgrades at the same or similar marginal cost levels. However, I can apply the same argument that renders effects through increased retailer-independent advertising spending unlikely. If the quality improves, the non-price part of consumers' utility will increase, but I would expect this increase to be rather similar across retailers, independent of the target's or acquirer's historical revenues to a retailer.

To summarize, out of the four possible mechanisms discussed, only two mechanisms seem to be able to explain the pattern described in my analysis. These two channels can also be subsumed under the portfolio power theory.

4.6 Conclusion

This paper documents the presence, direction, and size of portfolio effects by analyzing 57 consummated mergers of manufacturers in the US consumer packaged goods retail industry between 2006 and 2019. My analysis focuses on cross-category mergers where the merging firms have (almost) no overlap in their product portfolio prior to the merger. I exploit the large heterogeneity in the pre-merger bargaining positions of the targets and the acquirers at the different retailers (as measured by their pre-merger revenues at the respective retailers) and provide evidence that manufacturers with weaker pre-merger bargaining positions benefit from cross-category mergers through increases in revenues, while manufacturers with stronger pre-merger bargaining positions are harmed and experience revenue decreases. These increases (decreases) in revenues are almost entirely driven by increases (decreases) in the quantities sold and not by changes in prices. I show that these patterns

[26]As mentioned in Section 4.3.3, the estimated $\beta_{1\ell}$ coefficients from my main specification 4.2, which are intended to capture retailer-independent effects, are always close to 0 and negligible in magnitude.

can be rationalized within a structural model by changes in the perceived quality of the products of the merging firms. Changes in marginal costs do not seem to play a crucial role.

In the last section, I discussed two potential mechanisms related to the portfolio power theory that help explain these patterns. Both build on the idea that an increase in the sheer size of the product portfolio can impact the negotiations between the manufacturers and the retailers. The first channel is that bargaining breakdowns become increasingly costly for a retailer when the size of the manufacturer increases. This changes the incentives of the retailers to settle the negotiations with the manufacturers and allows the manufacturers to demand larger concessions (for instance, in the form of better or more shelf space). The second channel builds on improvements in logistics because the merging firms can operate a joint distribution network. The better logistics increase the incentives for the retailers to provide the products of the merging firms with more and better shelf space since stockouts (or similar problems) are less likely to occur. Finally, I argue that changes in advertising spending and efficiency gains are unlikely to explain the patterns.

An open question that I cannot answer at the moment is that of possible policy implications. To shed light on this question, I plan to use a structural model in future versions of this paper that will serve two purposes. On the one hand, it provides me with insights into the impact of portfolio effects on consumer surplus and welfare; on the other hand, it allows me to investigate whether portfolio effects can be regarded as pro- or anti-competitive in the merger cases I study.

Appendices

4.A Estimates of the Two-Way Fixed Effects Regressions

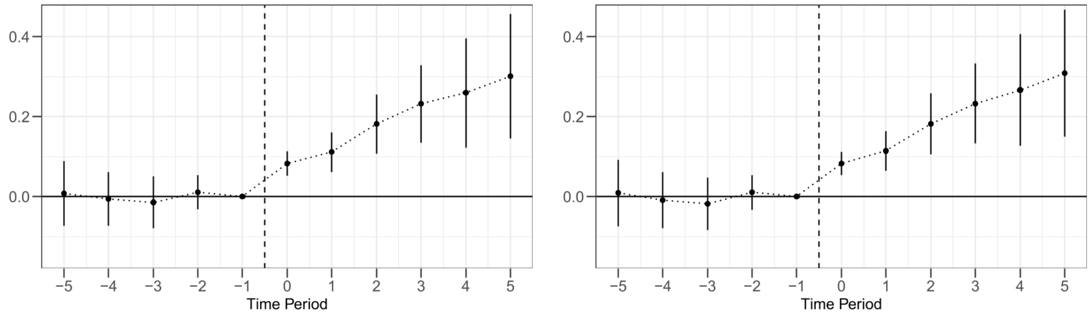
	log(Sales)	log(Market Share)	log(Quantity)	log(Price)	log(Marginal Cost)	Sd. Quality
Ratio in t = -5	0.012	0.009	0.022	-0.011***	-0.048**	0.003
	(0.039)	(0.041)	(0.039)	(0.003)	(0.016)	(0.023)
Ratio in t = -4	-0.005	-0.009	0.001	-0.005	-0.044*	0.009
	(0.034)	(0.035)	(0.035)	(0.003)	(0.020)	(0.016)
Ratio in t = -3	-0.015	-0.018	-0.013	-0.003	-0.030*	0.000
	(0.033)	(0.033)	(0.033)	(0.002)	(0.015)	(0.015)
Ratio in t = -2	0.010	0.011	0.011	-0.001	-0.022	0.000
	(0.022)	(0.022)	(0.022)	(0.001)	(0.018)	(0.009)
Ratio in t = 0	0.086***	0.083***	0.086***	0.000	-0.011	0.034***
	(0.015)	(0.015)	(0.014)	(0.001)	(0.014)	(0.008)
Ratio in t = 1	0.112***	0.114***	0.114***	-0.002	-0.046*	0.041***
	(0.024)	(0.025)	(0.025)	(0.003)	(0.020)	(0.011)
Ratio in t = 2	0.187***	0.182***	0.188***	-0.001	-0.008	0.075***
	(0.036)	(0.038)	(0.038)	(0.004)	(0.018)	(0.016)
Ratio in t = 3	0.235***	0.233***	0.239***	-0.004	-0.055	0.103***
	(0.048)	(0.050)	(0.049)	(0.005)	(0.035)	(0.020)
Ratio in t = 4	0.260***	0.267***	0.272***	-0.012*	-0.046	0.095**
	(0.066)	(0.070)	(0.066)	(0.005)	(0.027)	(0.030)
Ratio in t = 5	0.306***	0.309***	0.317***	-0.011*	-0.040	0.108**
	(0.074)	(0.079)	(0.073)	(0.005)	(0.028)	(0.034)
Ratio in t<-5	0.007	0.004	0.020	-0.013*	-0.058**	-0.007
	(0.051)	(0.053)	(0.050)	(0.006)	(0.022)	(0.024)
Ratio in t>5	0.288***	0.265***	0.301***	-0.014**	-0.003	0.097**
	(0.061)	(0.060)	(0.060)	(0.005)	(0.020)	(0.029)
Num.Obs.	62537	62537	62537	62537	53615	62468
R2	0.757	0.686	0.818	0.986	0.876	0.656
R2 Adj.	0.728	0.649	0.797	0.985	0.859	0.616
Retailer-Merger FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Period FE	X	X	X	X	X	X

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.B Effects on Revenue and Quantity Shares

Figure 4.6: Changes in Revenue and Quantity Shares

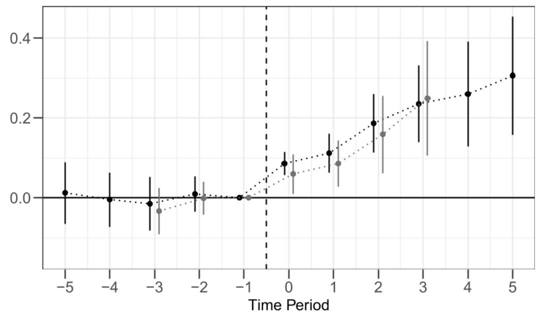
(a) Sales Shares (b) Quantity Shares



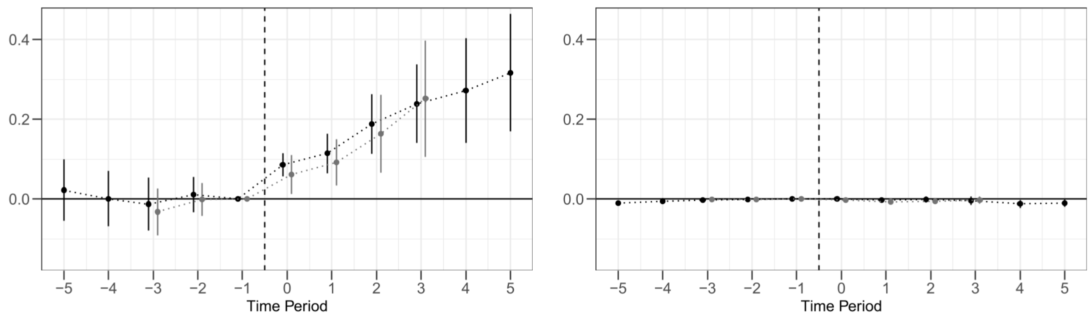
4.C Results with Balanced Panel

Figure 4.7: Changes in Revenues, Quantities, and Prices (Balanced Panel)

(a) Sales



(b) Quantities (c) Prices



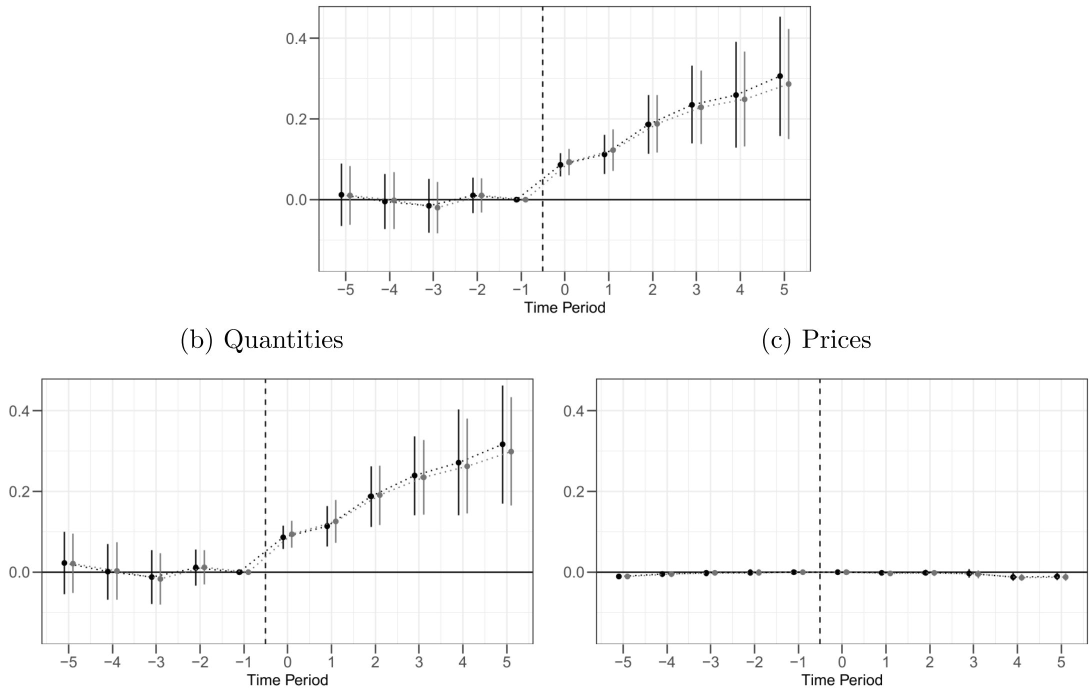
4.D Results without Time Period Fixed Effects

Figure 4.8: Changes in Revenues, Quantities, and Prices (without Time Period Fixed Effects)

(a) Sales

(b) Quantities

(c) Prices



Bibliography

Affeldt, P., T. Duso, K. Gugler, and J. Piechucka (2021): "Assessing EU Merger Control through Compensating Efficiencies," *DIW Berlin: Discussion Papers 1979*.

Anderson, E. T. and E. J. Fox (2019): "How price promotions work: A review of practice and theory," *Handbook of the Economics of Marketing*, 497–552.

Ashenfelter, O. and D. Hosken (2010): "The Effect of Mergers on Consumer Prices: Evidence from Five Mergers on the Enforcement Margin," *The Journal of Law & Economics*, 53.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Berry, S. T. and P. A. Haile (2020): "Nonparametric Identification of Differentiated Products Demand Using Micro Data," *NBER Working Paper 27704*.

Bhattacharya, V., G. Illanes, and D. Stillerman (2023): "Merger Effects and Antitrust Enforcement: Evidence from US Retail," *NBER Working Paper 31123*.

Brunner, D., F. Heiss, A. Romahn, and C. Weiser (2017): "Reliable Estimation of Random Coefficient Logit Demand Models," *DICE Discussion Paper No. 267*.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021): "Difference-in-Differences with a Continuous Treatment," *Mimeo (available via arXiv:2107.02637)*.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 225, 200–230.

Choi, J. P. (2001): "A Theory of Mixed Bundling Applied to the GE/Honeywell Merger," *Antitrust*, 16, 32–33.

Chunga, J. and S. Jeon (2014): "Portfolio effects in conglomerate mergers: the empirical evidence of leverage effects in Korean liquor market," *Applied Economics*, 46, 4345–4359.

Collard-Wexler, A., G. Gowrisankaran, and R. S. Lee (2019): ""Nash-in-Nash" Bargaining: A Microfoundation for Applied Work," *Journal of Political Economy*, 127, 163–195.

Conlon, C. and J. Gortmaker (2020): "Best practices for differentiated products demand estimation with PyBLP," *RAND Journal of Economics*, 51, 1108–1161.

——— (2023): "Incorporating Micro Data into Differentiated Products Demand Estimation with PyBLP," *Mimeo*.

Curhan, R. C. (1973): "Shelf Space Allocation and Profit Maximization in Mass Retailing," *Journal of Marketing*, 37, 54–60.

Dafny, L., K. Ho, and R. S. Lee (2019): "The price effects of cross-market mergers: theory and evidence from the hospital industry," *RAND Journal of Economics*, 50, 286–325.

de Chaisemartin, C. and X. D'Haultfœuille (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey," *Econometrics Journal*, 26, C1–C30.

Döpper, H., A. MacKay, N. Miller, and J. Stiebale (2023): "Rising Markups and the Role of Consumer Preferences," *Kilts Center at Chicago Booth Marketing Data Center Paper Series*.

Draganska, M., D. Klapper, and S. B. Villas-Boas (2010): "A Larger Slice or a Larger Pie? An Empirical Investigation of Bargaining Power in the Distribution Channel," *Marketing Science*, 29, 57–74.

Dubé, J.-P., A. Hortaçsu, and J. Joo (2021): "Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares," *Marketing Science*, 40, 637–660.

Emmelhainz, M. A. and J. R. Stock (1991): "Consumer Responses to Retail Stock-outs," *Journal of Retailing*, 67, 138–147.

Evans, D. S. and M. Salinger (2002): "Competition Thinking at the European Commission: Lessons from the Aborted GE/Honeywell Merger," *George Mason Law Review*, 10, 489–532.

Gandhi, A., Z. Lu, and X. Shi (2023): "Estimating demand for differentiated products with zeroes in market share data," *Quantitative Economics*, 14, 381–418.

Gilliver, A. and H. A. Gordon (1978): "An Analytic Information System for a Representative Sales Force – A Case Study," *Journal of the Operational Research Society*, 29, 719–730.

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225, 254–277.

Guadalupe, M., O. Kuzmina, and C. Thomas (2012): "Innovation and Foreign Ownership," *American Economic Review*, 102, 3594–3627.

Hendel, I. and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74, 1637–1673.

Lewis, M. S. and K. E. Pflum (2017): "Hospital systems and bargaining power: evidence from out-of-market acquisitions," *RAND Journal of Economics*, 48, 579–610.

MacKay, A. and N. H. Miller (2023): "Estimating Models of Supply and Demand: Instruments and Covariance Restrictions," *Harvard Business School Strategy Unit Work Paper No. 19-051 (first version: 2018, revised version: August 2023)*.

Majerovitz, J. and A. Yu (2023): "Consolidation on Aisle Five: Effects of Mergers in Consumer Packaged Goods," *Mimeo (available via SSRN).*

Noton, C. and A. Elberg (2018): "Are Supermarkets Squeezing Small Suppliers? Evidence from Negotiated Wholesale Prices," *Economic Journal*, 128, 1304–1330.

OECD (2001): "Portfolio Effects in Conglomerate Mergers," *OECD Journal: Competition Law and Policy*, 4.

Park, S. (2009): "An empirical testing of leverage effects via the common distribution network," *International Review of Law and Economics*, 29, 143–152.

Patterson, D. E. and C. Shapiro (2001): "Transatlantic Divergence in GE/Honeywell: Causes and Lessons," *Antitrust*, 16, 18–26.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110, 705–729.

Reynolds, R. J. and J. A. Ordover (2002): "Archimedian Leveraging and the GE/Honeywell Transaction," *Antitrust Law Journal*, 70, 171–198.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023): "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 235, 2218–2244.

Watson, P. (2003): "Portfolio effects in EC merger law," *Antitrust Bulletin*, 48, 781–806.

Williamson, O. E. (1968): "Economies as an Antitrust Defense: The Welfare Tradeoffs," *American Economic Review*, 58, 18–36.

Eidesstattliche Versicherung

Ich, Herr Hendrik Döpper, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Düsseldorf, der 21. November 2023 Unterschrift: ____________________

Statement of Contributions

Statement of contribution

My co-author Alexander Rasch contributed to the chapter

"Combinable Products, Price Discrimination, and Collusion"

of my dissertation

"Four Essays on the Analysis of Market Power and its Implications."

All authors contributed *equally* to this chapter.

Signature of Alexander Rasch: ______________________________

Statement of contribution

My co-authors Geza Sapi and Christian Wey contributed to the chapter

"A Bargaining Perspective on Vertical Integration"

of my dissertation

"Four Essays on the Analysis of Market Power and its Implications."

The starting point for this chapter was a discussion paper published by Geza Sapi in 2012 (DICE Discussion Paper No 61), which was also part of his dissertation. The chapter in its current form strongly differs from this discussion paper. Most notable are the following changes:

- We changed the structure. For instance, we dropped the sections 6 (market entry) and 7 (example). We also substantially extended the section on the comparison of horizontal and vertical merger incentives. For instance, we added an outside option to the auction model and a comparison of post- and pre-merger profits which allows to discuss incentives to preempt potentially harmful mergers of competitors. We also discuss the relationship to the result of Inderst and Wey (2003, RAND) in more detail.

- We substantially revised the remaining sections. For instance, we added an example to the introduction and modified the presentation of the model.

- We added two important robustness checks, namely the comparison with Nash-in-Nash bargaining and the introduction of downstream competition.

All authors contributed *equally* to the revision of the original discussion paper, with the exception of three mathematical extensions and robustness checks, which are reported hereafter. First, I extended the auction model by adding an outside option. Second, I added a robustness check that allows for downstream competition. Third, Christian Wey and I *jointly* developed an extension that compares our results to those that arise from a framework with Nash-in-Nash bargaining.

Signature of Geza Sapi: ______________________________

Statement of contribution

My co-authors Geza Sapi and Christian Wey contributed to the chapter

"A Bargaining Perspective on Vertical Integration"

of my dissertation

"Four Essays on the Analysis of Market Power and its Implications."

The starting point for this chapter was a discussion paper published by Geza Sapi in 2012 (DICE Discussion Paper No 61), which was also part of his dissertation. The chapter in its current form strongly differs from this discussion paper. Most notable are the following changes:

- We changed the structure. For instance, we dropped the sections 6 (market entry) and 7 (example). We also substantially extended the section on the comparison of horizontal and vertical merger incentives. For instance, we added an outside option to the auction model and a comparison of post- and pre-merger profits which allows to discuss incentives to preempt potentially harmful mergers of competitors. We also discuss the relationship to the result of Inderst and Wey (2003, RAND) in more detail.

- We substantially revised the remaining sections. For instance, we added an example to the introduction and modified the presentation of the model.

- We added two important robustness checks, namely the comparison with Nash-in-Nash bargaining and the introduction of downstream competition.

All authors contributed *equally* to the revision of the original discussion paper, with the exception of three mathematical extensions and robustness checks, which are reported hereafter. First, I extended the auction model by adding an outside option. Second, I added a robustness check that allows for downstream competition. Third, Christian Wey and I *jointly* developed an extension that compares our results to those that arise from a framework with Nash-in-Nash bargaining.

Signature of Christian Wey: ______________________

Statement of contribution

My co-authors Alexander MacKay, Nathan Miller and Joel Stiebale contributed to the chapter

"Rising Markups and the Role of Consumer Preferences"

of my dissertation

"Four Essays on the Analysis of Market Power and its Implications."

All authors contributed *equally* to this chapter.

Signature of Alexander MacKay: ____________________

Statement of contribution

My co-authors Alexander MacKay, Nathan Miller and Joel Stiebale contributed to the chapter

"Rising Markups and the Role of Consumer Preferences"

of my dissertation

"Four Essays on the Analysis of Market Power and its Implications."

All authors contributed *equally* to this chapter.

Signature of Nathan Miller: ______________________________

Statement of contribution

My co-authors Alexander MacKay, Nathan Miller and Joel Stiebale contributed to the chapter

"Rising Markups and the Role of Consumer Preferences"

of my dissertation

"Four Essays on the Analysis of Market Power and its Implications."

All authors contributed *equally* to this chapter.

Signature of Joel Stiebale: ______________________________