Three Essays on Competition Economics

DISSERTATION

zur Erlangung des akademischen Grades

Doctor Rerum Politicarum (Dr. rer. pol.)

im Fach Volkswirtschaftslehre

eingereicht an der Wirtschaftswissenschaftlichen Fakultät
Heinrich-Heine-Universität Düsseldorf

hhu Heinrich Heine Universität Düsseldorf

von: Jannika Leonie Schad, M.Sc.
geboren am 30.12.1989 in Dieburg

Erstgutachter: Prof. Dr. Justus Haucap
Zweitgutachter: Prof. Dr. Hans-Theo Normann

Abgabedatum: 14. August 2020

Acknowledgment

I would like to thank everyone who accompanied me through the time of my dissertation and stood by me. In particular, I would like to thank my supervisors Justus Haucap and Hans-Theo Normann for great support and a good supervision of my dissertation. I would like to thank my co-authors Olivia Bodnar, Melinda Fremerey, Matthias Hunold and Hans-Theo Normann for a great, always constructive and fair cooperation, which – at least in my opinion – has led to interesting projects. My office colleague Hendrik Döpper deserves special thanks for his great willingness to help and many interesting discussions – I have learned so much from you. I could not have imagined a better office colleague. I thank Daniel Kamhöfer for making me a better econometrician and for his amazing support. I am also very grateful for the opportunity to do my PhD at DICE, to be part of such a productive environment with mutual support and knowledge sharing.
I would like to thank my family for their constant support in all my life's work – without you I would never have made it this far. Last but not least I would like to thank my friends for their moral and spiritual support.

Coauthors and publications

Chapter 1: The Effects of Private Damage Claims on Cartel Stability: Experimental Evidence

- Coauthors:
  - Olivia Bodnar (DICE, Heinrich-Heine-Universität)
  - Melinda Fremerey (DICE, Heinrich-Heine-Universität)
  - Hans-Theo Normann (DICE, Heinrich-Heine-Universität)
- Status: DICE Discussion Paper No 315, June 2019, revise and resubmit in *The Journal of Law, Economics, and Organization*

Chapter 2: Single Monopoly Profits, Vertical Mergers and Downstream Entry Deterrence

- Coauthors:
  - Matthias Hunold (Universität Siegen)

Chapter 3: The Causal Effect of Product Reviews on Prices. Evidence from Amazon.com

- Coauthors: none

Contents

List of Figures

List of Tables

Introduction

My thesis consists of three essays in the field of competition economics. It is the product of three intensive years at the Düsseldorf Institute for Competition Economics (DICE) with its DFG Graduate Programme 1974 Competition Economics and not only reflects the institute's core research mission but also shows its diversity in terms of methods, research topics, and people.

The selection of research questions is motivated by my drive to learn about important current and ongoing issues in competition policy and to contribute to their economic understanding and settlement. All three chapters thereby link to (recent) policy debates of competition authorities. The first chapter studies possible undesired consequences of facilitating private damage actions against cartel members, which may result from the European Directive 2014/104/EU on antitrust damages actions (European Commission, 2014). The second chapter investigates the subject of foreclosure in the case of vertical integration, which is of ongoing interest for competition authorities and a key concern in merger control. Finally, the third chapter addresses online consumer reviews, which are covered by the European Directive 2005/29/EC and 2019/2161 (European Commission, 2005, 2019). The current sector inquiry on reviewing systems of the German competition authority also emphasizes the need to understand the mechanisms behind such systems and their effects on market outcomes (Bundeskartellamt, 2019).

Chapter 1 on "The Effects of Private Damage Claims on Cartel Stability: Experimental Evidence" is joint work with Olivia Bodnar, Melinda Fremerey and Hans-Theo Normann. We employ an experimental study to analyze the tensions between private damage actions against cartel members on the one hand and public cartel prosecution on the other hand. The topic is motivated by the facilitation of private damage compensation for breaches of competition law in Europe and the possible negative effects of private cartel damage claims on leniency: Whereas leniency applicants obtain full immunity regarding the public cartel fines, they have no or only restricted protection against private third-party damage claims. As a consequence, cartel damage claims could stabilize cartels. Thus, the main research question is whether private cartel damage claims make cartels more stable.

The theory of collusion is useful but often limited in its predictive power due to wide ranges of possible collusive equilibria and difficulties in modeling the process of coordination. As collusion is typically illegal and secret, empirical research with market data is difficult in this field. Laboratory experiments present a readily available testbed, which is unaffected by the sample-selection problem. Consequently, we make use of a laboratory experiment and base this on a theoretical framework.

In our experiment, the participants act as firms that choose whether to join a cartel, and – if they have done so – whether to apply for leniency afterward. We find that the implementation of private damage claims makes cartels indeed more stable, largely due to fewer leniency reports, but it decreases the instances of cartel formation. On balance, our main message for the European policymakers and their critics on the overall impact of private damage claims is indeed positive: Cartel prevalence declines. This comes, however, with the caveat of a negative effect on leniency.

Chapter 2 on "Single Monopoly Profits, Vertical Mergers, and Downstream Entry Deterrence" is joint work with Matthias Hunold. We contribute to a fundamental but still unsettled key research question in the field of merger control: Under what conditions does vertical integration of firms at different parts of the supply chain lead to the foreclosure of markets?

Proponents of the Chicago School argue that full vertical mergers enhance efficiency and at worst have neutral effects on competition (Bork, 1978; Posner, 1979). More recent theories based on richer models highlight both the pro-competitive and anti-competitive effects of vertical integration under certain conditions.[1] We show that vertical integration can foreclose markets even in a simple setting where these conditions are not fulfilled.

[1]Game theoretic models reveal anti-competitive effects of vertical integration under specific conditions, such as additional commitment power of the integrated firm (Ordover *et al.*, 1990), secret contract offers (Hart and Tirole, 1990), and upstream collusion (Normann, 2009).

We review the Chicago School's single monopoly profits theory whereby an upstream monopolist, which can use contracts to extract the monopoly profits from the downstream firms, cannot generate additional profits through vertical integration. For the case that the downstream firms cannot avoid sourcing from the upstream firm in order to be active on the market, our results are consistent with the Chicago School's theory. The upstream monopolist uses its contracts to extract all monopoly profits from the downstream firms. Thus, the downstream entrant's profit equals its outside option irrespective of vertical integration, such that vertical integration of the incumbents has no effect on the incentives to enter the market.

The main contribution is to show that entry deterrence may occur once there is the possibility of an alternative input supply, even though this is never used in equilibrium. The upstream firm thus has a market share of 100% in equilibrium and may earn high – monopoly like – margins. This is thus a setting where an observer, for instance, a competition authority, may conclude that the Chicago School's single monopoly profits theory would apply and vertical integration does not raise competitive concerns.

We show that with the possibility of alternative sourcing, the outside option value and thus the equilibrium profit of a downstream entrant does depend on whether there is vertical integration of the established downstream and upstream firms. Interestingly, the elimination of double marginalization acts as a commitment to intense downstream competition when the entrant does not source the inputs from the efficient upstream firm. We show that the incumbent firms can thereby use vertical integration to deter efficiency-enhancing entry. The results are not only relevant for traditional industries but might also apply to vertically integrated digital platforms where independent firms are both customers and competitors.

Chapter 3 on "The Causal Effect of Product Reviews on Prices. Evidence from Amazon.com" is a single-authored essay and presents my empirical study on product reviews on the Amazon marketplace. It is part of a growing and arguably highly important new strand of research in the field of industrial economics that deals with the digital economy and the influential firms therein. This includes Amazon as one of the largest e-commerce platforms with around 700 million users worldwide[2] and its well-known and extensive reviewing system.

User generated product reviews are frequently used by online shoppers to obtain information and have a substantial influence on purchase decisions. The British Competition and Markets Authority (CMA) estimates that 75% of online shoppers refer to product reviews (CMA, 2020).

The main research objective of this chapter is to identify the causal effect of product reviews on the pricing strategies of sellers on online platforms. In an extensive and time-consuming exercise, I collected information from Amazon and a price tracking page ("web scraping") to compile an appropriate data set for the econometric analysis.

Towards identifying the causal effect of product reviews on the prices, I develop an instrumental variable approach that addresses the problem of simultaneity of prices and ratings. To do so, I instrument a given star rating by a reviewer's past ratings for other products relative to these products' aggregated ratings. My estimations indicate that an increase in the aggregated star rating by one out of five stars significantly increases the prices of marketplace sellers by up to 6.5 percentage points.

My econometric findings reveal relevant economic insights. First, they show that marketplace sellers are aware of the product ratings and incorporate them into their pricing strategy. Second, the finding that better ratings lead to higher prices indicates that star ratings influence the consumers' willingness to pay – which the sellers extract through higher prices, at least partially. This suggests

[2]See chapter 3 for a detailed statement about the number of Amazon customers.

that the rating system is effective as consumers use the ratings when making their purchase decisions. Third, the finding that prices depend on the ratings – and thus presumably indirectly on the consumers' willingness to pay – indicates that pricing on the Amazon marketplace is not purely cost-based. This could indicate that marketplace sellers have a degree of market power, possibly due to captive customers. My analysis has brought up additional interesting findings, for instance, a higher share of non-verified reviews for Amazon than for pure third-party offers, that I aim to investigate against the backdrop of possible self-preference and review manipulation in a follow-up project.

References

BORK, R. H. (1978). The antitrust paradox: A policy at war with itself. *New York*.

BUNDESKARTELLAMT (2019). Bundeskartellamt / Verbraucherschutz, Bundeskartellamt leitet Sektoruntersuchung zu Nutzerbewertungen ein. https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Pressemitteilungen/2019/23_05_2019_SU_Nutzerbewertungen.pdf?__blob=publicationFile&v=2. Accessed: 2020-08-12.

CMA (2020). Facebook and ebay pledge to combat trading in fake reviews. https://www.gov.uk/government/news/facebook-and-ebay-pledge-to-combat-trading-in-fake-reviews. Accessed: 2020-07-23.

EUROPEAN COMMISSION (2005). Directive 2005/29/EC of the European Parliament and of the Council of 11 may 2005 concerning unfair business-to-consumer commercial practices in the internal market. *Official Journal of the European Union*.

EUROPEAN COMMISSION (2014). Directive on certain rules governing actions for damages under national law for infringements of the competition law provisions of the member states and of the European Union. *Official Journal of the European Union*.

EUROPEAN COMMISSION (2019). Directive (EU) 2019/2161 of the European Parliament and of the Council: of 27 november 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules. *Official Journal of the European Union*, **L 328**, 7–28.

HART, O. and TIROLE, J. (1990). Vertical integration and market foreclosure. *Brookings Papers on Economic Activity. Microeconomics*, pp. 205–286.

NORMANN, H.-T. (2009). Vertical integration, raising rivals' costs and upstream collusion. *European Economic Review*, **53** (4), 461–480.

ORDOVER, J. A., SALONER, G. and SALOP, S. C. (1990). Equilibrium vertical foreclosure. *The American Economic Review*, **80** (1), 127–142.

POSNER, R. A. (1979). The Chicago School of antitrust analysis. *University of Pennsylvania Law Review*, **127**, 925–948.

1

The Effects of Private Damage Claims on Cartel Stability: Experimental Evidence

Co-authored with Olivia Bodnar, Melinda Fremerey and Hans-Theo Normann

1.1 Introduction

In the airline-cargo cartel case, Lufthansa was the whistleblower and received full immunity from fines but was soon after sued privately by Deutsche Bahn for damages amounting to 1.76 billion euros.[1] Would Lufthansa have blown the whistle had they anticipated these damage claims? Do such private damages not provide a strong disincentive to report cartels and apply for leniency? In this paper, we try to answer these questions with evidence from laboratory experiments.

Largely driven by the introduction of leniency programs, cartel authorities can look back at successful years of public cartel enforcement.[2] Leniency policy offers companies involved in a cartel who self-report either total immunity from fines or a reduction in the fines which the authorities would have otherwise imposed on them (European Commission, 2006). As theoretical, empirical, and experimental work shows, leniency policy has a deterrent effect on cartel formation and, as it yields distrust among cartel members, it destabilizes the operations of existing cartels (see, for example, Bigoni *et al.*, 2012; Brenner, 2009; Harrington and Chang, 2009; Miller, 2009; Motta and Polo, 2003; Spagnolo, 2003). For a survey of the research on leniency programs, see Spagnolo (2008).

Damage claims — customers of a cartel may sue convicted wrongdoers for the loss they suffer in civil lawsuits — add an element of private enforcement to anti-cartel policy. Private damage claims have only recently gained attention in Europe. The European Commission started to consider private enforcement with its 2005 Green Paper (European Commission, 2005). It was signed into law in November 2014. In 2018 the last member states implemented the directive on antitrust damages actions into national law (European Commission, 2014, 2018). In

[1] See Kiani-Kreß and Schlesiger (2014) and Michaels (2014). At least initially, private damages far exceeded the fines which, eventually, summed up to 776 million euros (see European Commission (2017a)).

[2] For example, MAN revealed the EU-wide truck cartel (1997–2011) and received full immunity from the European Commission (EC). Further examples are the vitamins cartel (around 1985–1999) and the air cargo cartel (1999–2006), in which the EC and the US Department of Justice granted full immunity to Rhône-Poulenc, respectively Lufthansa, for revealing the cartel (Department of Justice, 2007, European Commission, 2001, No.(124), 2016, No.(31), 2017b, No.(28)).

the US, private damage claims have existed since the early 20th century. Private enforcement is viewed as an important and long-standing antitrust policy tool since public enforcement is restricted to litigation in order to impose fines on cartel members (Canenbley and Steinvorth, 2011).[3] Despite these differences, private damages now constitute a significant dimension of cartel policy both in the EU and the US.

At first sight, it seems that private damage claims nicely complement public enforcement. They raise the expected penalty for forming a cartel and therefore add to the deterrent effect of the fines imposed by antitrust authorities. Becker (1968) argues that increased sanctions decrease criminal activity.[4, 5] Private damage suits constitute a sanction and should accordingly reduce the criminal activity of explicit collusion.

There are, however, growing concerns about the negative effects of private enforcement. As the Lufthansa example shows, the detrimental impact that compensation payments for damaged parties have on the attractiveness of leniency programs are evident. Whereas penalties are waived or reduced for cooperating leniency applicants, the European Damages directive gives only restricted protection against third-party damage claims (European Commission, 2014).[6] The effect is aggravated by the fact that cartel members are jointly liable for the entire damage caused by the cartel, and compensation payments are not capped, in contrast to fines

[3]Private damage claims account for 90 to 95 percent of US cartel cases (Knight and Ste. Claire (2019)). US law incentivizes private lawsuits, for example, by making the infringer liable for treble damages and by admitting class action suits (§ 4 Clayton Act, 15 U.S.C. § 15; Jones, 2016).

[4] More recently, Bigoni *et al.* (2015) and Chowdhury and Wandschneider (2018) provide experimental evidence of the deterrent effect of penalties on cartels. See also below.

[5] An additional point in favor of private damages, raised by Knight and Ste. Claire (2019), is that private damages can reduce the profitability of sustained collusion. Cartels are no longer monitored by time- and money-constrained competition authorities only, but also by possible private plaintiffs. A higher detection probability reduces the profitability of a cartel, accordingly. This argument is also supported in the work by Land and Davis (2011).

[6] We will henceforth take a European perspective of this issue in that an existing leniency program was possibly weakened by the introduction of private damages. In the US, private damages predate leniency programs and so the existing anti-cartel policy was strengthened by the introduction of leniency. Nevertheless, the trade-off private damages also apply to US antitrust policy. This trade-off, however, might be weakened due to the US antitrust law's limitation of the leniency applicant's liability to single, instead of treble, damage compensation payments.

which may not exceed 10% of annual turnover (European Commission, 2011). With respect to private damage claims, the European legislation restricts the liability of leniency applicants for the harm caused to their own direct and indirect purchasers. In any event, applicants remain fully liable when non-applicants are not able to entirely compensate the injured parties (European Commission, 2014, Rn(38)). In comparison, the US antitrust law limits the liability of leniency applicants to single, instead of treble, damage compensation payments (Antitrust Criminal Penalty Enhancement and Reform Act of 2004, Sec. 213.).

The literature appears to largely acknowledge this artificially created trade-off between private damage claims and public leniency programs. Canenbley and Steinvorth (2011), Cauffman and Philipsen (2014), Knight and de Weert (2015), Migani (2014), Wils (2003), Wils (2009) argue informally, and Kirst and van den Bergh (2016) formally, that it is less desirable for firms to apply for leniency when they are liable for private damage claims. The higher the expected third-party claims, the lower the incentives to apply for leniency. As this is also anticipated by other cartel members, it could have a stabilizing effect on cartels as Hüschelrath and Weigand (2010) argue in a theory paper. Buccirossi *et al.* (2015) show in an experiment that a leniency applicant might become an easy target of damage suits due to its self-identification as guilty. This raises the question of whether applying for leniency remains attractive after the introduction of private damage claims.

In the end, it is an empirical question whether private damage claims strengthen or weaken the deterrence effects of public enforcement. On the one hand, higher fines should increase deterrence. On the other hand, they may render leniency ineffective. Somewhat surprisingly, we have not been able to find any sound empirical assessment of the effects of private enforcement. Figure 1.1 shows the number of EU cartel cases since 1990. Cartel cases rose sharply in 2000–2004 with the introduction of leniency programs but they are now in decline. This recent drop in cartel cases coincides with the EU's introduction of private damage claims in 2014. Could this decline have been triggered by private damages? The descriptive numbers in figure 1.1 cannot identify a causal effect of private damages as many factors are uncontrolled for; foremost, because there are no undetected cartels in the sample, of course.

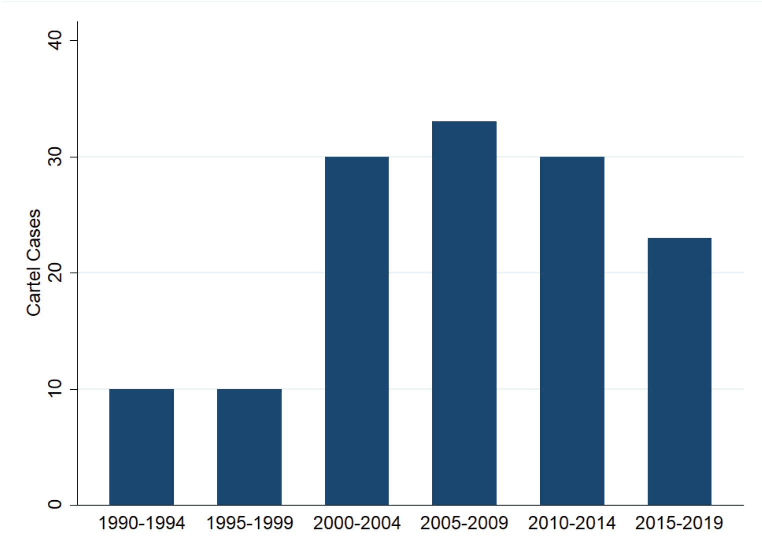


Figure 1.1: Cartel cases decided by the European Commission 1990–2019. Source, European Commission (2019, section 1.10).

We propose an experimental approach to study the effects of private damages empirically. Laboratory experiments present a readily available testbed which is unaffected by the sample-selection problems, which may bias field-data studies. Bigoni *et al.* (2012) mention that it is difficult to evaluate the deterrent or stabilizing effects of antitrust policies compared to other law enforcements because the number of cartels and changes in cartel formation is unobservable.[7] Experiments can be a useful instrument for the evaluation of new policy tools and for analyzing the effects of cartel stability *ceteris paribus*.

We build on – and extend – an established experimental literature on the effects of leniency programs. Apesteguia *et al.* (2007) examine the effect of leniency programs in one-shot Bertrand games. They find that the implementation of the leniency

[7] See Miller (2009) and Harrington and Chang (2009) for empirical identifications of policy effects on the number of detected cartels or cartel duration.

rule tends to increase self-reporting and decrease cartel formation, and leads to significantly lower market prices. Bigoni *et al.* (2012) and Hinloopen and Soetevent (2008) analyze the repeated game in Bertrand duopolies and triopolies, respectively.[8] The main result of this literature is that an introduction of leniency leads to a reduction in cartel formation.[9] This literature has not studied the effect of private damage claims on leniency programs.

A second dimension along which we extend the literature is that we compare structured and free chat-like communication between participants. Some experiments analyze structured communication in the form of price announcements among players where subjects have boilerplate messages available (Bigoni *et al.*, 2012; Hinloopen and Soetevent, 2008). In the context of cartels, both structured communication and chat seems plausible. Cheap talk is recognized as an important tool for the coordination of cooperative outcomes in experiments (Blume and Ortmann, 2007; Camera *et al.*, 2011; Cooper *et al.*, 1992). In the field of antitrust, experiments identify this kind of chat as a powerful device for fostering collusion (Brown Kruse and Schenk, 2000; Cooper and Kühn, 2014; Fonseca and Normann, 2012; Waichman *et al.*, 2014). While the comparison of chat to structured price announcements has been made for collusion experiments without leniency (recently, Harrington *et al.* (2016)), it seems promising to conduct this comparison with the inclusion of leniency. Likewise, Apesteguia *et al.* (2007) and Dijkstra *et al.* (2018)

[8] Bigoni *et al.* (2012) and Hinloopen and Soetevent (2008) differ in further various elements of the experimental design (for example, product differentiation and number of supergames). A significant difference to Hinloopen and Soetevent (2008) and our setup is that Bigoni *et al.* (2012) allow for reporting at any stage, even before prices are set. In this way, the Bigoni *et al.* (2012) design avoids a potential drawback: When firms can apply for leniency only after prices are observed, it becomes a dominant strategy for all firms to apply for leniency which may reduce the gains from deviating. We do not believe, though, that the drawback matters much because the simplification is constant across treatments in our paper, and any bias it may induce should not affect treatment differences. See the design section for details.

[9] Hinloopen and Onderstal (2014) study the effects of leniency on bidding rings in auctions. Bid-rigging is also analyzed in Luz and Spagnolo (2017) with a novel focus on the effect of corrupt officials involved in the cartelization. Feltovich and Hamaguchi (2018) find that leniency also has a pro-collusive effect due to the lower cost of forming a cartel. This effect is, however, offset by firms' reporting, so the overall effect on collusion is negligible. Clemens and Rau (2018) investigate leniency policies that discriminate against ringleaders and find that this, paradoxically, stabilizes collusion. Andres *et al.* (2019) add an innovative element to the experimental leniency literature by having participants play the role of the cartel authority. In a cartel experiment without leniency, Gillet *et al.* (2011) investigate how the managerial decision-making process affects cartel formation and pricing.

conduct leniency experiments with chat communication but do not compare to non-chat forms of communication.[10]

Our experiment is designed to analyze the effects of private damage claims on leniency, cartel formation, and cartel stability. We have the following main research questions. First, do we observe fewer cartels being established following the introduction of private damage claims? Second, is there a decreasing rate of leniency applications due to private damages? Third, what is the overall balance in terms of cartel prevalence?

The experimental design is largely based on Apesteguia *et al.* (2007), Bigoni *et al.* (2012) and Hinloopen and Soetevent (2008). Subjects play a repeated homogeneous-goods Bertrand triopoly game. They decide whether they want to engage in collusive behavior by communicating about prices, and we vary the communication format available to subjects. We investigate settings with and without private damage claims.

Our results – that are based on a comparison of exisiting private damages to a benchmark in which damage claims are not present at all – are as follows: We show that cartel formation at the individual and the group level is significantly lower with private damage claims. When private damage claims apply, leniency application rates are lower and, therefore, cartels are more stable. Overall, the balance is positive as there is an altogether significantly lower level of cartel prevalence. The effect on consumer welfare depends on the form of communication. Private enforcement significantly decreases average prices and therefore increases consumer surplus when communication is structured. In a treatment with chat communication, prices tend to significantly increase with private enforcement.

In an extension of our experiment, we show that leniency and damages can be complementary tools that reinforce cartel deterrence and maintain leniency incentives, provided the first leniency applicant is protected from damage claims. This

[10] Landeo and Spier (2009) demonstrate anticompetitive effects of chat-like communication in the context of exclusive dealing.

extension resembles the former Hungarian legislation, in which the first leniency applicant was the payer of last resort – liable only if other cartel members cannot cover their damages (see e.g., Buccirossi *et al.*, 2015). This gives a first hint that the conflict between leniency and damages can be removed by a change in the design of the current legislation.

The article is organized as follows: The subsequent section describes the experimental design and explains the treatments in detail. Section 1.3 presents our hypotheses which are the basis for our further analyses in section 1.4. Section 1.5 provides insights of an additional treatment that protects the leniency applicant from damage suits. We conclude in section 1.7.

1.2 The experiment

1.2.1 General setup

The market model underlying the experiment is a symmetric three-firm homogeneous-goods Bertrand oligopoly.[11] Demand is inelastic and {101, ..., 110} is the choice set of prices. Firms have constant marginal costs of $c = 100$. There is repeated interaction: the three players are grouped together in one market for the entire duration of the experiment (at least 20 periods).

In our experiment, firms can form cartels, report any existing cartel to a fictitious cartel authority in order to get immunity from leniency, and may face penalties and private damage claims. Our treatments vary with the implementation of private damage claims and the form of communication. The sequence of events in our experiment is as follows:

1. Decision whether to form a cartel; if all firms agree, communication is enabled and (non-binding) agreements on prices are possible,

[11] Dufwenberg and Gneezy (2000) show that the Bertrand solution is viable for randomly rematched markets with three and four firms but not for two. Huck *et al.* (2004) find that repeated Cournot markets with four or five firms do not behave collusively. See also Roux and Thöni (2015) for a more recent study.

2. Price decision,

3. Decision whether to report a cartel; unreported cartels may be detected by the cartel authority; in either case a fine is imposed,

4. Private damage claims.

We now explain these stages in turn.

1.2.2 Detailed account of the stages of the experiment

Stage 1. The three firms simultaneously and independently decide whether they want to establish a cartel. They press either the *discuss price* or the *do not discuss price* button on the computer screen. Only if all three firms decide to participate in price discussions is a cartel established, and a communication window opens. Depending on the treatment, firms have access to either structured or free chat communication (see section 1.2.3).

Stage 2. Firms simultaneously and independently choose an integer price from the set $\{101, ..., 110\}$. The lowest price among the three ask prices p_i with $i \in \{1,2,3\}$ is the market price, denoted by $\underline{p}$. Only firms that bid $\underline{p}$ are able to sell their product (Bertrand competition). The inelastic demand is normalized to one, so firm i's profit is:

$$\pi_i \quad = \quad \begin{cases} \dfrac{p_i - c}{n} & \text{if } p_i = \underline{p} \\ 0 & \text{if } p_i > \underline{p} \end{cases}$$

where $n \in \{1,2,3\}$ is the number of firms charging $\underline{p}$. Firms learn $\underline{p}$ and their own profit as feedback afterwards. Profit is the gain resulting from the market interaction, which may subsequently be reduced by penalties and private damage claims.

Stage 3. Firms decide whether to report any existing cartel to the authority and thereby apply for leniency. Reporting costs $r = 1$ point (the experimental currency unit) that represent legal fees for filing a leniency application. There is a "race to report": the first leniency applicant gets a 100% fine reduction and the second applicant gets 50%; the third applicant does not receive a reduction. If no participant reports the cartel, it may still be detected by the authority, namely with a probability of $\rho = 0.15$ in each period. If a cartel is detected (either through a whistleblower or the random draw of the authority), each cartel member has to pay a fine, F, equal to 10% of the current period revenue.[12], [13]

Stage 4. Private damage claims may occur after a cartel is detected. Since we do not include cartel customers in our experiment, this stage is not a decision. Rather, the damage claims are simply enforced with a probability of $\sigma = 0.95$.[14] If the private enforcement case is won in favor of the injured party, the cartel has to compensate 60% of the total damage.[15] The damage inflicted is the difference between the cartel price and the competitive (Nash equilibrium) price, 101 (European Commission, 2014, Rn(39)), summed over the number of periods, T, where the cartel exists. A cartel is established once all firms in one group decide to communicate by clicking the *discuss price* button. A cartel formally exists as long as it is not reported by a cartel member nor detected by the cartel authority in stage 3. In consequence, the cartel continues to exist even if one or more cartel members deviate from the price agreed upon during the communication phase. Similarly, a cartel continues

[12] The revenue is defined as the quotient of the market price and the number of firms that sell at market price, see 1.7.1

[13] These fines are consistent with European policy, including the "race to report" (European Commission, 2002, Rn(23)b). Leniency applicants are immune or eligible to reductions of fines levied on infringers by the commission (European Commission, 2006). Those who are first to report are fully relieved from cartel fines; "subsequent companies can receive reductions of up to 50% on the fine that would otherwise be imposed (European Commission, 2011)." In line with European competition law, fines shall not exceed a maximum of 10% of a firm's overall annual turnover when the respective firm is not eligible to reductions of fines (European Commission, 2011). These parameters are also used in Bigoni *et al.* (2012) and Hinloopen and Soetevent (2008).

[14] If damage claims are brought to court, the probability that a case is won is presumably relatively high because one goal of the Directive on antitrust damages actions (European Commission, 2014) is to make it easier for injured parties to get evidence (European Commission, 2015). A large share of private damage claims are also settled out of court (Bourjade *et al.*, 2009).

[15] For two reasons it is reasonable to assume that the total damage is not compensated. First, not all buyers will claim damages, for example, because the buyer structure might be fragmented or because it is costly to open a case. Second, it could be the case that part of the damage is passed on in the value chain. The passing-on argument can serve as a strategy of defense of the cartel members against a claim for damages (European Commission, 2014, Rn(39)).

to exist even if cartel members communicate only once in the very beginning of the cartel or stop communicating for any number of periods in-between. For each period in which a cartel formally exists, the cartel price is defined as the market price in the given period.

According to the European Commission (2014, Rn(37)) cartel members are jointly liable for the whole damage, therefore, each cartel member has to pay one third of the damage compensation. The per-firm per-period damage reads $D_i = \frac{1}{3}(p - 101) \cdot 0.6$ where p is the price the cartel charges in some period and 101 is the counterfactual (Nash) price. For example, fixing the cartel price at 110 (the maximum possible price), the compensation each cartel member has to pay for each period of the cartel's duration is $\frac{1}{3} \cdot (110 - 101) \cdot 0.6 = 1.8$. Table 1.1 summarizes the calculation for the damages and draws a comparison to fines.

	Fine	Private damage claims
Probability of imposition (if caught)	100%	95%
Basis	Current period firm revenue	Cumulated damage
Magnitude	10%	60% jointly
Possibility to reduce	Yes	No

Table 1.1: Fines and private damage claims of one firm.

1.2.3 Treatments

Our main treatment variable is the presence of private damage claims in stage 4. In the treatment labeled NOPDC, they are absent (there is no stage 4). In treatment PDC, they are potentially imposed. We conduct these two treatments *within subjects*: participants first play NOPDC and then PDC.[16]

[16]This within-subjects design allows us to observe cartels that were set up before the introduction of the PDC rule, such that the new PDC come unexpectedly for existing cartels. Empirically, it turns out there are only few such cases, so we refrain from exploiting this advantage of the experimental design.

Periods	1 ... 9	10	11 ... end
Treatment	NOPDC	NOPDC, introduce PDC after stage 2	PDC
Stages	1	1	1
	2	2	2
	3	3	3
		4	4

Table 1.2: Within-subjects variation of private damages.
Participants first play nine periods of NOPDC (stages 1–3). In period 10, the new PDC rule (stage 4) is announced after stage 2. Then, subjects play PDC (stages 1–4) for the remainder of the experiment.

Each experimental session consists of at least 20 rounds. From period 20 onwards, the session ends with 20% probability. Such a random termination rule is suitable for avoiding end-game effects (Normann and Wallace, 2012). As table 1.2 shows, subjects play nine periods of NOPDC. In period 10, the rules of the game change as we introduce private damage claims, after stage 2 (see table 1.2). From period 11 on, they play PDC for the rest of the experiment. The instructions mention that the rules might change during the course of the experiment, but they did not indicate when the change would occur nor what it would entail.[17]

In the field, private damage claims were introduced after and in addition to existing public enforcement, justifying the sequence NOPDC-PDC on which we focus in our experiment. For the sake of completeness, the reverse order PDC-NOPDC may seem warranted. We accordingly conduct sessions with the reverse order of treatments. Thereby, we can control for possible order effects by comparing the first 10 periods of each treatment sequence, for example, the first 10 periods of NOPDC–PDC with the first 10 periods of PDC–NOPDC. In the reverse-order variant, stage 4 is removed (rather than added) in period 10.

As mentioned, we also modify the communication format in two treatments.

[17] An alternative setup would have been to repeat the supergames in order to facilitate learning. This, however, would have precluded the within-subjects "before and after" evaluation of private damages which we considered essential for external validity.

This treatment variable is analyzed *between subjects*, that is, the treatment of different communication designs is done in separate experimental sessions. Potential carry-over effects (hysteresis) of the different communication formats make a within-subjects design unappealing in this case.

The communication formats are labeled CHAT and STRUC. (The procedure of structured communication (STRUC) closely follows Hinloopen and Soetevent (2008). It resembles experiments where subjects may announce prices non-bindingly but cannot communicate otherwise (Harrington *et al.*, 2016; Holt and Davis, 1990)). Hence, in sessions with STRUC, participants are only able to suggest a price range for which the good could be sold. Specifically, subjects can enter a minimum and a maximum price (within the range of $\{101, ..., 110\}$) in the communication window. In subsequent rounds of price discussions (in the same period), subjects can choose prices from the intersection of all three suggested price ranges from the preceding discussion. If no intersection exists, subjects can choose a price from the complete price range. This iterative process lasts until either the subjects (non-bindingly) agree on a common price or after 60 seconds have passed (which, according to Hinloopen and Soetevent (2008), is sufficient. After the communication phase has ended, subjects get feedback on the agreed upon price or the price interval.

In sessions with CHAT, subjects can freely communicate in a chat window. We allow for open communication, letting subjects exchange any information they want (except for offensive messages, or messages identifying participants). After 60 seconds, the chat window closes and subjects enter stage 2. Among others, Cooper and Kühn (2014), Fonseca and Normann (2012) and Harrington *et al.* (2016) have used similar chat devices in oligopoly experiments. Brosig *et al.* (2003) generally investigate the issue of the communication format on cooperation.

Table 1.3 summarizes our treatments. It also indicates the number of groups and participants for each treatment. We introduce and analyze an additional treatment, labeled PDC+ and also involving 48 subjects, in section 1.5.

Sequence	Communication	Number of indep. groups	Number of participants
NOPDC - PDC	STRUC	16	48
NOPDC - PDC	CHAT	16	48
PDC - NOPDC	STRUC	16	48
$\sum$		48	144

Table 1.3: Overview of treatments.

1.2.4 Procedures

The experimental sessions were conducted in the summer and fall of 2018 at the DICE-Lab of Duesseldorf University. We had a total of 192 participants. Subjects were students from all over campus. They had previously indicated their general willingness to participate in lab experiments by registering for our database and were then recruited for this experiment using ORSEE (Greiner, 2015).

Upon arrival at the DICE-Lab, subjects were welcomed and allocated to isolated computer cubicles. We used a randomization device to assign the cubicles. After all participants were seated, they were given written instructions. Subjects were given ample time to read the instructions and they had the opportunity to ask the experimenter questions (in private). Then the actual experiment began.

During period 10, the experiment was interrupted and a second set of written instructions (which explained the change regarding private damages) was distributed. The change of rules was also announced on the computer screen and was checked with control questions.

The experiment was programmed using z-Tree software (Fischbacher, 2007). Sessions lasted about one hour on average. Payments were as follows. Participants received an initial capital of 5 euros. Cumulated payoffs are added to or subtracted from the initial capital. The exchange rate was one point equal to 0.3 euros. The average payment was 13.08 euros.

1.3 Hypotheses

In this section, we will use the following notation (for a comprehensive overview of all variables and their numerical realizations in the experiment, see Appendix 1.7.1). The collusive profit per firm is denoted π_i^c. In the static Nash equilibrium, firms earn π_i^n. The profit of a defecting firm is denoted π^d. Reporting costs are r. Unless reported, a cartel is detected by the authority with a probability ρ and, if so, the authority imposes a fine F_i^j per firm i and outcome $j \in \{c, d, n\}$, with c for collusion, d for deviation and n for Nash. A busted cartel faces damage claims with probability σ. The per-firm per-period damage is denoted by D_i^j. Damages are cumulated over time. Fines and damages depend on the cartel price and thus differ in periods of collusion and defection.

We assume that the market game is repeated infinitely many times and that firms discount future profits with a discount factor δ. Firms collude on the maximum price (110) and use a simple Nash trigger to support collusion, such that the static Nash profit, π^n, is the punishment profit after a deviation.[18] For simplicity and following Bigoni *et al.* (2015), we assume that firms communicate once to establish successful collusion and collude tacitly after a detection by the authority. That is, firms risk being fined only once.[19] Formal proofs of the statements in this section can also be found in the Appendix 1.7.1.

Our first hypothesis is about cartel formation, that is, the number of newly formed cartels. The economic theory of crime predicts that criminal activity decreases in the expected costs of the activity (Becker, 1968). We derive this formally (see Appendix 1.7.1 for details) from the cartel's *participation constraint* which must necessarily be met, see also Bigoni *et al.* (2015) or Chowdhury and Wandschneider

[18] Colluding on the maximum price seems plausible as this maximizes joint profits. It is possible, however, to lower the threshold discount factor by choosing a lower collusive price. Since this effect is of minor magnitude and similar in all treatments (and hence does not affect our hypotheses), we refrain from exploring this issue in detail. We further note that punishments more severe than Nash are not feasible here because the Nash price is also the lowest price firms may charge.

[19] Alternatively, we could assume that successful cartels immediately resume the collusion after a detection. This leads to qualitatively similar results but implies a more cumbersome derivation of the damage payments.

(2018). The expected discounted profit from colluding minus the expected fine (left-hand side of the equation) must be at least as high as the expected discounted profit from competing à la Nash (right-hand side of the equation). For the NOPDC case, we have

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) \quad \geq \quad \frac{\pi_i^n}{1-\delta}$$

where $E(F_i^c)$ is the expected discounted fine. Private damage claims increase the expected costs of cartel formation because firms now need to cover the expected damages in addition to the fines. For PDC, the cartel participation constraint reads

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) - E(D_i t^c) \quad \geq \quad \frac{\pi_i^n}{1-\delta}$$

where $E(D_i t^c)$ is the expected, discounted cumulated, per-firm damage payment resulting from successful collusion. For our experimental parameters, both participation constraints are met, but, with private damages, the cartel participation constraint is more severe. We thus maintain:

Hypothesis 1. ***(Cartel formation)*** *Private damage claims reduce the number of cartels.*

The next hypothesis concerns the reporting behavior of firms: In which treatment will firms apply for leniency more often? We assume firms report a cartel only when they also deviate from the cartel price (reporting and not deviating makes no sense because the cartel will cease to exist after the report anyhow). Deviating from the cartel price happens only off equilibrium so, in theory, reporting behavior should never occur in any treatment. We can, however, compare the cost of reporting across treatments. In treatment NOPDC, reporting only involves r, the immediate cost of reporting. In treatment PDC, firms also incur r but they additionally need to pay damages σD_i^d. For the experimental parameters, it turns out that reporting costs are more than 2.5 times higher under PDC than under NOPDC (Appendix 1.7.1). As the cost of reporting and applying for leniency increases with private damages, we hypothesize:

Hypothesis 2. ***(Leniency)*** *Private damage claims reduce the frequency of leniency applications.*

We now analyze the dynamic incentives to collude. As mentioned, firms attempt to maximize joint profits with a trigger strategy involving Nash reversion. Cartel firms remain liable for the agreement in future periods, until detected or reported. The incentive constraints required for collusion to be a subgame perfect Nash equilibrium read as follows. Without private damages (NOPDC), sticking to the collusive agreement is (weakly) better than defecting if

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) \quad \geq \quad \pi_i^d - r + \frac{\delta \pi_i^n}{1-\delta}.$$

With private damages (PDC), colluding is better than defecting if

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) - E(D_{it}^c) \quad \geq \quad \pi_i^d - r - \sigma D_{it}^d + \frac{\delta \pi_i^n}{1-\delta}$$

where we note that damages have to be paid in either case, but they differ in magnitude (see Appendix 1.7.1 for details). Let the minimum δ that solves the NOPDC and PDC incentive constraints be δ_{min}^{NOPDC} and δ_{min}^{PDC}, respectively. We find that

$$\delta_{min}^{PDC} < \delta_{min}^{NOPDC}.$$

That is, private damage claims facilitate collusion. For the parameters in the experiment, we obtain $\delta_{min}^{NOPDC} = 0.664$ and $\delta_{min}^{PDC} = 0.655$. With a continuation probability of 0.8, both incentive constraints are met in the experiment and so collusion is an SGPNE in either case. We follow the frequently adopted interpretation that a lower minimum discount factor suggests that collusion is more stable. Hence, we state:

Hypothesis 3. ***(Cartel stability)*** *Cartels are more stable when private damage claims are possible.*

An interesting observation is that reporting costs and the incentive constraint under private damages become more severe over time because damages are cumulated. Deviations become more and more costly in later periods. Private damages accordingly have a self-enforcing effect on collusion. In theory, this effect is immaterial, though. All that matters is whether the incentive constraint is met in period zero when the incentive to deviate is at its maximum. The fact that the bill for reporting gets higher and higher could be important, though. For example, unanticipated shocks to collusion may be absorbed only with the high exit cost that the cumulated damages imply.

Our hypotheses suggest an overall ambiguous effect of private damage claims. On the one hand, there should be fewer cartels. On the other hand, cartels should be more stable and there may be less reporting in PDC. The overall balance in terms of cartel prevalence is ex ante not clear and we do not maintain a directed hypothesis here.

Statement 4 (Cartel prevalence) *The overall effect of private damage claims on cartel prevalence is ambiguous.*

As with cartel prevalence, we do not maintain a directed hypothesis about market prices (the measure for consumer welfare). Market prices (the lowest of the three ask prices) are affected by (at least) two channels. First, market prices may decrease because, according to hypothesis 1, fewer cartels are formed with private enforcement, leading to more competitive prices. Second, any existing cartels would suffer less from leniency (hypothesis 2) and may be more stable (hypothesis 3) and should therefore have higher market prices on average. The overall effect is ambiguous. Of course, we can look at the effect of PDC for cartelized markets only. But, even here, the effect is ex-ante ambiguous. On the one hand, cartels under PDC may collude more successfully due to a selection effect (only rather collusive-minded firms form a cartel despite the more severe constraints). On the other hand, cartel members could fear damage claims and therefore lower the prices.

Statement 5 (Market prices) *The overall effect of private damage claims on market prices is ambiguous.*

Our final hypothesis is about the impact of the different forms of communication. Existing experimental evidence (Cooper and Kühn, 2014; Fonseca and Normann, 2012) suggests cartels are more stable when subjects can communicate. It appears that open communication fosters trust between players (Brosig *et al.*, 2003). Also, subjects can communicate entire strategies rather than just price targets. Furthermore, chat communication can enhance the understanding of the mutual benefits of collusion in their group. Brown Kruse and Schenk (2000) observe that only one group member has to understand the profit-maximizing strategy and can use the chat to convince its group members to comply.

Hypothesis 6. ***(Impact of communication)*** *Compared to structured communication, unrestricted communication increases cartel formation and stability.*

1.4 Results

To analyze the impact of private damage claims, we foremost analyze the data within subjects. That is, we compare the first 10 periods (NOPDC) to the subsequent 10 periods (PDC). We restrict the analysis to observations from periods 1 to 20 in order to exclude potential end-game effects. With the help of the reverse-order control treatment, we then compare the data between subjects to exclude possible order effects (both PDC and NOPDC data from periods 1 to 10). We use non-parametric tests like the Wilcoxon matched-pairs test (WMP) for the within-subjects analysis and the Mann-Whitney-U test (MWU) for the between-subjects analysis. With the WMP-Test, we match the NOPDC with the PDC observations of each group. For all analyses, we first take the average per group as one observation and aggregate across groups afterward. In total, we have 16+16 observations. When we analyze the share of firms that report a cartel, we generally have fewer observations because the analysis is conditional on having a cartel in the first place which is not the case for all groups.

We complement the non-parametric tests with linear regression models (ordinary least squares), often linear probability models, with and without time fixed effects. We run the estimations separately for each communication treatment. Due to the fixed group structure, we cluster standard errors at the group level. We bootstrap the standard errors with 1,000 replications. Statistical significance levels are indicated by an asterisk, where $^{+}$ ($p < 0.15$), * ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.01$). We report two-sided p-values throughout.

An overview of the summary statistics of our main results is displayed in table 1.4. The exact definition of each variable can be found in table 1.12 in the Appendix. The exact values underlying figures 1.2 to 1.9 can be obtained from table 1.4.

	STRUC		CHAT	
	NOPDC	PDC	NOPDC	PDC
Propensity to collude	0.619 (0.142)	0.394 (0.192)	0.578 (0.288)	0.225 (0.289)
Share cartel	0.207 (0.153)	0.019 (0.054)	0.271 (0.373)	0.063 (0.250)
Share report	0.462 (0.230)	0.296 (0.339)	0.103 (0.214)	0.000 (0.000)
Cartel stability	1.000 (0.000)	2.167 (0.866)	6.556 (3.522)	8.000 (1.441)
Cartel prevalence	0.238 (0.178)	0.063 (0.163)	0.325 (0.380)	0.163 (0.359)
Market price	102.706 (2.009)	101.681 (2.095)	105.913 (3.969)	107.038 (4.227)

Table 1.4: Summary statistics of the results in the treatments NOPDC–PDC (STRUC and CHAT); average results per treatment (standard deviations in parentheses).

1.4.1 Cartel formation

Hypothesis 1 states that cartel formation decreases when private damage claims are introduced. Consider the individual level first: how often do subjects press the *discuss price* button when they are not already in a cartel? (Table 1.12 in the Appendix provides a more detailed definition of the variable.) Without private damages, the average propensity to collude in STRUC (CHAT) is 61.9% (57.8%), see figure 1.2 and table 1.4. With PDC, the average propensity to collude decreases to 39.4% (22.5%), and the reduction is significant (STRUC: WMP, $p - value = 0.0007$; CHAT: WMP, $p - value = 0.0015$). For both communication treatments, the individual propensity to form a cartel declines by about 35–22 percentage points when PDC are possible. The estimation results of the linear probability model in table 1.5 are also consistent with hypothesis 1. We see that the dummy variable PDC is highly significant and economically substantial.

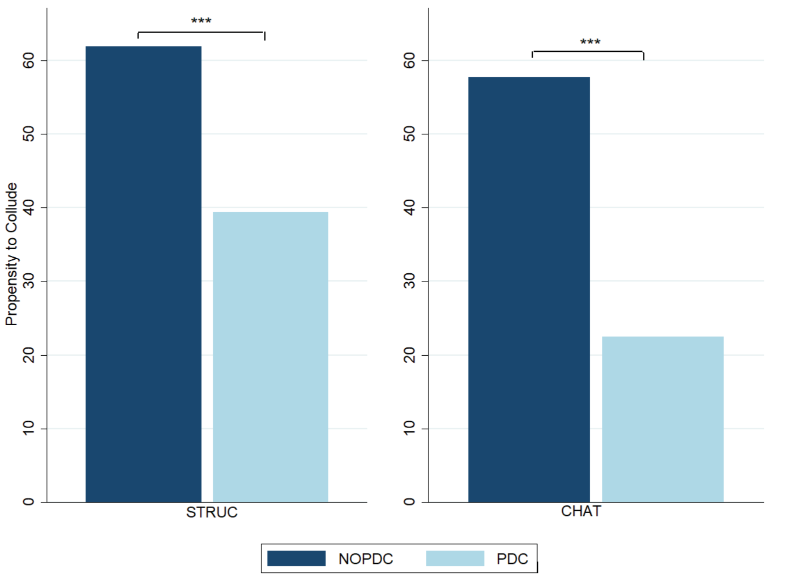


Figure 1.2: The impact of PDC on the individual propensity to collude in STRUC (left) and CHAT.

	(1)	(2)	(3)	(4)
	Collude	Collude	Collude	Collude
PDC	-0.225***	-0.219***	-0.208***	-0.604***
	(0.0353)	(0.0482)	(0.0497)	(0.0926)
constant	0.592***	0.381***	0.583***	0.729***
	(0.0350)	(0.0605)	(0.0537)	(0.0648)
Time FE	No	No	Yes	Yes
Sample STRUC	Yes	No	Yes	No
Sample CHAT	No	Yes	No	Yes
N	960	960	960	960
R^2	0.051	0.060	0.063	0.106

Standard errors in parentheses

$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Individual decisions to communicate – linear regression (standard errors in parentheses).

Next, consider the market (or group) level. Here, we ask the question, how often is a cartel actually established? This is the case when all three group members press the *discuss price* button, given they are not already in a cartel (for this and all other variable definitions consult table 1.12 in the Appendix). Figure 1.3 and table 1.4 show the results. We observe that, with PDC, the share of newly formed cartels is strongly and significantly reduced (STRUC: WMP, $p - value = 0.0007$; CHAT: WMP, $p - value = 0.0087$). As above, this holds for both communication treatments, STRUC and CHAT. The regressions in table 1.6 confirm that the effect is significant.

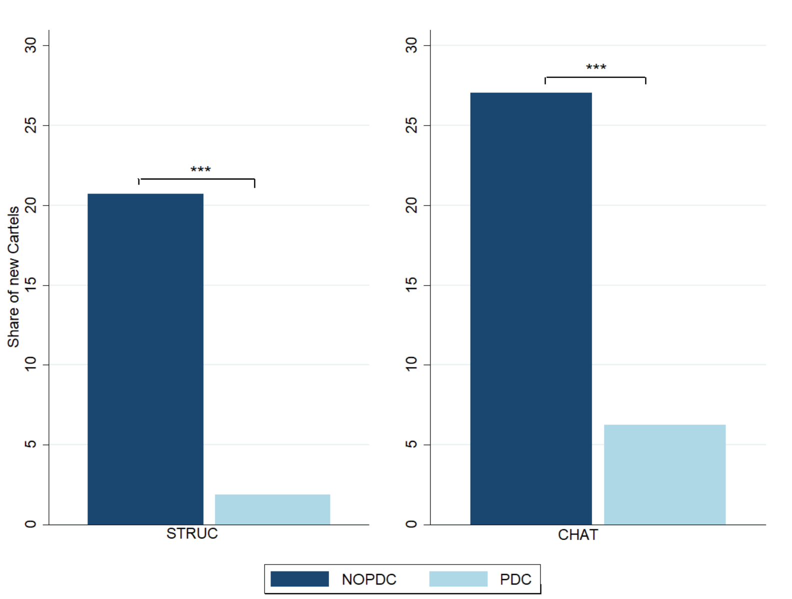


Figure 1.3: The impact of PDC on the number of cartels in STRUC (left) and CHAT.

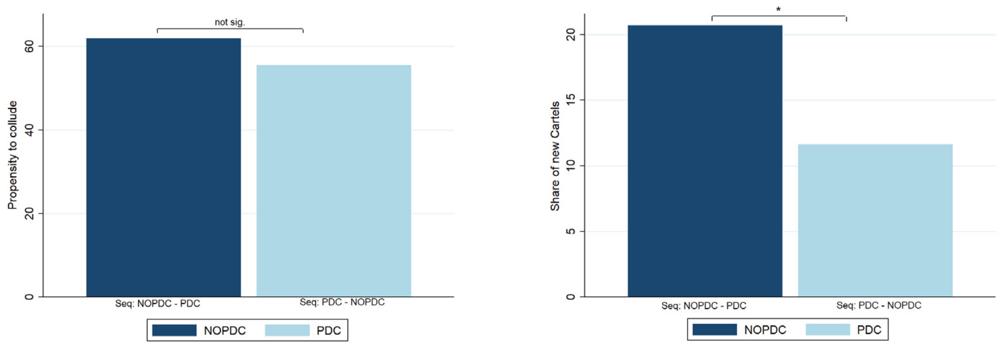
	(1) Collusion	(2) Collusion	(3) Collusion	(4) Collusion
PDC	-0.181***	-0.0813***	-0.125+	-0.375***
	(0.0311)	(0.0130)	(0.0817)	(0.116)
constant	0.194***	0.0875***	0.125+	0.375***
	(0.0344)	(0.0172)	(0.0817)	(0.116)
Time FE	No	No	Yes	Yes
Sample STRUC	Yes	No	Yes	No
Sample CHAT	No	Yes	No	Yes
N	320	320	320	320
R^2	0.089	0.037	0.119	0.183

Standard errors in parentheses

$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: Group decisions to communicate – linear regression (standard errors in parentheses).

Exploiting the treatment with the reverse sequence PDC-NOPDC with structured communication, we compare the first 10 periods of the NOPDC-PDC sequence with the first 10 periods of PDC-NOPDC sequence. This allows us to additionally conduct the comparison NOPDC and PDC between subjects, thereby excluding order effects.[20] For the sake of completeness, results of the PDC-NOPDC session analyzed within subjects can be found in the Appendix in section 1.7.6. Figure 1.4 shows that the possibility of PDC reduces cartel formation in STRUC both at the individual (a) and at the group (b) level. The reduction is statistically significant at the market level ((a) MWU, $p - value = 0.153$ (b) MWU, $p - value = 0.0899$).[21]



(a) Propensity to collude

(b) Share of cartelized markets

Figure 1.4: Cartel formation in STRUC.
Between-subjects comparison with PDC data from treatment with reverse order (PDC-NOPDC).

Result 1 (Cartel formation) With PDC, there are significantly fewer attempts to form a cartel (individual level) and significantly fewer successfully formed cartels (group level).

1.4.2 Leniency applications and cartel stability

Hypotheses 2 and 3 are about leniency behavior and cartel stability. For these analyses cartels need to have actually been formed in the first place. We compare

[20] Due to bankruptcy we exclude one group in the reverse-order treatment from our analysis.
[21] Linear regressions, available upon request, yield the same result.

the first nine periods NOPDC and period 11 to 19 PDC.[22]

Leniency applications

Hypothesis 2 suggests that there will be fewer leniency applications with PDC. We first analyze the share of individual reporting decisions by each group, that is, we consider the sum of subjects of each group revealing the cartel over all periods that any cartel exists by treatment (see table 1.12 in the Appendix for a detailed explanation of the variable *share report*).

Figure 1.5 and table 1.4 show that PDC significantly decreases the share of leniency applications in each group in STRUC (STRUC: WMP, $p - value = 0.101$; CHAT: WMP, $p - value = 0.3173$). In the case of STRUC, the effect is economically substantial.

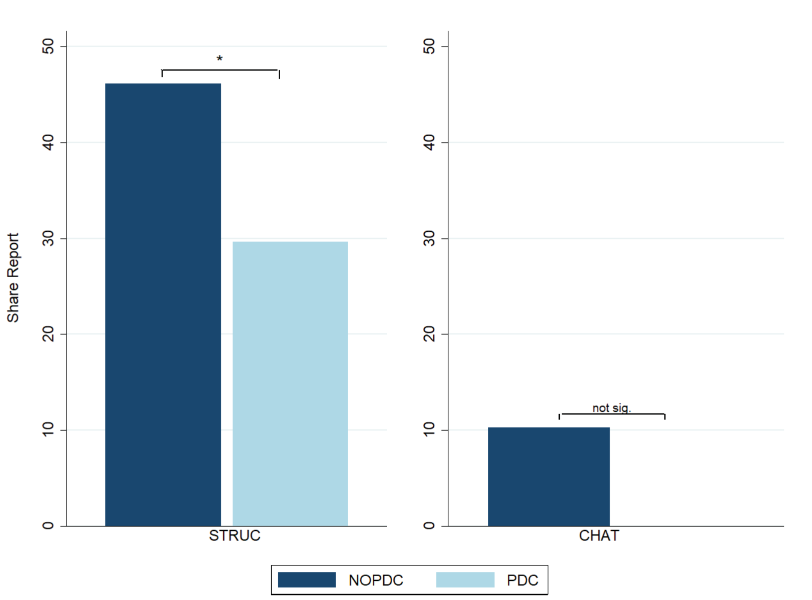


Figure 1.5: The impact of PDC on the individual reporting decision in STRUC (left) and CHAT.

Table 1.7 reports a linear regression of PDC on the individual decision to report

[22] For the analysis of leniency applications and cartel stability, we exclude period 10. Subjects decide whether to report a cartel after private damage claims are introduced. Thus, period 10 belongs to neither PDC nor NOPDC. For the analysis of variables other than stability this problem does not exist because decisions about cartel formation or price setting were made before the introduction of private damage claims. For symmetry, we also exclude period 20.

a cartel. In STRUC as well as in CHAT the number of cartel members applying for leniency decreases as PDC occur. However, this effect is only significant in the STRUC regressions without time fixed effects. The between-subjects comparison indicates that the share of leniency applications does not differ between NOPDC and PDC. Our interpretation is that subjects may have had too little time – only one repetition of the supergame – to learn the impact of private damages and are thus not more disinclined to report than in NOPDC.

	(1)	(2)	(3)	(4)
	Report	Report	Report	Report
PDC	-0.264^{+}	-0.0347	-0.167	-0.0556
	(0.178)	(0.0250)	(0.128)	(0.0494)
constant	0.412***	0.0347	0.167	0.0556
	(0.0674)	(0.0250)	(0.128)	(0.0494)
Time FE [Period 1-19, without 10]	No	No	Yes	Yes
Sample STRUC	Yes	No	Yes	No
Sample CHAT	No	Yes	No	Yes
N	129	216	129	216
R^2	0.050	0.012	0.138	0.077

Standard errors in parentheses

$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Individual decision to report a cartel – linear regression (standard errors in parentheses).

Result 2 (Leniency rate) Compared to NOPDC, there are fewer leniency applications with PDC.

Cartel stability

Hypothesis 3 is that cartels become more stable as we introduce private damage claims. In order to analyze cartel stability, we compare the mean number of periods when a cartel was stable,[23] in NOPDC and PDC, conditional on cartel existence. Cartels that are formed and uncovered in the same period count as stable for one period (see also table 1.12 in the Appendix.) Descriptive results show that the mean of cartel stability roughly doubles in STRUC (in NOPDC 1.0 stable period compared to 2.2 in PDC). In CHAT, stable periods increase from 6.6 in NOPDC to 8.0 in PDC (see table 1.4). Whereas this result is in line with hypothesis 3, we cannot make any statement about significance because there are too few groups forming a cartel in NOPDC and PDC. For the same reason, we cannot conduct survival estimates.

Result 3 (Cartel stability) With PDC, cartels are more stable.

In connection with hypothesis 3, we noted above that private damages have an enforcing effect on stability over time because damages cumulate. Cartels should, accordingly, be more strongly discouraged from reporting the longer they exist.

1.4.3 Cartel prevalence

We finally look at cartel prevalence, defined as the percentage of periods where a stable cartel existed (table 1.12 in the Appendix). Result 1 on the one hand, and results 2 and 3 on the other, suggest an overall ambiguous effect of PDC on cartel prevalence: fewer cartels are formed but these remaining cartels are more stable. (Due to this ex-ante ambiguity, statement 5 in section 1.3 is not a directed hypothesis about prevalence.) What is the overall balance?

Figure 1.6 and table 1.4 show the results. For the communication treatment STRUC, we find cartel prevalence present in 23.8% (NOPDC) and 6.3% (PDC) of all groups over all periods. In CHAT, we see 32.5% (NOPDC) and 16.3% (PDC) of periods where a stable cartel existed. That is, there is a strong and significant reduction in cartels due to PDC in both communication treatments (STRUC: $p -$

[23] A cartel is stable until it is reported or detected by the authority. Of course, cartels may continue to set a high price after being reported or detected. For such pricing behavior, they cannot be penalized.

$value = 0.0051$ and CHAT: WMP, $p - value = 0.0139$). The linear regressions in table 1.8 confirm this.

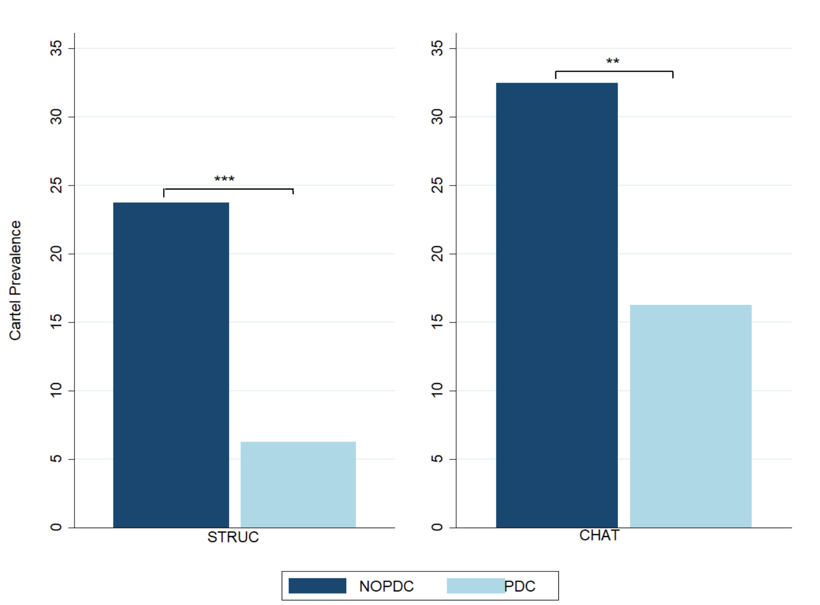


Figure 1.6: The impact of PDC on cartel prevalence in STRUC (left) and CHAT.

	(1) Prevalence	(2) Prevalence	(3) Prevalence	(4) Prevalence
PDC	-0.175***	-0.163**	-0.0625	-0.250**
	(0.0484)	(0.0797)	(0.106)	(0.105)
constant	0.237***	0.325***	0.125+	0.375***
	(0.0413)	(0.0915)	(0.0817)	(0.116)
Time FE	No	No	Yes	Yes
Sample STRUC	Yes	No	Yes	No
Sample CHAT	No	Yes	No	Yes
N	320	320	320	320
R^2	0.060	0.036	0.096	0.061

Standard errors in parentheses

$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: Cartel prevalence – linear regression (standard errors in parentheses).

We again analyze the treatment with the reverse order, PDC-NOPDC and compare the first 10 periods in NOPDC to those in PDC. The results are similar: the between-subjects test is significant (MWU, $p - value = 0.0842$).

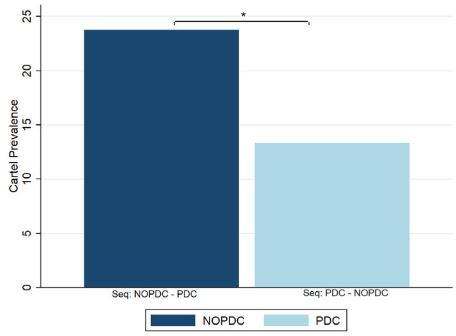


Figure 1.7: Cartel prevalence in STRUC.
Between-subjects comparison with PDC data from treatment with reverse order (PDC-NOPDC).

Result 4 (Cartel prevalence) There are significantly fewer cartelized periods with PDC.

1.4.4 Prices and consumer welfare

To complete the analysis of cartel behavior, we examine the market price. This is the lowest price of the three individually entered prices in stage 2.[24] The market price is the relevant factor for consumer welfare (see statement 5 in section 1.3).

	STRUC		CHAT	
	NOPDC	PDC	NOPDC	PDC
Market price non-cartels	102.049 (1.897)	101.589 (2.089)	104.566 (3.807)	106.621 (4.373)
Market price cartels	104.654 (2.570)	103.278 (1.669)	109.250 (2.050)	109.967 (0.058)
Market price all markets	102.706 (2.009)	101.681 (2.095)	105.913 (3.969)	107.038 (4.227)

Table 1.9: Market price – averages per treatment (standard deviations in parenthesis). Seq: NOPDC–PDC.

[24]For an analysis of individual ask prices see Appendix 1.7.5.

We compare the average market price with and without private damage claims across the CHAT and STRUC treatments as shown in table 1.9 and figure 1.8. We see that PDC reduce prices in STRUC, but CHAT shows the opposite pattern. This concerns the overall average ("all markets") as well as the market prices of cartelized and non-cartelized markets. The differences are statistically significant in the structured treatment (STRUC: WMP, $p - value = 0.0034$; CHAT: WMP, $p - value = 0.2513$). In order to control for possible order effects, we conduct the between-subjects comparison based on PDC data from the treatment with the reversed order PDC-NOPDC. Figure 1.9 verifies the lower overall market prices in PDC with STRUC communication (WMU, $p - value = 0.0511$).

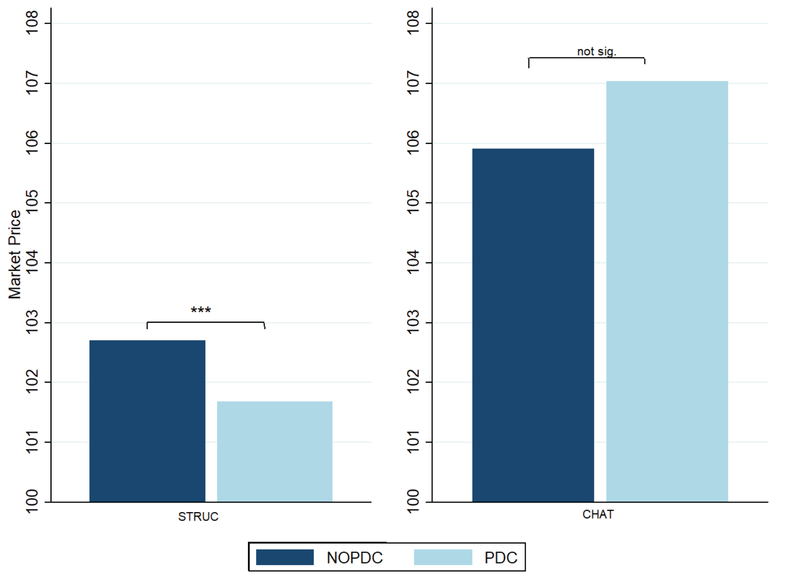


Figure 1.8: The impact of PDC on market prices in STRUC and CHAT.

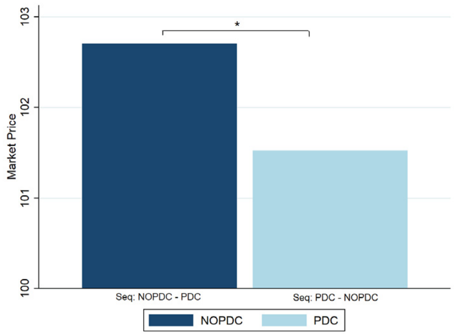


Figure 1.9: Market price in STRUC: between-subjects comparison with PDC data from the treatment with reverse order (PDC-NOPDC).

Table 1.10 reports the results from a regression analysis on the dependent variable *MarketPrice*. The results confirm previous observations that market prices significantly decrease in the subsample of STRUC if private damage claims are introduced (table 1.10, column 1). They significantly increase in CHAT.

	(1)	(2)	(3)	(4)
	MarketPrice	MarketPrice	MarketPrice	MarketPrice
PDC	-1.025***	1.125*	-1.563***	1.750+
	(0.256)	(0.588)	(0.468)	(1.174)
constant	102.7***	105.9***	102.8***	104.5***
	(0.482)	(0.957)	(0.415)	(0.981)
Time FE	No	No	Yes	Yes
Sample STRUC	Yes	No	Yes	No
Sample CHAT	No	Yes	No	Yes
N	320	320	320	320
R^2	0.044	0.017	0.060	0.031

Standard errors in parentheses

$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.10: Market price – linear regression (standard errors in parentheses).

Result 5 (Market prices) With STRUC communication, PDC significantly decrease average market prices and therefore increase consumer surplus. With CHAT communication, PDC increase market prices and therefore decrease consumer surplus.

What could be the intuition for the contradicting effects in CHAT and STRUC? Recall that statement 5 in section 1.3 is not a directed hypothesis in the first place. Prices could be lower when private damage claims apply because there are fewer cartels and remaining cartels might be reluctant to set higher prices because of the risk of paying damage claims. This is what might be going on in STRUC. We suggest that the counter-intuitive result in CHAT is triggered by a hysteresis effect (see also 1.7.4 in the Appendix). In CHAT, subjects have the chance to coordinate their behavior even beyond a cartel breakdown.

Since CHAT allows for threats, cartels are more stable and cartel members stick to the cartel price even after cartels break down. According to our definition, cartels that break down represent a competitive market although the market price is equal to the former collusive price. The number of periods covering this behavior is higher in the private damage claim treatment. Therefore, we can conclude that hysteresis explains the higher competitive and overall market prices in CHAT as well as the increasing prices with the treatment of private damage claims. Due to hysteresis the competitive prices are biased upwards in the PDC and CHAT treatment.

1.4.5 Structured vs. chat communication

Our experimental design enables us to analyze not only the effect of private damage claims but also the impact of different types of communication designs on cartel formation and stability. As expected from hypothesis 6, we see quadrupled stability in CHAT compared to STRUC across both treatments, NOPDC and PDC (see table 1.4). This is also emphasized by the result that infringers apply less often for leniency ($p - value = 0.0011$) (see figure 1.5). These results are in line with the literature observing that CHAT communication helps to better coordinate (for example, Fonseca and Normann, 2012; Fonseca *et al.*, 2018), or generally, that communication facilitates collusion (see e.g., Bigoni *et al.*, 2019; Cooper *et al.*, 1992; Cooper and Kühn, 2014; Waichman *et al.*, 2014).

Perhaps surprisingly, the propensity to collude—new attempts to collude at the subject level—is significantly higher in STRUC compared to CHAT ($p - value = 0.0150$) (see figure 1.2). There are two explanations for this seemingly counterintuitive result. First, CHAT communication facilitates trust among group members and makes group members stick to the agreements more often and, as seen above, report the cartel less frequently. As a result, subjects in CHAT need to press the *discuss price* less often. Secondly, the lower fraction of subjects deciding in favor of a new price discussion in CHAT is explained by agreements to stick to the collusive price after cartel breakdown. Subjects in CHAT are able to agree on setting the same price as under collusion after they have been detected and without renewing their price discussion. This is not possible in the STRUC design. This can be seen from the following excerpts of communication (translated from the original German), groups agree to communicate only once:

- `Without in future rounds without [sic] communication then?` (group 5, period 1)
- `When rules change communicate again` (group 7, period 1)
- `Yes but not more communication in the next rounds` (firm 3)
  `Ok, no more communication and 110` (firm 2)
  `Alright. Yes. Always 110, no more communication and no reports.` (firm 1, group 13, period 1).

Market prices are higher in CHAT compared to STRUC across all types of markets ($p - value : 0.0218$) (see table 1.9). As already mentioned, higher prices in CHAT can be explained by an hysteresis effect that keeps prices high even after cartels break down. In line with that, we see much less variation in collusive market prices in CHAT compared to STRUC ($p - value : 0.0001$) (see table 1.9).

To conclude, CHAT allows subjects to better coordinate their practice compared to STRUC, which leads to an increased stability and hysteresis of cartel prices.

Result 6 (CHAT vs. STRUC) Cartel stability is higher and there are significantly fewer leniency reports in CHAT. The propensity to collude is significantly lower in CHAT.

1.5 Protection from damage claims for leniency applicants

Although in an overall assessment of PDC we find a decreasing cartel prevalence in PDC, the results of the preceding section 1.4 also suggest that private damage claims may lower leniency application rates so that cartels are more stable. This negative effect of PDC on leniency and cartel stability suggests a careful reconsideration of the tool of private enforcement.

Better protection of whistleblowers is an obvious option. Kersting (2014) proposes an approach in which the leniency applicant can obtain full compensation for damage payments from its co-infringers. This should remove the tension between private and public enforcement. As formally demonstrated by Buccirossi *et al.* (2015), damage claim actions and leniency programs can reinforce each other when the first leniency applicant's liability is minimized (or even eliminated) also with respect to damage claims. This scheme corresponds to the former Hungarian legislation before the implementation of the directive on antitrust damage actions (Buccirossi *et al.* (2015); European Commission (2014)). In a related piece of experimental evidence, Mechtenberg *et al.* (2017) analyze whistleblowing in the context of corporate fraud. They find that an increase in reports can be triggered by better whistleblower protection.

In order to test such a potential improvement of current European legislation, we introduce a new treatment called PDC+. In this new treatment, the first leniency applicant is fully protected from private damages. Instead, the remaining two cartel firms jointly pay the damage payment (which remains at 60% of excess Nash industry profit). That is, the remaining cartel members, no matter whether they also reported the cartel, have to pay half of the per-period damage compensation, $D_i = \frac{1}{2}(p - 101) \cdot 0.6$. By contrast, in our standard PDC treatment, all three cartel members pay one third of the damage. Private damage claim actions in PDC+ are enforced with a probability of $\sigma = 0.95$ and they are cumulated over time, as in PDC. If no reporting takes place or cartel authority detects the cartel by probability $\rho = 0.15$, the design follows the PDC treatment as explained in section 1.2. The extension of the experiment is also conducted within subjects. Participants first play nine periods with private damages as above, followed by PDC+ in the remaining periods. Again, the rules of the experiment change in period 10 and PDC+ is introduced after stage 2 (price decision).

The extension of the experiment was conducted in the structured communication setting and was programmed using z-Tree software (Fischbacher, 2007). The sessions took place in January and July 2020 covering 48 participants.

What are our hypotheses for PDC+? First, the participation constraints in PDC+ and PDC are the same because fines and damages for successful collusion do not change compared to PDC (only deviation and reporting change). We thus do not expect an impact on the frequency of cartels. The costs of reporting are much lower in PDC+ as no damages have to be paid; merely reporting costs r occur. Second, the incentive constraint in PDC+ changes compared to PDC because damages have to be paid only in the case of stable collusion. In the case of a deviation, the deviator will report (which costs r) but pays no fine and no damages (because of the damage-leniency of PDC+). The incentive constraint thus becomes

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) - E(D_{it}^c) \quad \geq \quad \pi_i^d - r + \frac{\delta \pi_i^n}{1-\delta}$$

which is more severe than the constraint obtained above for PDC, so $\delta_{min}^{PDC+} > \delta_{min}^{PDC}$.

For the parameters in the experiment, we obtain $\delta_{min}^{PDC+} = 0.723$ whereas $\delta_{min}^{PDC} = 0.655$. That is, PDC+ hinders collusion as intended by the new policy. For all statements, see Appendix 1.7.1 for details.

Hypothesis 7. ***(Protection from damages for leniency applicants)*** *More cartels will be reported in PDC+ than in PDC.*

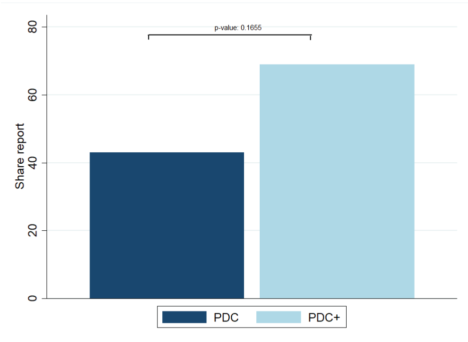


Figure 1.10: Share report in STRUC: within-subjects comparison from the treatment PDC-PDC+.

The results support the notion that PDC+ results in lower cartel stability. Cartels break down more often due to a higher share of reports by individuals. The within-subjects design results based on group level can be seen in figure 1.10. We see a reporting share of 43% in the PDC treatment and a reporting share of 68.9% in the PDC+ treatment, resulting in an increase of 25.9 percentage points. The same holds for the number of stable cartel periods. In the PDC+ treatment, cartels are, on average, 0.33 periods less stable compared to the PDC treatment. Whereas this result is in line with hypothesis 1 for the PDC treatment, we cannot make any statement about significance because there are too few groups forming a cartel in PDC and PDC+.

Results also hold in a between-subjects analysis. In the PDC treatment, we observe a reporting share of 29.6% and 68.9% in the PDC+ treatment, which is significantly higher in PDC+ (see figure 1.11) ($p - value = 0.0929$).

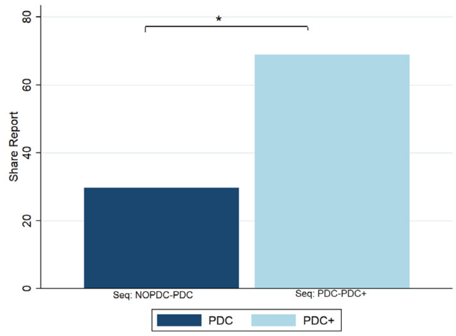


Figure 1.11: Share report in STRUC: between-subjects comparison with PDC data from NOPDC-PDC and PDC+ data from PDC-PDC+.

1.6 Conclusion

Private damage claims, introduced into European law through Directive 2014/104/EU (European Commission, 2014), are controversially discussed. This is especially the case when it comes to the adverse effects private damages may inflict on the well-established and successful tool of leniency. A leniency applicant's fines are waived or reduced, but their damage claim payments are not (completely reduced or only capped to a certain degree). Private enforcement may therefore decrease incentives to apply for leniency and may result in more stable cartels.

Our work contributes to the existing literature in two ways. The main goal of our paper is to provide a first quantification of the trade-off between leniency and private damage claims in an experiment. Our design builds on the literature on leniency experiments (Apesteguia *et al.*, 2007; Bigoni *et al.*, 2012; Dijkstra *et al.*, 2018; Hinloopen and Soetevent, 2008). We analyze a repeated cartel game where firms can discuss prices and may later apply for leniency. We extend the literature by allowing for private damages when a cartel is uncovered. Our treatments further vary the form of communication by analyzing structured price announcements vs. unrestricted chat.

The results are as follows. First, we show that the propensity of cartel formation decreases as private enforcement is introduced. Second, when private damage claims exist, the number of leniency applications is reduced. Third, the implementation of damage claims has a stabilizing effect on cartels. Fourth, and perhaps most importantly, overall there are fewer stable cartels with private damage claims. Fifth, we find ambiguous results regarding consumer surplus depending on the type of communication. Private enforcement decreases prices in a structured communication treatment yielding a rise in consumer surplus, whereas prices tend to increase when subjects are not restricted in communication, implying a decrease in consumer welfare. Sixth, chat-type communication not only lowers the incentives for leniency applications, it also increases cartel stability.

Since overall cartel prevalence is lower with private damages, our main take on the new instrument is positive: private damages have a beneficial impact. Nevertheless, the fact that they involve a negative effect on leniency and cartel stability suggests a careful reconsideration of the tool of private enforcement. As suggested by Buccirossi *et al.* (2015), improved protection from damages for whistleblowers may avoid the negative impact that private damages have on leniency. We take a first step in this direction and analyze the new policy in an additional treatment variation. The data indeed suggest that firms report cartels more often in the treatment where leniency applicants are additionally protected from private damages.

One disclaimer is that we only analyze one set of parameters for the damages. Different magnitudes and likelihoods of the damages may lead to different results. Further experiments along this line are promising for future research. Another aspect of private enforcement that is not captured in our experimental design is that buyers have higher incentives to uncover cartels themselves when damage claims are possible. This is a likewise interesting question for future research.

References

ANDRES, M., BRUTTEL, L. and FRIEDRICHSEN, J. (2019). The effect of a leniency rule on cartel formation and stability: Experiments with open communication. *Preliminary draft*.

APESTEGUIA, J., DUFWENBERG, M. and SELTEN, R. (2007). Blowing the whistle. *Economic Theory*, **31** (1), 143–166.

BECKER, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, **76** (2), 169–217.

BIGONI, M., FRIDOLFSSON, S.-O., LE COQ, C. and SPAGNOLO, G. (2012). Fines, leniency, and rewards in antitrust. *The RAND Journal of Economics*, **43** (2), 368–390.

—, —, — and SPAGNOLO, G. (2015). Trust, leniency, and deterrence. *Journal of Law, Economics, and Organization*, **31** (4), 663–689.

—, POTTERS, J. and SPAGNOLO, G. (2019). Frequency of interaction, communication and collusion: An experiment. *Economic Theory*, **68** (4), 827–844.

BLUME, A. and ORTMANN, A. (2007). The effects of costless pre-play communication: Experimental evidence from games with pareto-ranked equilibria. *Journal of Economic Theory*, **132** (1), 274–290.

BOURJADE, S., REY, P. and SEABRIGHT, P. (2009). Private antitrust enforcement in the presence of pre-trial bargaining. *The Journal of Industrial Econommics*, **57** (3), 372–409.

BRENNER, S. (2009). An empirical study of the european corporate leniency program. *International Journal of Industrial Organization*, **27** (6), 639–645.

BROSIG, J., WEIMANN, J. and OCKENFELS, A. (2003). The effect of communication media on cooperation. *German Economic Review*, **4** (2), 217–241.

BROWN KRUSE, J. and SCHENK, D. J. (2000). Location, cooperation and communication: An experimental examination. *International Journal of Industrial Organization*, **18** (1), 59–80.

BUCCIROSSI, P., MARVAO, C. and SPAGNOLO, G. (2015). Leniency and damages. *CEPR Discussion Paper No. DP10682*.

CAMERA, G., CASARI, M. and BIGONI, M. (2011). Communication, commitment, and deception in social dilemmas: Experimental evidence. *Quaderni - Working Paper DSE No. 751*.

Canenbley, C. and Steinvorth, T. (2011). Effective enforcement of competition law: Is there a solution to the conflict between leniency programmes and private damages actions? *Journal of European Competition Law & Practice*, **2** (4), 315–326.

Cauffman, C. and Philipsen, N. J. (2014). Who does what in competition law: Harmonizing the rules on damages for infringements of the EU competition rules? *Maastricht European Private Law Institute Working Paper No. 2014/19*.

Chowdhury, S. M. and Wandschneider, F. (2018). Antitrust and the Beckerian proposition: The effects of investigation and fines on cartels. In *Handbook of Behavioral Industrial Organization*, Cheltenham: Edward Elgar Publishing.

Clemens, G. and Rau, H. A. (2018). Do discriminatory leniency policies fight hard–core cartels? *Journal of Economics & Management Strategy*, **28** (2), 336–354.

Cooper, D. J. and Kühn, K.-U. (2014). Communication, renegotiation, and the scope for collusion. *American Economic Journal: Microeconomics*, **6** (2), 247–278.

Cooper, R., DeJong, D. V., Forsythe, R. and Ross, T. W. (1992). Communication in coordination games. *The Quarterly Journal of Economics*, **107** (2), 739–771.

Department of Justice (2007). British Airways PLC and Korean Air Lines Co. LTD. agree to plead guilty and pay criminal fines totaling USD600 million for fixing prices on passenger and cargo flights. `https://www.justice.gov/archive/atr/public/press_releases/2007/224928.htm`. Accessed: 2019-05-19.

Dijkstra, P. T., Haan, M. A. and Schoonbeek, L. (2018). Leniency programs and the design of antitrust: Experimental evidence with free-form communication. *Working Paper*.

Dufwenberg, M. and Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, **30** (2).

European Commission (2001). Case COMP/E-1/37.512 - Vitamins. *Official Journal of the European Communities*.

European Commission (2002). Commission notice on immunity from fines and reduction of fines in cartel cases (2002/C 45/03). *Official Journal of the European Communities*.

European Commission (2005). Green paper - damages actions for breach of the EC antitrust rules. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52005DC0672&from=en`. Accessed: 2019-06-19.

European Commission (2006). Commission notice on immunity from fines and reduction of fines in cartel cases (2006/C 298/11). *Official Journal of the European Communities*.

European Commission (2011). Fines for breaking EU competition law. http://ec.europa.eu/competition/cartels/overview/factsheet_fines_en.pdf. Accessed: 2019-06-19.

European Commission (2014). Directive on certain rules governing actions for damages under national law for infringements of the competition law provisions of the member states and of the European Union. *Official Journal of the European Union*.

European Commission (2015). The damages directive - towards more effective enforcement of the EU competition rules. *Competition policy brief*.

European Commission (2016). Case AT.39824 - trucks. http://ec.europa.eu/competition/antitrust/cases/dec_docs/39824/39824_6567_14.pdf. Accessed: 2019-06-19.

European Commission (2017a). Antitrust: Commission re-adopts decision and fines air cargo carriers EUR 776 million for price-fixing cartel. *Press Release IP/17/661*.

European Commission (2017b). Case AT.39258 — Airfreight. *Official Journal of the European Union*.

European Commission (2018). Directive on antitrust damages actions: Transposition of the directive in member states. http://ec.europa.eu/competition/antitrust/actionsdamages/directive_en.html. Accessed: 2019-03-25.

European Commission (2019). Cartel statistics. http://ec.europa.eu/competition/cartels/statistics/statistics.pdf. Accessed: 2019-04-01.

Feltovich, N. and Hamaguchi, Y. (2018). The effect of whistle-blowing incentives on collusion: An experimental study of leniency programs. *Southern Economic Journal*, **84** (4), 1024–1049.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **10**, 171–178.

Fonseca, M. A., Li, Y. and Normann, H.-T. (2018). Why factors facilitating collusion may not predict cartel occurrence - experimental evidence. *Southern Economic Journal*, **85** (1), 255–275.

— and Normann, H.-T. (2012). Explicit vs. tacit collusion—the impact of communication in oligopoly experiments. *European Economic Review*, **56** (8), 1759–1772.

Gillet, J., Schram, A. and Sonnemans, J. (2011). Cartel formation and pricing: The effect of managerial decision-making rules. *International Journal of Industrial Organization*, **29** (1), 126–133.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association*, **1** (1), 114–125.

Harrington, J. E. and Chang, M.-H. (2009). Modeling the birth and death of cartels with an application to evaluating competition policy. *Journal of the European Economic Association*, **7** (6), 1400–1435.

—, Gonzalez, R. H. and Kujal, P. (2016). The relative efficacy of price announcements and express communication for collusion: Experimental findings. *Journal of Economic Behavior & Organization*, **128**, 251–264.

Hinloopen, J. and Onderstal, S. (2014). Going once, going twice, reported! Cartel activity and the effectiveness of antitrust policies in experimental auctions. *European Economic Review*, **70**, 317–336.

— and Soetevent, A. R. (2008). Laboratory evidence on the effectiveness of corporate leniency programs. *The RAND Journal of Economics*, **39** (2), 607–616.

Holt, C. A. and Davis, D. (1990). The effects of non-binding price announcements on posted-offer markets. *Economics Letters*, **34**, 307–310.

Huck, S., Normann, H.-T. and Oechssler, J. (2004). Two are few and four are many: Number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, **53** (4), 435–446.

Hüschelrath, K. and Weigand, J. (2010). Fighting hard core cartels. *ZEW Discussion Papers, No. 10-084*.

Jones, A. (2016). Private enforcement of EU competition law: A comparison with, and lessons from, the US. In *Harmonising EU competition litigation*, Swedish studies in European law, London: Hart Publishing.

Kersting, C. (2014). Removing the tension between public and private enforcement: Disclosure and privileges for successful leniency applicants. *Journal of European Competition Law & Practice*, **5** (1), 2–5.

Kiani-Kress, R. and Schlesiger, C. (2014). Die Bahn rächt sich am Cargo-Kartell: Milliardenklage gegen Lufthansa. *WirtschaftsWoche from 2014-12-01*.

Kirst, P. and van den Bergh, R. (2016). The European directive on damages actions: A missed opportunity to reconcile compensation of victims and leniency incentives. *Journal of Competition Law and Economics*, **12** (1), 1–30.

Knight, T. and de Weert, W. (2015). On implementing private damages in European competition cases. *Working Paper*.

— and Ste. Claire, C. (2019). Reconciling the conflict: Antitrust leniency programs and private enforcement. *Working Paper*.

Land, R. H. and Davis, J. P. (2011). Comparative deterrence from private enforcement and criminal enforcement of the U.S. antitrust laws. *BYU Law Review 315*.

Landeo, C. M. and Spier, K. E. (2009). Naked exclusion: An experimental study of contracts with externalities. *American Economic Review*, **99**, 1850–1877.

Luz, R. D. and Spagnolo, G. (2017). Leniency, collusion, corruption, and whistleblowing. *Journal of Competition Law and Economics*, **13** (4), 729–766.

Mechtenberg, L., Gerd, M. and Roider, A. (2017). Whistle-blower protection: Theory and experimental evidence. *IZA DP No. 10607*.

Michaels, D. (2014). Deutsche Bahn to claim damages of more than USD 3 billion over air-cargo cartel. *The Wall Street Journal from 2014-11-30*.

Migani, C. (2014). Directive 2014/104/EU: In search of a balance between the protection of leniency corporate statements and an effective private competition law enforcement. *Global Antitrust Review,* (7), 81–111.

Miller, N. H. (2009). Strategic leniency and cartel enforcement. *American Economic Review*, **99** (3), 750–768.

Motta, M. and Polo, M. (2003). Leniency programs and cartel prosecution. *International Journal of Industrial Organization*, **21** (3), 347–379.

Normann, H.-T. and Wallace, B. (2012). The impact of the termination rule on cooperation in a prisoner's dilemma experiment. *International Journal of Game Theory*, **41** (3), 707–718.

Roux, C. and Thöni, C. (2015). Collusion among many firms: The disciplinary power of targeted punishment. *Journal of Economic Behavior & Organization*, **116**, 83–93.

Spagnolo, G. (2003). Divide et impera optimal deterrence mechanisms against cartels and organized crime. *mimeo, Mannheim*.

— (2008). Leniency and whistleblowers in antitrust. In *Handbook of antitrust economics*, Cambridge, Massachusetts and London: MIT.

Waichman, I., Requate, T. and Siang, C. K. (2014). Communication in Cournot competition: An experimental study. *Journal of Economic Psychology*, **42**, 1–16.

Wils, W. P. (2003). Should private antitrust enforcement be encouraged in Europe? *World Competition: Law and Economics Review*, **26** (3), 473–488.

— (2009). The relationship between public antitrust enforcement and private actions for damage. *World Competition*, **32** (1), 3–26.

1.7 Appendix

1.7.1 Variable definitions and theoretical model of the experimental setup

To prove the statements in the main text for the experimental parameters and equilibrium realization of the variables, consider the parameters in table 1.11. We analyze treatments NOPDC, PDC, and PDC+ in turn.

Definition	Variable	Numerical realization in experiment
Detection probability	ρ	0.15
Damage liability probability	σ	0.95
Discount factor & continuation probability	δ	0.8
Reporting cost	r	1
Marginal cost of production	c	100
Nash price	p_i^n	101
Collusive price	p_i^c	110
Deviation price	p_i^d	109
Nash revenue	R_i^n	$101/3$
Collusive revenue	R_i^c	$110/3$
Deviation revenue	R_i^d	deviator: 109, others: 0
Nash equilibrium profit	π_i^n	$(101-100)/3 = 1/3$
Collusive profit	π_i^c	$(110-100)/3 = 10/3$
Deviation profit	π_i^d	deviator: $(109-100)/1 = 9$, others: 0
Fine under collusion	F_i^c	$0.1 \cdot R^c = 11/3$
Fine under deviation	F_i^d	deviator: $0.1 \cdot R^d = 10.9$, others: 0
Fine under Nash pricing	F_i^n	$0.1 \cdot R_i^n = 101/30$
Damage payments collusion	D_i^c	$0.6 \cdot (110-101)/3 = 1.8$
Damage payments deviation	D_i^d	$0.6 \cdot (109-101)/3 = 1.6$

Table 1.11: Definition of variables and values realized in the experiment.

NOPDC

Following Bigoni et al. (2015, Appendix A.1), we assume that firms communicate once to establish successful collusion, but are able to collude tacitly following a detection by the competition authority. This implies that cartel firms risk being fined only once on the collusive path.

With a probability of detection of ρ, a general fine F_i^j per firm i and outcome $j \in \{c, d, n\}$, with c for collusion, d for deviation and n for Nash, and a discount factor of δ, the net present value of the fine is obtained as follows. In each period, the cartel is either detected and has to pay F_i^j (happens with probability ρ), or the cartel is not detected (which happens with probability $1-\rho$) but might have to pay the fine in the next period (and accordingly this potential fine has to be discounted by δ). If the next period is reached, the same contingencies arise again, and so on. The stream of potential fine payments reads:

$$E(F_i^j) = \rho F_i^j + (1-\rho)\rho\delta F_i^j + (1-\rho)^2\rho\delta^2 F_i^j + (1-\rho)^3\rho\delta^3 F_i^j + \dots .$$

Multiplying both sides of the equation with $\delta(1-\rho)$, we have

$$\delta(1-\rho)E(F_i^j) = (1-\rho)\rho\delta F_i^j + (1-\rho)^2\rho\delta^2 F_i^j + (1-\rho)^3\rho\delta^3 F_i^j + \dots$$

and therefore we obtain

$$E(F_i^j) = \frac{\rho F_i^j}{1-\delta\,(1-\rho)}$$

as an expression for the discounted expected fine, $E(F_i^j)$.

The *participation constraint* in NOPDC states that colluding must be more profitable than competing (static Nash equilibrium)

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) \;\; \geq \;\; \frac{\pi_i^n}{1-\delta}.$$

Using the numerical values of the experiment, we find

$$14.948 \quad \geq \quad 1.667.$$

So the participation constraint is met for our experimental setup.

Before analyzing the incentive constraint, we need to analyze whether or not a deviator will report the cartel to the authorities. Reporting incurs cost of r and no fine because of leniency. Not reporting saves the reporting cost but involves the risk of the cartel being fined due to detection. The authority may detect the cartel during the period of the deviation (resulting in fine F_i^d) or in a later period when firms play the Nash price as a punishment for the deviation (a cartel formally exists until a cartel member reports or the cartel is uncovered by the cartel authority). Comparing reporting versus not reporting, we get

$$r = 1 \; < \; \rho F_i^d + \delta(1-\rho)E(F_i^n) = 2.708.$$

That is, a deviator will report.

The *incentive constraint* in NOPDC requires that there should be no incentive to deviate from collusion, given such deviation triggers a return to the static Nash equilibrium price. The incentive constraint accordingly reads

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) \quad \geq \quad \pi_i^d - r + \frac{\delta \pi_i^n}{1-\delta}.$$

Using the experimental parameters, we solve for the minimum discount factor required for collusion and obtain

$$\delta_{min}^{NOPDC} \geq 0.664.$$

This implies that colluding at the highest price of 110 is a subgame perfect Nash equilibrium in our setup. Alternatively, we can plug $\delta = 0.8$ into the incentive constraint and obtain

$$14.948 \geq 9.333$$

with the same implication.

PDC

In the treatment of PDC, the expected fine remains the same; it has to be paid at most once. The expected private damages also have to be paid only once (when the cartel busts), but the analysis differs because damages are cumulated over time. The stream of discounted potential damage payments is

$$\begin{aligned} E(D_{it}^j) &= \rho\sigma D_{it}^j + (1-\rho)\delta\rho\sigma 2D_{it}^j + (1-\rho)^2\delta^2\rho\sigma 3D_{it}^j + (1-\rho)^3\delta^3\rho\sigma 4D_{it}^j + ... \\ \delta(1-\rho)E(D_{it}^j) &= (1-\rho)\delta\rho\sigma D_{it}^j + (1-\rho)^2\delta^2\rho\sigma 2D_{it}^j + (1-\rho)^3\delta^3\rho\sigma 3D_{it}^j + ... \end{aligned}$$

where $j \in \{c, d\}$ denotes the outcome on which the damages are based. Taking the difference $E(D_{it}^j) - \delta(1-\rho)E(D_{it}^j)$ yields

$$\begin{aligned} (1 - \delta(1-\rho))E(D_{it}^j) = \rho\sigma D_{it}^j + (1-\rho)\delta\rho\sigma D_{it}^j + (1-\rho)^2\delta^2\rho\sigma D_{it}^j + \\ (1-\rho)^3\delta^3\rho\sigma D_{it}^j + ... \end{aligned} \tag{1.1}$$

and therefore (proceeding as above with steady fines)

$$E(D_{it}^j) = \frac{\rho\sigma D_{it}^j}{(1-\delta(1-\rho))^2}$$

which, for the experimental parameters, becomes $E(D_{it}^j) = 1.3916D_{it}^j$.

The *participation constraint* in PDC reads

$$\begin{aligned}\frac{\pi_i^c}{1-\delta} - E(F_i^c) - E(D_{it}^c) &\geq \frac{\pi_i^n}{1-\delta} \\ 12.443 &\geq 1.667.\end{aligned}$$

This participation constraint is also met for the experimental parameters, but it is more severe than the one above under NOPDC since it has less slack. We conclude that private damages deter more cartels.

The *incentive constraint* in PDC is obtained as follows. First, we have to compare the report vs. not report cases. A deviator who reports has to pay the reporting cost, r, and damages σD_{it}^d whereas a deviator who does not report faces the fine F_i^d and damages σD_{it}^d, with detection probability ρ as well as the expected Nash fine $E(F_i^n)$. For our experimental parameters, we see that off-equilibrium reporting is better than not reporting:

$$r + \sigma D_{it}^d = 2.52 \; < \; \rho F_i^d + \frac{\rho \sigma D_{it}^d}{(1-\delta(1-\rho))} + \delta(1-\rho)E(F_i^n) = 3.421.$$

The incentive constraint reads

$$\frac{\pi_i^c}{1-\delta} - E(F_i^c) - E(D_{it}^c) \;\; \geq \;\; \pi_i^d - r - \sigma D_{it}^d + \frac{\delta \pi_i^n}{1-\delta}.$$

Solving for the minimum discount factor required for collusion obtains

$$\delta_{min}^{PDC} \geq 0.655.$$

That is, $\delta_{min}^{NOPDC} > \delta_{min}^{PDC}$. Or, applying $\delta = 0.8$, yields

$$12.443 \geq 7.813.$$

The incentive constraint in PDC has less slack (namely 4.630) than the one in NOPDC

(5.615) and is thus more *severe*. We conclude that PDC makes collusion more stable than NOPDC.

The calculations of the incentives to report are based on the assumption that deviations take place in the first period. For NOPDC, the incentive to report does not change over time as the fine remains unchanged when reporting takes place in later periods. However, in PDC the incentive to report does change. It decreases with the duration of the cartel as damages are cumulated. The highest incentive to deviate is, nevertheless, present in period zero, so the repeated-game incentive constraint above is the one that is relevant when solving the overall game.

PDC+

In the PDC+ case the *participation constraint* remains the same

$$\frac{\pi^c}{1-\delta} - E(F_i^c) - E(D_{it}^c) \geq \frac{\pi^n}{1-\delta}$$

because fines and damages for successful collusion do not change compared to PDC (only deviation and reporting change).

The *incentive constraint* in PDC changes as follows. Damages have to be paid only in the case of stable collusion. In the case of a deviation, the deviator will report (which costs r) but pays no fine and no damages (because leniency applies to damages, too, in PDF+).

We obtain that, in the case of a deviation, reporting again is cheaper than not reporting. The incentive constraint becomes

$$\frac{\pi^c}{1-\delta} - E(F_i^c) - E(D_{it}^c) \quad \geq \quad \pi^d - r + \frac{\delta\pi^n}{1-\delta}.$$

In terms of the minimum discount factor required for collusion, we get

$$\delta_{min}^{PDC+} \quad \geq \quad 0.723.$$

Taking the continuation probability of 0.8 into account yields

$$12.443 \geq 9.333.$$

As expected, PDC+ makes collusion more demanding than PDC and NOPDC. That is, PDC+ hinders collusion as intended by the new policy.

1.7.2 Definitions of variables

Variable	Definition
Propensity to collude	Number of periods in which a subject chooses to enter the communication stage when a cartel does not already exist over the total number of periods in which a cartel does not exist.
Share cartel	Number of periods in which all three subjects of a group choose to enter the communication stage when a cartel does not already exist over the total number of periods in which a cartel does not exist.
Share report	Number of active reports of a cartel (click 'report button') by a group member over all periods that a cartel existed (active cartel formation or liability from an older cartel). We exclude periods 10 and 20.
Cartel stability	The number of periods when a cartel was stable divided by the number of cartels of the group. A cartel is stable until it is reported or detected by the authority. We exclude periods 10 and 20.
Cartel prevalence	Number of periods in which a cartel exists (all three subjects of a group choose to enter the communication stage or are liable from an older cartel) over all periods of a treatment (10 periods).
Ask non-cartel markets	Average price when a cartel does not exist.
Ask cartel market	Average price when a cartel does exist (active communication or liability form an older cartel).
Ask all markets	Average price in both non-cartel and cartelized markets.
Market price non-cartel markets	Lowest price of a group when a cartel does not exist.
Market price cartel market	Lowest price of a group when a cartel does exist (active communication or liability form an older cartel).
Market price all markets	Lowest price of a group in both non-cartel and cartelized markets.

Table 1.12: Definition of the main variables.

1.7.3 Group dynamics over time

Figures 1.12 and 1.13 give an overview of the cartelizing behavior of each group in STRUC and CHAT. The blue line plots the binary group dependent variable *collusion*, which becomes one when a group forms a cartel and zero when at least one group member decides against cartelization. The red line shows the course of the market price. The dots mark the reason for a cartel breakdown: while the black dot indicates a breakdown because of leniency application by at least one group member, the green dot characterizes a breakdown due to discovery by the cartel authority. Consequently, a cartel is stable for more than one period if the blue line moves along its upper boundary without being interrupted by any dots.

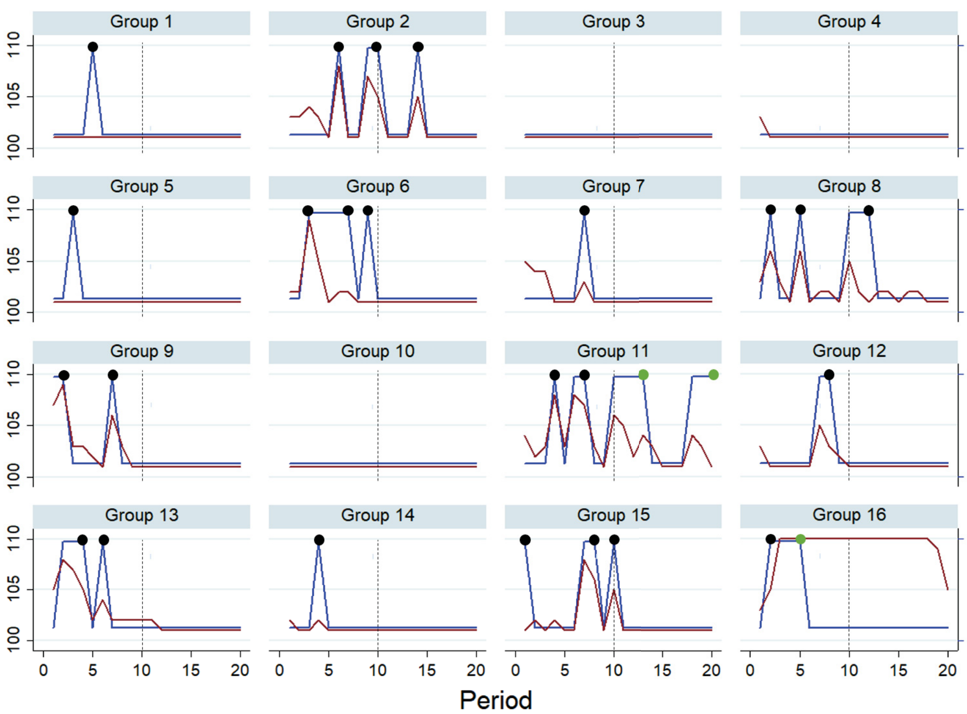


Figure 1.12: Collusive activity and market price by group for the treatment in STRUC.

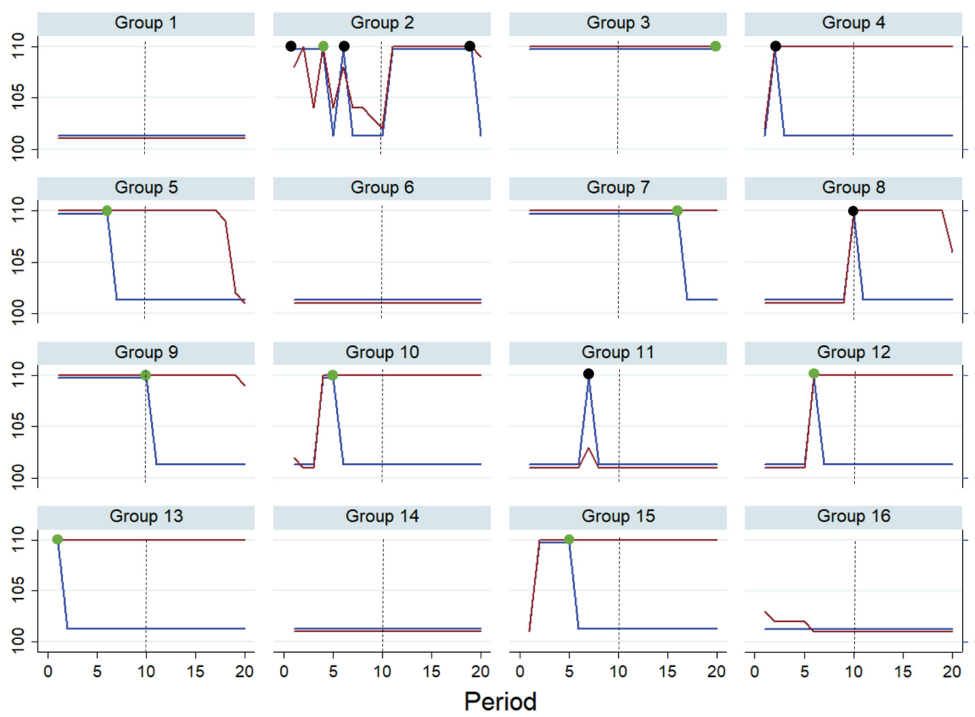


Figure 1.13: Collusive activity and market price by group in CHAT.

1.7.4 Deviations from agreed price

Figures 1.14 and 1.15 give an overview of the agreed-upon price during the communication stage and the (independently set) ask price. If subjects decide to discuss prices and agree on a single price, this is displayed by the blue line. In STRUC, price discussion can result in an interval of agreed prices. Figure 1.14 indicates this by the upper and lower bound of agreed prices (see e.g., group 9).

In figure 1.15, we can observe a more stable price setting following the agreed price even in periods without a cartelized market in CHAT. Figure 1.14, which considers STRUC, provides an indication of lack of trust in collusive markets (this does not apply to group 16). For example, although group 2 in STRUC agrees on setting a price of 110, all three subjects never simultaneously set the agreed price as their individual ask price, instead they continuously undercut the agreed price. In contrast to that, in figure 1.15 group 7 gives a perfect example of subjects sticking to the agreed price although price discussion has not taken place in this period. This behavior emphasizes our explanation of hysteresis regarding subjects not communicating but setting high prices.

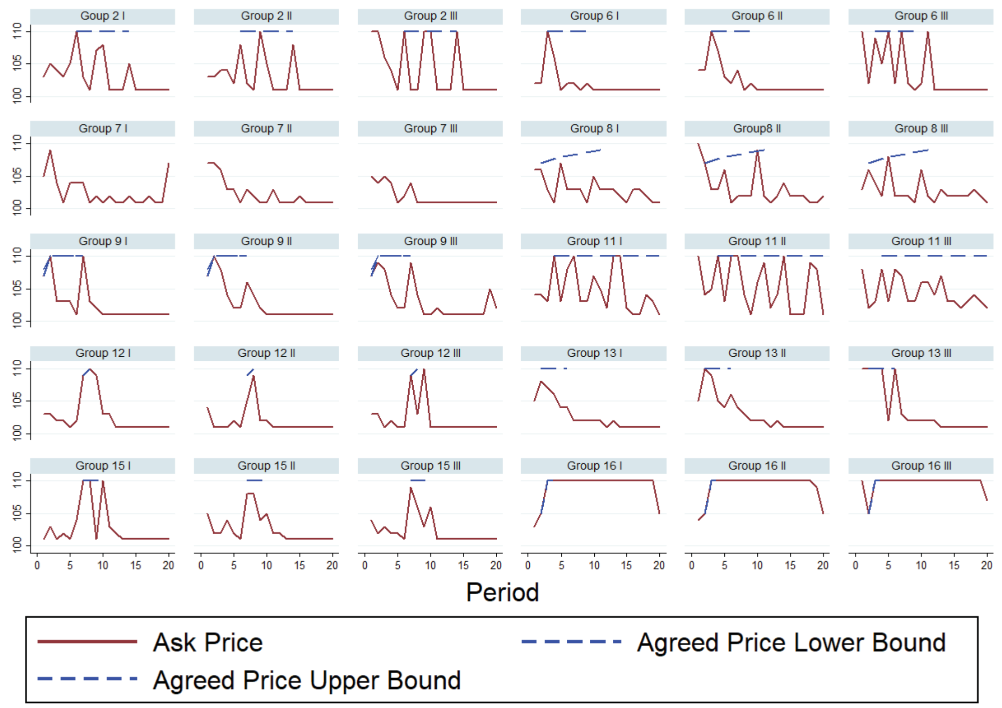


Figure 1.14: Agreed price and set price by subject in STRUC.

Note: Groups that do not discuss prices or could not agree on an interval other than 101 to 110 are excluded.

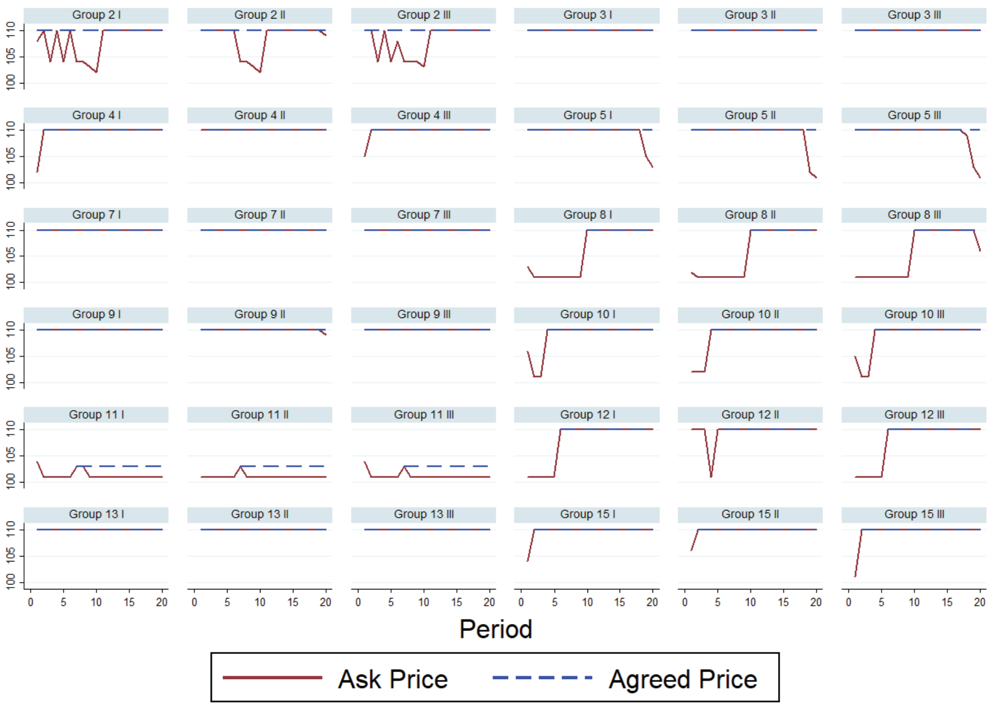


Figure 1.15: Agreed price and set price by subject in CHAT.
Note: Groups that do not discuss prices or could not agree on an interval other than 101 to 110 are excluded.

1.7.5 Ask Prices

In this section we investigate the ask (or offer) price. The ask price is the price firms individually demand in stage 2. Figure 1.16 (and the bottom line in table 1.13) illustrate the overall change in ask prices. We see the same pattern as in the above analysis of overall market prices. It shows for treatment STRUC an average overall ask price of 103.67 in NOPDC and 101.94 in PDC. This is statistically significantly different (STRUC: WMP, $p - value = 0.0011$). The difference in ask prices of NOPDC and PDC in CHAT is not statistically significant (CHAT: WMP, $p - value = 0.6033$).

	STRUC		CHAT	
	NOPDC	PDC	NOPDC	PDC
Ask price non-cartels	102.885 (1.899)	101.835 (2.125)	105.036 (3.727)	106.700 (4.351)
Ask price cartels	106.158 (2.537)	104.852 (2.727)	109.328 (2.016)	109.989 (0.019)
Ask price all markets	103.669 (2.062)	101.938 (2.162)	106.277 (3.803)	107.110 (4.203)

Table 1.13: Ask price – averages per treatment (standard deviations in parenthesis).

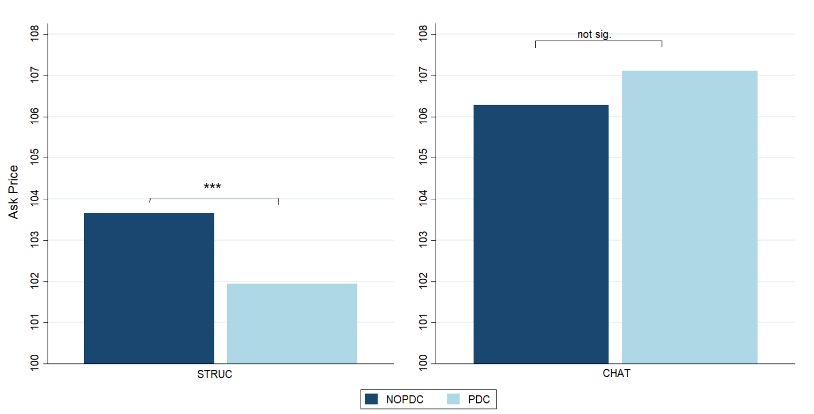


Figure 1.16: The impact of PDC on ask prices in STRUC (left) and CHAT.

In table 1.14 we estimate an ordinary least squares (OLS) model with the dependent variable *Askprice* (all markets). The results show that PDC have a negative effect on ask prices in the subsample of STRUC (table 1.14, column 1), whereas PDC have a positive impact on ask prices in CHAT on a 15% level (table 1.14, column 2).

	(1)	(2)	(3)	(4)
	Price	Price	Price	Price
PDC	-1.731***	0.833+	-3.542***	0.458
	(0.317)	(0.573)	(0.460)	(1.046)
constant	103.7***	106.3***	105.0***	106.1***
	(0.492)	(0.916)	(0.417)	(0.748)
TIME FE	No	No	Yes	Yes
Sample STRUC	Yes	No	Yes	No
Sample CHAT	No	Yes	No	Yes
N	960	960	960	960
R^2	0.084	0.010	0.116	0.014

Table 1.14: Ask price – linear regression (standard errors in parentheses).

Figure 1.17 shows the analysis of the sequence of reverse order PDC-NOPDC in STRUC. The robustness check confirms the significantly lower ask prices in PDC (WMU, $p - value = 0.0785$).

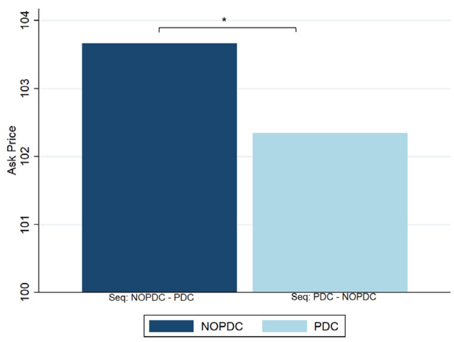


Figure 1.17: Ask price in STRUC: between-subjects comparison with PDC data from treatment with reverse order (PDC-NOPDC).

1.7.6 Within-subjects results reverse-order treatment (PDC-NOPDC)

For the robustness check of our main analysis we only use the PDC data from the session PDC-NOPDC (see chapter 1.4). This allows us to explore any potential order effects, because we only analyze the first 10 periods, for both the NOPDC and PDC treatment. For the sake of completeness table 1.15 shows an overview of the summary statistics of our reverse-order treatment within-subjects. There are basically no differences between PDC and NOPDC in the within analysis of the reverse-order treatment.

	STRUC		Test
	NOPDC	PDC	p-value
Propensity to collude	0.573 (0.193)	0.555 (0.120)	0.7114
Share cartel	0.134 (0.222)	0.117 (0.137)	0.2264
Share report	0.383 (0.267)	0.451 (0.263)	0.5176
Cartel stability	1.611 (0.656)	1.333 (0.476)	–
Cartel prevalence	0.180 (0.283)	0.133 (0.150)	0.4956
Market price	101.2 (0.314)	101.527 (0.680)	0.0364

Table 1.15: Summary statistics of the results in treatments PDC–NOPDC (STRUC); average results per treatment (standard deviations in parentheses).

1.7.7 Instructions

Instructions for the experiment with structured communication (translated from German):

Welcome to our experiment.
Please read these instructions carefully. Please do not talk to your neighbor and be quiet throughout the experiment. If you have any questions, please raise your hand. We will come to your place and answer your question in private. In this experiment, you have to take decisions repeatedly. In the end, you can earn money. How much you earn depends on your decisions and the decisions of two other participants who are randomly assigned to you. At the end of the experiment, you will receive your earnings in cash. All participants receive (and are reading) the

same instructions. You remain completely anonymous for us and for the other participants. We do not store any data connected with your name.

Overview:

The experiment lasts for at least 20 periods, each period consists of seven steps. These steps are the same in each period. Below you will find an overview of the experiment as well as an explanation of all seven steps of each period.

At the beginning of the experiment, all participants will be randomly distributed into groups of three. The group composition does not change during the experiment. Group members remain anonymous. During the experiment you will have no contact to participants of the experiment outside your group.

You can collect points in any period of the experiment. At the end of the experiment these points will be converted into euros, where: $1\ point = 0.3\ euros$. At the beginning of the experiment you will receive a starting capital of 15 points. At the end of each period, all the points collected during that period will be credited to your account. If you score a negative number of points in a period, this number of points will be deducted from your starting capital.

Like the other two group members, you are a supplier of the same good in a market. In each period you must choose a price for the good. This price must be one of the following: 101, 102, 103, 104, 105, 106, 107, 108, 109 or 110. You and the other two group members choose the price at the same time.

You only earn points if your price is the lowest of the three prices. Your profit will then be equal to your price minus the cost of 100. However, if one or both other group members have chosen the same lowest price, you must share the profit with them.

It is possible to discuss the price you want to set. Price discussion is only possible if all group members agree to discuss the prices. If there has been a communication about prices, you might risk that points will be deducted later, either through reports from the group members (step 5) or a random move (step 6).

Each period has seven steps. Below is a more detailed explanation of each step.

In step 1 of each period the following question is asked: ”Do you want to discuss the price with your group members? To answer this question, press the ”DISCUSS PRICE” or ”DO NOT DISCUSS PRICE” button. The other two group members will make the same decision at the same time.

Only if all group members press the button ”DISCUSS PRICE,” a communication window opens and step 2 (the communication phase) will begin. If one or more group members click on the button ”DO NOT DISCUSS PRICE” there will be no communication. In this case step 2 (the communication phase) will be skipped and you will proceed to step 3 (the pricing phase).

If a communication has taken place, there is a risk that points will be subtracted from your account in step 5 or 6. See below.

Step 2: Communication. After opening the communication window, you can talk about the price as explained in the following: You can choose a minimum price and a maximum price that is acceptable to you from the following price range: 101, 102, 103, 104, 105, 106, 107, 108, 109, 110. If only one price is acceptable to you, choose the same value for the minimum price and the maximum price.

If all group members have chosen their minimum price and maximum price, each group member is informed about the overlap of the three price ranges. If the overlap consists of one price, this is the agreed price and step 2 is completed.

If there is no overlap, this procedure is repeated until the overlap consists of only one price or 60 seconds have passed. If no price agreement is reached after 60 seconds, the discussion screen closes. In this case, the last overlap is the agreed price interval.

Communication about anything other than the price is not possible.[25]

[25] The instructions for the OPEN treatment differ from the CLOSED-instructions with respect to step 2. The OPEN-instructions read the following: After opening the communication window, you can discuss the price with your group members by entering a text in the communication field and

Step 3: Pricing phase. You chose your market price. You are again restricted to prices from 101 to 110. The other two group members make the same decision at the same time. Results of any communication are not binding.

Step 4: Market price. In step 4, you learn the market price that has been set in your group. The market price corresponds to the lowest entered price in step 3 in your group. You only earn points if your price is the lowest of the three prices.
The turnover corresponds to the market price without a reduction of costs (100):

- If your price is the lowest price and no other group member has chosen the same price: Turnover = market price.
- If the price you chose is the lowest price and one other group member has set the same price: Turnover = market price / 2.
- If the price you have chosen is the lowest and the other two group members have set the same price: Turnover = market price / 3.
- If your price is not the lowest price: Turnover = 0.

Your profit corresponds to the market price after the deduction of costs (100):

- If your price is the lowest and no other group member has chosen the same price: Profit = market price - 100, i.e., you alone get the profit.
- If the price you chose is the lowest and one other group member has set the same price: Profit = (market price - 100) / 2, i.e., you both share the profit.
- If the price you chose is the lowest and the other two group members have set the same price: Profit = (market price - 100) / 3, i.e., you share the profit with the two other group members.
- If your price is not the lowest: Profit = 0 points.

The experiment continues with step 5 (reporting decision) when a communication

pressing Enter. During the communication you remain anonymous. The communication window closes after 60 seconds. After the communication window has closed, communication in the current period is no longer possible.

about prices in step 1 has taken place. If not all group members have agreed to a communication in step 1, the experiment will continue with step 7 (end of period).

Step 5: Point deduction through reporting. If communication has taken place, you must decide in this step whether you want to report the communication. You can report price discussion by pressing the ”REPORT” button. If you do not want to report, press the ”DO NOT REPORT” button. The other group members must take the same decision at the same time. Reporting always costs one point.
Step 5 only takes place if (i) there was a communication in the current period or (ii) there was a communication in one or more of the previous periods and since then none of the group members pressed the REPORT button and no point deduction by a random move (step 6) has taken place.
After a communication has been reported by you or one of your group members, the ability to report in future periods will expire until the communication about prices is renewed.
In the event of one or more group members reporting the communication, each group member will receive a point deduction of the following amount: The point deduction generally is 10% of your revenue in that period.
If you report the communication, your point deduction can be prevented or reduced in the following:

- You will not receive a point deduction if you are the first to press the REPORT button.
- If you are the second to press the REPORT button, your point deduction is cut by half.
- If you are the third to press the REPORT button, your point deduction will not be reduced.

The experiment will continue with step 6 (Random Points) if all group members have pressed the ”DO NOT REPORT” button. If one or more group members have reported the communication, the experiment continues to step 7.

Step 6: Points deducted by random draw. In this step, a random draw decides whether points will be deducted from you and your group members' account. The probability of a point deduction is 15%; with an 85% probability no points will be deducted.
Step 6 will only take place if (i) there has been communication about prices in the current period and there has been no random point deduction, or (ii) there has been communication in one or more of the previous periods and since then none of the group members pressed the REPORT button and no random point deduction has taken place so far.
After the random draw you will be informed whether you and your group members received any point deductions in that period.
If there is a point deduction by chance, the point deduction will be 10% of your current period revenue.
If the random draw results in point deduction, there will be no further point deductions again until communication is renewed and (i) and (ii) are fulfilled (see above).

Step 7: Period End. In this step you will receive the information of your accumulated points from the current period and from previous periods. The total score (the sum of the points from all periods played) is also displayed. Your accumulated points in the current period correspond to your profit after possible point deductions:

Accumulated points in a period = profit - possible deduction of points

The points are calculated in the same way for each group member. Your points will be credited to your point account after each period. If there has been a deduction of points, the reason for the deduction of points (report or random draw) is shown for all group members.

Next step: Sudden change of rules. In the course of the experiments, there may be a rule change. You will be informed of such a change at the appropriate point.
New period: You play at least 20 periods. From period 20 the experiment ends at the end of each period with 20% probability. With a probability of 80% the next period will start with step 1.

Instructions for the change of rules in period 10 (translated from German):

Introduction of step 8: In addition to the point deduction in step 6, there is now a 95% probability that there will be another point deduction if:

1.) you or some other of your group members have reported the communication, or
2.) in step 6, chance decides that you and your group members will receive a deduction of points.

This point deduction is in addition to the point deduction from step 6 which covers 10% of your current period revenue. The additional point deduction for each group member is 20% of the difference between the group's market price and 101 (the lowest price to choose). The point deduction is added up over all periods in which you communicated but the communication was not discovered or reported.

Instructions for the change of rules in period 10 for the extension of the experiment (translated from German):

Change of the second point deduction in step 5: The second point deduction can now be reduced:
The amount of the second point deduction can now be either 20% or 30% of the difference between the market price and 101 (the lowest price to be chosen). The

second point deduction differs in the cases of random draws and reporting by a group member as follows:

1.) if the random draw decides in step 6 that points will be deducted from you and your group members' account, the second point deduction will still be 20% of the difference between the market price and 101 for all group members. The point deduction is added up over all periods in which you communicated but the communication was not discovered or reported.
2.) if you or one or more of your group members reported the communication in step 5, the second point deduction will be different for each group member. The point deductions for group members due to reporting are as follows:
    - Points will not be deducted from your account if you are the first group member to press the REPORT button.
    - If you are the second or third group member to press the REPORT button or do not press the REPORT button at all, the second point deduction in step 5 is 30% of the difference between the market price and 101. The point deduction is added up over all periods in which you communicated but the communication was not discovered or reported.

The reduction of the first point deduction by reporting in step 5 remains unchanged.

Acknowledgments:

We thank the editor of the *The Journal of Law, Economics, and Organization* and three anonymous referees for very helpful comments and suggestions. We are further grateful to Maria Bigoni, Peter Dijkstra, Bernhard Ganglmair and Giancarlo Spagnolo for valuable comments. Seminar participants and conference audiences at the following venues gave useful feedback on the paper: International Conference on Competition and Regulation (CRESSE) Rhodos, European Association for Research in Industrial Economics (EARIE) Barcelona, Competition and Innovation Summer School (CISS) Montenegro, German Economic Association (Committee for Social Sciences, Essen), Université de Lausanne, Max-Planck-Institute for Tax Law and Public Finance Munich, OLIGO Workshop Nottingham, Universität Nürnberg, and Swiss IO Day Bern. Thanks to Maria Bigoni for providing us with her z-Tree code. Finally, funding by the Deutsche Forschungsgemeinschaft (DFG) -– project 235577387/GRK1974 — is gratefully acknowledged.

2

Single Monopoly Profits, Vertical Mergers, and Downstream Entry Deterrence

Co-authored with Matthias Hunold

2.1 Introduction

The effects of vertical integration on competition are a key issue in competition policy. Proponents of the Chicago School argue that full vertical mergers enhance efficiency and, at worst, have neutral effects on competition (e.g., Bork, 1978; Posner, 1979). More recent theories based on richer models highlight both the pro-competitive and anti-competitive effects of vertical integration. Game theoretic models reveal anti-competitive effects of vertical integration under specific conditions, such as additional commitment power of the integrated firm (Ordover *et al.*, 1990), secret contract offers (Hart and Tirole, 1990), and the cost of switching suppliers (Chen, 2001).[1] We show that vertical integration can also deter entry in a basic setting where these conditions do not apply. Our analysis of entry deterrence thereby contributes to the literature of market foreclosure.[2]

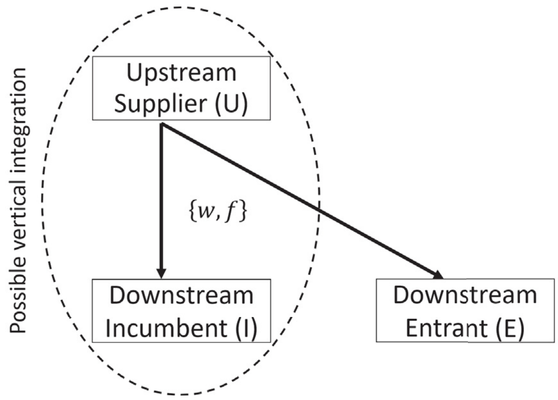


Figure 2.1: Model framework.
Our framework with an incumbent upstream firm and downstream firm as well as a potential downstream entrant. The circle indicates the possible merger of the incumbents, w the per-unit price, and f the upfront fee of the two-part tariff.

Our model features an efficient upstream firm with market power, an established

[1]Other notable assumptions include input choice specifications (Choi and Yi, 2000), two-part tariffs (Sandonís and Faulí-Oller, 2006), exclusive dealing contracts (Chen and Riordan, 2007), upstream collusion (Normann, 2009), only integrated upstream firms (Bourreau *et al.*, 2011) and information leakages (Allain *et al.*, 2010).

[2]Foreclosure refers to the situation that actual or potential rivals' access to supplies or markets is hampered or eliminated, thereby reducing these companies' ability and/or incentive to compete (European Commission, 2008).

downstream firm, and a symmetric (potential) downstream entrant. The upstream firm offers contracts with two-part tariffs to the active downstream firm(s). Figure 2.1 provides an overview.

Absent entry, the upstream firm offers the established downstream firm a contract with a unit price equal to its marginal costs and a fixed fee that leaves the downstream firm indifferent to not buying the input from the upstream firm. The downstream firm accepts the tariff in equilibrium and sets the downstream price or quantity based on the true marginal costs of the supply chain. This leads to the same market outcome as the vertical integration of the two firms, so that vertical integration appears to be innocuous here.

Adverse effects of vertical integration on entry incentives are not obvious either. As the upstream firm can extract the downstream profits with the fixed fee of the two-part tariff, the downstream entrant will only get a profit equal to its outside option, irrespective of whether the upstream firm is vertically integrated with the established downstream firm. The incentives to enter the market thus seem to be unaffected by the vertical integration.

A vertical merger may even be pro-competitive by reducing double marginalization in the case of observable[3] two-part tariff offers when there is downstream competition.[4] With two-part tariffs offered by the upstream firm, the downstream entrant's profit equals its outside option irrespective of whether there is vertical integration of the established firms, such that an asymmetry in the input costs should not affect the entrant's profits and thus the entry incentives.

We contribute by showing that once we add competition through a less efficient

[3]Contract observability means that, when deciding, which supply contract to accept, even the non-integrated downstream firms know what input contracts have been offered to their competitors. Observability results from restricting the upstream firm's offers to be uniform (non-discriminatory) between the independent downstream firms but can also be assumed explicitly for the case of (possibly) non-uniform contract offers.

[4]The upstream firm optimally charges unit prices above costs to account for the competitive downstream margins. With vertical integration, the upstream firm can only increase the unit price for the non-integrated firm. Within the integrated firm, downstream pricing is always based on the true upstream costs. Vertical integration can thus reduce the downstream price level compared to a situation without vertical integration.

fringe at the upstream level, the incumbent firms can vertically integrate to deter entry. This result seems counter-intuitive as the outside option should restrain the upstream supplier in its input pricing. However, with an alternative input supply, the outside option value of the downstream entrant does depend on whether there is vertical integration of the established downstream and upstream firms. The outside options are thus *endogenous* and depend on the upstream market structure and prices.

The twist is that our results also hold even if the alternative supply sources are never actually used in the market. There may merely be the possibility of the downstream firms to ramp up an alternative supply (e.g. less efficient in-house production) if they cannot agree on a contract with the efficient supplier. The setup may therefore look equivalent to a setting with a "pure" upstream monopoly. In particular, the market outcome in this setting is all downstream firms source exclusively from the efficient upstream supplier who consequently has a market share of 100%. Moreover, the alternative source may be relatively inefficient, such that the efficient upstream firm may be able to earn high margins – like a monopolist. This is thus a setting where an observer – for instance a competition authority – might think that Chicago School's *single monopoly profits* theory could apply, such that vertical integration would not raise competitive concerns. However, in this setting, integration could deter entry.

When there is the possibility of an alternative input supply, we present the following theory of harm for the case of *observable* two-part tariff offers: Vertical integration eliminates the double marginalization of the integrated firms, such that the entrant, when not purchasing from the integrated unit, competes against a firm that sets downstream prices based on the true input costs. Instead, without vertical integration of the established firms, the entrant competes against a firm that has a unit input price above the true upstream costs.[5] Consequently, the entrant faces a more

[5]We demonstrate that this is always the case for imperfect price competition with linear demand and also the case for a large parameter range in the case of quantity competition.

aggressive competitor and thus makes lower profits with vertical integration of the established firms. This can deter entry.

Interestingly, the elimination of double marginalization acts as a commitment to intense downstream competition when the entrant does not source the inputs from the efficient upstream firm. While the elimination of double marginalization is often seen as an important pro-competitive effect of vertical integration, it is actually the reason for anti-competitive concerns in the present case.

The article proceeds as follows. We review the related literature in 2.2. We present the model in section 2.3 and solve it for the case of observable two-part tariff offers with both quantity as well as price competition in section 2.4 and conduct a welfare analysis. In section 2.5, we provide additional analyses. We study secret contracting and cover the case of downstream competition for a given market structure. Section 2.6 concludes with a summary and policy conclusions.

2.2 Related literature

In this section we present related literature in addition to the foreclosure literature that we already cited in the introduction.

Related framework. Our article is related to Sandonís and Faulí-Oller (2006) who analyze the competitive effects of vertical integration by a research laboratory. Our model is similar to theirs and they also relate their analysis to market foreclosure (we look at the special case of entry deterrence). However, while they consider two downstream incumbents, we focus on the downstream entry into a monopoly market. The differences in the framework result in opposite effects and different policy implications. We discuss the differences in more detail in section 2.5.1.

Chicago School. Our article links to theories formulated by proponents of the (Post-) Chicago School. See Riordan (2008) for a summary. According to the Chicago School's *single monopoly profits* theory, an upstream monopolist, which can use contracts to extract all monopoly profits from the downstream firms, cannot generate additional profits through vertical integration (e.g., Bork, 1978). Vertical integration would, thus, not have the objective of *leveraging monopoly power* and thus should also not foreclose markets.

In our framework, the efficient supplier and the downstream monopolist can jointly obtain monopoly profits when no entry occurs. As in the *single monopoly profits* theory, vertical integration does not change the total profit obtained by the involved firms. In line with the Chicago School*'s eliminating markups* theory, the efficient upstream firm chooses a contract that prevents the emergence of excessive double marginalization.

In our framework, once we add the possibility of alternative sourcing in the downstream market,[6] vertical integration becomes an instrument to retain monopoly profits through entry deterrence. Note that double marginalization occurs with observable contract offers, even if two-part tariffs are used, as with downstream competition there is a strategic incentive for double marginalization. However, vertical integration eliminates double marginalization for the integrated chain of firms. In contrast to the general perception that the elimination of double marginalization is pro-competitive, we show that the elimination of double marginalization can also be anti-competitive as it leads to more aggressive downstream competition, which can deter entry of an efficiency-enhancing firm.

Secret contracting and opportunism. Whereas our main analysis focuses on the case of non-secret tariff offers, we also study secret contracting and the opportunism problem in section 2.5.2. This relates to the theory summarized under the name *restoring monopoly power* in Riordan (2008). This theory mainly deals with the

[6]This means that firms can produce the inputs less efficiently in-house or purchase them from a less-efficient competitive fringe.

commitment problem whereby an upstream monopolist cannot extract the monopoly profits from the downstream firms due to its opportunistic behavior (Hart and Tirole, 1990).[7] Crucial assumptions for the commitment problem are secret bilateral contracts, two-part tariffs, and no multilateral commitment power of the upstream firm.

It is well known, at least since Hart and Tirole (1990), that in the simple framework with a downstream duopoly that produces homogeneous goods, vertical integration restores the monopoly power of the upstream firm by fully foreclosing the separated firm. When products are differentiated, a vertically integrated firm does not fully foreclose the separated firm. However, vertical integration now solves the commitment problem by creating a situation with complete information; it becomes common knowledge that the monopolist sells input at marginal cost to its subsidiary (compare Rey and Tirole, 2007, p. 32–33, for the case of homogeneous goods).

Rey and Tirole (2007) provide a short analysis of upstream competition through a less efficient firm that offers inputs at marginal cost for the case of secret contracting.[8] Vertical integration now leads to partial foreclosure of the separated downstream firm and, depending on the form of competition, decreases or increases the separated firm's outside option profit. This result is complementary to our analysis of non-secret two-part tariffs. As Rey and Tirole (2007) only analyze the case of quantity competition and interim-unobservability in their overview article, we extend their analysis to price competition as well as interim observability in section 2.5 and demonstrate that the foreclosure results crucially depend on these assumptions. In that section, we also highlight how the foreclosure effects differ between secret and observable two-part tariff offers.

[7] That is, to capture residual demand, the monopolist has an incentive to give a variable cost advantage to the last firm that enters into a contract. The firms with the cost disadvantage anticipate the opportunistic behavior of the monopolist and refuse to accept the monopolizing contracts. Instead of monopoly prices, the monopolist charges marginal cost.

[8] See section 2.2 on page 32 in Rey and Tirole (2007).

Exclusive dealing. Apart from vertical integration, other vertical restraints like exclusive dealing can result in entry deterrence as well (Aghion and Bolton (1987) and Fumagalli and Motta (2006)). In Aghion and Bolton (1987), two vertically-related firms use exclusive dealing to decrease the entry incentives and make themselves better off. While in some sense similar in spirit to our article, there are also other essential differences. Our incumbents use vertical integration to prevent entry in the downstream market and retain the monopoly profit. Vertical integration in our case does not entail any exclusivity. In Aghion and Bolton (1987), the incumbents employ exclusive dealing to mainly obtain a share of the surplus that the entrant generates if entering the supplier's market. The vertical restraint in our model prevents entry as it decreases post-entry profits, such that entry will only occur when entry costs are small enough. In Aghion and Bolton (1987), the exclusive contract creates entry costs and entry depends on the efficiency of the entrant.[9]

Fumagalli and Motta (2006) also look at entry deterrence in the upstream market under exclusive contracts. They show that if competition is fierce enough in the downstream market, the incumbent upstream firm will not employ exclusive contracts to prevent entry.

2.3 Model

There are two downstream firms with index $i \in \{I, E\}$, one is the incumbent firm I and the other a potential entrant E. They produce substitutes. The downstream firms need homogeneous inputs; they transform the input 1:1 into output at zero marginal cost. Supplier U produces the inputs at marginal costs of zero. The downstream firms can alternatively obtain the inputs at marginal costs of $c > 0$. One can think of this less efficient source as either in-house production or a competitive fringe supply. For a small enough c, U is restricted in its pricing. For a large enough c, U

[9]Entry costs either take the form of waiting costs, namely the entrant needs to wait until the contract expires or the entrant pays liquidated damages if the retailer breaks the contract to trade with the entrant.

is an unconstrained monopolist.

The timing is as follows:

1. U and I decide whether to merge or to stay separate.

2. E decides whether to become active at a fixed cost $\theta > 0$.

3. U offers a uniform two-part tariff, comprising a fixed fee f and a unit input price w, to all independent and active downstream firms and, in case of vertical integration, provides the input at marginal cost to firm I.

4. Each active and separate downstream firm either accepts or rejects the contract offer.

5. All active downstream firms set their downstream price p_i or quantity q_i (we study both cases).

For the main analysis in section 2.4, we assume that actions of the previous stages are common knowledge and solve the game by backward induction. The assumption of a uniform (that is: non-discriminatory) two-part tariff in stage 3 already implies observability of the contract terms. This is because in stage 4, when downstream firms decide about contract acceptance, each firm knows the contract terms offered to the competitor.[10] As all independent downstream firms are symmetric post-entry, assuming a uniform two-part tariff is equivalent to allowing for discriminatory tariffs and explicitly assuming contract observability in stage 4.[11] Thus, the assumption of a uniform tariff is – in our case – without loss of generality and simplifies the notation.

[10]The assumptions of uniform pricing and the observability of contract acceptance and rejection are also used in related models, such as Caprice (2006) and Hunold (2020).

[11]The equilibrium offers turn out to be symmetric. Exclusion of one downstream firm by making asymmetric offers with one being effectively no offer is not optimal as (i) either there is an alternative supply source that this downstream firm would use or (ii) if there is no efficient enough alternative supply source, the upstream firm can extract all profits from the downstream firm. The assumption of observable two-part tariffs has been used in similar models. See, for instance, Hunold and Stahl (2016).

We denote equilibrium outcomes with the superscripts M for monopoly, D for duopoly, I for integration, and S for separation. For instance, $\{w^{DS}, f^{DS}\}$ denotes the equilibrium tariff for the case of duopoly and vertical separation and $\{w^{DI}, f^{DI}\}$ the equilibrium tariff for a duopoly in the vertically integrated case.

Profits. For the profit of the up- and downstream firms, we use the following notation:

- $\Pi_U(w)$ denotes the profit of supplier U in case of vertical separation as a function of the unit input price w.
- $\Pi_i(p_i, p_{-i})$ denotes the profit of downstream firm i as function of its own downstream price p_i and the price p_{-i} of its rival $-i$, with $i \neq -i \in \{I, E\}$.
- Π_{UI} denotes the profit of the vertically integrated firms U and I.

The downstream firms are symmetric apart from E's cost of entry. The entrant E only enters when its post-entry profit is larger than its entry cost of θ. One can think of the entry cost as a random variable from the perspective of the established firms, such that entry is more (less) likely when the post-entry profits of the entrant are higher (lower). We use the following notation for the outcomes of stage 5.

- Denote by $\widetilde{q}(x, y)$ and $\widetilde{p}(x, y)$ the reduced-form quantities and prices as a function of the downstream firm's own unit input cost x and the unit input cost y of its competitor.
- Denote by $\pi(x, y)$ a downstream firm's profit before the fixed fee with x and y defined as above.

For these functions, we denote the case where the entrant is not active by setting its unit input costs to ∞.

We focus on reduced-form profits that are well-behaved in the players' actions, such that the relevant first-order conditions characterize the equilibria. A sufficient condition for this is linear demand and imperfect downstream competition, as specified below. In particular, for comparative statics, we also assume that

Assumption 1. *The profit of a downstream firm depends negatively on its own input cost* $(\partial\pi(x,y)/\partial x < 0)$ *and positively on the input cost of its competitor* $(\partial\pi(x,y)/\partial y > 0)$.

Parametric demand based on a linear quadratic utility function. We study the game for both downstream price and quantity competition and denote the general demand and inverse demand functions by $q_i(p_i, p_{-i})$ and $p_i(q_i, q_{-i})$. We characterize results in terms of general demand and profit functions as far as possible. We sometimes use a parametric demand function, which we derive from a linear-quadratic utility function as in Sandonís and Faulí-Oller (2006):

$$u(q_I, q_E) = q_I + q_E - \frac{q_I^2}{2} - \frac{q_E^2}{2} - \gamma q_I q_E. \tag{2.1}$$

Parameter γ measures the degree of substitutability between the products and ranges from $\gamma = 0$ for independent products to $\gamma = 1$ for perfect substitutes.[12] The representative consumer maximizes $u(q_i, q_{-i}) - \Sigma_{i=I,E}\, p_i q_i$ with respect to q_i, where q_i is the amount that the consumer purchases from firm i and p_i the respective price, with $i \in \{I, E\}$ and $i \neq -i \in \{I, E\}$.[13] Utility maximization implies the inverse linear demand function for product i of

$$1 - q_i - \gamma q_{-i} \tag{2.2}$$

and the parametric demand function for product i of

$$\frac{1 - p_i - \gamma + \gamma p_{-i}}{1 - \gamma^2}. \tag{2.3}$$

Welfare is given by

$$W(q_I, q_E) = u(q_I, q_E) - \theta \cdot I(\text{entry}), \tag{2.4}$$

[12]The value of $\gamma = 1$ is excluded for price competition. In this case, the entry incentives do not depend on vertical integration because the downstream entrant would make zero profits in any case because of perfect price competition.

[13]The demand functions derived from the underlying utility function in equation (2.1) allows for a consistent analysis of entry where another, possibly differentiated product becomes available. They also allow for an expansion of demand thereby preventing an underestimation of the entrant's incentive to enter (Höffler, 2008; Levitan and Shubik, 1971).

where the indicator $I(entry)$ is one in the case of downstream entry and zero otherwise.[14]

2.4 Main analysis

2.4.1 Pricing and output decisions (stages 3 to 5)

Downstream monopoly (no entry) and vertical separation. Absent entry, the independent downstream incumbent maximizes its profit

$$\Pi_I(p_I, q_I) = (p_I - w)\, q_I - f$$

by setting either the monopoly price $p_I = \widetilde{p}(w, \infty)$ or the quantity $q_I = \widetilde{q}(w, \infty)$, which depend on the unit input price w.

Under vertical separation, supplier U maximizes

$$\Pi_U(w) = \{w \cdot \widetilde{q}(w, \infty) + f\} \tag{2.5}$$

with respect to w and f, subject to the downstream incumbent's participation constraint

$$\underbrace{\pi(w, \infty)}_{\text{operational profit}} - f \geqslant \underbrace{\pi(c, \infty)}_{\text{outside option profit}}.$$

The profit on the left – that the incumbent obtains when buying from U – has to be larger or equal to its outside option on the right, which is the profit that I obtains when sourcing alternatively at marginal costs of c. Recall that we set the entrant's cost to ∞ when the entrant is not active.

For a given unit price w, supplier U chooses a fixed fee f, such that the above participation constraint binds:

[14] To be precise, the quantity of input produced by the alternative supply source at the inefficiently high cost of c also enters the welfare function. However, in all equilibria, we will obtain that this quantity is zero. We therefore abstract from it in the welfare function.

$$f = \pi(w, \infty) - \pi(c, \infty). \tag{2.6}$$

The equilibrium profit of I thus equals its outside option profit of $\pi(c, \infty)$. The efficiency of the alternative therefore determines the share of the monopoly profit that the downstream incumbent can keep.

As U extracts the residual downstream profit through the fixed fee, it maximizes the industry profit. This profit is maximized when the unit price equals marginal cost. Consequently, the equilibrium two-part tariff is given by

$$\left\{w^M, f^M\right\} = \{0, \pi(0, \infty) - \pi(c, \infty)\}.$$

Lemma 1. *Under vertical separation and without downstream entry, the downstream monopolist obtains a profit $\pi(c, \infty)$ equal to its outside option.*

Downstream monopoly (no entry) and vertical integration. Vertical integration is profitable for U and I when their joint profit post-merger exceeds their joint profits prior to the merger. Absent entry, the incumbents jointly earn monopoly profits both under vertical separation and vertical integration and cannot increase their joint profits through integration.

Lemma 2. *Absent entry, the incumbents jointly earn monopoly profits irrespective of their integration decision and are indifferent between integration and separation.*

Downstream duopoly (entry) and vertical integration. Supplier U offers a two-part tariff that the entrant accepts. Given the unit price w, the supplier sets f such that the downstream firm's profit equals its outside option profit. The entrant's outside option profit is $\pi(c, 0)$ under vertical integration as U cannot commit to charging an input price above its marginal cost to its subsidiary I.

Supplier U now maximizes

$$\Pi_U = p_I \widetilde{q}(0,w) + w \cdot \widetilde{q}(w,0) + \underbrace{\pi(w,0) - \pi(c,0)}_{\text{fixed fee of entrant}} \tag{2.7}$$

with respect to w. Let (w^{DI}, f^{DI}) denote the resulting equilibrium tariff for the downstream duopoly in the vertically integrated case, where w^{DI} is defined by $\partial\Pi_U/\partial w = 0$ and f^{DI} equals $\pi\left(w^{DI},0\right) - \pi(c,0)$.

Lemma 3. *With vertical integration of the incumbents U and I, the downstream entrant E obtains a profit of* $\pi(c, 0)$.

Downstream duopoly (entry) and vertical separation. Supplier U extracts all profits from the two downstream firms through the fixed fee, except for their outside option values. For this case the assumption of a uniform input tariff matters as it implies that the tariff offers are observable to the downstream firms in stage 4. This eliminates the opportunism problem, which we study in subsection 2.5.2.
With downstream competition, a downstream firm's profit when sourcing alternatively equals $\pi(c,w)$ and depends on the input price of the competitor. The unit input cost of the competitor is w. The outside option profit is reached through a unilateral deviation from the equilibrium path where both downstream firms buy from supplier U based on the uniform two-part tariff (w, f).[15]
Supplier U maximizes

$$\Pi_U = \sum_{i\in\{I,E\}} \left(w \cdot \widetilde{q}(w,w) + \underbrace{\pi(w,w) - \pi(c,w)}_{\text{fixed fee}} \right) \tag{2.8}$$

with respect to w.
Without a relevant outside option (c large enough), the outside option profits are zero ($\pi(c,w) = 0$) and the profit in equation (2.8) equals the industry profit. With

[15] Recall that a uniform tariff implies that the contract offers are observable.

downstream competition, the industry profit is maximized at a unit input price above marginal costs in order to induce the downstream firms to set prices at the monopoly level.

With relevant outside options, U can extract more profits from the downstream firms when their outside option profit $\pi(c, w)$ is low. The supplier thus has an incentive to set w below the industry maximizing level. Let (w^{DS}, f^{DS}) denote the resulting equilibrium tariff, where w^{DS} is defined by $\partial\Pi_U/\partial w = 0$ and $f^{DS} = \pi\left(w^{DS}, w^{DS}\right) - \pi\left(c, w^{DS}\right)$.

Lemma 4. *Each downstream firm obtains a profit $\pi\left(c, w^{DS}\right)$ when both downstream firms are vertically separated.*

2.4.2 Entry incentives and vertical integration (stage 2)

The entrant earns a profit $\pi\left(c, w^{DS}\right)$ under separation of U and I and $\pi(c, 0)$ under integration (lemmas 4 and 3). Recall that the profit increases as the unit input costs of the competitor increase (assumption 1). Integration decreases the post-entry profits and thus the incentives to enter when the entrant's profit is larger under separation than integration:

$$\pi\left(c, w^{DS}\right) > \pi(c, 0).$$

This holds if U finds it optimal to set the unit input price with a downstream duopoly under separation above marginal costs:

$$w^{DS} > 0.$$

Instead, for $w^{DS} < 0$ the entrant's incentive to enter is larger under integration.

Proposition 1. *Vertical integration of U and I yields a lower post-entry profit for the entrant than separation if and only if $w^{DS} > 0$, and a higher profit iff $w^{DS} < 0$.*

We conclude that for vertical integration to affect entry, the entry costs must be in-between the profits of the entrant with vertical separation and integration of the incumbents (at least with positive probability):

$$\min\left(\pi\left(c,0\right),\pi\left(c,w^{DS}\right)\right)<\theta\leq\max\left(\pi\left(c,0\right),\pi\left(c,w^{DS}\right)\right). \tag{2.9}$$

Proposition 2. *Provided (i) it is optimal for the supplier to charge a unit input price above marginal costs under vertical separation ($w^{DS}>0$), and provided (ii) entry costs are in an intermediate range according to 2.9, then a merger between U and I decreases the entrant's post entry profit and thus the likelihood of entry.*

The above propositions make clear that it is crucial to determine whether the unit input price in the case of entry and vertical separation is above the supplier's marginal costs. To assess the conditions under which this is the case, we employ the linear demand functions specified in equations (2.2) and (2.3).

Proposition 3. *Suppose that the downstream firms compete in prices and demand is given by equation (2.3). The unit input price under separation always exceeds U's marginal cost: $w^{DS}>0$. Vertical integration of the established firms thus implies lower post-entry profits of the entrant than separation.*

Proof. See Appendix A. □

While the result is unambiguous with price competition, the findings are more differentiated with quantity competition.

Proposition 4. *Suppose the downstream firms compete in quantities and demand is given by equation (2.2). The unit input price under separation exceeds U's marginal cost ($w^{DS}>0$) if the efficiency advantage of the supplier U over the alternative is sufficiently large:*

$$c>\hat{c}(\gamma). \tag{2.10}$$

The threshold is higher when the products are closer substitutes: $\hat{c}'(\gamma)>0$.

Proof. See Appendix A. □

Under price competition, vertical integration of the incumbents always leads to lower post-entry profits of the entrant than separation. With quantity competition, this is the case under condition (2.10), which implies that the unit wholesale price is above the supplier's marginal cost ($w^{DS} > 0$).

Vertical integration does not decrease the entrant's profits under quantity competition if the supplier has too strong incentives to decrease the downstream firms' outside option profits in the case of vertical separation and downstream duopoly. See equation (2.8) where c only enters the outside option profit $\pi(c, w)$. As can be seen in Figure 2.2, this incentive dominates when U's cost advantage is small, such that the downstream firms can keep a relatively large share of their flow profits. The effect is stronger when the products are less differentiated, which implies relatively intense downstream competition. A lower unit price w decreases the downstream firms' outside option profits $\pi(c, w)$. As a result, the supplier can extract a larger share of downstream profits through the fixed fees.

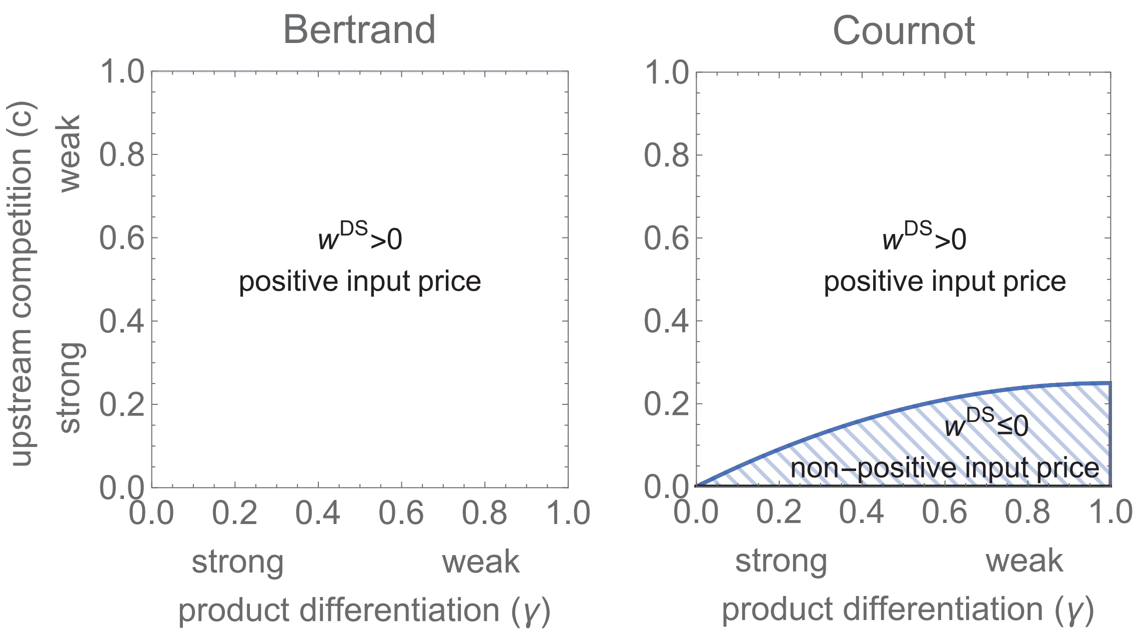


Figure 2.2: Input price in relation to marginal cost.
Input price w^{DS} under Bertrand (left) and Cournot (right) and separation as a function of product differentiation γ and the cost of sourcing alternatively c.

Discussion: below-cost pricing. Although the (unrestricted) equilibrium unit price is below marginal cost for quantity competition if $c < \hat{c}(\gamma)$, this is not necessarily the most likely real market outcome. Negative unit prices can be implausible for various reasons. Prices below marginal cost may be considered anti-competitive and might be prohibited, especially when a firm has a strong or even dominant position. Intel's fidelity rebates provide an example of below marginal cost pricing of a dominant firm that was ruled to be anti-competitive by the European Commission (European Commission, 2009).[16] The case that vertically integrated incumbents separate in order to achieve unit wholesale pricing below marginal cost in the case of entry may thus be of little practical relevance. We are therefore cautious in drawing conclusions from the case where the (unrestricted) equilibrium price is below marginal cost.

2.4.3 Profitability of vertical integration and welfare (stage 1)

We focus on the case where

$$\pi\left(c, 0\right) < \theta < \pi\left(c, w^{DS}\right), \tag{2.11}$$

which implies $w^{DS} > 0$ and arises under both price and quantity competition for a large parameter range (propositions 3 and 4).[17] Figure 2.3 illustrates our focus on medium entry costs.

[16]Intel awarded rebates to major original equipment manufacturers under the condition that the manufacturers purchase at least 80% of their supply needs for x86 CPUs from Intel. The EU's Court of Justice ruled that Intel's behavior tied purchasers and thereby diminished the ability of competitors to compete for the respective product. The Commission furthermore ruled that Intel's use of fidelity rebates establishes an abuse of its dominant position.

[17]For larger entry costs ($\theta > \pi\left(c, w^{DS}\right)$), entry never occurs. For smaller, entry costs ($\pi\left(c, 0\right) > \theta$), entry always occurs. We study the latter case in section 2.5.

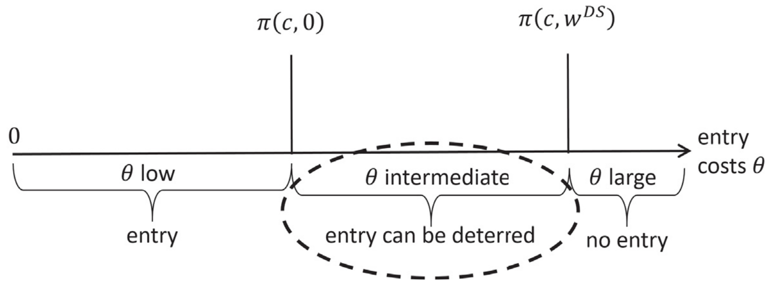


Figure 2.3: Entry costs.
Entry always occurs for low enough entry costs and never occurs for large enough entry costs. The circle highlights our focus on intermediate entry costs and cases where $w^{DS} > 0$, which together implies that vertical integration deters entry.

Merger incentives. When the incumbents U and I are separate, market entry unleashes two countervailing effects that both enhance welfare and affect the incumbents in different ways:

- The market expands when E's product is differentiated and attracts new consumers. As U can capture some of the additional profit through the sale of its input, the *market expansion effect* makes an entry-deterring merger less profitable. The magnitude of this effect depends on the upstream margins and thus on the efficiency advantage c of the incumbent over the alternative supply source.
- Market entry creates competition in the downstream market, which leads to lower equilibrium prices and decreases profits. The *competition effect* makes an entry-deterring merger more profitable.

Proposition 5. *Suppose condition (2.11) holds, such that vertical integration of the incumbents deters entry. A vertical merger that deters entry yields higher profits than separation for the incumbents if the competition effect of entry dominates its market expansion effect. With linear demand (equations (2.2) and (2.3)), vertical integration leads to higher profits than vertical separation when the supply alternative is relatively efficient:*

$$c < \bar{c}_k(\gamma)$$

with $k \in \{Bertrand, Cournot\}$ *and* $\bar{c}_k'(\gamma) > 0$.

Proof. See Appendix A. □

The conditions under which entry-deterring vertical integration is profitable for the incumbents are qualitatively the same under price and quantity competition, although the exact parameter range differs slightly in the illustrative case with linear demand (see Figure 2.4):

- Vertical integration is profitable (the *competition effect* prevails) when the fringe is relatively efficient (c small) and the products are relatively homogeneous (γ large).
- Vertical integration is unprofitable (the *market expansion effect* prevails) when the fringe is relatively inefficient (c large) and the products are relatively differentiated (γ small). In this case, the incumbents favor separation and E enters the market.

Figure 2.4 depicts the merger incentives of vertical integration as a function of c and γ. Entry-deterring vertical integration yields lower profits for the incumbents than vertical separation in the north-west of the dashed line and higher profits in the south-east.

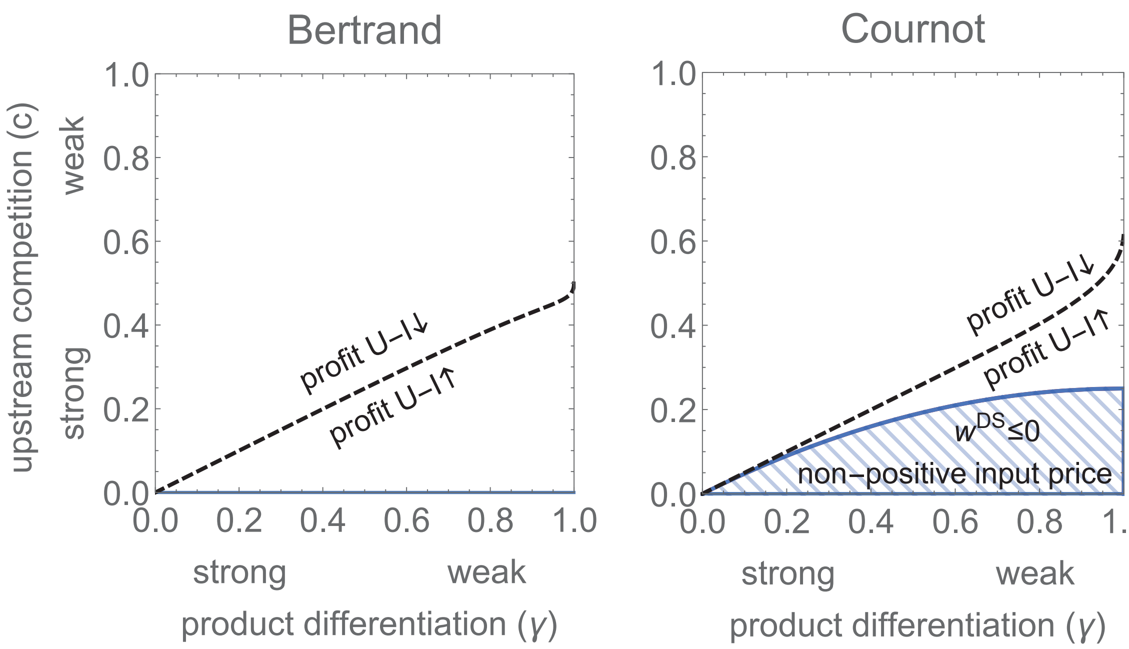


Figure 2.4: Profitability of vertical integration.
Profitability of entry-deterring vertical integration for the incumbents U and I. Vertical integration deters entry except where $w^{DS} < 0$ under Cournot. Entry deterring vertical integration always lowers welfare.

Welfare. For the welfare analysis we also focus on intermediate entry costs and positive unit input prices under vertical separation and duopoly (condition (2.11)).[18] To analyze the effect of vertical integration and entry deterrence on welfare, we compare welfare of the cases

- integration and downstream monopoly[19] and
- separation, entry and downstream duopoly.

Entry under vertical separation affects total welfare in three ways:

- The entry costs decrease welfare;
- The lower price level under duopoly increases welfare (competition effect);

[18] In section 2.5.1, we analyze welfare, taking into account low entry costs. In doing so, we consider a duopoly in the downstream market for the welfare comparison under both integration and separation.

[19] Recall that with a downstream monopoly the market outcome and thus welfare is the same under vertical integration and vertical separation (lemma 2).

- The additional variety under duopoly (for differentiated products) increases demand and welfare for a given price level (market expansion effect).

Using the welfare function $W(q_I, q_E)$ of equation (2.4) that measures total surplus based on the linear-quadratic utility, entry under vertical separation yields higher welfare than vertical integration and no entry if

$$u\left(q_I = \widetilde{q}(w^{DS}, w^{DS}), \widetilde{q}(w^{DS}, w^{DS})\right) - \theta > u\left(\widetilde{q}(w^{M}, \infty), 0\right). \tag{2.12}$$

For the next proposition, we evaluate condition (2.12) at the upper bound of intermediate entry cost of $\bar{\theta} = \pi\left(c, w^{DS}\right)$, see inequality (2.11).

Proposition 6. *The welfare function (equation (2.4)), attains a strictly higher value under vertical separation and entry than under entry-deterring vertical integration for any value of intermediate entry costs as defined by condition (2.11).*

Proof. See Appendix A. □

For intermediate entry costs, the optimal merger policy is simple and summarized in

Corollary 1. *Whenever the unit input prices under vertical separation and duopoly are (expected to be) above costs ($w^{DS} > 0$ in the model), a vertical merger restricts potential competition with detrimental effects on welfare and thus should be prohibited absent other efficiencies (which are not modeled here).*

One may wonder whether this policy is still optimal if the entry costs are possibly not "intermediate". For larger entry costs, entry never occurs and vertical integration has no effects on welfare in the model. In that sense the policy does not yield a welfare loss if it is applied to cases of larger entry costs.

For entry costs that are smaller than intermediate, entry always takes place. The question is thus no more whether potential competition is restricted but whether actual competition suffers from vertical integration. We discuss this case and compare it to the case of intermediate entry costs in section 2.5.1.

Finally, if one expects unit input prices at the level of marginal costs ($w^{DS} = 0$), the interval of intermediate entry costs is empty (see condition (2.11)) and the analysis of either small or large entry costs applies. As the case of $w^{DS} = 0$ may only arise for certain parameter combinations (c, γ) under quantity competition (see figure 2.2), the implication for these cases apply.[20]

2.5 Additional analyses

2.5.1 Small entry costs and Sandonís and Faulí-Oller (2006)

Our analysis of merger profitability and welfare in section 2.4.3 focuses on intermediate entry cost, defined by

$$\pi(c, 0) < \theta < \pi\left(c, w^{DS}\right),$$

such that entry takes place under vertical separation whereas vertical integration deters entry. For intermediate entry cost, we therefore derive our insights on the profitability of a vertical merger and its welfare effects from the comparison of the cases of vertical integration when there is a downstream monopoly (MI) and vertical separation when there is a downstream duopoly (DS) and the entry cost materialize (as depicted in table 2.1).[21]

[20]See the discussion at the end of section 2.4.2 for the case of unit input prices below costs.

[21]Albeit the cases MI and MS are equivalent in terms of profits and welfare. The point is that we compare a monopoly situation with a duopoly situation under vertical separation, whereas Sandonís and Faulí-Oller (2006) compare vertical integration and vertical separation both under duopoly.

		vertical link between U and I	
		integration	separation
# of active downstream firms	1 (only I)	MI	MS
	2 (I and E)	DI	DS

Table 2.1: Combinations of entry and vertical ownership.

For small enough entry cost, entry takes place irrespective of whether the incumbents are vertically integrated. Comparing the effects of vertical integration and vertical separation in these cases therefore boils down to comparing the market outcome with a duopoly under vertical integration to that with a duopoly under vertical separation (cases DI and DS in table 2.1).
This relates to Sandonís and Faulí-Oller (2006) who analyze the competitive effects of vertical integration by a research laboratory. They also briefly relate their analysis to market foreclosure, although this is not their focus. For the case of small entry cost, our simplified framework indeed resembles theirs.[22] For vertical separation and duopoly, they also focus on non-negative unit input prices under Cournot competition and unit prices weakly below the alternative cost of c under price competition.[23]

Welfare analysis. With small entry cost, the analysis of total surplus essentially boils down to comparing the downstream prices between vertical separation and vertical integration. There is a trade-off between

1. eliminated double marginalization with vertical integration and

[22] Sandonís and Faulí-Oller (2006) do not consider entry costs at all. However, when the entry cost are small enough, they are fixed sunk cost in both ownership cases (vertical integration and vertical separation) and therefore do not affect comparisons of profits and welfare across the cases (at least in absolute terms).

[23] Please see the discussion at the end of section 2.4.2 for further details.

2. the outside option profits $\pi(c, w)$ under vertical separation disciplining the wholesale price level of both downstream firms.

The welfare analysis for intermediate entry cost, which we conduct in section 2.5.2, differs as it essentially compares a downstream monopoly and duopoly, which includes the following effects:

- Entry costs decrease welfare under duopoly;
- The lower price level under duopoly increases welfare;
- The additional variety under duopoly (for differentiated products) increases demand and welfare for a given price level.

For intermediate entry costs, an entry-deterring vertical integration decreases welfare for all upstream efficiency differentials (proposition 6), for quantity as well as for price competition. Both with a total surplus and with a consumer surplus standard, the optimal merger policy is simple: When entry and potential competition is of concern in a market, a vertical merger is welfare decreasing and should be prohibited absent further efficiencies as it tends to deter entry.[24] Our analysis thus suggests the presumption of anti-competitive vertical mergers in the model at hand.
This differs partly in the case of small entry costs as studied by Sandonís and Faulí-Oller (2006). For quantity competition, welfare is only higher in the case of vertical integration when the alternative supply is relatively inefficient (c large enough). For price competition, Sandonís and Faulí-Oller (2006) welfare is higher with vertical integration either when the alternative supply is rather inefficient or highly efficient but lower in an intermediate range of efficiency differentials (c). See Figure 2.5 for an overview.
The different results should find consideration when assessing a vertical merger.

[24]This holds strictly for price competition and when unit input prices above marginal costs are expected under vertical separation and downstream duopoly. We exclude below-cost pricing in our analysis, as do Sandonís and Faulí-Oller (2006). For the case of unit input prices at costs, the results of the case with small entry costs apply if entry is feasible. When entry is not feasible, the merger is welfare-neutral.

Analysis of entry deterrence. For intermediate entry costs, we compare the duopoly profit, which the entrant would obtain with either vertical integration or vertical separation of the incumbents. The case of vertical integration and duopoly never materializes for medium entry costs when vertical integration deters entry but is rationally anticipated by the entrant when deciding whether to enter the market.

With small entry costs, there is no relevant entry decision as entry is always profitable. The welfare effects of vertical integration depend purely on how vertical integration affects the downstream prices relative to vertical separation but not on whether the independent firm makes higher or lower profits with vertical integration.

Optimal merger policy. Comparing the figures 2.4 and 2.5 shows that the profitability of vertical integration depends on whether entry always occurs or whether it depends on the vertical ownership (it also depends on the parameters c and γ). It is noteworthy that a vertical merger may generate other synergies (e.g. a reduction in some forms of fixed costs) that may make a merger profitable even if it is not profitable according to the price effects implied by vertical integration which are depicted in the figures. A vertical merger may thus be proposed to a competition authority of any parameter combination in terms of upstream constraint c, downstream substitutability γ and entry costs θ.

As described above, the welfare effects of vertical integration depend on whether the downstream market features a duopoly (low entry costs) irrespective of vertical ownership or whether there is potential downstream entry (intermediate entry costs). A competition authority should thus take this distinction into account when assessing the likely effects of the merger. Our analysis of intermediate entry costs thus complements the analysis of Sandonís and Faulí-Oller (2006) in the arguably highly relevant dimension of potential competition.

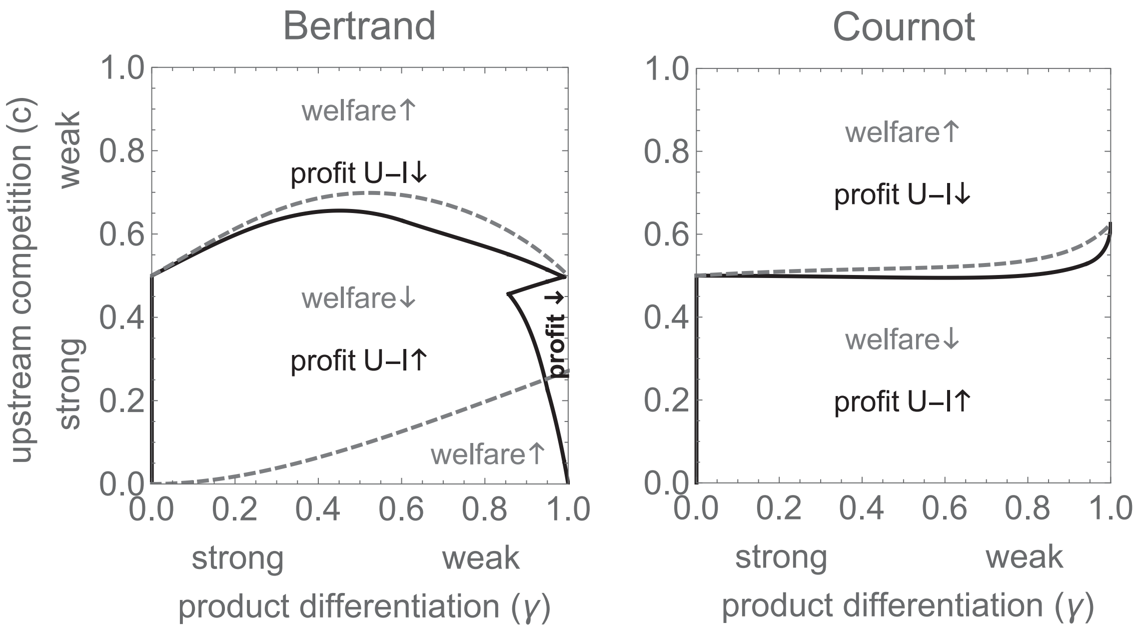


Figure 2.5: Merger profitability and welfare with small entry costs.
Effects of vertical integration compared to separation when there is always a downstream duopoly (small entry costs).

2.5.2 Secret contracting and interim observability

The analysis in the previous section 2.4 relies – for the case of vertical separation and downstream competition – on the assumption of uniform and thus observable contract offers. With these contract offers, there is no opportunism problem as described by (Hart and Tirole, 1990). Let us now allow for firm-specific contract terms (f_i, w_i) with $i \in \{I, E\}$. We will look into the following two cases where the competitors have limited information about the rival's supply source and contract terms:

- **Full secrecy.** The (non-integrated) downstream firms do not know the contract that U offers to the rival. Moreover, they do not know whether the rival accepts the offer or sources alternatively when setting the price or quantity.
- **Interim observability**. The contract terms of U remain secret but the acceptance/rejection decisions of stage 4 (see section 2.3) and the decision where to

source are observable when the downstream firms set their prices/quantities.[25]

For the relevant case of a downstream duopoly and vertical separation, we focus on equilibria with passive beliefs where, in the case of an unexpected offer, the downstream firms believe that the competitor receives an equilibrium contract offer and accepts it. The resulting symmetric equilibrium, when it exists, features unit cost pricing ($w_I = w_E = w^{DS} = 0$).[26] We start from the case of full secrecy and quantity competition as analyzed in Rey and Tirole (2007) and explain how the results differ with price competition under full secrecy as well as under both price and quantity competition under interim observability.[27]

Full secrecy (interim unobservability). What matters for the entrant's profits is whether the downstream incumbent reacts when the entrant rejects the contract offer of U and procures the input alternatively at the higher unit input cost of c instead of $w^{DS} = 0$.

With vertical separation, the downstream firms cannot observe their rival's actual supply choice and base their strategy on the belief that their rival purchases input from the efficient supplier – which indeed happens in equilibrium. As a consequence, the downstream incumbent does not choose the best response quantity or price when the entrant deviates by sourcing input at a higher input price from the alternative.

For **quantity competition**, $\widetilde{q}(0,0)$ is the equilibrium quantity per downstream firm. With vertical separation, the expected deviation profit of the entrant in the case of separation is given by

$$\max_q \left(p_E\left(q, \widetilde{q}(0,0)\right) - c\right) \cdot q$$

where q is the best response of the entrant when expecting an output of $\widetilde{q}(0,0)$ of

[25] See, for instance, Caprice (2006) for a model with interim observability. Caprice (2006) uses this assumption in conjunction with a ban on price discrimination but does not consider vertical integration.

[26] See Rey and Verge (2004) for details. For the case of price competition, the equilibrium in passive beliefs may only exist if the degree of downstream substitution is not too high.

[27] Their analysis of the independent downstream firms is analogous to our analysis of the entrant in terms of post-entry profits.

the rival and having marginal costs of c.

With vertical integration, the entrant, when deviating, knows that the integrated downstream rival knows that the entrant has rejected the contract offer. The downstream rival thus plays a best response to the entrant's actual unit input costs of c. Given strategic substitutes, this implies a higher quantity for the incumbent and a lower quantity, and thus lower deviation profits, for the entrant, when compared to the case of vertical separation described above. As U sets the entrant indifferent to its outside option profit by means of the fixed fee, vertical integration yields lower profits for the entrant than vertical separation under quantity competition. See Rey and Tirole (2007) on p. 33 for a more detailed analysis of this case.

Suppose now that there is downstream **price competition** with strategic complementarity, such that both downstream firms benefit from an increase in the downstream prices. The outside option profit of the entrant now turns out to be higher under vertical integration compared to vertical separation.

To understand why this is true, note that the equilibrium prices under vertical separation equal $\widetilde{p}(0,0)$ as $w^{DS} = 0$. The resulting deviation profit of the entrant under vertical separation equals

$$\max_{p} \, (p - c) \, q_E(p, \widetilde{p}(0,0)),$$

where p is the best response of the entrant when expecting a price of $\widetilde{p}(0,0)$ of the rival and having marginal costs of c. With vertical integration, the entrant's profit when sourcing alternatively equals $\pi(c,0)$ as in the case of observable tariffs because the vertically integrated entity of U and I has complete information about the downstream entrant's input costs and will therefore charge higher prices. As prices are strategic complements, the entrant benefits from vertical integration as, in case of alternative sourcing, the rival's price under vertical integration is above that under vertical separation: $\widetilde{p}(0,c) > \widetilde{p}(0,0)$.[28]

[28]This follows under standard assumptions on the profits and strategic complementarity $(\partial^2 \Pi_I / (\partial p_I \partial p_I) > 0)$, as it is the case with linear demand.

We summarize the results in the first column of table 2.2.

Interim observability. What matters for the entrant's equilibrium profits are the deviation profits when sourcing alternatively. As discussed above, with vertical integration the downstream incumbent I knows the true costs of the entrant in the case of a deviation. This yields a deviation profit for the entrant of $\pi(c, 0)$.
Interim observability implies that even with vertical separation the downstream incumbent I knows when the entrant rejects the offer of U and sources alternatively. Firm I thus plays a best response to the rival having marginal cost of c. This drives the difference in results compared to fully secret contracts.
However, relative to the case of full secrecy, there is no additional information in the stage of contract acceptance when there is interim observability. Hence, there still exists an equilibrium in passive beliefs under vertical separation where U charges each downstream firm a unit price of $w^{DS} = 0$ both under price and under quantity competition.[29] For $w^{DS} = 0$, there is no difference between vertical separation and vertical integration for the entrant's profit.

Discussion of results. Table 2.2 compares the entrant's profits with secret contract offers to our main results with observable contract offers.[30] One can see that the result of Rey and Tirole (2007) whereby an independent downstream firm's profits are lower under vertical integration crucially depends on each of the following two assumptions:

1. Quantity competition: With price competition and full secrecy, vertical integration leads to higher profits of the entrant due to strategic complementarity in the case of price competition instead of strategic substitutability in the case of quantity competition.

[29] See Caprice (2006) for a formal proof in the case of quantity competition. The argument is analogous for price condition with the caveat that equilibrium existence is subject to conditions as under full secrecy, see fn. 26. Moreover, marginal cost pricing may no longer exist in the case of wary beliefs, see Rey and Verge (2004).

[30] The results hold for $w^{DS} \geq 0$ when the contract offers are observable.

2. Full secrecy – even the downstream firms' sourcing decisions are not observable by the rivals under vertical separation.

    - With interim observability, vertical integration has no negative effect on the entrant's profits (see column ii).
    - Fully observable tariff offers imply lower profits of the entrant with vertical integration and quantity competition in certain cases and equal or possibly higher profits (see the bottom of column iii).

The economics for the different profits of the entrant under vertical integration and vertical separation of the incumbents differ between observable and unobservable contract offers. When the contracts are secret or interim observable, the opportunism problem prevents the efficient upstream firm under vertical separation from charging unit prices above its marginal cost ("strategic double marginalization"). Such strategic double marginalization instead occurs when the contracts are observable. In the latter case, vertical integration eliminates the double marginalization of the incumbents and yields more aggressive downstream competition when the entrant sources alternatively.[31]

Summary. With full secrecy, vertical integration hurts the entrant with strategic substitutes (competition in quantities) as it provides the integrated rival with knowledge about the entrant's deviation to higher marginal costs, which leads to a more aggressive action of the rival. This result is due to Rey and Tirole (2007). Focusing on equilibria with passive beliefs, we have shown that with price competition and strategic complementarity, instead, this knowledge leads to the accommodating action of a higher price, which benefits the entrant. Moreover, vertical integration does not affect the entrant's profits when the downstream firm's sourcing decisions are observable (interim observability).

[31]This holds for price competition and a large parameter range of quantity competition (propositions 3 and 4).

		contract observability		
		(i) full secrecy	**(ii) interim observable**	**(iii) observable**
competition in	**prices**	$\Pi_E^{VS} < \Pi_E^{VI,32}$	$\Pi_E^{VS} = \Pi_E^{VI}$	$\Pi_E^{VS} > \Pi_E^{VI}$
	quantities	$\Pi_E^{VS} > \Pi_E^{VI}$	$\Pi_E^{VS} = \Pi_E^{VI}$	$\Pi_E^{VS} <> \Pi_E^{VI,33}$

Table 2.2: The table states the relationship of the entrant's profits (Π_E^j) between separation (j = VS) and integration (j = VI) of the incumbents.

2.6 Conclusion

We review the Chicago School's *single monopoly profits* theory whereby an upstream monopolist, which can use contracts to extract all monopoly profits from the downstream firms, cannot generate additional profits through vertical integration. For this, we employ a model where the upstream firm uses two-part tariffs to sell inputs to a downstream incumbent and – in the case of entry – an entrant.

For the case that the downstream firms cannot avoid sourcing from the upstream firm in order to be active in the market, our results are consistent with the Chicago School's *single monopoly profits* theory. The upstream monopolist, which can use contracts to extract all monopoly profits from downstream firms, cannot generate additional profits through vertical integration. The downstream entrant's profit equals its outside option irrespective of vertical integration, such that the vertical integration of the incumbents has no effect on the incentives to enter the market.

[32]See fn. 26 regarding equilibrium existence.

[33]The case $\Pi_E^{VS} > \Pi_E^{VI}$ occurs for a large range of parameters with linear demand. The case $\Pi_E^{VS} < \Pi_E^{VI}$ only occurs if negative marginal input prices are feasible. See proposition 4.

The result is different when the downstream firms can alternatively produce the inputs less efficiently in-house or purchase them from a competitive fringe supply. With an alternative input supply, the outside option value of the downstream entrant does depend on whether there is a vertical integration of the established downstream and upstream firms.

We show by means of linear demand that with downstream price competition and – for a large parameter range – quantity competition, the entrant faces a more aggressive competitor when obtaining the inputs alternatively (the outside option) and thus makes lower profits with the vertical integration of the established firms. If the entry costs are in a range that a vertical merger can deter entry, the following trade-off arises for the incumbent firms: On the one hand, vertical integration would deter entry and retain the downstream monopoly. On the other hand, the incumbents can only capture their share of the additional profits generated through entry and market expansion when firms remain separate and entry occurs. We show that the incumbents only merge if the loss of profits due to competition that results from entry exceeds the additional profits generated through market expansion.

Our parametric computations with the linear demand function reveal that entry deterrence through vertical integration is always to the detriment of welfare in the present setting for the case of observable two-part tariffs. Our abstract model suggests the following optimal merger policy for the analyzed setting when potential entry is a possibility and wholesale contracting is in non-discriminatory or, at least, observable tariffs, such that no opportunism problem as described by Hart and Tirole (1990) is relevant: Absent further efficiencies, a vertical merger should be prohibited based on the theory of harm that potential competition is restricted.

This finding is complementary to the case that entry has already taken place. For this case, a merger assessment needs to take the effects on actual instead of potential competition into account. Sandonís and Faulí-Oller (2006) have shown for this case that vertical integration can also be profitable, depending on different competitive parameters.

Moreover, our results are also complementary to the findings of Rey and Tirole (2007) whereby vertical integration can reduce an independent downstream firm's profit in the case of secret contracting with a competitive fringe supply. We show that the finding crucially depends on the assumptions of quantity competition and full secrecy and illustrate that opposite implications may arise either under price competition or interim observability of the sourcing decisions.

The main take-away of this article is that vertical integration can also restrict potential competition in settings where an educated observer who is aware of the previous economic literature on foreclosure may think that the classic Chicago School's *single monopoly profits* theory could apply and vertical integration does not raise competitive concerns. Our analysis highlights that an in-depth review of the likely effects of a proposed vertical merger on potential competition needs to incorporate a careful market investigation. To draw reliable policy conclusions based on the complex theories of vertical relations, one needs to obtain insights on the type of wholesale tariffs and the contracting process, along with other market information.

References

Aghion, P. and Bolton, P. (1987). Contracts as a barrier to entry. *The American Economic Review*, **77** (3), 388–401.

Allain, M.-L., Chambolle, C. and Rey, P. (2010). Vertical integration, innovation and foreclosure. *Preliminary Work*, pp. 1–60.

Bork, R. H. (1978). The antitrust paradox: A policy at war with itself. *New York*.

Bourreau, M., Hombert, J., Pouyet, J. and Schutz, N. (2011). Upstream competition between vertically integrated firms. *The Journal of Industrial Economics*, **59** (4), 677–713.

Caprice, S. (2006). Multilateral vertical contracting with an alternative supply: The welfare effects of a ban on price discrimination. *Review of Industrial Organization*, **28** (1), 63–80.

Chen, Y. (2001). On vertical mergers and their competitive effects. *RAND Journal of Economics*, **32** (4), 667–685.

— and Riordan, M. H. (2007). Vertical integration, exclusive dealing, and expost cartelization. *American Economic Review*, **38** (1), 1–21.

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

European Commission (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**.

European Commission (2009). Summary of commission decision of 13 may 2009 relating to a proceeding under article 82 of the ec treaty and article 54 of the eea agreement (case comp/c-3/37.990 - intel). *Official Journal of the European Union*, **C 227/13**.

Fumagalli, C. and Motta, M. (2006). Exclusive dealing and entry, when buyers compete. *The American Economic Review*, **96** (3), 785–795.

Hart, O. and Tirole, J. (1990). Vertical integration and market foreclosure. *Brookings Papers on Economic Activity. Microeconomics*, pp. 205–286.

Höffler, F. (2008). On the consistent use of linear demand systems if not all varieties are available. *Economics Bulletin*, **4** (14), 1–5.

Hunold, M. (2020). Non-discriminatory pricing, partial backward ownership, and entry deterrence. *International Journal of Industrial Organization*, **70**, 102615.

— and STAHL, K. (2016). Passive vertical integration and strategic delegation. *The RAND Journal of Economics*, **47** (4), 891–913.

LEVITAN, R. and SHUBIK, M. (1971). Price variation duopoly with differentiated products and random demand. *Journal of Economic Theory*, **3**, 23–39.

NORMANN, H.-T. (2009). Vertical integration, raising rivals' costs and upstream collusion. *European Economic Review*, **53** (4), 461–480.

ORDOVER, J. A., SALONER, G. and SALOP, S. C. (1990). Equilibrium vertical foreclosure. *The American Economic Review*, **80** (1), 127–142.

POSNER, R. A. (1979). The chicago school of antitrust analysis. *University of Pennsylvania Law Review*, **127**, 925–948.

REY, P. and TIROLE, J. (2007). A primer on foreclosure. *Handbook of industrial organization*, **3**, 2145–2220.

— and VERGE, T. (2004). Bilateral control with vertical contracts. *The RAND Journal of Economics*, **35** (4), 728.

RIORDAN, M. H. (2008). Competitive effects of vertical integration. In *Handbook of antitrust economics*, Cambridge, Mass. and London: MIT, pp. 145–182.

SANDONÍS, J. and FAULÍ-OLLER, R. (2006). On the competitive effects of vertical integration by a research laboratory. *International Journal of Industrial Organization*, **24** (4), 715–731.

2.7 Appendix A: Proofs

Proof of proposition 3. As shown in proposition 1, the post-entry profit for the entrant under price competition is lower under integration if and only if $w^{DS} > 0$. Let us now evaluate $w^{DS} > 0$ for the case of price competition and linear demand $q_i\left(p_i, p_{-i}\right) = \left(1 - p_i - \gamma + \gamma p_{-i}\right) / \left(1 - \gamma^2\right)$ as defined in equation (2.3).
To compute the supplier's reduced profit

$$\Pi_U^{DS} = \sum_{i \in \{I,E\}} \left(w \cdot \widetilde{q}\left(w, w\right) + \pi\left(w, w\right) - \pi\left(c, w\right)\right)$$

from equation (2.8) for linear demand, first compute $\pi\left(w, w\right)$, which is the operational profit that firm i obtains when both firms source from U. Each firm $i \in \{I, E\}$ maximizes $(p_i - w) q_i\left(p_i, p_{-i}\right)$ with respect to p_i. This yields

$$\widetilde{p}\left(w, w\right) = \frac{\gamma - w - 1}{\gamma - 2} \tag{2.13}$$

and

$$\widetilde{q}\left(w, w\right) = \frac{w - 1}{(\gamma - 2)(\gamma + 1)}. \tag{2.14}$$

The downstream flow profit equals

$$\pi\left(w, w\right) = \frac{(1 - \gamma)(1 - w)^2}{(\gamma - 2)^2(\gamma + 1)}. \tag{2.15}$$

To compute the outside option profit, we have to find the equilibrium when firm i maximizes $\pi_i = \left(p_i - w\right) q_i$ with respect to p_i, while firm $-i$ maximizes $\pi_{-i} = \left(p_{-i} - c\right) q_{-i}$ with respect to p_{-i}. This yields

$$\widetilde{p}\left(w, c\right) = \frac{\gamma^2 + \gamma - \gamma c - 2w - 2}{\gamma^2 - 4}, \ \widetilde{p}\left(c, w\right) = \frac{\gamma^2 + \gamma - 2c - \gamma w - 2}{\gamma^2 - 4}$$

and

$$\widetilde{q}\left(w,c\right)=\frac{-\gamma^2+\gamma(c-1)+\left(\gamma^2-2\right)w+2}{\gamma^4-5\gamma^2+4},\;\widetilde{q}\left(c,w\right)=\frac{-\gamma^2+\left(\gamma^2-2\right)c+\gamma(w-1)+2}{\gamma^4-5\gamma^2+4},$$

provided that c is small enough, such that they take on positive values. The outside option profit is given by

$$\pi\left(c,w\right)=\frac{\left(-\gamma^2-\gamma+\left(\gamma^2-2\right)c+\gamma w+2\right)^2}{\left(4-\gamma^2\right)^2\left(1-\gamma^2\right)}. \tag{2.16}$$

Plugging $\widetilde{q}\left(w,w\right)$, $\pi\left(w,w\right)$ and $\pi\left(c,w\right)$ into Π_U^{DS} yields

$$\Pi_U^{DS}=\frac{2\left(\left(\gamma^2-2\right)^2c^2-2\left(\gamma^2-2\right)c\left(\gamma^2+\gamma-\gamma w-2\right)+w\left(\gamma^2\left(\gamma^2+\gamma-2\right)-\left(\gamma^3+2\gamma^2-4\right)w\right)\right)}{\left(\gamma^2-4\right)^2\left(\gamma^2-1\right)}.$$

To obtain the equilibrium unit price w^{DS}, we differentiate the supplier's profit in the separated duopoly case with respect to w and solve the resulting FOC $\Pi_U^{DS}/\partial w=0$.[34] This yields

$$w^{DS}=\frac{\gamma\left(2c\left(2-\gamma^2\right)+\gamma\left(2-\gamma-\gamma^2\right)\right)}{2\left(4-2\gamma^2-\gamma^3\right)}. \tag{2.17}$$

As the numerator and the denominator are positive for all relevant values of c and γ: w^{DS} is always positive for $0<c<1$ and $0\leq\gamma<1$. Moreover, when c becomes large, $\pi\left(c,0\right)$ eventually becomes 0. In these cases, the optimal price w^{DS} is also positive, as this is effectively the case of an unconstrained upstream monopolist.
If one wants to exclude negative fixed fees, w^{DS} is restricted to not be above c. This is only relevant when the unrestricted solution of w^{DS} is positive and thus does not affect this proof.
When the firms compete in prices, the parametric solution thus yields $w^{DS}>0$ in the relevant range of c and γ. □

[34] If one wants to exclude negative fixed fees, w^{DS} is restricted to not be above c. This is only relevant when the unrestricted solution of w^{DS} is positive and thus does not affect this proof.

Proof of proposition 4. As shown in proposition 1, the post-entry profit for the entrant is lower under integration if and only if $w^{DS} > 0$. Let us now evaluate $w^{DS} > 0$ for the case of quantity competition and the inverse linear demand $p_i(q_i, q_{-i}) = 1 - q_i - \gamma q_{-i}$ as defined in equation (2.2).

To compute the supplier's reduced profit

$$\Pi_U^{DS} = \sum_{i \in \{I,E\}} \left(w \cdot \widetilde{q}_i(w,w) + \pi(w,w) - \pi(c,w) \right)$$

as defined in equation (2.8) for the case of separation and a downstream duopoly, we first compute its parts separately.

The symmetric quantities as a function of the input price w are given by

$$\widetilde{q}(w,w) = \frac{1-w}{2+\gamma} \tag{2.18}$$

and the symmetric price is given by

$$\widetilde{p}(w,w) = \frac{\gamma w + w + 1}{\gamma + 2}. \tag{2.19}$$

The downstream flow profit thus equals

$$\pi(w,w) = \frac{(w-1)^2}{(\gamma+2)^2}. \tag{2.20}$$

When one firm sources alternatively and the other firm from supplier U, the downstream quantities as a function of w are given by

$$\widetilde{q}(w,c) = \frac{\gamma - \gamma c + 2w - 2}{\gamma^2 - 4}, \ \widetilde{q}(c,w) = \frac{\gamma + 2c - \gamma w - 2}{\gamma^2 - 4}$$

and the prices by

$$\widetilde{p}(w,c) = \frac{\gamma + \gamma(-c) + \left(\gamma^2 - 2\right) w - 2}{\gamma^2 - 4}, \ \widetilde{p}(c,w) = \frac{\gamma + \left(\gamma^2 - 2\right) c - \gamma w - 2}{\gamma^2 - 4},$$

provided that all values are positive. The outside option profit is then given by

$$\pi\left(c,w\right)=\frac{\left(\gamma+2c-\gamma w-2\right)^{2}}{\left(\gamma^{2}-4\right)^{2}}. \tag{2.21}$$

Plugging these expressions into Π_U^{DS} yields

$$\Pi_U^{DS}=\frac{8c(\gamma(w-1)+2)-8c^{2}+2w\left((\gamma-2)\gamma^{2}-\left(\gamma^{3}-2\gamma^{2}+4\right)w\right)}{\left(\gamma^{2}-4\right)^{2}}.$$

Differentiating Π_U^{DS} with respect to w and solving the FOC yields

$$w^{DS}=\frac{\gamma((\gamma-2)\gamma+4c)}{2\left(\gamma^{3}-2\gamma^{2}+4\right)}, \tag{2.22}$$

which is positive for $c\geq\hat{c}$, with $\hat{c}(\gamma)=\frac{1}{4}\gamma\left(2-\gamma\right)$, which implies $\hat{c}'(\gamma)>0$ in the relevant parameter range.

If c becomes too large, the outside option profit eventually becomes 0. In this case, U is an unconstrained monopolist and the optimal unit input prices are positive in the case of downstream competition.

Moreover, if one wants to exclude negative fixed fees, w^{DS} is restricted to be not above c. This is only relevant when the unrestricted solution of w^{DS} is positive and thus does not affect this proof.

□

Proof of proposition 5. To determine the merger incentives, we focus on the case in which supplier U sets an input price above marginal costs $\left(w^{DS}>0\right)$ and in which entry costs are in an intermediate range: $\pi\left(c,0\right)<\theta<\pi\left(c,w^{DS}\right)$. We now demonstrate that, given linear demand as defined in equations ((2.2) and (2.3)), the incumbents can profitably integrate when the supply alternative is relatively efficient.

First, calculate the joint equilibrium profit of supplier U and downstream firm I for (i) the monopoly case, denoted Π^{M} (here it does not matter whether firms are integrated or separated) and (ii) the separated downstream duopoly case, denoted Π_{UI}^{DS}, to evaluate the condition

$$\Pi^M > \Pi_{UI}^{DS}. \tag{2.23}$$

(i) Monopoly. When E is not active, the demand function from equation (2.3) reduces to $p_I = 1 - q_I$. The monopoly profit is given by $\Pi^M = p_I q_I = (1 - q_I) q_I$ (recall that U's marginal costs are normalized to zero). The monopoly profit is the same for price and quantity competition. The monopolist maximizes $\Pi^M (q_I) = (1 - q_I) q_I$ with respect to q_I. The equilibrium quantity is $q^M = 1/2$ and the equilibrium price is $p^M = 1/2$. The joint profit of U and I without downstream entry and irrespective of integration or separation is thus given by

$$\Pi^M = \frac{1}{4}.$$

(ii) Separation and downstream duopoly. The joint profit is given by

$$\Pi_{UI}^{DS} = \widetilde{p}\left(w^{DS}, w^{DS}\right) \widetilde{q}\left(w^{DS}, w^{DS}\right) + w^{DS} \widetilde{q}\left(w^{DS}, w^{DS}\right) + \left(\pi\left(w^{DS}, w^{DS}\right) - \pi\left(c, w^{DS}\right)\right).$$

We distinguish between price competition (ii.a) and quantity competition (ii.b). Given the restriction $w^{DS} > 0$ applies, we distinguish between three different cases:

1. The fringe is relatively efficient: $c < c^e$, where the latter is a threshold that we define below, such that we need a restriction $w^{DS} \leq c^e$ to ensure that the fixed fees do not become negative in the case of price competition and that the unit input price does not become negative in the case of quantity competition ($w^{DS} \geq 0$).

2. Effectively no alternative supply exists as the unit cost of the alternative are so large that the outside option profit is zero: $c > c^m$, with $\pi(c, w) = 0$ and $\pi(c, 0) = 0$.

3. U is unrestricted in its choice of an unit input price, such that we derive w^{DS} from the first-order condition.

(ii.a) Price competition. As described above, the unit input price can now take three values (the third value was derived in equation (2.17) in the proof of proposition 3). For $c < c^e \equiv \gamma^2/4$ the alternative imposes a constraint on U's choice of w, such that $w^{DS} = c$. For $c > c^m \equiv (\gamma^2 + 2\gamma - 4)/(2\gamma^2 - 2)$ effectively no alternative sourcing exists for the downstream firms:

$$w^{DS} = \begin{cases} c & \text{if } c < c^e \equiv \frac{\gamma^2}{4}, \\ \frac{\gamma}{2} & \text{if } c > c^m \equiv \frac{\gamma^2+2\gamma-4}{2(\gamma^2-2)}, \\ \frac{\gamma\left(\gamma\left(\gamma^2+\gamma-2\right)+2\left(\gamma^2-2\right)c\right)}{2(\gamma^3+2\gamma^2-4)} & \text{if } c^e < c < c^m. \end{cases}$$

Plugging w^{DS} in $\widetilde{p}(w,w)$ and $\widetilde{q}(w,w)$ as given by equation (2.13) and (2.19) yields the following equilibrium prices and quantities:

$$\widetilde{p}\left(w^{DS},w^{DS}\right) = \begin{cases} \frac{\gamma-c-1}{\gamma-2} & \text{if } c < c^e, \\ \frac{1}{2} & \text{if } c > c^m, \\ \frac{\gamma^4-2\gamma^2+\gamma^3(1-2c)+4\gamma(c-2)+8}{2\left(\gamma^4-4\gamma^2-4\gamma+8\right)} & \text{if } c^e < c < c^m; \end{cases}$$

$$\widetilde{q}\left(w^{DS},w^{DS}\right) = \begin{cases} \frac{c-1}{(\gamma-2)(\gamma+1)} & \text{if } c < c^e, \\ \frac{1}{2\gamma+2} & \text{if } c > c^m, \\ \frac{\gamma^4-\gamma^3-6\gamma^2+2\left(\gamma^2-2\right)\gamma c+8}{2(\gamma-2)(\gamma+1)(\gamma^3+2\gamma^2-4)} & \text{if } c^e < c < c^m. \end{cases} \tag{2.24}$$

Plugging w^{DS} in $\pi(w,w)$ and $\pi(c,w)$ from equations (2.15) and (2.16) yields the equilibrium profits:

$$\pi\left(w^{DS},w^{DS}\right) = \begin{cases} \frac{(1-\gamma)(1-c)^2}{(\gamma-2)^2(\gamma+1)} & \text{if } c < c^e, \\ \frac{1-\gamma}{4\gamma+4} & \text{if } c > c^m, \\ \frac{(1-\gamma)\left(\gamma^4-\gamma^3-6\gamma^2+2\left(\gamma^2-2\right)\gamma c+8\right)^2}{4(\gamma+1)\left(\gamma^4-4\gamma^2-4\gamma+8\right)^2} & \text{if } c^e < c < c^m; \end{cases}$$

$$\pi\left(c, w^{DS}\right) = \begin{cases} \frac{(1-\gamma)(1-c)^2}{(\gamma-2)^2(\gamma+1)} & \text{if } c < c^e, \\ 0 & \text{if } c > c^m, \\ \frac{(1-\gamma)(\gamma+2)^2\left(\gamma^2+2\gamma-2\left(\gamma^2-2\right)c-4\right)^2}{4(\gamma+1)\left(\gamma^4-4\gamma^2-4\gamma+8\right)^2} & \text{if } c^e < c < c^m. \end{cases} \tag{2.25}$$

Consequently, the joint profit Π_{UI}^{DS} is

$$\begin{cases} \frac{(c-1)(\gamma+(\gamma-3)c-1)}{(\gamma-2)^2(\gamma+1)} & \text{if } c < c^e, \\ \frac{1}{2\gamma+2} & \text{if } c > c^m, \\ \frac{2\left((\gamma-1)\gamma^4+4\left(\gamma^2-2\right)^2c^2-4\left(\gamma^4+2\gamma^3-6\gamma^2-4\gamma+8\right)c\right)}{4(\gamma-2)^2(\gamma^3+2\gamma^2-4)(\gamma+1)} - \frac{(\gamma-1)(\gamma+2)^2\left(\gamma^2+2\gamma-2\left(\gamma^2-2\right)c-4\right)^2}{4(\gamma^4-4\gamma^2-4\gamma+8)^2(\gamma+1)} & \textit{else}. \end{cases}$$

We can now evaluate condition (2.23) to find out for which values of γ and c integration is profitable. Straightforward calculations show that integration is profitable when

$$c < \bar{c}_{Bertrand}(\gamma),$$

with

$$\bar{c}_{Bertrand}(\gamma) \equiv \frac{\gamma^2+2\gamma-4}{2(\gamma^2-2)} - \frac{1}{2}\sqrt{\frac{\gamma^9-\gamma^8-8\gamma^7+40\gamma^5-80\gamma^3-16\gamma^2+128\gamma-64}{(\gamma^2-2)^2(\gamma^3+\gamma^2-4)}}.$$

(ii.b) Quantity competition. Also for quantity competition, the unit cost price can now take three values (the third value was derived in equation (2.22) in the proof of proposition 4). For $c < c^e \equiv (1/4)(2\gamma-\gamma^2)$ the alternative imposes a constraint on U's choice of w, such that $w^{DS} = 0$. For $c > c^m \equiv \left(4+2\gamma-\gamma^2\right)/4(\gamma+1)$ no outside option exists for the downstream firms:

$$w^{DS} = \begin{cases} 0 & \text{if } c < c^e \equiv \frac{1}{4}(2\gamma-\gamma^2), \\ \frac{\gamma}{2\gamma+2} & \text{if } c > c^m \equiv \frac{4+2\gamma-\gamma^2}{4(\gamma+1)}, \\ \frac{\gamma((\gamma-2)\gamma+4c)}{2(\gamma^3-2\gamma^2+4)} & \text{if } c^e < c < c^m. \end{cases}$$

Plugging w^{DS} in $\widetilde{p}(w,w)$ and $\widetilde{p}(w,w)$ as given by equations (2.19) and (2.18) yields the following equilibrium price and quantity:

$$\widetilde{p}\left(w^{DS}, w^{DS}\right) = \begin{cases} \frac{1}{\gamma+2} & \text{if } c < c^e, \\ \frac{1}{2} & \text{if } c > c^m, \\ \frac{\gamma^4+\gamma^3-6\gamma^2+4(\gamma+1)\gamma c+8}{2\left(\gamma^4-4\gamma^2+4\gamma+8\right)} & \text{if } c^e < c < c^m, \end{cases}$$

$$\widetilde{q}\left(w^{DS}, w^{DS}\right) = \begin{cases} \frac{1}{\gamma+2} & \text{if } c < c^e, \\ \frac{1}{2\gamma+2} & \text{if } c > c^m, \\ \frac{\gamma^3-2\gamma^2-4\gamma c+8}{2\gamma^4-8\gamma^2+8\gamma+16} & \text{if } c^e < c < c^m. \end{cases} \tag{2.26}$$

Plugging w^{DS} in $\pi(w, w)$ and $\pi(c, w)$ from equations (2.20) and (2.21) yields the equilibrium profits

$$\pi\left(w^{DS}, w^{DS}\right) = \begin{cases} \frac{1}{(\gamma+2)^2} & \text{if } c < c^e, \\ \frac{1}{4(\gamma+1)^2} & \text{if } c > c^m, \\ \frac{\left(\gamma^3-2\gamma^2-4\gamma c+8\right)^2}{4\left(\gamma^4-4\gamma^2+4\gamma+8\right)^2} & \text{if } c^e < c < c^m, \end{cases}$$

$$\pi\left(c, w^{DS}\right) = \begin{cases} \frac{(\gamma+2c-2)^2}{\left(\gamma^2-4\right)^2} & \text{if } c < c^e, \\ 0 & \text{if } c > c^m, \\ \frac{(\gamma-2)^2\left(\gamma^2-2\gamma+4(\gamma+1)c-4\right)^2}{4\left(\gamma^4-4\gamma^2+4\gamma+8\right)^2} & \text{if } c^e < c < c^m. \end{cases} \tag{2.27}$$

The joint profit Π_{UI}^{DS} is thus given by

$$\begin{cases} \frac{(\gamma-2)^2-4c^2-4(\gamma-2)c}{\left(\gamma^2-4\right)^2} & \text{if } c < c^e, \\ \frac{1}{2\gamma+2} & \text{if } c > c^m, \\ \frac{1}{4}\left(\frac{2\left(\gamma^4-16(\gamma+1)c^2-8\left(\gamma^2-2\gamma-4\right)c\right)}{(\gamma+2)^2(\gamma^3-2\gamma^2+4)} + \frac{(\gamma-2)^2\left(\gamma^2-2\gamma+4(\gamma+1)c-4\right)^2}{\left(\gamma^4-4\gamma^2+4\gamma+8\right)^2}\right) & \text{if } c^e < c < c^m. \end{cases}$$

As under price competition, we compare the integration profit with the separation profit to infer about the merger profitability. More precisely, we evaluate the following condition for the case of quantity competition:

$$\Pi^M > \Pi^{DS}_{UI}.$$

Straightforward calculations yield that entry-deterring vertical integration is profitable when

$$c < \bar{c}_{Cournot}(\gamma),$$

with

$$\bar{c}_{Cournot}(\gamma) \equiv \frac{4+2\gamma-\gamma^2}{4(\gamma+1)} - \frac{1}{4}\sqrt{\frac{-\gamma^9+\gamma^8+8\gamma^7-16\gamma^6-24\gamma^5+64\gamma^4+16\gamma^3-112\gamma^2+64}{(\gamma+1)^2(\gamma^3-\gamma^2+4)}}. \tag{2.28}$$

□

Proof of proposition 6. To compute welfare, we plug the equilibrium quantities of 1/2 and 0 in the monopoly case and the quantities in equation (2.24) for Bertrand and equation (2.26) for Cournot in the duopoly case into the welfare function $W(q_I, q_E) = u(q_I, q_E) - \theta \cdot I(\text{entry})$ from equation (2.4). We thus compare welfare for the case that no entry occurs (downstream monopoly) W^M and for the case that E enters the market W^{DS}:

$$W^M > W^{DS}(\theta), \theta \in \{\underline{\theta}, \overline{\theta}\}. \tag{2.29}$$

For vertical integration and the downstream monopoly, we get

$$W^M = \frac{3}{8}.$$

Bertrand. For Bertrand, we get $W^{DS}(\theta)$

$$\begin{cases} \frac{(1-c)(-2\gamma+c+3)}{(\gamma-2)^2(\gamma+1)} - \theta & \text{if } c < c^e, \\ \frac{3}{4\gamma+4} - \theta & \text{if } c > c^m, \\ \frac{\left(3\gamma^4-10\gamma^2+\gamma^3(1-2c)+4\gamma(c-4)+24\right)\left(\gamma^4-\gamma^3-6\gamma^2+2\left(\gamma^2-2\right)\gamma c+8\right)}{4(\gamma+1)\left(\gamma^4-4\gamma^2-4\gamma+8\right)^2} - \theta & \text{if } c^e < c < c^m. \end{cases}$$

As entry costs are defined within a range of $\pi(c,0) < \theta < \pi\left(c, w^{DS}\right)$ – as described in the beginning of the proof of proposition 6 – we compute a lower bound with the highest possible entry costs $\left(\bar{\theta} = \pi\left(c, w^{DS}\right)\right)$ for welfare. We know $\overline{\theta}$ from equation (2.25) in the proof of proposition 5:

$$\bar{\theta} = \begin{cases} \frac{(1-\gamma)(1-c)^2}{(\gamma-2)^2(\gamma+1)} & \text{if } c < c^e, \\ 0 & \text{if } c > c^m, \\ \frac{(1-\gamma)(\gamma+2)^2\left(\gamma^2+2\gamma-2\left(\gamma^2-2\right)c-4\right)^2}{4(\gamma+1)\left(\gamma^4-4\gamma^2-4\gamma+8\right)^2} & \text{if } c^e < c < c^m. \end{cases}$$

The lower bound of welfare under entry and separation is given by

$$W^{DS}\left(\theta = \overline{\theta}\right).$$

A comparison of the integration welfare with the lower bound of the separation welfare, as in equation (2.29), shows that separation yields the higher welfare for all relevant values of c and γ. This finding is illustrated in figure 2.4.

For **Cournot**, we get

$$
W^{DS}(\theta) = \begin{cases} \frac{\gamma+3}{(\gamma+2)^2} - \theta & \text{if } c < c^e, \\ \frac{3}{4\gamma+4} - \theta & \text{if } c > c^m, \\ \frac{\left(\gamma^3-2\gamma^2-4\gamma c+8\right)\left(\gamma^4+\gamma^3-6\gamma^2+4(\gamma+1)\gamma c+8\right)}{2\left(\gamma^4-4\gamma^2+4\gamma+8\right)^2} - \theta & \text{if } c^e < c < c^m. \end{cases}
$$

The lower bound of welfare under vertical separation and duopoly is given by

$$
W^{DS}\left(\theta = \bar{\theta}\right),
$$

with

$$
\bar{\theta} = \begin{cases} \frac{(\gamma+2c-2)^2}{\left(\gamma^2-4\right)^2} & \text{if } c < c^e, \\ 0 & \text{if } c > c^m, \\ \frac{(\gamma-2)^2\left(\gamma^2-2\gamma+4(\gamma+1)c-4\right)^2}{4\left(\gamma^4-4\gamma^2+4\gamma+8\right)^2} & \text{if } c^e < c < c^m, \end{cases}
$$

where we know $\bar{\theta}$ from equation (2.27).

When comparing welfare under integration with welfare under separation, we learn that the integration welfare is larger than the separation welfare, when $c < \widetilde{c}(\gamma)$ and $\gamma > \widetilde{\gamma}$, with

$$
\widetilde{\gamma} = \frac{2}{3} \text{ and } \widetilde{c}(\gamma) = \frac{2-\gamma}{2} - \frac{\sqrt{-3\gamma^4+8\gamma^3+16\gamma^2-64\gamma+48}}{4\sqrt{2}},
$$

and $\widetilde{c}'(\gamma) > 0$ for $\gamma > \widetilde{\gamma}$.

For the parameter range where welfare is lower under vertical separation, the unit input price is non-positive. Formally: $\widetilde{c}(\gamma) > \hat{c}(\gamma)$ in the relevant range, where the latter is defined in the proof of proposition 4. This means that these cases are excluded from the current proposition as vertical integration does not deter entry. Consequently, for the relevant parameter range, welfare is strictly higher under vertical separation and duopoly than under vertical integration and monopoly. □

Acknowledgments:

We thank participants at the DICE PhD Research Workshop and at the Winter School on Applied Microeconomics, and in particular Justus Haucap, Hans-Theo Normann, Hendrik Döpper, Paul Heidhues, Salvatore Piccolo, Frank Schlütter, and Nicolas Schutz for valuable comments and suggestions. Funding by the Deutsche Forschungsgemeinschaft (DFG) — project 235577387/GRK1974 — is gratefully acknowledged.

3

The Causal Effect of Product Reviews on Prices. Evidence from Amazon.com

Single-authored

3.1 Introduction

E-commerce is a large and growing sector, even more so with the COVID-19 pandemic that is expected to induce a long-term shift in demand from traditional retail to online platforms (Mattioli and Herrera, 2020; Reintjes, 2020). Online consumers use peer reviews as a substitute for the physical experience of handling a product and presale advice. The British Competition and Markets Authority (CMA) estimates that 75% of online shoppers refer to product reviews (CMA, 2020). Peer reviews are easily accessible and offer a wide range of information about personal experiences with products and brands (ACCC, 2013).[1] Reviews thereby have the potential to increase efficiencies and to reduce the uncertainty about product quality, which may help to overcome problems of information asymmetries (e.g. Fan *et al.*, 2016; Ghose and Ipeirotis, 2011).

However, it is not self-evident that review systems pay off for platform operators. Poor reviews may, for example, cause sustained damage to a product's reputation or brand. Manipulated reviews impose an additional threat on review systems. Nevertheless, Amazon is one of the largest online stores with more than 700 million users worldwide[2] and has been using an extensive review system for many years.

This article uses Amazon data to analyze the platform's review system and, in particular, the effects of product reviews on prices of both marketplace sellers and Amazon. The results indicate that the Amazon review system is indeed effective – I find that changes in the aggregate star ratings induce sellers to adjust their product prices.

The article uses a self-collected product-time panel with information based on data from Amazon.com and a price tracking page. The panel data consists of 136,038 observations of reviewing activities between 1997 and 2020 from 908 different

[1] Appendix 3.9.1 provides a list of definitions for review related terms (such as peer review) and Amazon features.

[2] Amazon itself states that it has 300 million customers worldwide (Amazon, 2020a). Own research yields a larger number (at least when combining data about the number of unique users and monthly visitors). Appendix 3.9.2 displays the number of vistors / unique users for the markets in which Amazon is active.

products that were randomly drawn from twelve different product categories. To use full information, my main analysis is based on an unbalanced panel.The longest product panel starts in 1997 and the shortest in 2019. All panels end either in late 2019 or early 2020.

The article contributes to the existing literature in three ways: First, it tackles the endogeneity problem of reviews and prices by employing an instrumental variable strategy (as explained below) to estimate causal effects of product reviews on pricing decisions. Second, it distinguishes the pricing strategies of Amazon as a direct seller on the one hand and third party sellers on the Amazon marketplace on the other hand.[3] The estimations show that the aggregated star rating significantly affects the prices of marketplace sellers where an improvement of the aggregated star rating by one star increases the price by up to 6.5 percentage points. In contrast, there seems to be no robust effect on the prices of Amazon's direct offers. Third, (what many consumer's might not know) Amazon's aggregated star rating is not a raw but weighted average of the product reviews. This article contributes to the understanding of how Amazon aggregates its star rating by estimating the underlying weights. The results indicate that, for example, low star ratings (1 to 2 stars) receive a higher weight in Amazon's aggregated star rating compared to high star ratings (3 to 5 stars) reviews.[4]

An instrumental variables approach addresses the problem of a potential simultaneity: Peer reviews may not only affect prices but prices may also affect peer reviews. I instrument a given star rating by a reviewer's past ratings for other products relative to these products' aggregated ratings. I call this the reviewer's tendency to rate. The basic idea of the instrument is that each reviewer has a general tendency to rate products such as fair, unfair, or friendly. If a product has a true aggregated star rating of 4 stars, a fair reviewer tends to award indeed 4 stars,

[3]Appendix 3.9.1 provides a list of definitions that also defines marketplace sellers.

[4]A reason might be that reviews with low star ratings differ systematically from reviews with high star ratings in other dimensions than the number of stars, for instance, the length of the review or indicators that suggest trustworthiness.

whereas a friendly reviewer tends to award even 5 stars and an unfair reviewer at most 3 stars. Presumably, a reviewer's general tendency to rate (other) products will affect his rating for a given product but is otherwise independent from the price of a given product.

The econometric findings reveal interesting economic insights about Amazon's reviewing system.

- First, they show that marketplace sellers are aware of the product ratings and incorporate them into their pricing strategy.
- Second, the finding that better ratings lead to higher prices indicates that star ratings influence the consumers' willingness to pay – which the sellers extract through higher prices, at least partially. This suggests that the rating system is effective as consumers use the ratings when making their purchase decisions.
- Third, the finding that prices depend on the ratings – and thus presumably indirectly on the consumers' willingness-to-pay – indicates that pricing on the Amazon marketplace is not purely cost based. This could indicate that marketplace sellers have a degree of market power, possibly due to captive consumers.

The remainder of the paper proceeds as follows: Section 3.2 reviews the related literature. Section 3.3 presents the existing hypotheses. Section 3.4 explains the process of data collection and section 3.5 the empirical model as well as the identification strategy. Section 3.6 presents the results for a simple and weighted aggregated star rating and section 3.7 for extensions of the model. Section 3.8 concludes the article.

3.2 Literature review

There is a vast literature that deals with reviews on online platforms. However, to the best of my knowledge, there is no literature dealing with the *causal* effect of product reviews on the prices of consumer goods. Moreover, most of the existing articles analyze the effects of peer reviews on a single product type.[5] Instead, I conduct a broad range analysis with randomly chosen products from several product categories sold on Amazon.com. This section summarizes three strands of literature to which this article relates: Reviews and prices, reviews and seller reputation, and, last but not least, reviews and the demand-side. The relationship between prices and reviews is divided into two causal directions, both of which are discussed in this chapter.

Effect of reviews on prices. Most importantly to this study is the literature on the effects of reviews on prices. Lawani *et al.* (2019) infer quality measures from reviews of Airbnb room offers in Boston to estimate the direct effect on room prices. In addition, Lawani *et al.* (2019) use a spatial model to estimate the indirect effect that stems from an increase in price, triggered by the price increases of hosts located nearby in response to quality improvements. Jiang and Wang (2008) theoretically study the effects of product reviews on prices with a focus on the distinction of competitive and monopolistic markets as well as high and low quality firms. In a brief empirical extension, they study the correlated relationship of reviews and prices. Yu *et al.* (2016) and Wang *et al.* (2011) formally model dynamic pricing to derive pricing strategies of online sellers that receive product reviews. More specifically, Yu *et al.* (2016) study the waiting incentives of firms and consumers – for consumer-generated information (through peer reviews), respectively a price decrease – given that two-sided learning is possible. They find that firms price lower at the time of product introduction to prevent strategic waiting for more product information generated through reviews. Wang *et al.* (2011) find that prices increase as more product information is generated through reviews.

Effect of prices on reviews. Two articles examine the reverse causality to my article – the effect of prices on reviews. Lee *et al.* (2016) study the reactions of two price drops on four features of product reviews: star rating, review depth, positive and

[5] Anderson and Magruder (2012) and Luca (2016) investigate restaurants, Berger *et al.* (2010) and Chevalier and Mayzlin (2006) books, Duan *et al.* (2008) movies, Lee *et al.* (2016) Amazon kindle and Li and Hitt (2010) digital cameras.

negative emotion. They find significant changes in all four review features for a price decrease of Amazon kindle – both in the short and long run. The effects of the first round of price decreases is mainly negative, while reactions are moderate for the second-round price decreases. Li and Hitt (2010) find that an unidimensional review system[6] – as on Amazon – can be substantially biased by price effects.

Effects of seller reputation on prices. While this article addresses product reviews, a related strand of literature investigates the effect of seller reviews (reputation) on prices. Fan *et al.* (2016) find that new sellers on Taobao decrease prices as an investment in future reviews when their reputation increases.[7] Established sellers behave in reverse. Jolivet *et al.* (2016) find a strong and significantly positive effect of seller reputation on prices on PriceMinister.com. Other articles draw their conclusions from Ebay: Ba and Pavlou (2002) and Dewally and Ederington (2006) show that sellers with a positive reputation can set higher product prices and have a higher probability of selling (Dewan and Hsu, 2004; McDonald and Slawson, 2002; Resnick *et al.*, 2000).

Effect of reviews on the demand-side. An extensive part of research on product reviews focuses on the demand-side – the effect on revenues and sales. All of these articles, with the exception of Magnusson (2019), do not consider prices at all. The article that does consider prices uses a regression discontinuity design to study the causal effect of reviews on revenue on the online platform Wayfair.[8] Magnusson (2019) finds that an increase in revenues on Wayfair is purely caused by an increase in sales, and not by prices. Anderson and Magruder (2012) and Luca (2016) identify the causal effect of Yelp reviews on restaurant reservations. Chevalier and Mayzlin (2006) use the variation in relative book sales and reviews on Barnesandnoble and Amazon to analyze the effect of reviews on sales.[9] All four papers find that an

[6]In contrast to unidimensional review systems, which integrate all evaluation aspects in a single rating, multidimensional systems separate quality and value ratings (Li and Hitt, 2010).

[7]Taobao is the largest e-Commerce platform in China (Fan *et al.*, 2016).

[8]Wayfair is an e-commerce platform that mainly sells home goods (Magnusson, 2019).

[9]As Chevalier and Mayzlin (2006) use publicly available data, they approximate sales with the sales rank.

improvement in product ratings increases sales. Berger *et al.* (2010) show that even negative reviews can increase sales of books with low awareness. Hu *et al.* (2012) find a positive correlation between manipulated reviews and sales.

3.3 Hypotheses

The main research question of this article accounts for the effect of a change in the aggregated star rating on prices. A high aggregated star rating for a product signals a high product quality to consumers (Jiang and Wang, 2008). Kostyra *et al.* (2016) experimentally show that an improvement in the aggregated star rating increases customers' willingness-to-pay.[10] Except for extreme cases, such as perfect price competition, strategic sellers should take this process into account and increase their prices once the aggregated star rating improves. The first hypothesis states:

Hypothesis 1. *Sellers account for an increase in the aggregated star rating by increasing the product price.*

The next hypothesis accounts for differences in sellers' pricing strategies. I distinguish between the pricing strategies of marketplace sellers and Amazon's direct sales. Presumably, Amazon in its double function of a platform and seller does not engage in review manipulation, while some of the marketplace sellers might. Amazon prices might therefore react to a lesser extent to changes in the aggregated star rating than marketplace prices.

Hypothesis 2. *Amazon prices react less to changes in the aggregated star rating than marketplace prices.*

The next hypothesis refers to Amazon's composition of the aggregated stars. Amazon uses a machine-learned algorithm to determine weights for each review. Ac-

[10] In their experiment, Kostyra *et al.* (2016) find that a one-star improvement in reviews (rated on a five-star scale) increases the willingness-to-pay for an eBook reader – depending on the number of reviews – by €49 (6 reviews) and €66 (200 reviews). The prices of the eBook readers ranged from €99 to €139.

cording to Amazon these weights depend on characteristics like the age and trustworthiness of a review.[11] I hypothesize that young and trustworthy reviews have higher weights than old and untrusted reviews. Likewise, low star ratings (1 or 2 stars) could have higher weights than high star ratings (3 stars or more). This seems plausible, as low star ratings are usually quantitatively outnumbered by high star ratings and therefore often receive more attention.

Hypothesis 3. *Young, and trusted reviews as well as reviews with low star ratings have higher weights in Amazon's aggregated star rating than their old, untrusted and highly rated counterparts.*

The final hypothesis addresses verified and non-verified reviews. On Amazon, reviews can be labeled as "verified purchase", which implies that Amazon has confirmed the purchase of the reviewer. A review might lack the label because the customer received the product at a "deep" discount.[12] These reviews could be used to proxy untrustworthy reviews.[13] In case hypothesis 1 cannot be rejected, sellers would increase their prices when the aggregated star rating improves. It seems implausible to assume that sellers react differently to e.g. an untrustworthy five-star rating (e.g., that they bought to improve their aggregated product rating) than to an authentic five-star rating. My initial hypothesis is therefore that sellers either react or do not react to changes in the aggregated star rating and do not distinguish between untrustworthy and authentic reviews.

Hypothesis 4. *There is no difference in the effect of a change in the aggregated star rating on prices between manipulated and authentic reviews.*

[11] Amazon's statement of the calculation of ratings can be found here: "How Are Product Star Ratings Calculated?", `www.Amazon.com/gp/help/customer/display.html?nodeId=GQUXAMY73JFRVJHE`, accessed on 7 Aug 2020.

[12] See "About Amazon Verified Purchase Reviews", `www.amazon.com/gp/help/customer/display.html?nodeId=202076110`, accessed on 10 Aug 2020.

[13] Luca and Zervas (2016) base their empirical investigation on reviews identified as suspicious by Yelp.

3.4 Data collection process

My data collection started in November 2019 and lasted until July 2020. I collected all of the data with self-written scrapers that download relevant information from the American website of Amazon.com (henceforth Amazon) and the price tracking page camelcamelcamel.com (camel). The data collection process can be divided into three main steps (appendix 3.9.4 shows examples of all scraped web-pages.):

1. Collect all product IDs in relevant categories on the American webpage of Amazon.com.

2. Draw a random sample of the product IDs collected in step 1.

3. For each drawn product ID, scrape the following:

   (a) Amazon product page,

   (b) Amazon review page,

   (c) reviewers' Amazon-profile page and

   (d) price panel data from camelcamelcamel.com.

Product IDs (steps 1 and 2). In the first step, I scraped all product IDs from twelve different product categories on Amazon.com.[14] In the second step, I drew a random sample of products from the more than 4.5 million product IDs scraped in the first step.

Product and review data collection (step 3.(a) to (b)). In the third step, I collected (where possible) detailed information for all products in the sample. I collected product page information for a total of 908 different products (step 3.(a)). For these 908 products, I collected 136,038 product reviews with detailed information (step 3.(b)). On Amazon, customers can rate products with integer stars of one to five,

[14]These product categories include: "Clothing, Shoes, Jewelry & Watches", "Books & Audible", "Electronics, Computers & Office", "Smart Home", "Home, Garden & Tools", "Pet Supplies", "Beauty & Health", "Toys, Kids & Baby", "Sports & Outdoors", "Automotive & Industrial", "Amazon Warehouse" and "Amazon Launchpad".

where five stars are the best possible rating. Platforms like Tripadvisor and Yelp allow everyone to write reviews – without verification of their actual purchase. In contrast, on Amazon only customers who bought the product on Amazon can review the product there.[15] The review data includes stars, title, text, date, and the number of helpful votes. I furthermore extracted information that specifies whether the review stems from a ”verified purchase”[16] or has any badges like "Top 10 Reviewer".[17]

Since 2019, Amazon has implemented a one-tap review possibility for a test group of reviewers. These one-tap reviews enable reviewers to rate a product with stars only. The one-tap reviews are included in the aggregated star rating without any further information. Therefore, I cannot collect any additional data on these ratings. As Amazon implemented the one-tap possibility only in (late) 2019, the number of one-tap reviews should not be too large and should have no effect on the results.[18]

Profile data collection (step 3.(c)). For each review, I scraped the reviewer's profile page on Amazon.com (step 3.(c)). The profile displays all the reviewer's past reviews including the reviewer's star rating, date of the review, review title, a short summary of the review text as well as the name, picture, and aggregated star rating of the reviewed article. This scraping exercise resulted in around 7.5 million reviews.

Price data collection (step 3.(d)). In step 3.(d), I scraped price graphs from the price-tracking page camel that presents price history graphs for Amazon products.

[15] Here you can access the statement from Amazon customer service about reviews: "About Comments, Feedback, & Ratings", `www.Amazon.com/gp/help/customer/display.html?nodeId=201889150`, accessed on 07 Aug 2020.

[16] The badge ”verified purchase” implies that Amazon has confirmed the purchase of the reviewer. A review could lack the badge because the customer received the product at a "deep" discount. For more information see: "About Amazon Verified Purchase Reviews", `www.Amazon.com/gp/help/customer/display.html?nodeId=202076110`, accessed on 07 Aug 2020.

[17] Amazon awards some reviewers with a badge for their engagement in the reviewing activity. These badges comprise *#1 Reviewer, Top 10 Reviewer, Top 50 Reviewer, Top 100 Reviewer, Top 500 Reviewer, Top 1000 Reviewer, Hall of Fame Reviewer, Top Contributor, Amazon Verified Profile, THE, Amazon Official, Author, Artist, Manufacturer* and *Vine Voice*. See "Badges", `www.Amazon.com/gp/help/customer/display.html?nodeId=GED7RL944YMQ8CE3`, accessed on 07 Aug 2020, for a complete list of badges with explanations.

[18] From different references, I get the impression that one-tap reviews were implemented in the end of 2019 (see e.g., Fruncillo, 2019; Perez, 2019). Unfortunately, no reliable information about the exact implementation date exists.

Camel collects prices for products sold directly by Amazon and by marketplace sellers,[19] which sell new and used products.[20] Camel compiles the collected prices in a graph; the underlying data is not available. I only collect prices for new products sold through Amazon.com and marketplace sellers, as I expect that prices for used products strongly depend on the condition of the product. An image analysis of the camel price graphs resulted in around 1 million daily prices for marketplace sellers (767 products) and around 0.4 million daily prices for 247 products sold directly through Amazon. A total of 193 products in my sample are sold by both Amazon and marketplace sellers.

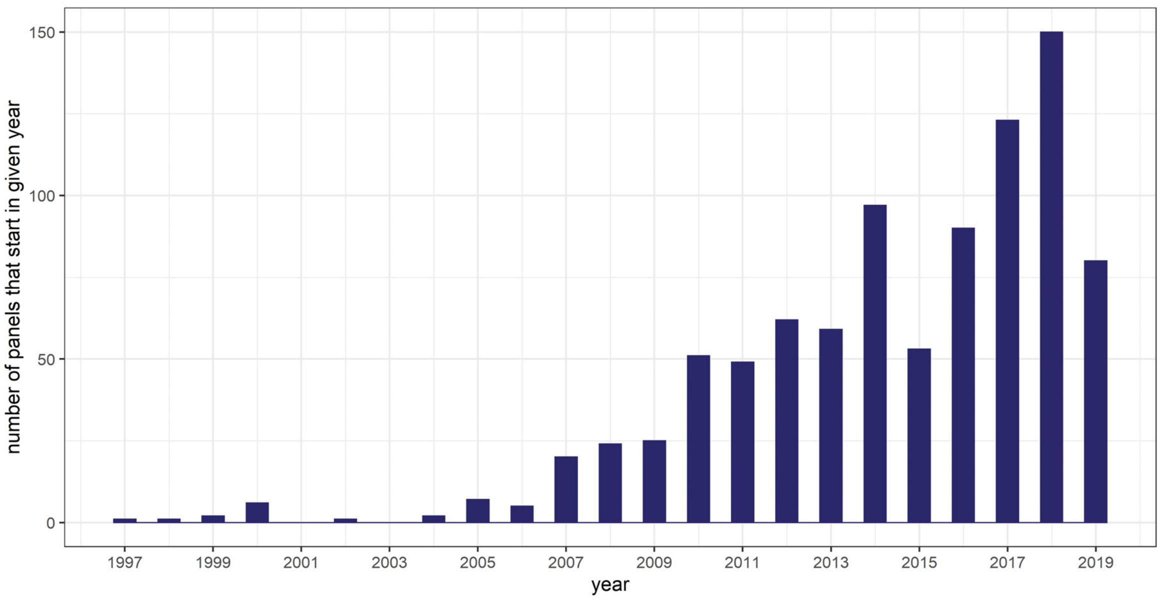


Figure 3.1: Overview of time-panel.
The figure shows how many of the product-panels start in a given year. The longest panel starts in 1997, while the first observations for the shortest panel are in 2019.

To use all of the review data, I work with an unbalanced product time panel. Figure 3.1 shows an overview of the number of product panels starting in a given year.

[19] Amazon features an integrated platform for third-party sellers. This platform is called Amazon marketplace and distributes new, refurbished, and used items (Amazon, 2020b).

[20] Camel tracks more than 4 million products on Amazon.com. They chose products to track according to searches on their site and through their browser extensions (Camelcamelcamel, 2011). If several marketplace sellers offer the same product, camel displays the lowest of all the marketplace prices.

The earliest observation starts in 1997, the latest in 2019. The mode for the start of the time panel is in 2018 (around 10% of observations).

	Obs.	Range	1st Qu.	Median	Mean	3rd Qu.	Sd.dev
Summary marketplace retailers							
Prices (USD)	767	0.29 – 3290.87	11.50	23.30	64.94	56.47	188.88
Average stars	767	1.28 – 5	4.09	4.45	4.33	4.72	0.57
Number of reviews	767	2 – 5000	10	33.00	140.25	114.5	361.26
Available days	767	81 – 4265	572	1159	1439.76	2101.5	1019.37
Days betw. reviews	767	0.19 – 1673	10.17	32.47	99.32	94.25	198.89
Summary Amazon retailer							
Prices (USD)	247	2.97 – 1384.44	13.80	27.28	78.90	77.72	155.02
Average stars	247	1.28 – 5	3.99	4.42	4.26	4.71	0.65
Number of reviews	247	2 – 4405	11.50	30.00	156.35	119.5	413.39
Available days	247	69 – 4265	895	1793	1841.62	2749	1150.80
Days betw. reviews	247	0.56 – 1573	12.98	41.50	120.48	113.21	239.44

Table 3.1: Descriptive summary statistics.
The table contains descriptive summary statistics for marketplace (top) and Amazon (bottom) products by product mean, respectively product maximum for the number of reviews, and the available days.

Summary statistics. Table 3.1 presents individual summary statistics for products sold on the marketplace (top) and directly by Amazon (bottom). The statistics are based on product means, except for the number of reviews and available days, for which I first calculated the product maximum. The table shows that the mean marketplace product has a price of USD 64.94, an aggregated star rating of 4.33 stars and 140 product reviews. The mean marketplace product is reviewed every 99 days and is available for 1,440 days.

The mean Amazon product has a price of USD 78.90, a rating of 4.26 stars and 156 product reviews. On average, an Amazon product is reviewed every 120 days and is available for around 1,842 days. Appendix 3.9.5 presents joint summary statistics based on product means and without taking product means. Appendix 3.9.5 furthermore contains a summary statistic for the market in which marketplace sellers and Amazon are jointly active.

3.5 Empirical framework

To identify the effect of a one-star change in the aggregated star rating of a product on the change in prices, I specify the following regression equation:

$$\begin{aligned}\Delta price_{i,t,x} \equiv \frac{price_{i,t+x}}{price_{i,t}} =& \beta_1 \left((aggregated\ star\ rating)_{i,t} - \left(\overline{rating}\right)_i \right) + \\ & \beta_2 (availability)_{i,t} + \beta_3 (availability)^2_{i,t} + \omega_i + \tau_t + \varepsilon_{i,t}.\end{aligned} \tag{3.1}$$

The left hand side of the equation displays the change in the price of a product i at the time $t + x$ relative to the price at the time of a new review t, where x is measured in days and $x \in \{1, 31\}$. The right-hand side takes the difference of the *aggregated star rating* of product i at time t and the constant $\overline{rating}$ of product i over all time periods where the product is available. The variable $\overline{rating}_i$ reflects whether the new rating is a relatively good or relatively bad rating for product i. As $\overline{rating}_i$ is constant for one product, it drops out when including product fixed effects. A further regressor is the product availability on the online platform scaled on a yearly basis. As prices might react stronger to product availability in the beginning of the life-cyle, I include the linear and quadratic term: $availability_{i,t}$ and $availability^2_{i,t}$.[21] The variable ω_i controls for unobserved product heterogeneity and τ_t for any monthly seasonality effects.

The empirical strategy tackles various identification issues, which are explained in detail below:

- **First:** Lagged price reaction;

[21] The true date of product launch is unknown. I approximate the launch date with the date on which I first observe the product. Depending on what happens first, this can either be the date when the product was first reviewed or the date when the price of the product was first tracked by camel.

- **Second:** Product-specific price trends;
- **Third:** Endogeneity of reviews in the price equation.

Firstly, to investigate a reaction in prices, I consider various price lags. Specifically, I study how the price on day $t + x$ changes in relation to the price on day t, which is the day of a new review. The lag is indicated by x and ranges from one to 31 days.[22] Secondly, the observed products might have heterogeneous price trends, which may make it difficult to precisely estimate any potential price effects caused by reviews. To analyze price trends, I regress the price change within one year ($price_{i,t+365}/price_{i,t}$) on product i. Figure 3.2 shows that there are different price trends at the product level. To eliminate product-specific price trends I use product fixed effects and the change in prices as a dependent variable ($\Delta price_{i,t,x}$).

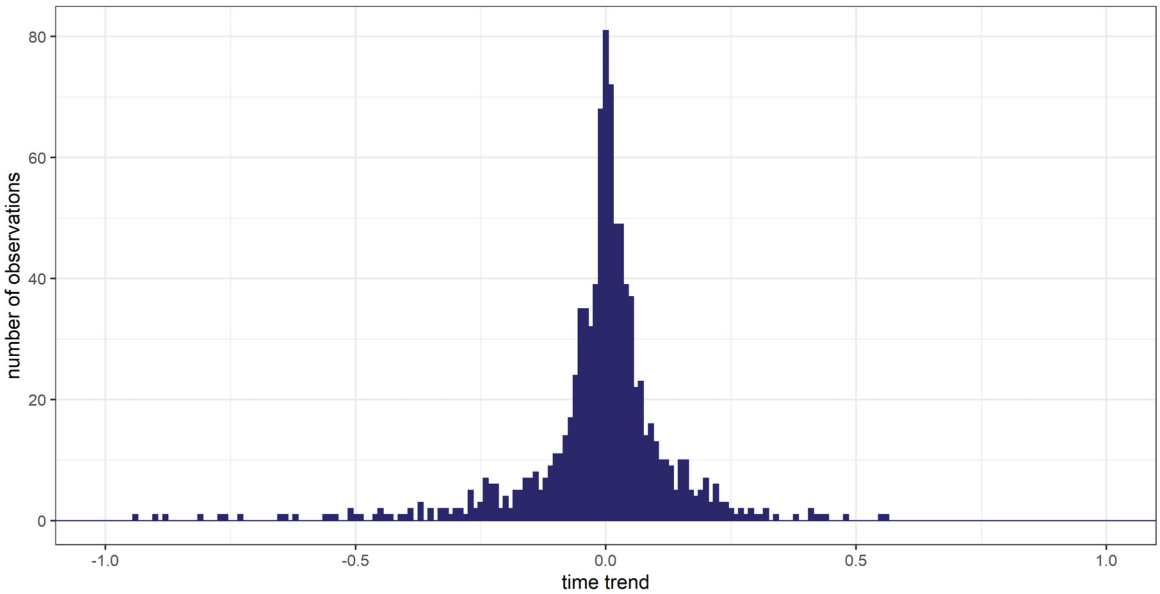


Figure 3.2: Average price changes year on year.
The figure shows the average price changes year on year (that is the coefficient estimates of the product specific time trends).

Thirdly, the analysis of the price-review relationship might not only exhibit an

[22]In a robustness check, explained in section 3.7.3, I increase the possible lag to 100 days.

effect of reviews on prices, but also a reverse effect of prices on reviews.[23] This simultaneity leads to an endogeneity problem. Due to autocorrelation in prices, the simultaneity problem also persists when the regression equation contains lagged prices.

To tackle endogeneity, I use a two-stage least squares (2sls) regression analysis and instrument the star rating of a single reviewer with the reviewer's tendency to rate products in general. To calculate the tendency to rate, I visit each reviewer's Amazon page and collect each of their star ratings as well as the aggregated star rating for each reviewed product. I then compute the difference between the star rating of the observed reviewer and the weighted aggregated star rating for each reviewed product. The tendency to rate is defined as the mean over all differences between the individual and aggregated star rating for a reviewer. The subsequent example will help to understand the instrument.

Product	Tom's rating	Aggregated rating	Difference: Tom's rating - star rating
Watch	3 stars	3.9 stars	- 0.9
Mug	4 stars	4.2 stars	- 0.2
Backpack	3 stars	4.9 stars	- 1.9
Tom's tendency to rate			$(-0.9-0.2-1.9)/3=-1$

Table 3.2: Illustration of the instrumental variables approach. Calculation of a reviewer's tendency to rate.

Say, Tom reviewed four products, a book, a watch, a mug, and a backpack. The book is the product of interest and Tom's mean tendency to rate shall instrument Tom's rating of the book. For that purpose Tom's rating of the book is excluded from Tom's mean tendency to rate. As table 3.2 illustrates, Tom rates the watch with 3

[23]E.g., Jiang and Wang (2008) and Lawani *et al.* (2019) study the effects of reviews on prices and Lee *et al.* (2016) as well as Li and Hitt (2010) the effects of prices on reviews.

stars, the mug with 4 stars, and the backpack with 3 stars. The overall star rating is 3.9 for the watch, 4.2 for the mug, and 4.9 for the backpack. Tom's tendency to rate is -1, which is the mean over all differences between Tom's rating and the respective product's aggregated star rating. Tom's tendency to rate implies that he awards, on average, 1 star less than the average reviewer. The idea of the IV strategy is that Tom's tendency to rate below the aggregated rating will also affect his rating of the book and will thereby influence the aggregated star rating of the book. However, Tom's rating of the watch, the mug or the backpack does not affect the price of the product of interest – the book. It therefore seems plausible to assume that the tendency to rate is exogenous.

To obtain a causal effect of the aggregated star rating on the price, the instrument needs to fulfill four criteria: Relevance, independence, exclusion, and monotonicity. The *relevance criteria* requires that the tendency to rate affects whether a reviewer rates friendly, fair, or unfriendly and thereby affects the aggregated star rating of the product. The *independence assumption* requires the instrument to be independent from the potential outcomes. The independence restriction might not hold if consumers with a specific tendency to rate would prefer certain price levels. In this case, a consumer's tendency to rate might be correlated with the price of a product. For example, some consumers might only buy products when prices are reduced. Such consumers might be friendly reviewers such that low prices and positive tendencies to rate correlate. The *exclusion restriction* requires that the instrument affects the price only through the reviewer's rating but not through other channels. Given the setting at hand, it is however hard to think about such a channel. An example might be price discrimination through exclusive discounts or surcharges for a consumer on the basis of their previous ratings. Appendix 3.9.6 provides additional evidence that the instrument is significant and relevant for various subsamples, such as different price levels. It additionally includes some summary statistics about the tendency to rate.

There are two possible approaches of how to employ the instrument: (I) The first approach is to instrument the aggregated star rating of product i at time t with the average tendency to rate of all reviewers that reviewed product i before or at time t. As the average tendency to rate converges to zero, this approach offers little variation. (II) The second strategy is to instrument the individual star rating of a reviewer with the reviewer's tendency to rate. As I want to regress the change in price on the aggregated star rating (as displayed in equation 3.1) and not the individual star rating in the second-stage of the 2sls analysis, a manual computation of the stages is necessary for this approach. The second approach offers greater variation than the first. I consequently choose to instrument the star rating of an individual reviewer k of product i at time t with their tendency to rate. The first-stage regression equation looks as follows:

$$\begin{aligned} \textit{individual star rating}_{k,i} = &\beta_1(\textit{tendency to rate})_k + \beta_2(\textit{availability})_{i,t} + \\ &\beta_3(\textit{availability})^2_{i,t} + \omega_i + \tau_t + \varepsilon_{k,i,t}. \end{aligned} \tag{3.2}$$

To obtain consistent and at the product level clustered standard errors,[24] I bootstrap the standard errors with 500 replications. I use the fitted star rating to compute the aggregated star rating for the second-stage. Table 3.3 displays the first-stage regression results. The regression results show that the star rating increases by roughly 0.8 stars when the tendency to rate of a reviewer increases by one. Table 3.3 additionally shows that the instrument is significant and relevant as the F-statistic for the instrument is larger than 700 (Stock and Watson, 2015).

[24]Clustered standard errors are common in the panel data literature (Bertrand *et al.*, 2004).

	Stars		
	Combined	Marketplace	Amazon
	(1)	(2)	(3)
availability (in years)	-0.0507^{**}	-0.0584^{**}	-0.0799^{*}
	(0.0196)	(0.0236)	(0.0467)
availability^2 (in years)	$1.29e{-5}^{**}$	$1.77e{-5}^{***}$	$2.13e{-5}^{*}$
	($5.4e{-6}$)	($5.88e{-6}$)	($1.11e{-5}$)
tendency to rate	0.807^{***}	0.792^{***}	0.857^{***}
	(0.023)	(0.027)	(0.032)
Product FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	136,038	82,511	23,552
F Statistic (proj model)	435.1 (df = 907)	302.8 (df = 754)	309.5 (df = 241)
F Statistic (excl instr.)	1,220.9 (df = 907)	886.5 (df = 754)	702.9 (df = 241)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3.3: First-stage regression results.

The table contains the regression results of the first-stage of a two-stage least squares analysis (see equation 3.2). Column (1) shows the effect on the individual star rating for all offers available on Amazon – marketplace and Amazon –, column (2) displays the effect for marketplace and column (3) for Amazon offers. Robust standard errors (in parentheses) are clustered at product level.

3.6 Results

3.6.1 Baseline regression - simple aggregated star rating

I use the regression specifications presented in equation 3.1 to quantify the effect of an increase in the aggregated star rating of a product on product prices both for marketplace sellers and Amazon offers on Amazon with an instrumental variables approach. As marketplace sellers and Amazon react with different lags to a change in the aggregated star rating, I perform the baseline regression with different time lags. I use a price lag of 14 and 21 days for marketplace prices and a price lag of 3 days for Amazon prices. I will later show graphical results for further time lags. Columns (1) and (4) of table 3.4 show the OLS regression results for the marketplace and Amazon offers. An improvement in the aggregated star rating by one star is associated with a significant change in marketplace prices after 14 days of roughly one percentage point. For Amazon products, an improvement in the aggregated star rating by one star is associated with a price change after three days of roughly 0.5 percentage points.

The OLS estimation most likely suffers from an endogeneity problem and thus only displays a correlation of prices and the aggregated star rating. As previously described, simultaneity leads to the most severe issue of endogeneity in the estimation. The simultaneous effects are likely to work in opposite directions: A product's star rating is likely to affect prices positively; a better product rating increases the perceived quality of a product and thereby increases a consumer's reservation price. Instead, prices are likely to affect the product rating negatively. A high product price increases a consumer's expectation and makes disappointment and a negative star rating more likely compared to low priced products, where consumers are more forgiving of imperfections. The coefficient of the aggregated star rating captures the negative effect of prices on product ratings and is thus likely to be underestimated. To obtain a causal relationship, I instrument the individual star rating with the reviewer's tendency to rate. Thereafter, I compute the fitted aggregated star rat-

	$\Delta \text{price}_{i,t,x}$				
	OLS (x = 14) Mktplc	IV (x = 14) Mktplc	IV (x = 21) Mktplc	OLS (x = 3) Amazon	IV (x = 3) Amazon
	(1)	(2)	(3)	(4)	(5)
availability (in years)	0.002	0.0027*	0.0035*	0.0007	0.0009^{+}
	(0.0014)	(0.0015)	(0.002)	(0.0006)	(0.0006)
availability^2 (in years)	$-$9.6e$-$7	$-$1.08e$-$6*	$-$1.40e$-$6*	$-$1.27e$-$7	$-$1.65e$-$7
	(6.06e$-$7)	(6.13e$-$7)	(8.29e$-$7)	(2.14e$-$7)	(2.17e$-$7)
aggregated star rating	0.013**			$-$0.0005	
	(0.006)			(0.0009)	
aggregated star rating (fit)		0.031***	0.039***		0.005*
		(0.012)	(0.012)		(0.003)
Product FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	78,515	78,515	78,023	22,622	22,622
Residual Std. Error	0.145	0.145	0.165	0.028	0.028
DoF	77,747	77,747	77,256	22,367	22,367
Note:			$^{+}$p<0.15; *p<0.1; **p<0.05; ***p<0.01		

Table 3.4: Baseline regression results.
The table contains the regression output based on equation 3.1 for marketplace (column (1) to (3)) and Amazon (column (4) to (5)) products. Columns (1) and (4) display the OLS results. All other columns display the second-stage results of the IV regression. The numbers in parentheses behind the dependent variable indicate the lag of the price change. The standard errors (in parantheses) are bootstrapped to correct for a manual implementation of 2sls and are clustered on product level.

ing based on the instrumented individual star rating as explained in section 3.5. Columns (2) and (3) of table 3.4 present the regression results for the instrumental variable estimator for a lag of 14 days (column 2) and 21 days (column 3) for marketplace products. As expected, the effect of the star rating is larger in the instrumental variable regression compared to OLS. If the aggregated star rating improves by one star, the change in the marketplace price increases by roughly 3 percentage points after 14 and by roughly 4 percentage points after 21 days. Table 3.4 furthermore suggests that the relationship between product availability and product prices is non-linear in time. Initially, product availability increases the product prices. The quadratic term of product availability indicates, that each additional year of product availability decreases the change in product prices. However, this effect is negligibly small.

Turning to the IV estimation for Amazon offers (column 5), table 3.4 reveals that the coefficient of the aggregated star rating is also higher compared to the OLS estimation. When the aggregated star rating of a product improves by one star, the price change increases by 0.5 percentage point after three days. The effect for Amazon products is only significant at the 10% level and figure 3.4 will show that the effect is not robust. The coefficients that indicate product availability are not significant.

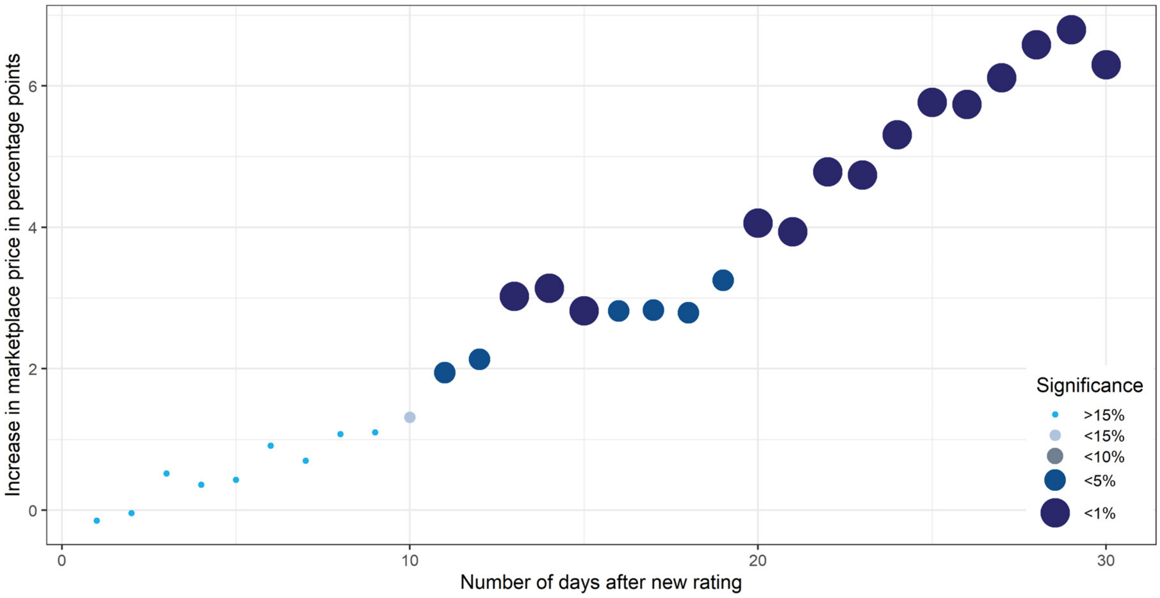


Figure 3.3: Baseline regression results for the marketplace.
Evolution of the change in marketplace prices from 1 to 31 days after an improvement of the simple aggregated star rating. The size of the dots indicates the significance-level.

The Amazon results should be interpreted with caution, especially in comparison to the marketplace results. In total there are only 247 independent observations (different products) for Amazon, while there exist 767 independent observations

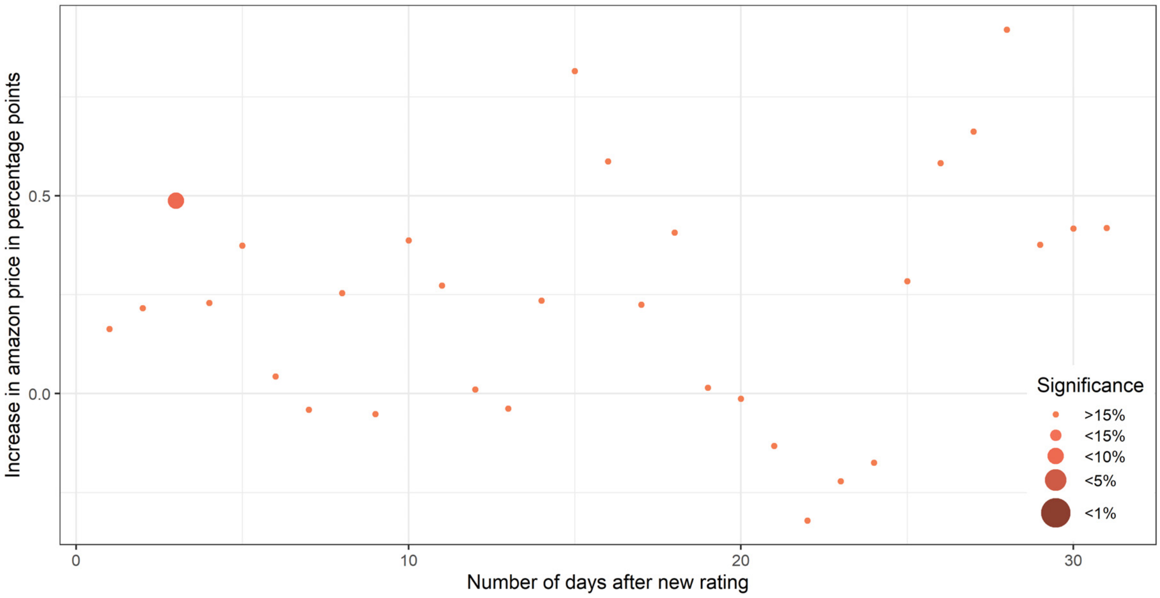


Figure 3.4: Baseline regression results for Amazon.
Evolution of the change in Amazon prices from 1 to 31 days after an improvement of the simple aggregated star rating. The size of the dots indicates the significance-level.

for marketplace sellers. The mean Amazon product has 156 reviews but 50% of products (median) have 30 reviews or less. Consequently, there might be too few products with enough variation to obtain statistically significant results.

Figures 3.3 and 3.4 display the evolution of the price change due to an improvement of the aggregated star rating for the marketplace (figure 3.3) and Amazon (figure 3.4) product offers for each, of 31 days after an additional review. Figure 3.3 shows that prices indeed react with a lag. The prices for marketplace offers start to react after ten days and increase by up to 6.5 percentage points after 29 days. In sharp contrast to the evident marketplace price effect that is gradually increasing, figure 3.4 indicates that Amazon prices do not react to an improvement in the aggregated star rating. Additionally, there seems to be no pattern in the effect. Taken together, this suggests that the price effect after three days is merely statistical coincidence. Remember, that the lack of significance in the Amazon effect could be caused by to few observations.

Result 1 (Effect of reviews on prices.) Marketplace sellers increase their prices up to 6.5 percentage points when the aggregated star rating improves. There is no robust statistical evidence for a price increase of Amazon offers in response to an improvement of the aggregated star rating.

In order to take a closer look at hypothesis 2 that refers to the different price reactions of Amazon and marketplace offers, I consider the following regression equation:

$$
\begin{aligned}
\frac{price_{i,t+x}}{price_{i,t}} =& \beta_1 \left((aggregated\ star\ rating)_{i,t} - \left(\overline{rating}\right)_i \right) + \beta_2 (Amazon)_i + \\
& \beta_3 \left((aggregated\ star\ rating)_{i,t} - \left(\overline{rating}\right)_i \right) * (Amazon)_i + \\
& \beta_4 (availability)_{i,t} + \beta_5 (availability)^2_{i,t} + \omega_i + \tau_t + \varepsilon_{i,t}.
\end{aligned}
\tag{3.3}
$$

In contrast to the baseline regression equation 3.1, equation 3.3 includes the binary variable $Amazon_i$, which is one for Amazon offers and zero otherwise. Note that I consider Amazon and marketplace offers as different products (even if the offer is for the same product). In consequence, the indicator variable drops out as the regression equation includes product fixed effects. Additionally there is an interaction term of the aggregated star rating with the $Amazon_i$ indicator. The interaction term allows to draw a direct conclusion about the difference in the price effects of marketplace and Amazon offers. Table 3.5 displays the regression results. There is no significant effect for an improvement in the aggregated star rating three and five days after a new review, neither for marketplace nor for Amazon offers. As already displayed in figure 3.3 marketplace prices significantly increase by 3 (4) percentage points 14 days (21 days) after an improvement in the aggregated star rating by one star. For both days (14 and 21), the effect is significantly smaller for Amazon offers. The regression results also show that there is basically no effect of an improvement in the aggregated star rating on Amazon prices.

Result 2 (Different effects on marketplace and Amazon prices.) There is basically no effect of an improvement in the aggregated star rating on Amazon prices. The effect on Amazon prices is significantly smaller than the effect on marketplace prices.

	$\Delta\text{price}_{i,t,x}$			
	(x = 3)	(x = 5)	(x = 14)	(x = 21)
	(1)	(2)	(3)	(4)
availability (in years)	0.0006^{+}	0.001^{+}	0.00271^{**}	0.0034^{**}
	(0.0004)	(0.0006)	(0.0011)	(0.0014)
availability^2 (in years)	$-2.27e{-7}^{+}$	$-4e{-7}^{*}$	$-9.32e{-7}^{**}$	$-1.18e{-6}^{**}$
	(4.2e−7)	(5.87e−7)	(1.47e−7)	(2.39e−7)
star rating (fit)	0.0061	0.005	0.0322^{***}	0.0399^{***}
	(0.0049)	(0.0057)	(0.0113)	(0.0129)
star rating (fit) * Amazon	−0.0036	−0.0032	-0.0354^{***}	-0.0457^{***}
	(0.0055)	(0.0067)	(0.0127)	(0.01585)
Product FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	102,902	102,028	100,098	99,330
Residual Std. Error	0.065	0.085	0.131	0.151
DoF	101,891	101,017	99,090	98,322
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3.5: Regression results with an interaction of the Amazon indicator and the aggregated rating.
The table shows the regression results based on regression equation 3.3 that tests the different effect in price reactions by marketplace sellers and Amazon. Column (1) displays the results for a lag of 3 days after a new review. Column (2), (3), (4) display the results for a lag of 5, 14, 21 days – as indicated in parentheses behind the price variable. The standard errors (in parantheses) are bootstrapped to correct for a manual implementation of 2sls and are clustered on product level.

3.6.2 Weighted aggregated rating

The aggregated star rating used in the baseline regressions presented in subsection 3.6.1 is the simple mean of all star ratings. According to Amazon, the website uses a machine-learned algorithm to determine the weighted average of all product reviews.[25] A reviewer's trustworthiness and timeliness establish the weights. Amazon prominently displays the weighted aggregated star rating of a product in yellow stars at the top of a products page, as can be seen in figure 3.5. Henceforth, I will refer to Amazon's weighted aggregated star rating as the *true stars*.

Figure 3.5: Amazon product page.
The figure shows the top of a product page on Amazon. A product's star rating is prominently displayed in yellow stars at the top of the page.

For each product, I observe the true stars once – at the time of scraping the data. Thus, the data does not include a panel of the true stars. The observed true stars are the true aggregated star rating of all the observed reviews of a product. If I wanted to know the aggregated star rating for all reviews except the latest five reviews, I would have to compute the aggregated star rating myself, as I do not observe the true stars at that point of time. The simple average of all reviews already yields a

[25]Here you can access Amazons description of the calculation of product star ratings: "How Are Product Star Ratings Calculated", `www.Amazon.com/gp/help/customer/display.html?nodeId=GQUXAMY73JFRVJHE`, accessed on 28 Feb 2020).

good approximation of the true stars. Column (1) of table 3.7 shows that the simple average of all product reviews explains 99.5% of the variation in true stars.[26]

I exploit all of the collected review data[27] to establish a regression that estimates weights for review characteristics and yields a more realistic approximation of the true stars. As Amazon claims that the weights of the star rating depend on a review's age and trustworthiness, I partition the data into three age categories – younger, and older one year, and older three years. The "trust" indicator summarizes all variables that can determine a trustworthy review. The "trust" indicator takes a value of one if

- Amazon labels a review as "Verified Purchase",
- Amazon awards the reviewer with a badge[28] for its remarkable reviewing activity,
- the reviewer is a "Vine Voice",[29]
- the review contains an image or video[30] or
- if it has at least one helpful vote and is zero otherwise.

I furthermore control for the number of stars awarded in a review. The respective indicator variable equals one if the reviewer awards one or two stars ("low rating") and is zero otherwise. Interacting the four indicators yields 12 categories (see table 3.6). For each category, I compute the sum of stars awarded by all reviews in the category.

---

[26]The regression equation looks as follows: $(true\ stars)_i = \beta_1 (aggregated\ star\ rating)_i + \epsilon_i$.

[27]The data comprises 185,375 observations of 1,734 products. I do not include all of these observations in my final regression, as several observations lack the data needed for the instrument.

[28]As noted before, Amazon awards some reviewers with badges such as "#1 Reviewer" or "Hall of Fame Reviewer" for their extraordinary reviewing activity. This site displays all badges with explanations: "Badges", `www.Amazon.com/gp/help/customer/display.html?nodeId=GED7RL944YMQ8CE3`, accessed on 07 Aug 2020.

[29]Amazon invites trustworthy reviewers to participate in their vine program. The program enables participants to select free products in return for an "unbiased" review. The review is considered as unbiased, as sellers are not able to influence, modify or edit the review. For more information see: "What is Amazon Vine?", `www.Amazon.com/gp/vine/help`, accessed on 07 Aug 2020.

[30]An image or video of the product indicates that the reviewer has at least received the product.

Category	Trust	Low Rating	Younger 1 Year	Older 3 Years
younger3y	0	0	0	0
trust younger3y	1	0	0	0
trust low younger3y	1	1	0	0
low younger3y	0	1	0	0
young	0	0	1	0
trust young	1	0	1	0
trust low young	1	1	1	0
low young	0	1	1	0
old	0	0	0	1
trust old	1	0	0	1
trust low old	1	1	0	1
low old	0	1	0	1

Table 3.6: Overview of category variables.
The weighted star regression as described in equation 3.4 is based on twelve category variables that take into account the age, trustworthiness and star rating of a review. The table offers an overview of all category variables.

To determine the weights for each category, I regress the observed true star rating of each product i on the sum of stars in each category c of product i:

$$true\ stars_i = \sum_{c=1}^{c=C} \beta_c \frac{sum\ stars_{c,i}}{num\ reviews_i} + \varepsilon_i. \tag{3.4}$$

I standardize the regressors by dividing them by the total number of product reviews. The standardization could also be achieved by multiplying the true stars with the number of reviews of the respective product in order to have sums on both sides of the equation. Albeit, taking the ratio on the right-hand side has the advantage that it returns coefficients around one, which are easier to interpret.

To estimate equation 3.4, I use OLS and the least absolute shrinkage and selection operator (Lasso). OLS has the drawback that it does not account for the predictive power of single regressors. OLS will therefore return estimates with low bias but large variance. Lasso can increase the predictive performance by reducing some of the variance in return for more bias (Tibshirani, 1996). Columns (2) and (3) of table 3.7 compare the results of both estimators. I display the Lasso results without standard errors, as it is difficult to compute accurate estimates for the standard errors of the Lasso regression (Tibshirani, 1996). Lasso especially shrinks those coefficients with few observations and low variance, namely $low\ young$, $low\ younger3y$ and $low\ older3y$. Table 3.7 shows that the OLS performance is slightly better in terms of the mean squared error (MSE) and the coefficient of determination (R^2).

As expected, the coefficients of both estimators are higher for more recent reviews than for older reviews. Moreover, trusted reviews have higher weights than their counterparts. Perhaps most interestingly, both estimators – OLS and Lasso – show rather clearly that low ratings (one or two stars) obtain higher weights than good reviews. It is, however, counterintuitive that untrusted low reviews obtain higher weights than their trusted counterparts in the OLS estimation. As noted before, Lasso penalizes these coefficients relatively strongly and leads to arguably more realistic estimates.

Result 3 (Weights of the true star rating.) Amazon assigns lower weights to untrustworthy, old, and highly rated reviews compared to their trustworthy, young, and low rated counterparts.

	true stars		
	OLS	OLS	Lasso
	(1)	(2)	(3)
average stars	0.989***		
	(0.000)		
trust young		0.992***	0.991
		(0.003)	
trust low young		1.140***	1.076
		(0.041)	
trust younger3y		0.982***	0.982
		(0.005)	
trust low younger3y		1.175***	1.123
		(0.054)	
trust old		0.946***	0.945
		(0.005)	
trust low old		1.180***	1.094
		(0.063)	
young		0.989***	0.944
		(0.024)	
low young		1.151***	0.704
		(0.213)	
younger3y		0.971***	0.911
		(0.034)	
low younger3y		1.360***	0.928
		(0.183)	
old		0.928***	0.866
		(0.032)	
low old		1.432***	0.984
		(0.202)	
Observations	1,734	1,734	1,734
R^2	0.995	0.996	0.836
MSE	0.105	0.079	0.081
Note:		*p<0.1; **p<0.05; ***p<0.01	

Table 3.7: Estimation of weights for Amazon's aggregated star rating.
The table contains the results from 3 different regressions. Column (1) shows the explained variation in Amazon's true rating through the aggregated rating. Column (2) and (3) show the weight estimates for each category variable as proposed in equation 3.4. Column (2) shows the results for OLS and column (3) for Lasso. (Clustered standard errors in parentheses.)

I use the OLS and Lasso estimates to calculate a weighted aggregated star rating. I rerun the regression described in equation 3.1 with the weighted instead of the simple aggregated rating. An important difference to the results from section 3.6.1 is that I now directly instrument the weighted aggregated star rating and not the individual star rating (as described in approach one in section 3.5).[31] The reason is simple, the weighted rating does not include the individual star rating but only the sum of stars in each category.

Figures 3.6 and 3.7 display the regression results for marketplace sellers and Amazon product offers. The figures are identical for the OLS and Lasso estimates, for which reason there is only one figure per seller group. For marketplace sellers, the trajectory of the effect is virtually identical to the course of the effect with a simple aggregated rating. However, the magnitude of the effect is weaker (up to only 5 percentage points instead of 6.5) and the significance level is partly lower in the regression with the weighted rating compared to the simple rating.

For Amazon offers, there now exist weakly significant effects in the first five days after a new review and after 15 to 19 days. The higher significance level for some observations is likely caused by the different employment of the instrument for the weighted rating compared to the simple rating. Figure 3.20 in appendix 3.9.7 displays the results for a regression analysis with the simple rating and the same use of the instrument as for the weighted rating. Figure 3.20 shows a very similar pattern as figure 3.7 and gives reason to believe that the use of the instrument and not the weighted rating causes the higher significance level of some observations. Again, I caution that the results could be different if I had more observations for Amazon offers.

[31] Appendix 3.9.7 displays the results when the *simple* aggregated star rating is instrumented.

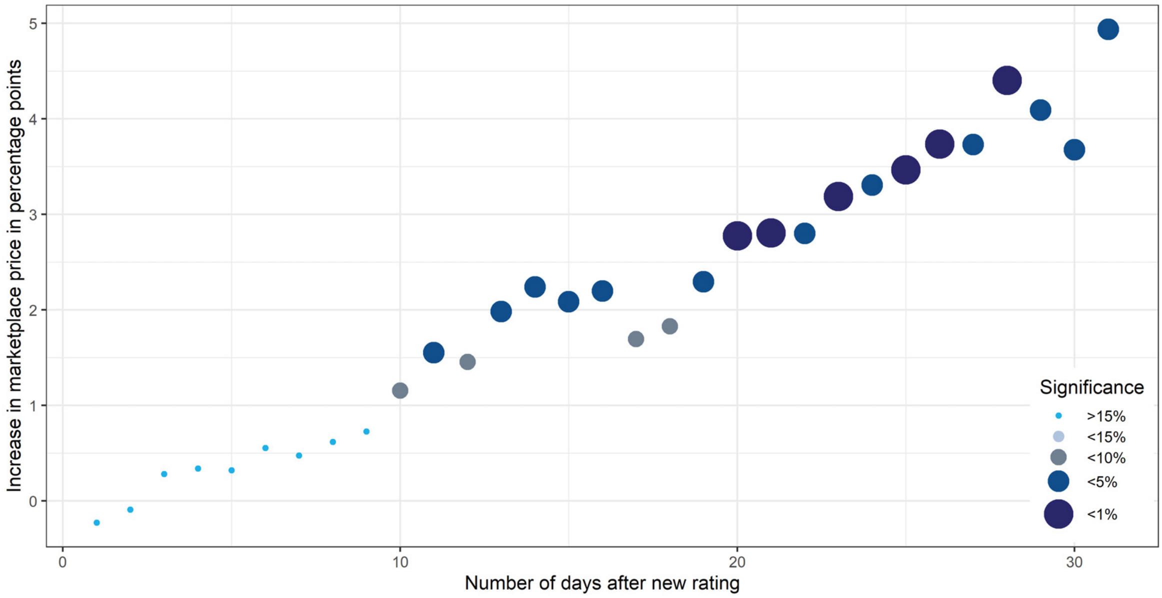


Figure 3.6: Regression results with the weighted aggregated star rating and marketplace offers.
Evolution of the change in marketplace prices from 1 to 31 days after a new rating based on the weighted aggregated star rating. The size of the dots indicates the significance level of the effect.

3.7 Further analyses and robustness tests

I perform the following additional analyses:

1. As it seems reasonable to assume that consumers choose to buy the offer with the lowest price, I analyze how the lowest price is affected by reviews, I create a “lowest” price variable and rerun the baseline regression.

2. Non-verified reviews are a proxy for untrustworthy reviews. I delete all non-verified reviews to learn whether the price effects stem from non-verified reviews.

3. To learn how long the marketplace price effect prevails, I rerun the main regression for price changes within 100 days after a new review.

Afterwards, I will explain why I do not perform a regression discontinuity design.

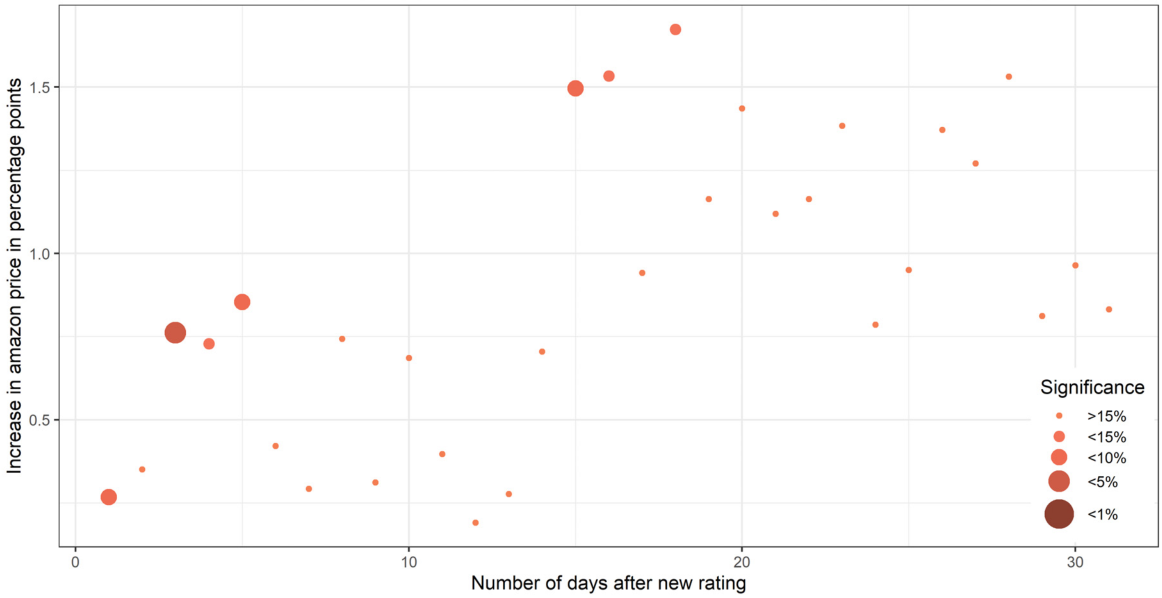


Figure 3.7: Regression results for the weighted aggregated star rating and Amazon offers.
Evolution of the change in Amazon prices from 1 to 31 days after a new rating based on the weighted aggregated star rating. The size of the dots indicates the significance level of the effect.

3.7.1 Lowest price and consumer surplus

In economics, we commonly assume that consumers buy the product with the lowest price. Especially on Amazon, prices for different offers of a product with the same identification number are transparent and easily comparable.[32] Starting from the assumption that consumers choose the offer with the lowest price, it seems reasonable to analyze the impact on consumer surplus on the basis of the lowest price. The *lowest price* is defined as the lower price of Amazon and marketplace prices when both prices exist. In case only one of the prices is observed, the *lowest price* is the observed price. I use the specification from equation 3.1 with the price relation of the *lowest price* as the dependent variable for the regression

[32]There is still reason to believe that consumers on Amazon do not search for the offer with the lowest price but choose the offer in the buy box. The buy box is the fastest way to shop on Amazon – only a click on "Buy Now" is required. The buy box is – if available – always displayed on the right hand side at the top of a product page, as can be seen in figure 3.12. The buy box price might not necessarily be the lowest price.

analysis and again instrument the individual star rating, as in the baseline analysis. The results might furthermore shed some more light on the reasons for finding no effect of an improvement in the aggregated star rating on Amazon prices. Is there no effect on Amazon prices, because there is just no effect? Or do I observe no effect because there is not enough data to analyze? Merging the marketplace price with the Amazon price naturally yields more observations and it is conceivable that the marketplace effect is enforced when there indeed is an effect on Amazon prices. It might, however, not be as easy as that. The lowest price is a hybrid of marketplace and Amazon prices and therefore might react slower to price changes or might react early but for a shorter duration.

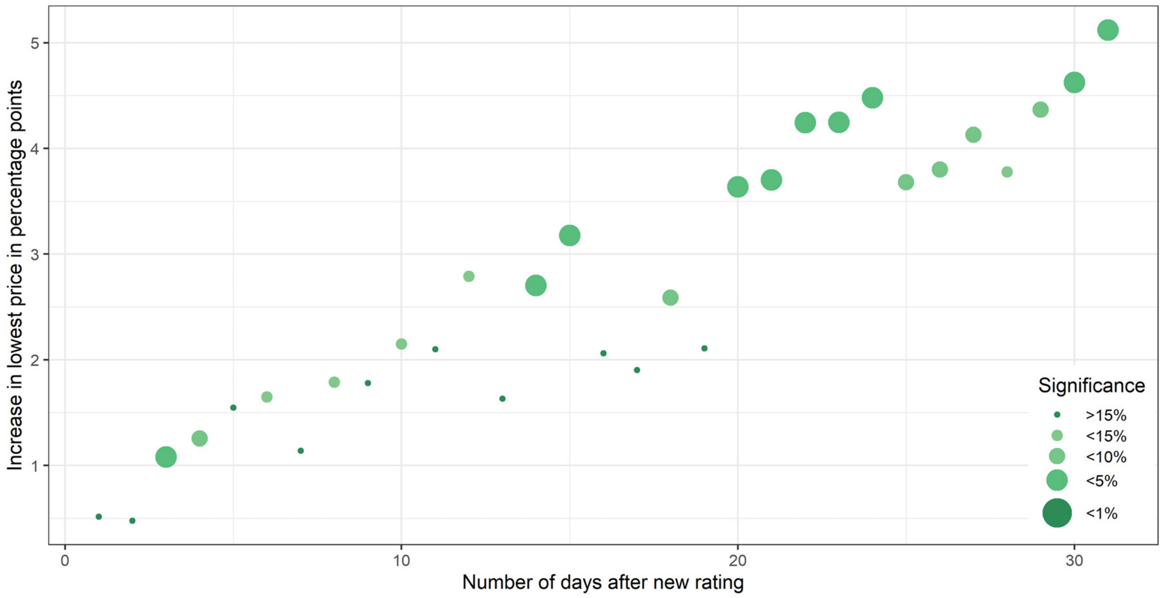


Figure 3.8: Regression results for the lowest price.
Evolution of the lowest price of marketplace and Amazon prices from 1 to 31 days after a new rating. The results are based on the simple aggregated rating. The size of the dots indicates the significance level of the effect.

Figure 3.8 displays the regression results. The prices significantly increase from day three until day 31 and up to around 5 percentage points. Just as the lowest price is a hybrid of marketplace and Amazon prices, the effect is a hybrid of the two effects.

The price effect of the lowest price is thus smaller than the effect of marketplace prices and larger than the effect of Amazon prices. Interestingly, the effect is now significant from day three onwards – with some exceptions. Based on the lowest price, the effect of an improvement of the aggregated star rating by one star affects the price channel of consumer surplus negatively starting three days after a new review. This can be regarded as a first indication that manipulated reviews – when bought to improve the own product rating – hurt consumers not only by distorting their consumption decisions but also through higher prices.
Figure 3.8 yields no indication of the existence of a price effect on Amazon offers. At least, the effect in the first five days after a new review is not more pronounced as, for example in figure 3.7 where the effect of the weighted rating was analyzed. Also the marketplace effect that starts to be prevalent from day ten onwards is not boosted in the low price analysis.

3.7.2 Removal of non-verified observations

In its Community Guidelines,[33] Amazon states that it will take action against members who contribute "false, misleading, or inauthentic content". This action comprises the deletion of content, denial of access to community content or exclusion from the community. A further attempt of Amazon to take action against misleading reviews is the "verified purchase" badge. When a review lacks the "verified purchase" badge, Amazon cannot verify that the product was purchased on Amazon or it has reason to believe that the product was purchased with a "deep discount".[34] A "deep discount" might be granted to a reviewer to provide an incentive for a good review. Section 3.6.2 shows that Amazon penalizes non-verified reviews with a lower weight in the aggregated star rating.

[33]See: "Guidelines for Amazon.com Community participation", `www.Amazon.com/gp/help/customer/display.html/ref=amb_link_1?ie=UTF8&nodeId=201602680&pf_rd_i=customer-reviews-guidelines`, accessed on 14 Jul 2020.

[34]Amazon's statement about verified purchases can be found under here: "About Amazon Verified Purchase Reviews", `www.Amazon.com/gp/help/customer/display.html?nodeId=202076110`, accessed on 07 Aug 2020. A definition of "deep" discount is not available.

I use non-verified reviews as a proxy for fake reviews.[35] To check whether the price effects, as identified in the previous sections, are driven by fake reviews, I delete all reviews that are not marked as "verified" by Amazon. Non-verified reviews comprise about 7% of all observations.[36] I rerun the regression proposed in equation 3.1 and 3.2 based on the simple aggregated star rating.[37]

Figure 3.9 displays the effect of an improvement in the simple aggregated rating by one star on the marketplace price. As in figure 3.3 – the regression results that include non-verified reviews – the effect on prices increase gradually with each additional day. The magnitude of the effect remains basically unchanged. Interestingly, the effect now already prevails earlier (starting on day six instead of ten). Another difference is the lower significance level of the effect, which is presumably caused by the lower number of observations.

Figure 3.10 displays the effect of an improvement in the simple aggregated rating by one star on the price of Amazon offers. The results are all insignificant and again show that the significant effects found, e.g. on day three in the baseline regression or in the first five days of the weighted rating regression are not robust. Interestingly, the regression based solely on verified reviews shows for the first time a pattern in the course of the coefficients.

To test, whether the effect for verified reviews is statistically different from non-verified reviews, I consider the following regression equation:

[35] To use reviews identified from a platform as misleading as a proxy for review manipulation is not uncommon. E.g. Luca and Zervas (2016) base their empirical analysis of review fraud on Yelp's identification of suspicious reviews.

[36] Surprisingly, on average, an Amazon product has more non-verified reviews than a marketplace product. On average, 8% of marketplace product reviews are non-verified, while 10% of Amazon product reviews are non-verified. Considering products that are exclusively distributed by marketplace sellers reveals that 7.5% of all reviews are non-verified, while it is 11% for Amazon-exclusive products. This could be an interesting start for future research.

[37] As explained, only 7% of reviews are non-verified. This leaves to few observations per product to run a regression analysis on non-verified reviews. An average product contains around 11 non-verified reviews.

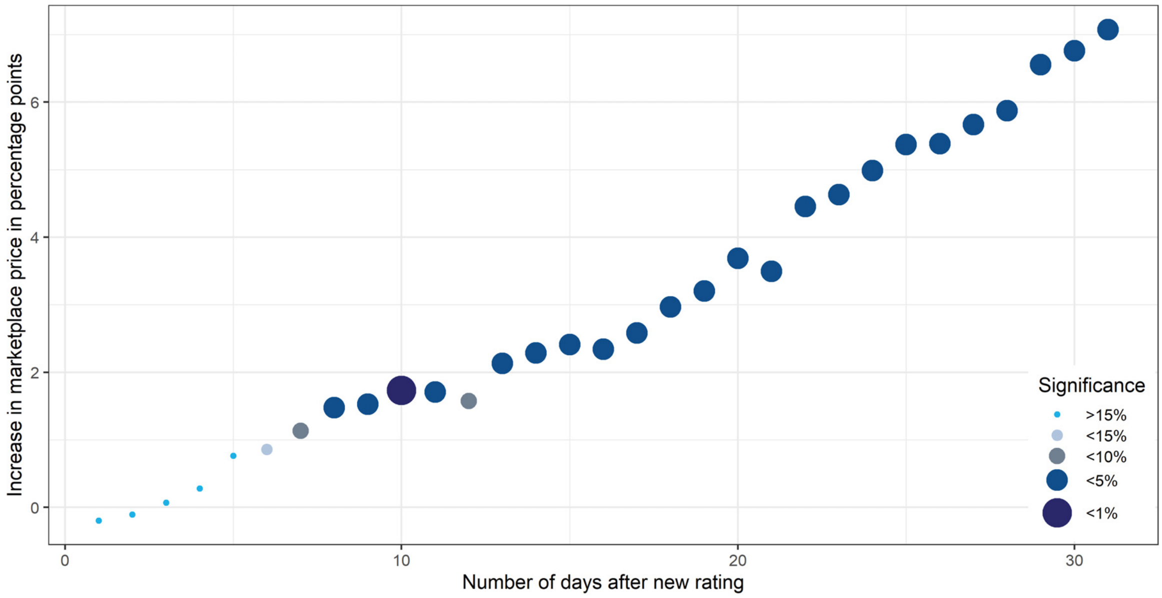


Figure 3.9: Regression results for a sample that excludes non-verified reviews and for marketplace offers.
Evolution of marketplace prices from 1 to 31 days after a new rating based on the simple aggregated rating and excluding non-verified reviews. The size of the dots indicates the significance level of the effect.

$$\begin{aligned}\frac{price_{i,t+x}}{price_{i,t}} =&\beta_1 \left((aggregated\ star\ rating)_{i,t} - \left(\overline{rating}\right)_i \right) + \beta_2 (verified)_{i,t} + \\ &\beta_3 \left((aggregated\ star\ rating)_{i,t} - \left(\overline{rating}\right)_i \right) * (verified)_{i,t} + \\ &\beta_4 (availability)_{i,t} + \beta_5 (availability)^2_{i,t} + \omega_i + \tau_t + \varepsilon_{i,t}.\end{aligned} \tag{3.5}$$

The results for equation 3.5 are displayed in table 3.8. They show that there is no significant statistical difference in the effect of an improvement in the aggregated star rating between verified and non-verified reviews for marketplace offers. Instead, there is a statistically different effect for Amazon prices. Amazon increases its prices slightly more in response to verified reviews.

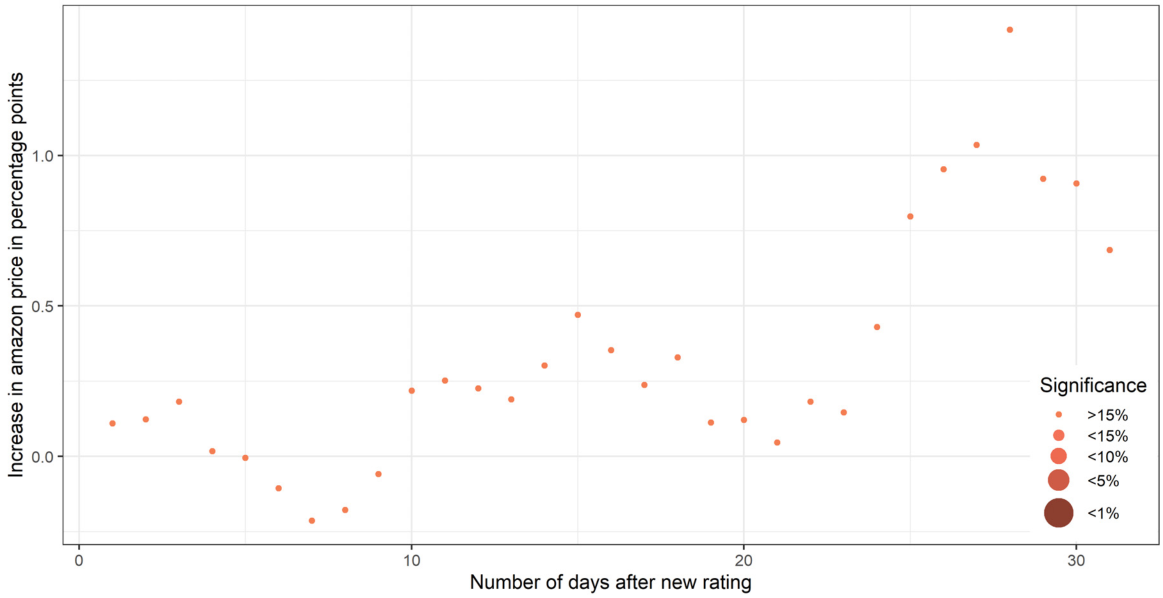


Figure 3.10: Regression results for a sample that excludes non-verified reviews and for Amazon offers.
Evolution of Amazon prices from 1 to 31 days after a new rating based on the simple aggregated rating and excluding non-verified reviews. The size of the dots indicates the significance level of the effect.

Result 4 (Review manipulation.) There is no significant statistical difference in the effect of an improvement in the aggregated star rating between verified and non-verified reviews on marketplace prices. Amazon increases its prices stronger in response to verified reviews.

3.7.3 100-days analysis

Section 3.6.1 presents results for the effect of a new review on the price within 31 days after a new review. While the effect of an improvement in the aggregated rating on Amazon prices is barely existent, the effect on marketplace prices persists from day 10 to day 31 after a new review. To test how long the prices increase after a new review, I perform the baseline regression analysis from equation 3.1 for 100 days after a new review. Figure 3.11 shows that marketplace prices increase up to

	$\Delta price_{i,t,x}$				
	Mktplc (x = 8)	Mktplc (x = 14)	Mktplc (x = 21)	Amz (x = 3)	Amz (x = 5)
	(1)	(2)	(3)	(4)	(5)
availability	0.0027^{**}	0.0024^{*}	0.0034^{*}	0.0009^{***}	0.0013^{***}
(in years)	(0.0011)	(0.0014)	(0.002)	(0.0003)	0.0002
$availability^2$	$-1.06e{-6}^{**}$	$-9.49e{-7}^{*}$	$-1.3e{-6}^{+}$	$-1.6e{-7}$	$-3.78e{-7}^{***}$
(in years)	(4.34e−7)	(5.69e−7)	(8.25e−7)	(1.25e−7)	(7.39e−8)
star rating (fit)	0.0122^{+}	0.0197^{*}	0.0377^{**}	0.0034^{***}	−0.0001
	(0.0079)	(0.0111)	(0.0164)	(0.0013)	(0.0016)
verified	−0.0116	−0.0283	−0.0252	−0.0039	-0.0129^{***}
	(0.0197)	(0.0292)	(0.0273)	(0.0030)	(0.0043)
star rating (fit) *	0.0025	0.0068	0.0065	0.0012^{*}	0.0033^{***}
verified	(0.0044)	(0.0067)	(0.0062)	(0.0007)	(0.0009)
Product FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Observations	79,524	78,902	78,412	22,686	22,395
Residual Std. Error	0.114	0.144	0.165	0.0278	0.032
DoF	78,753	78,132	77,643	22,429	22,138
Note:	$^{+}p<0.15$; $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$				

Table 3.8: Regression results for an interaction of the star rating with verified reviews.
The table displays the regression results that test whether sellers react differently to verified reviews compared to non-verified reviews. The regression is based on equation 3.5. Column (1) to (3) display the results for marketplace sellers with a lag of 8, 14 and 21 days in the price change. Column (4) and (5) display the regression results for Amazon offers for 3 and 5 days after a new review. The standard errors (in parantheses) are bootstrapped to correct for a manual implementation of 2sls and are clustered on product level.

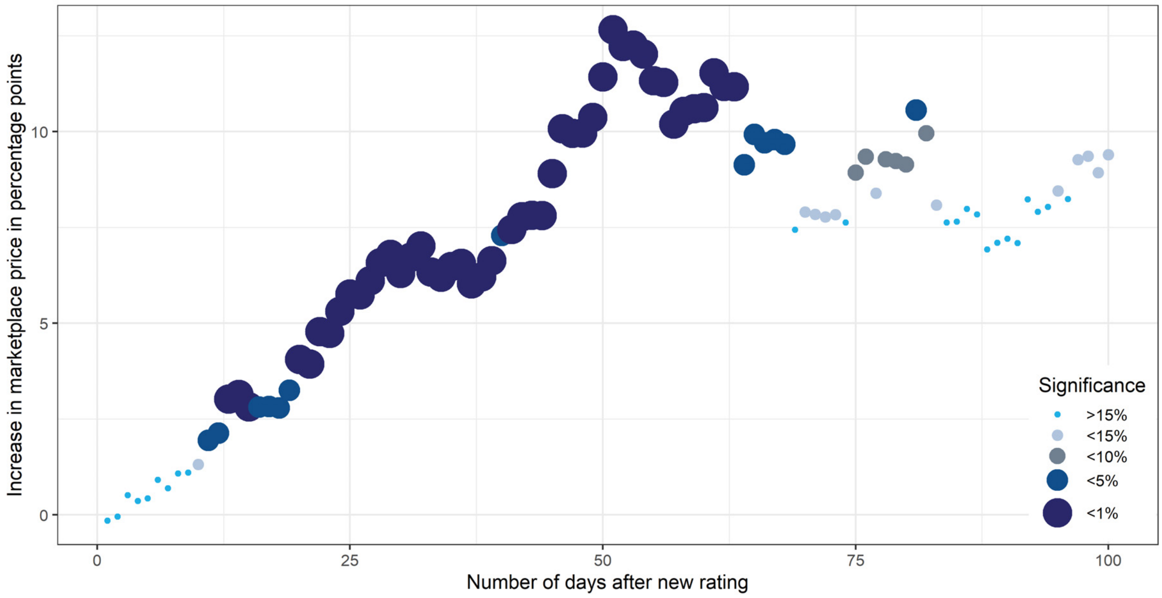


Figure 3.11: Regression results for the 100 day analysis.
Evolution of the change in marketplace prices from 1 to 100 days after a new rating based on the simple aggregated rating. The size of the dots indicates the significance level of the effect.

around 13 percentage points after around 50 days. An effect is prevalent for up to 68 days after a new review. For marketplace products, the interval between two reviews is, on average, 99 days long (see table 3.1). This implies that the effects displayed in figure 3.11 should not be the result of a confounding of effects from further new reviews.

3.7.4 Regression discontinuity design

To answer the empirical research question of this article it might seem natural to use a regression discontinuity (rd) design.[38] Due to imperfections of the data used in this article an rd design cannot be implemented for my analysis. The idea of the rd analysis is that the aggregated stars – prominently displayed at the top of each product's page or the product search page – are rounded to the nearest half-star. As

[38]Some commentators have raised this point in seminars.

a consequence, products with a very similar actual aggregated star ranking might appear to be of very different quality, depending on the rounding cutoff. A product that actually has 4.24 stars will thus be displayed as a 4-star product, whereas a product with 4.26 stars will be displayed as a 4.5-star product.

The cutoff of aggregated stars is the main treatment variable in an rd analysis. As explained in section 3.6.2, I have no data on the true stars, but only approximations. The regression analyses employed in this paper can cope with a degree of imprecision in the rating variable. A downside is, of course, that this may make it more difficult to estimate significant coefficients. In the rd analysis, however, the true stars are crucial for the design of the analysis. Articles that implement an rd-design to identify the causal effects of reviews use data from Yelp (Anderson and Magruder, 2012; Luca, 2016) or Wayfair (Magnusson, 2019), where the aggregated stars represent the average stars of the reviews. This is unfortunately not possible with the data set used for the present article.

3.8 Conclusion

Online product reviews are a frequently used information tool on digital shopping platforms. Peer reviews might help to decrease information asymmetries and to build trust. However, online product reviews might as well mislead consumers and bias competition in the case of review manipulation.[39] Understanding product reviews and their effects on market outcomes are therefore a relevant topic for economists studying digital markets.

This article analyzes the causal effect of online product reviews on prices of both Amazon and marketplace sellers for a wide range of products from twelve different categories. The empirical analysis exploits the variation in prices and reviews of self-collected product-time panel data. To approach a causal understanding of product

[39]Section 3.9.3 in the appendix gives an overview of the actions (and cases) of some competition authorities against fake reviews.

reviews on prices, I instrument a reviewer's star rating with the reviewer's general tendency to rate products. Finally, I use a linear regression model to approximate the weights of the star rating that Amazon displays for each product.
My analysis yields three main results:

1. Marketplace sellers on Amazon.com reflect reviews in their prices. I find that an increase in the aggregated star rating by one star increases the product's price, on average, by up to 6.5 percentage points. In contrast, there is no robust effect that indicates an increase in prices of Amazon offers in response to an improvement in the aggregated star rating.

2. The price effect occurs with a lag. Marketplace prices, on average, start to react to new ratings after ten days.

3. Amazon's true star rating can be approximated as a weighted sum of the individual ratings of customers. The weights of individual ratings depend systematically on the characteristics of the ratings. For instance, verified and low (one or two stars) ratings enter with higher weights than non-verified and/or high ratings.

On the one hand, the results suggests that Amazon's review system is effective – sellers react to changes in the aggregated star rating and presumably to a change in consumer's willingness-to-pay. On the other hand, the results might suggest that sellers could, on average, have an incentive to manipulate ratings, as good ratings can be translated into higher prices.
An additional analysis that excludes reviews from verified purchases indicates that marketplace sellers also increase their prices in response to these less reliable 'unverified' reviews while Amazon reacts stronger to verified reviews. This might indicate that marketplace sellers are, on average, more willing to benefit from fake reviews than Amazon. However, it is important to acknowledge that this is only a very tentative finding and more research in this regard is necessary.
Furthermore it needs to be noted that – due to the immense scraping effort – the

results regarding Amazon's position as a seller are based on a relatively small sample. The sample of Amazon offers contains 247 products, from which 50% have 30 reviews or less. An analysis of a larger sample could therefore yield further evidence.

References

ACCC (2013). *What you need to know about: Online reviews—a guide for business and review platforms*. https://www.accc.gov.au/system/files/Online%20reviews%E2%80%94a%20guide%20for%20business%20and%20review%20platforms.pdf. Accessed: 2020-07-23.

Amazon (2020a). Become an amazon seller. https://sell.amazon.com. Accessed: 2020-07-03.

— (2020b). Bestellen bei amazon marketplace. https://www.amazon.de/gp/help/customer/display.html/ref=video_mp_general/?nodeId=886412. Accessed: 2020-07-03.

Anderson, M. and Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, **122** (563), 957–989.

Ba, S. and Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, **26** (3), 243–268.

Berger, J., Sorensen, A. T. and Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, **29** (5), 815–827.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, **119** (1), 249–275.

Bundeskartellamt (2020). Bundeskartellamt identifiziert Probleme bei Nutzerbewertungen. https://www.bundeskartellamt.de/SharedDocs/Meldung/DE/Pressemitteilungen/2020/18_06_2020_SU_Nutzerbewertungen_Konsultation.html?nn=3591568. Accessed: 2020-07-02.

Camelcamelcamel (2011). How our price checking system works. https://camelcamelcamel.com/blog/how-our-price-checking-system-works. Accessed: 2020-05-17.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: online book reviews. *Journal of Marketing Research*, **63**, 345–354.

CMA (2016). CMA takes enforcement action against fake online reviews. https://www.gov.uk/government/news/cma-takes-enforcement-action-against-fake-online-reviews. Accessed: 2020-05-17.

— (2020). Facebook and ebay pledge to combat trading in fake reviews. https://www.gov.uk/government/news/facebook-and-ebay-pledge-to-combat-trading-in-fake-reviews. Accessed: 2020-07-23.

Dewally, M. and Ederington, L. (2006). Reputation, certification, warranties, and information as remedies for seller–buyer information asymmetries: Lessons from the online comic book market. *The Journal of Business*, **79** (2), 693–729.

Dewan, S. and Hsu, V. (2004). Adverse selection in electronic markets: evidence from online stamp auctions. *The Journal of Industrial Economics*, **53** (4), 497–516.

Duan, W., Gu, B. and Whinston, A. B. (2008). Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems*, **45** (4), 1007–1016.

European Commission (2005). Directive 2005/29/EC of the European Parliament and of the Council of 11 may 2005 concerning unfair business-to-consumer commercial practices in the internal market. *Official Journal of the European Union*.

European Commission (2019). Directive (EU) 2019/2161 of the European Parliament and of the Council: of 27 november 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules. *Official Journal of the European Union*, **L 328**, 7–28.

Fan, Y., Ju, J. and Xiao, M. (2016). Reputation premium and reputation management: Evidence from the largest e-commerce platform in china. *International Journal of Industrial Organization*, **46**, 63–76.

Federal Trade Commission (February 26, 2019). FTC brings first case challenging fake paid reviews on an independent retail website: Commission also alleges weight-loss supplement claims were deceptive. `https://www.ftc.gov/news-events/press-releases/2019/02/ftc-brings-first-case-challenging-fake-paid-reviews-independent`. Accessed: 2020-07-23.

Fruncillo, L. (2019). New amazon rating system causes rise in reviews. `https://tamebay.com/2019/12/new-amazon-rating-system-causes-rise-in-reviews.html`. Accessed: 2020-07-11.

Ghose, A. and Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, **23** (10), 1498–1512.

Hu, N., Bose, I., Koh, N. S. and Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, **52** (3), 674–684.

Jiang, B.-J. and Wang, B. (2008). Impact of consumer reviews and ratings on sales, prices, and profits: Theory and evidence. *Twenty Ninth International Conference on Information Systems, Paris*, pp. 1–17.

Jolivet, G., Jullien, B. and Postel-Vinay, F. (2016). Reputation and prices on the e-market: Evidence from a major french platform. *International Journal of Industrial Organization*, **45**, 59–75.

Kostyra, D. S., Reiner, J., Natter, M. and Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing*, **33** (1), 11–26.

Lawani, A., Reed, M. R., Mark, T. and Zheng, Y. (2019). Reviews and price on online platforms: Evidence from sentiment analysis of airbnb reviews in boston. *Regional Science and Urban Economics*, **75**, 22–34.

Lee, K. Y., Jin, Y., Rhee, C. and Yang, S.-B. (2016). Online consumers' reactions to price decreases: Amazon's kindle 2 case. *Internet Research*, **26** (4), 1001–1026.

Li, X. and Hitt, L. M. (2010). Price effects in online product reviews: An analytical model and empirical analysis. *MIS Quarterly*, **34** (4), 809–831.

Luca, M. (2016). Reviews, reputation, and revenue: The case of yelp.com. *Harvard Business School NOM Unit Working Paper 12-016*, pp. 1–39.

— and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, **62** (12), 3412–3427.

Magnusson (2019). Unboxing the causal effect of ratings on product demand: Evidence from wayfair.com. *Working Paper*, pp. 1–70.

Mattioli, D. and Herrera, S. (2020). Amazon's sales jump as coronavirus prompts surge in online shopping. *The Wall Street Journal*, (30.04.2020).

McDonald, C. G. and Slawson, C., JR. (2002). Reputation in an internet auction market. *Economic Inquiry*, **40** (3), 633–650.

Miranda, C. (2019). Cure encapsulations' misleading claims and fake reviews. `https://www.consumer.ftc.gov/blog/2019/02/cure-encapsulations-misleading-claims-and-fake-reviews`. Accessed: 2020-05-19.

Perez, S. (2019). Amazon tests a one-tap review system for product feedback. `https://techcrunch.com/2019/09/13/amazon-tests-a-one-tap-review-system-for-product-feedback`. Accessed: 2020-07-11.

Reintjes, D. (2020). Amazon wird Kerngewinner der Ladenschließungen sein. *WirtschaftsWoche*, (16.03.2020).

Resnick, P., Kuwabara, K., Zeckhauser, R. and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, **43** (12), 45–48.

Stock, J. H. and Watson, M. W. (2015). *Introduction to econometrics*. The Pearson series in economics, Harlow: Pearson Education Limited, updated third edition, global edition edn.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58** (1), 267–288.

Valant, J. (2015). Briefing, october 2015 – online consumer reviews: The case of misleading or fake reviews. *European Parliament Research Service*, https://www.eesc.europa.eu/resources/docs/online-consumer-reviews---the-case-of-misleading-or-fake-reviews.pdf. Accessed: 2020-07-23.

Wang, H., Zhang, W. and Zheng, L. (2011). Dynamic pricing in B2C based on online product reviews. *Procedia Engineering*, **23**, 270–275.

Wu, Y., Ngai, E. W., Wu, P. and Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, **132**, 113280.

Yu, M., Debo, L. and Kapuscinski, R. (2016). Strategic waiting for consumer-generated quality information: Dynamic pricing of new experience goods. *Management Science*, **62** (2), 410–435.

3.9 Appendix

3.9.1 Definitions

Term	Definition
Aggregated star rating	Describes the average of all star ratings. The aggregated star rating is (an approximation of) the star rating that is presented at the top of an Amazon product page (true star rating) – for an example see figure 3.12. If not noted otherwise the aggregated star rating is based on the simple average. The weighted aggregated star rating takes the weighted average of all star ratings. The weights are estimated in section 3.6.2.
Fake review	A review is considered as faked or misleading when it does not represent the true opinion or experience of the reviewer (see e.g. Valant, 2015).
Individual review	Refers to a single product review as displayed at the bottom of figure 3.13.
Individual star rating	Refers to the star rating of an individual review. The individual star rating is an integer number between 1 and 5.
Manipulated review	Is a more cautious expression for fake review.
Peer reviews	User-generated review. Reviews by other consumers that write about their experience with a product. In contrast to peer reviews, expert reviews are written by professionals who test a product specifically for the review.
Product review	Used as a synonym for peer reviews.
Seller review	Reviews written by consumers about sellers of a product or service. Seller reviews might establish the reputation of a seller.
Star rating	Describes the number of stars that ranges from one to five awarded in an individual review.
True stars	The term refers specifically to the exact number of yellow stars that are displayed at the top of each Amazon' product page. This is the "true" aggregated star rating of an Amazon product. The true stars specifically exclude any approximation of these stars.

Table 3.9: Defintions reviews.
Definitions in the review context, ordered alphabetically. The definitions have no general applicability and are defined to fit the usage of the terms in this article.

Term	Definition
Badges	Amazon awards some reviewers with a badge for their engagement in the reviewing activity. These badges comprise *#1 Reviewer, Top 10 Reviewer, Top 50 Reviewer, Top 100 Reviewer, Top 500 Reviewer, Top 1000 Reviewer, Hall of Fame Reviewer, Top Contributor, Amazon Verified Profile, THE, Amazon Official, Author, Artist, Manufacturer*
Buy box	The buy box is the fastest way to shop on Amazon – only a click on "Buy Now" is required. The buy box is – if available – always displayed on the right hand side at the top of a product page, as can be seen in figure 3.12.
Marketplace	Amazon features an integrated platform for third-party sellers. This platform is called Amazon marketplace and distributes new, refurbished, and used items (Amazon, 2020b).
Verified purchase	The badge "verified purchase" implies that Amazon has confirmed the purchase of the reviewer. A review could lack the badge because the customer received the product at a "deep" discount. More information can be accessed under this link: `www.Amazon.com/gp/help/customer/display.html?nodeId=202076110`
Vine Voice	Amazon invites trustworthy reviewers to participate in their vine program. The program enables participants to select free products in return for an "unbiased" review. The review is considered to be unbiased, as sellers are not able to influence, modify or edit the review. (`www.Amazon.com/gp/vine/help`)

Table 3.10: Definitions of Amazon features.
Definitions features of the Amazon platform, ordered alphabetically.

3.9.2 Number of Amazon visitors and unique users

Country	Million visitors	Reference
USA	95 monthly	Amazon (2020), www.amazon.com/b?ie=UTF8&language=en_US&node=10560941011
Canada	127 monthly	Statista (2020), www.statista.com/statistics/1047699/canada-websites-ranking-by-average-monthly-traffic/
Mexico	63 monthly	PracticalEcommerce (2020), www.practicalecommerce.com/mexicos-ecommerce-matures
Brasil	only started with own marketplace in 2019	
UK	~ 58	MobileMarketing (2019), www.mobilemarketingmagazine.com/amazon-uk-user-base-prime-mintel
Germany	44	WirtschaftsWoche (2016), https://blog.wiwo.de/look-at-it/2016/05/18/amazon-in-deutschland-44-millionen-kunden-davon-17-millionen-nutzer-von-prime/
France	15 monthly	Ecommerce News (2020), www.ecommercenews.eu/ecommerce-in-europe/ecommerce-france/
Italy	22	Statista (2018), www.statista.com/statistics/885521/amazon-unique-users-in-italy/
Spain	7.7	Expansión (2016), www.expansion.com/empresas/tecnologia/2016/02/13/56bf500fe2704e296a8b45e8.html
Netherlands	only started with own marketplace in 2020	
Turkey	only started with own marketplace in 2019	
Japan	> 50	Travel Voice (2019), www.travelvoice.jp/english/amazon-users-exceed-50-million-in-japan-boosted-by-an-increase-in-mobile-users
India	150	The New York Times (2018), www.nytimes.com/2018/09/04/technology/amazon-hindi-india.html
Singapore	only started with own marketplace in 2019	
China	market closed since August 2019	
Australia	33.6 monthly	SimilarWeb (2020), www.similarweb.com/website/amazon.com.au/#overview
UAE	45 monthly	BBC (2017), www.bbc.com/news/business-39416636

Table 3.11: (Monthly) Number of Amazon visitors / unique users for the different Amazon marketplaces.
If not marked with "monthly" the number in column two marks the number of unique users of Amazon for the given marketplace and year. All websites were accessed on 27 October 2020.

3.9.3 Policy background on online product reviews

In this section, I will give an overview of policy actions taken against fake reviews. Reviews reduce the uncertainty about product quality and help to overcome problems of information asymmetries (e.g., Fan *et al.*, 2016; Ghose and Ipeirotis, 2011). Thereby, reviews have the potential to strengthen competition (ACCC, 2013). At the same time, product reviews can deteriorate information asymmetries if they are falsified and send the wrong signals about product quality. This harms competition and consumers, who might form wrong beliefs about product quality. Wrong beliefs can alter the willingness-to-pay of consumers or drive them to buy products that they would not have bought otherwise. This might bias competition.

Major (competition) authorities have recognized the problem of fake reviews and are now taking action against them. The Federal Trade Commission (FTC) in the USA regulates fake reviews and endorsements in its "Guides Concerning the Use of Endorsements and Testimonials in Advertising" (16 CFR, Part 255). The FTC has already processed multiple cases in relation to paid fake reviews.[40] In February 2019, the FTC filed its first case on an independent retail website. A seller of weight loss supplements on Amazon had bought fake reviews (Federal Trade Commission, February 26, 2019; Miranda, 2019).

The European Commission addresses fake reviews in its Directive 2005/29/EC and 2019/2161 that prohibits traders from submitting or endorsing fake reviews and false representation as a consumer (European Commission, 2005, 2019).

In May 2019, the German federal cartel office started a sector inquiry on online reviews.[41]

The CMA regulates fake reviews in "The Consumer Protection from Unfair Trading Regulations 2008", Part 2, Regulation 5. In 2016, the CMA filed a case, in which a marketing firm posted fake reviews for their clients (CMA, 2016).

[40] This link www.ftc.gov/news-events/media-resources/truth-advertising/advertisement-endorsements displays all FTC cases in relation with advertisement endorsement.

[41] A report on the sector inquiry will be available in fall 2020 (Bundeskartellamt, 2020).

The Australian Competition and Consumer Commission (ACCC) prohibits fake or misleading consumer reviews in its “Competition and Consumer Act 2010”, Part VIA. The ACCC has prosecuted an infringement by a moving company that published false consumer reviews on its website (ACCC, 2013).
Also, China takes actions against fake reviews with its “E-commerce Law” that prohibits deception through false or misleading reviews Wu *et al.* (2020).

3.9.4 Examples of pages from data collection

Figure 3.12: Example of an Amazon product page.
The yellow stars below the product title are the true stars that are a weighted average over all reviews for a product. On the right-hand side, the figure shows the buy box that allows a quick purchase without browsing through all offers.

Figure 3.13: Example of an Amazon review page.
The figure shows the aggregated star rating at the top and the distribution of ratings. Below, the figure displays two out of 21,063 reviews for the product.

Figure 3.14: Example of an Amazon profil page of a reviewer.
The profile page of a reviewer can be accessed by clicking on the name of the reviewer on the review page (see figure 3.13 for an example of the review page). The profile page shows an overview of all the existing reviews of a reviewer.

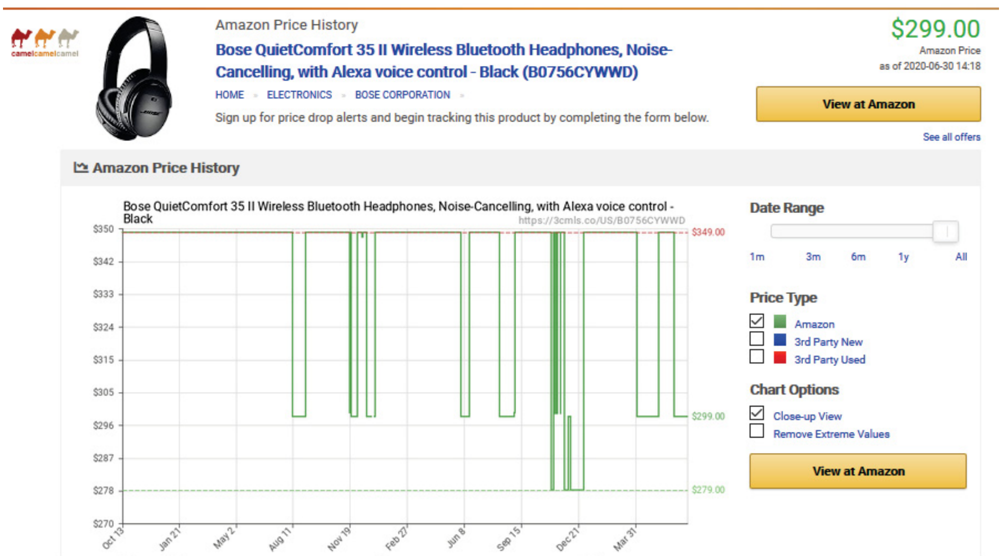


Figure 3.15: Example of the price tracking page camelcamelcamel.com.
The graph shows the course of Amazon prices. I use image analysis to translate the price graphs of the page into daily price data. The panel on the right-hand side enables the user, i.e., to choose the displayed price type.

3.9.5 Summary statistics

	Obs.	Range	1st Qu.	Mean	3rd Qu.	Sd.dev
Summary statistics without taking the product mean						
Prices 3rd (USD)	1,065,617	0.25 – 4,995.56	10.01	66.33	55.09	200.35
Price Amazon (USD)	405,506	1.66 – 1,431.26	11.40	69.29	62.8	131.59
Aggregated stars	136,038	1 – 5	4.17	4.32	4.55	0.37
Number of reviews	136,038	1 – 5,000	71	662.37	773	980.05
Available days	1,375,959	0 – 4,265	384	1,089.01	1625	863.08
Days betw. reviews	135,130	0 – 3,855	0	10.64	5	56.51
Summary statistics by product mean						
(resp. product maximum for # of reviews & available days)						
Prices 3rd (USD)	767	0.29 – 3290.87	11.50	64.94	56.47	188.88
Prices Amz (USD)	247	2.97 – 1384.44	13.80	78.90	77.72	155.02
Aggregated stars	908	1.28 – 5	4.09	4.32	4.72	0.57
Number of reviews	908	2 – 5000	10	149.82	114.5	419.61
Available days	908	69 – 4265	600.75	1471.09	2150.25	1030.78
Days betw. reviews	908	0.19 – 1788.0	10.66	109.32	102.70	213.43

Table 3.12: Joint summary statistics.

The table presents joint summary statistics for all marketplace and all Amazon offers. The top rows show the statistics without taking a mean at product level. As a large panel for one product might strongly influence these statistics, the bottom rows present the statistics by product mean, respectively product maximum for the number of reviews and the number of available days.

	Obs.	Range	1st Qu.	Mean	3rd Qu.	Sd.dev
Prices 3rd in USD	193	1.27 – 886.12	11.31	66.46	64.38	132.30
Prices Amz in USD	193	2.97 – 1054.80	12.23	63.66	59.94	129.94
Aggregated stars	193	1.28 – 5	3.97	4.24	4.71	0.66
Number of reviews	193	2 – 2229	12	147.02	136	306.73
Available days	193	136-4265	984	1892.97	2729	1106.38
Days betw. reviews	193	0.56 – 1564	12.04	110.32	103.38	221.41

Table 3.13: Summary statistics for products that are offered by marketplace sellers and Amazon.
The summary statistics only consider products, which are offered by both retailers on the Amazon platform – marketplace sellers and Amazon. The statistics are based on the product mean, respectively product maximum for the number of reviews and the number of available days.

3.9.6 Additional evidence for instrument

I exploit the publicly available data on Amazon profile pages to instrument a reviewer's star rating. The profile pages display all past reviews of a reviewer. Based on this data, I calculate a tendency to rate for each reviewer. On average, a reviewer reviews 422 products. Only 0.7% of the reviewers have a tendency to rate that is exactly equal to 0 and 1.62% have a tendency to rate close to zero, that means larger -0.01 but smaller than 0.01. Almost 60% of reviewers are friendly, i.e., have a positive tendency to rate. Consequently, almost 40% of reviewers are unfair, i.e., have a negative tendency to rate. Figure 3.16 depicts the distribution of the tendencies to rate and shows that they range from -3.8 to 3. The mean tendency to rate is -0.0056.

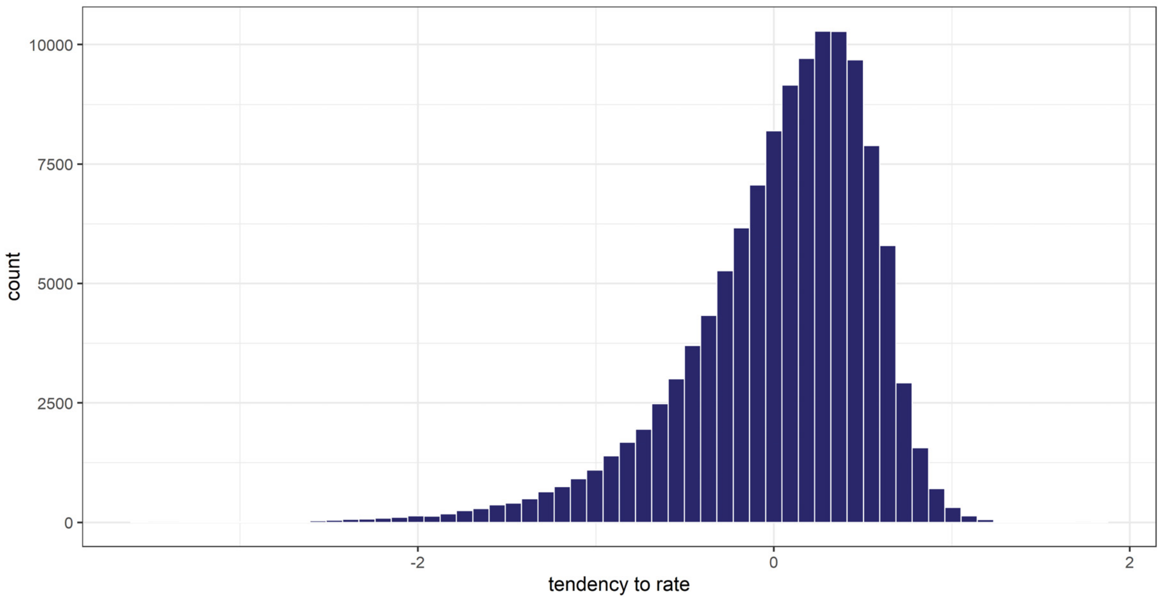


Figure 3.16: Histogram of the reviewers' tendency to rate.

Robustness check. As a robustness check, I delete all tendencies to rates that are based on very few reviews, i.e., less than five reviews. This causes some of the extreme tendencies to rate drop out such that the tendency to rate now ranges from -3.5 to 1.8. This also affects the mean, which now is positive at 0.0331.

The distribution of positive and negative reviews remains unchanged. I base the regression analysis on equation 3.1 and I instrument the star rating of a reviewer with the same reviewer's tendency to rate as proposed in equation 3.2. Figure 3.17 and figure 3.18 show that the results are robust to instrument less extreme tendencies to rate. The effect of an improvement in the aggregated star rating on market place prices is still gradually increasing and significant from eleven days after a new review onwards. The magnitude of the effect now even increases to a maximum of 8 percentage points, while the maximum of the baseline regression was 6.5 percentage points.

While there was a significant effect of an improvement of the aggregated star rating on Amazon prices on day three in the baseline regression, the effect is now prevalent in the first three days and insignificant afterwards.

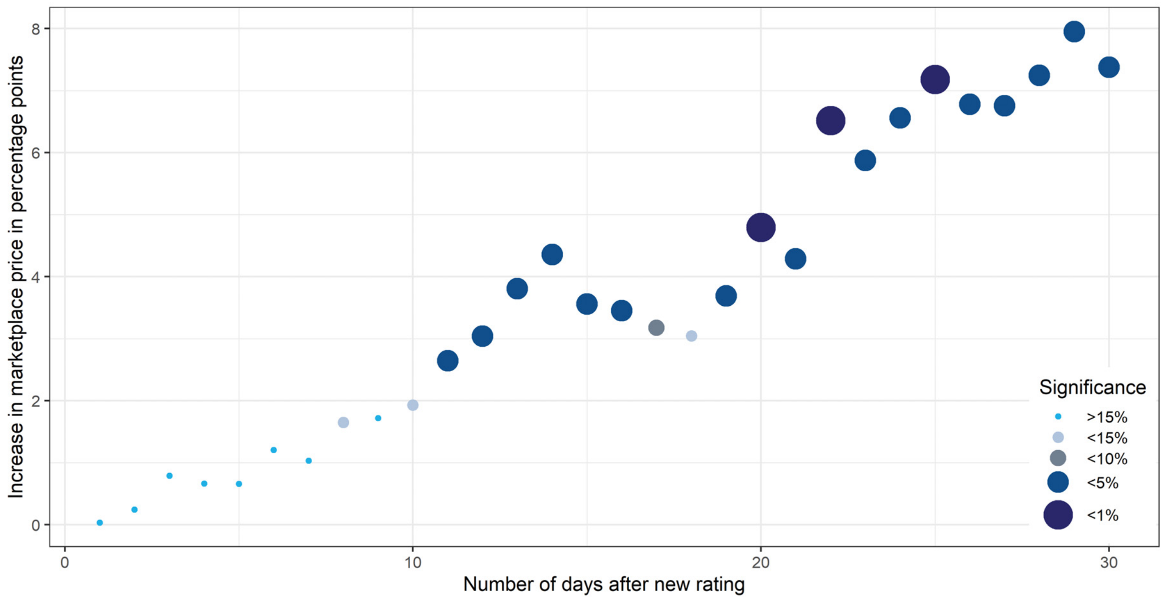


Figure 3.17: Regression results for the marketplace for a robustness check that excludes extreme tendencies to rate.
Evolution of the change in marketplace prices in response to an improvement in the simple aggregated star rating from 1 to 31 days after a new rating. The tendency to rate for reviewers with less than 5 reviews is dropped as a robustness check.

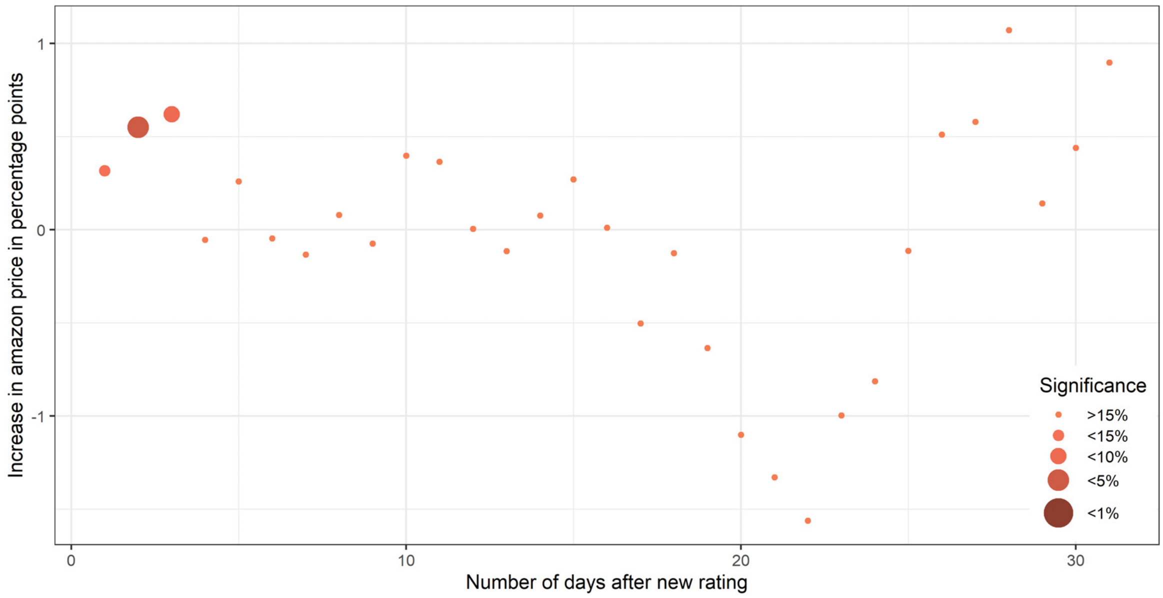


Figure 3.18: Regression results for Amazon for a robustness check that excludes extreme tendencies to rate.
Evolution of the change in Amazon prices in response to an improvement in the simple aggregated star rating from 1 to 31 days after a new rating. The tendency to rate for reviewers with less than 5 reviews is dropped as a robustness check.

Additional evidence. To provide additional evidence for the employed instrumental variable approach in which I instrument the individual star rating of a reviewer with the reviewer's tendency to rate, I perform the first-stage regression as proposed in equation 3.2 with subsamples. I first split the original sample into four parts according to the price quartiles. The subsamples include the following marketplace prices (p^{mktplc}) and Amazon prices (p^{amz}):

- **Q1**: p^{mktplc} <10.01; $p^{amz} < 11.4$;
- **Q2**: $10.01 \leq p^{mktplc} < 21.49$; $11.4 \leq p^{amz} < 23.10$;
- **Q3**: $21.49 \leq p^{mktplc} < 55.09$; $23.1 \leq p^{amz} < 65.97$;
- **Q4**: $p^{mktplc} > 55.09$; $p^{amz} > 65.97$.

As the first-stage regression does not include prices, I do not create different subsamples for Amazon and marketplace offers. Table 3.14 displays the first-stage regression results. For all subsamples, the instrument is significant on the one percent level and relevant (F-Statistic "excl instr" is above 170 for all subsamples). For the complete sample the coefficient for the tendency to rate ranges from 0.8 (for the combined and marketplace sample) to 0.85 (for the Amazon sample). In comparison to the coefficient range from the complete sample the coefficients from the subsample are not conspicuous. They range from 0.74 to 0.82. This implies that the star rating of a reviewer increases by around 0.8 stars when the reviewer's tendency to rate increases by one star. Table 3.14 shows that the effect of the instrument on the endogenous variable does not strongly depend on the price level.

Secondly, I partition the data into four subsamples according to their product category: *Home*, *Electronics*, *Apparel* and *Miscellaneous*. Table 3.15 shows the regression results. Again, the instrument is relevant (F-Statistic "excl instr" is above 83 for all subsamples) and significant on the one percent level for all subsamples. The coefficient for the tendency to rate is a bit lower for electronic products (here the tendency to rate, on average, only increases the star rating by around 0.7 stars when the tendency to rate increases by one star) than for other products.

	Individual star rating			
	Q1	Q2	Q3	Q4
	(1)	(2)	(3)	(4)
availability	-0.0546	-0.0643^{**}	-0.0004	-0.0355
(in years)	(0.0556)	(0.0288)	(0.0268)	(0.0315)
availability^2	$1.34e-5$	$1.86e-5^{**}$	$3.58e-6$	$1.01e-5$
(in years)	($1.29e-5$)	($7.23e-6$)	($8.83e-6$)	($1.24e-5$)
tendency to rate	0.798^{***}	0.738^{***}	0.793^{***}	0.824^{***}
	(0.051)	(0.056)	(0.044)	(0.034)
Product FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	20,503	35,409	20,171	15,078
F Statistic (proj model)	160.9 (df = 249)	64.96 (df = 341)	117 (df = 342)	204.1 (df = 255)
F Statistic (excl instr.)	245.0 (df = 249)	171.6 (df = 341)	318.8 (df = 342)	590.8 (df = 255)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3.14: First stage regression results for a split sample according to price quartiles.
The table shows the first-stage regression results for split samples. The sample is divided according to the marketplace/Amazon price quartiles. Robust standard errors (in parentheses) are clustered at product level.

	Individual star rating			
	Home	Electronics	Apparel	Misc
	(1)	(2)	(3)	(4)
availability	-0.0539^{**}	-0.0543	-0.0712	-0.035^{***}
(in years)	(0.0256)	(0.0356)	(0.0588)	(0.0126)
availability^2	$1.41e{-5}^{*}$	$1.78e{-5}^{*}$	$1.42e{-5}$	9.2199e-06*
(in years)	($8.31e{-6}$)	($1.02e{-5}$)	($1.19e{-5}$)	($5.09e{-6}$)
tendency to rate	0.819^{***}	0.681^{***}	0.890^{***}	0.788^{***}
	(0.090)	(0.068)	(0.036)	(0.026)
Product FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	19,188	9,798	27,389	79,663
F Statistic (proj model)	33.65 (df = 139)	33.68 (df = 85)	245.2 (df = 166)	346.3 (df = 514)
F Statistic (excl instr.)	83.2 (df = 139)	99.4 (df = 85)	622.0 (df = 166)	888.3 (df = 514)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3.15: First stage regression results for a split sample according to categories. The table shows the first-stage regression results for split samples of marketplace offers. The sample is divided in four different product categories: Home, Electronics, Apparel, and Miscellaneous. Robust standard errors (in parentheses) are clustered at product level.)

3.9.7 Instrument the aggregated star rating

This section presents the regression results when the aggregated star rating is instrumented by the mean tendency to rate. The baseline results in section 3.6.1 instrument a reviewer's individual star rating by the same reviewer's tendency to rate. The first-stage regression equation for strategy I – as used in this section – looks as follows:

$$\begin{aligned} \textit{aggregated star rating}_{i,t} =& \beta_1(\textit{average tendency to rate})_{i,t} + \beta_2(\textit{availability})_{i,t} + \\ & \beta_3(\textit{availability})^2_{i,t} + \omega_i + \tau_t + \varepsilon_{i,t}. \end{aligned} \tag{3.6}$$

As in section 3.6.1, I use the regression specifications presented in equation 3.1 to quantify the effect of an improvement in the aggregated star rating of a product on product prices for the second-stage. Columns (1) and (2) of table 3.16 show the regression results for the marketplace and columns (3) and (4) for Amazon products when the aggregated star rating is instrumented. An improvement in the aggregated star rating by one star is associated with a significant change in marketplace prices after 14 days of roughly 3 percentage points and after 21 days of roughly 6 percentage points. For Amazon products, an improvement in the aggregated star rating by one star is associated with a significant change in prices after three and five days of roughly one percentage point. As for the baseline regression, the relationship between product availability and product prices is non-linear for marketplace offers. Contrary to the baseline regression, the relationship is now also significant for Amazon offers (for a lag of five days). Initially, product availability slightly increases the product prices by 0.3 percentage points for marketplace offers and by 0.1 percentage points for Amazon offers. Each additional year of product availability decreases the change in product prices.

	$\Delta \text{price}_{i,t,x}$			
	IV	IV	IV	IV
	Mktplc	Mktplc	Amz	Amz
	(14 d)	(21 d)	(3 d)	(5 d)
	(2)	(3)	(5)	(6)
availability	0.0037**	0.0038**	0.0009	0.0014***
(in years)	(0.0019)	(0.0019)	(0.0006)	(0.0005)
availability^2	−1.67e−06**	−1.67e−06*	−1.89e−07	−4.2e−07***
(in years)	(9.05e−07)	(9.05e−07)	(2.12e−07)	(1.47e−07)
star rating (fit)	0.0336**	0.0568***	0.01*	0.0114
	(0.0133)	(0.0207)	(0.006)	(0.0076)
Product FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	76,499	76,033	22,291	22,004
Residual SE	0.141	0.163	0.027	0.031
DoF	75,732	75,267	22,036	21,749

Note:

Table 3.16: Regression results when the aggregated star rating is instrumented. Second-stage regression results for marketplace (columns (1) and (2)) and Amazon (columns (3) and (4)) offers of a two-stage least squares regression analysis with when the aggregated star rating are instrumented as displayed in equation 3.6 . The numbers in parentheses behind the dependent variable indicate the lag of the price change. Robust standard errors (in parentheses) are clustered at product level.

Figures 3.19 and 3.20 show the course of the price effect within 31 days after a new review for marketplace and Amazon offers. While the different employment of the instrument has only a minor influence on the significance of the price effect and the pattern of the effect for marketplace products, it does affect the magnitude of the effect. While the strongest effect was 6.5 percentage points when the individual star rating is instrumented, it increases to roughly 10 percentage point when the aggregated star rating is instrumented. For Amazon offers, the effect becomes slightly more significant by instrumenting the aggregated star rating instead of the individual star rating. However, there is still no visible pattern in the effect that indicates a robust result.

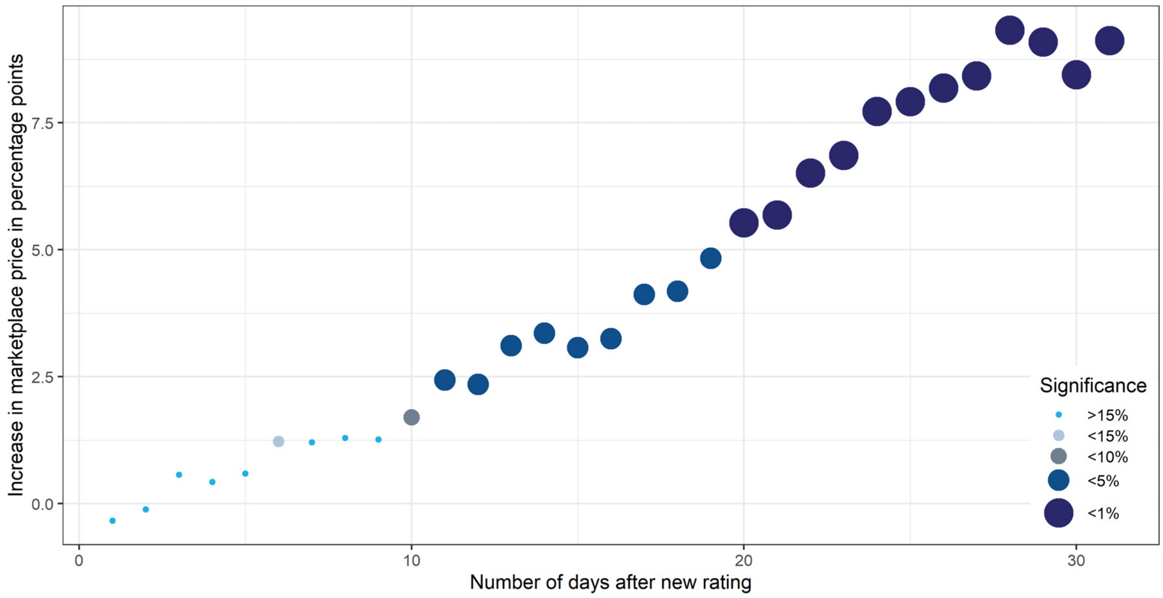


Figure 3.19: Regression results for the marketplace when the aggregated star rating is instrumented.
Evolution of the change in marketplace prices from 1 to 31 days after a new rating.

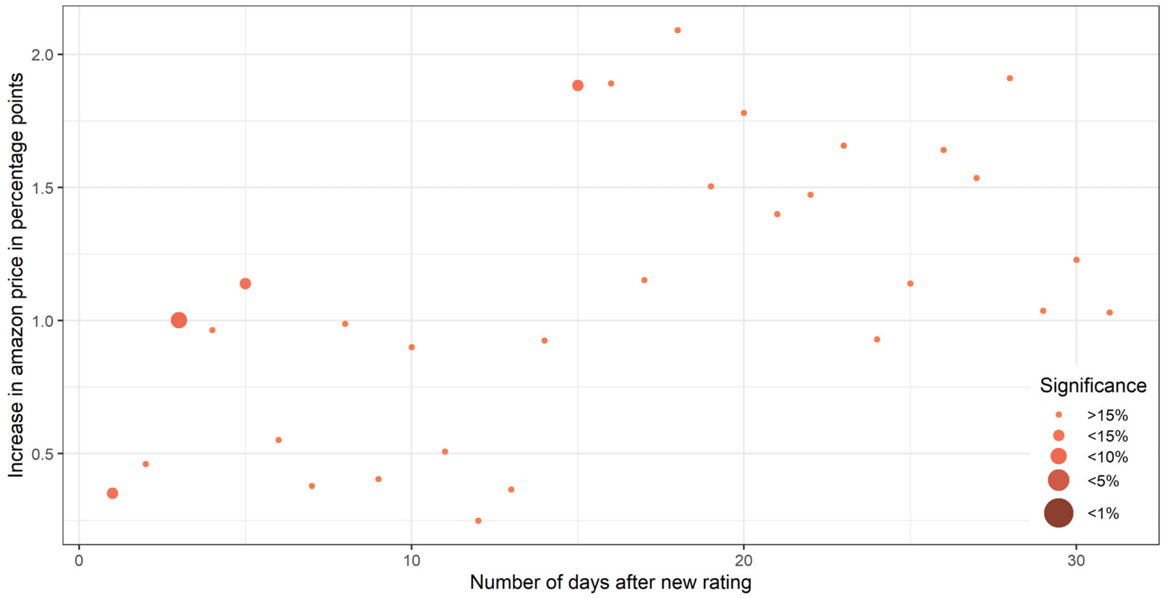


Figure 3.20: Regression results for Amazon when the aggregated star rating is instrumented.
Evolution of the change in Amazon prices from 1 to 31 days after a new rating.

Acknowledgments:

I am grateful to Hendrik Döpper, Justus Haucap, Matthias Hunold, Daniel Kamhöfer, Simon Martin, Hans-Theo Normann, Hannah Schildberg-Hörisch, Joel Stiebale and seminar participants at the DICE PhD Research Workshop and Brown-Bag Seminar for their comments and suggestions. Funding by the Deutsche Forschungsgemeinschaft (DFG) -– project 235577387/GRK1974 — is gratefully acknowledged.