Five Empirical Essays on Antitrust and Regulatory Affairs

DISSERTATION

zur Erlangung des akademischen Grades

Doctor Rerum Politicarum (Dr. rer. pol.)

im Fach Volkswirtschaftslehre

eingereicht an der Wirtschaftswissenschaftlichen Fakultät
Heinrich-Heine-Universität Düsseldorf

von:	Olivia Alexandra Bodnar, M.Sc. geboren am 21.01.1993 in Leverkusen
Erstgutachter:	Prof. Dr. Annika Herr
Zweitgutachter:	Prof. Dr. Justus Haucap
Abgabedatum:	20.08.2020

Acknowledgment

I would like to express my gratitude towards numerous people, but I would like to especially single-out Annika Herr. She has already aroused my interest in applied economics and especially the regulation of health care markets in my undergraduate studies. Besides, Annika Herr has taught me research and also constantly shared her network with me. I am especially thankful for the opportunity to visit the Center for Health Economics at the University of York during my Ph.D. studies. She constantly encouraged and challenged me, for which I am very grateful. I would also like to thank Justus Haucap. He arouses my interest in competition economics in general, and constantly shared his research environment and network with me.

I am also grateful to Hans-Theo Normann who always supported me during my time at DICE, especially as my co-author. In addition, I am very grateful to Melinda Fremerey, as well as Jannika Schad who are not only co-authors, but I am very happy to call both my friends now. Ph.D. life would have been much less satisfactory without both of you.

Moreover, I am grateful to (former) employees of the Monopolies Commission, who already started research projects with me during my master's program, paved my way for doing research and encouraged me to pursue doctoral studies. I would like to especially single-out Marc Bataille, Achim Buchwald, Susanne Thorwarth, and John Weche.

Besides, I would like to thank Hugh Gravelle and Nils Gutacker for having me at the Center for Health Economics at the University of York and for collaborating with me. During my stay I gained a lot of experience and received valuable input from great researchers in the area of health economics.

Special thanks go to my great colleagues at DICE. Research discussions, seminars, conferences, lunch, and ice cream breaks would have not been the same without a great team. In particular, I would like to mention Daniel Kamhöfer and Hendrik Döpper who both always offered helpful opinions on different kind of things. In addition, I would like to thank Judith Schwartz, a colleague, and friend who accompanies me since my undergraduate studies.

Funding by the Deutsche Forschungsgemeinschaft (DFG) - project 235577387/ GRK1974 - is gratefully acknowledged.

Finally, I like to thank my family – especially my parents Dorothea and Andre Bodnar–, my boyfriend Alexander Mönch, and my friends. Thank you for supporting me on my way.

Coauthors and publications

Chapter 1: The Return on Withholding Capacity Index

- Coauthors:
  - Marc Bataille (Monopolies Commission)
  - Alexander Steinmetz (University of Applied Sciences Würzburg-Schweinfurt)
  - Susanne Thorwarth (DICE Consult)
- Status: published in *Energy Economics* Vol. 81, Jun 2019, pp. 227-237

Chapter 2: Effect of Minority Shareholdings

- Coauthors:
  - Achim Buchwald (Federal University of Applied Administrative Sciences)
  - John Weche (Monopolies Commission)
- Status: published in *Journal of Economics and Statistics*, Vol. 237(6), Dec 2017, pp. 499-525

Chapter 3: Effect of Private Damage Claims: Experimental Evidence

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

Chapter 4: Effect of Physician Dispensing – Cross-sectional setting

- Coauthors:
  - Hugh Gravelle (Centre for Health Economics, University of York)
  - Nils Gutacker (Centre for Health Economics, University of York)
  - Annika Herr (Institute of Health Economics, Leibniz University Hannover)
- Status: to be made available as working paper

Chapter 5: Effect of Physician Dispensing – Dynamic setting

- Coauthors: none
- Status: to be made available as working paper

Contents

List of Figures

List of Tables

Introduction

"We can speculate on what's likely, but what's needed is an investigation. And speculation is no substitute for facts."

– Richard Blumenthal

Competition ensures offers to consumers at most favorable prices, quality, and full variety. This requires a market structure which is among others free of illegal agreements between firms. However, sometimes incentives might be in contradiction to independent, non-strategic behavior. The abuse of a dominant market position, foreclosure of a rival or agreeing on prices might be a profit-maximizing strategy, leading to a non-competitive market. To refocus incentives and to protect competition, competition policy is a necessary and powerful tool. When markets work competitively and given no market failure, there is usually no need for government intervention. Consumers benefit through higher quality, lower prices and better service. If competitive systems do not work, it is important to have laws and tools to protect competition. This is where antitrust law comes into place (U.S. Department of Justice, 2015).

The European antitrust policy is based on two pillars. The first pillar is Article 101 of the Treaty on the Functioning of the European Union, which states that every agreement between firms that restricts competition is prohibited. This includes agreements between competitors (horizontal agreement), as well as agreements along the supply chain (vertical agreement). Such agreements, for instance, include prices, territories, quantities or quality standards. The second pillar is Article 102 of same Treaty which deals with a dominant market position of a firm. If a firm holds a dominant market position, it is prohibited to abuse this position, for example, with charging excessive prices or limiting the variety of products (European Commission, n.d.). The abusive use of a dominant position is likely to happen in markets with essential facilities, for example, in telecommunication, rails services or the electricity sector (OECD, 1996).

My dissertation consists of several case studies in the field of antitrust and regulation. Figure 1 shows the analysed case studies on a time line. The first three case studies focus on 'antitrust'. It starts with the first energy sector inquiry by the European Commission in January 2007. This is followed by the European Commissions White Paper "Towards a more effective EU merger control" in 2013 seven years later. In this White Paper, the EU Commission discusses the effect of minority

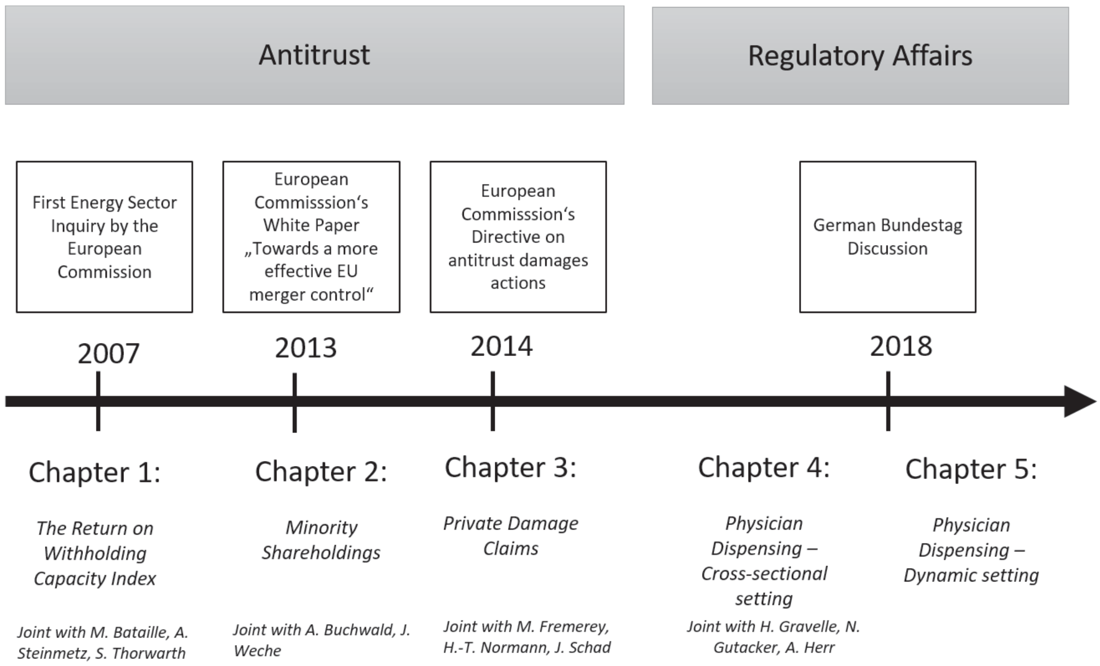


Figure 1: Timeline on competition policies

shareholdings on the incentive of foreclose strategies. One year later in November 2014, the European Commission signed private enforcement into law. This makes it possible for customers of a cartel to sue convicted wrongdoers for the loss they suffer in civil lawsuits. The second part of the thesis deals with 'regulatory affairs' in a broader sense. A discussion in the German parliament, the Bundestag, in 2018 deals with drug dispensing by physicians. The Chairman of the German General Practitioners' Association asked for a blurring in the prescribing and dispensing process and requested a dispensing rule for physicians in rural areas in Germany.

The first case study analyses a market with essential facilities. These markets with essential facilities are typically subject to particular attention in competition policy since their cost structure favors lower number of competitors[1]. However, some markets with essential facilities were liberalized to introduce competition. One of

[1]Important examples of a market monitoring include the European Commission (2016), Final Report of the Sector Inquiry on Capacity Mechanisms, COM(2016) 752 final, Nov 2016, `https://ec.europa.eu/energy/sites/ener/files/documents/com2016752en_.pdf`. and the German Federal Cartel Office (2011), Sector Inquiry into Electricity Generation and Wholesale Markets, Report in Accordance with Section 32e (3) of the German Act against Restraints of Competition-ARC, January 2011, Bonn.

these markets is the electricity market, which was liberalized in many countries in the 1990s. Nevertheless, incumbents often still hold a dominant market position. High energy prices on the Californian markets at the turn of the millennium initially triggered a large number of scientific studies where abusive market practice, such as capacity withholding, was investigated (Faruqui et al., 2001; Joskow and Kohn, 2002; Kwoka and Sabodash, 2011). This strategy is profitable because a supplier does not provide units of electricity although they could be offered at a price that exceeds its marginal costs. The rational behind this strategy is profitable because the supplier expects that the shortage of supply volume will lead to a shift in the merit order. As a result of the price mechanism in the wholesale electricity market prices rise, which leads to additional profits through higher contribution margins for the remaining power plants in the portfolio. However, the profit margin must exceed the loss margin resulting from the failure of a power plant (German Federal Cartel Office, 2011). In Germany, the by far highest price for a wholesale hour of electricity on the electricity exchange was measured in 2006 and amounted 2,437 euro[2]. In response to market developments at the time, the European Commission published its Energy Sector Inquiry in 2007 (DG Competition, 2007). Four years later the German Federal Cartel Office followed with its own sector inquiry. Since then, the market has continued to be monitored, e.g., by the Monopolies Commission's biennial reports. This high frequency in market observation is important since electricity wholesale trade is of great economic importance as electricity is a major input resource in other industries. At the same time, abusive market practice is not unlikely as it was seen in the California electricity crisis (Faruqui et al., 2001; Joskow and Kohn, 2002; Kwoka and Sabodash, 2011). However, it is a fundamental methodological challenge to identify potential market power of electricity producers and to distinguish it from competitive market developments. To address this challenge, Marc Bataille, Alexander Steinmetz, Susanne Thorwarth and I introduce a new market monitoring instrument that was solely developed for the special characteristics of the electricity sector in chapter 1. Right now, the Residual Supply Index (RSI) is the most important instrument for market monitoring (Sheffrin, 2002.). However, a major drawback of this index is its focus on just one specific aspect of market power in wholesale electricity markets whereas different consequences of market power are possible. Hence, markets could be distorted in several ways and we propose the "Return on Withholding Capacity Index" (RWC) as a complementary

[2]Own calculation from historical data from Energate.

index to the RSI. The index is a measure of the firms' incentive to withhold capacity. The benefits and practicability of the RWC are shown by an application on data for the German-Austrian electricity wholesale market in 2016. In the recently published Market Power Report, the Federal Cartel Office presents its current findings on the electricity market, in which the question of an appropriate combination of several market power indicators also plays an important role. At this point "according to the Federal Cartel Office current assessment [...] the RWC could therefore usefully supplement the RSI as a screening instrument in the future if properly implemented" (German Federal Cartel Office, 2019).

Following the first Energy Sector Inquiry, another important policy regulation was discussed seven years later. On the 9th of July 2014, the European Commission launched a consultation on a possible reform of the European merger control and published a White Paper entitled "Towards a more Effective EU Merger Control" to protect competition non-industry specific. The Commission has the task of implementing an effective and efficient competition policy but believes that this is not yet fully possible. The main reform proposal resulting from the consultation concerns the assessment of minority shareholdings. So far, the European Commission lacks the instrument and the adequate resources to examine them in more detail (European Commission, 2014a). However, since theory predicts possible negative anti-competitive effects from minority shareholdings (Gans and Wolak, 2008; Gilo and Spiegel, 2011; Gilo et al., 2018; Hunold and Stahl, 2016), the European Commission considers this tool to be important. For instance, the acquisition of a minority shareholding in the context of a non-horizontal merger may lead to input foreclosure. In some cases, minority shareholding may even make foreclosure more likely than in the case of a full integration process as only a part of the losses of the foreclosure strategy needs to be internalized (European Commission, 2014a). One main problem of an adjustment of the merger control are high implementation costs. Moreover, national authorities and companies have expressed their concerns about an adjustment and requested a sound empirical assessment of the effects of minority shareholdings (European Commission, 2015). However, there are some challenges with an empirical assessment. One of these challenges is the correct classification of firms as horizontal, vertical or conglomerate. Chapter 2 deals with this challenge and tries to give some empirical assessment about the effect of non-controlling minority shareholdings. Achim Buchwald, John Weche and I present a method for identifying up- and downstream industries in inter-industry datasets via input-output tables. We apply this approach to aggregated European input-output data and present results on identified industry links and their sensitivity to threshold definitions. We furthermore test the time consistency of the up- and downstream assignments based on input-output tables, and discuss the limitations of this method. Finally, the method is used to test the anti-competitive effects of non-controlling minority shareholdings. The results do not suggest anti-competitive effects on average.

Following the discussion about a stricter merger control, another major policy regulation occurred one year later. The European Commission made a large adjustment in European antitrust law and signed private enforcement into law in November 2014. The European Commission already started to consider private enforcement with its 2005 Green Paper (European Commission, 2005). In 2018 the last member states implemented the directive on antitrust damages actions into national law (European Commission, 2014b, 2018). However, the effect of private damage claims is seen contradictory in theoretical work (Canenbley and Steinvorth, 2011; Cauffman and Philipsen, 2014; Knight and de Weert, 2015; Migani, 2014; Wils, 2003, 2009). Private damage claims in cartel cases may have negative effects on leniency, a prime tool to uncover cartels. Chapter 3 looks at these possible contradictory effects. The trade-off between public and private enforcement arises because whistleblowers obtain no or only restricted protection against third-party damage claims. This may stabilize cartels. Melinda Fremerey, Hans-Theo Normann, Jannika Schad and I run an experiment to study this trade-off. 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. Experiments can be a useful instrument for the evaluation of new policy tools and for analyzing the effects of cartel stability ceteris paribus. In our experimental setting firms choose whether to join a cartel and may apply for leniency afterwards. Our design extends existing leniency experiments by adding a stage where private damages may occur after a cartel has been uncovered (either by a whistleblower or the cartel authority). We further investigate two communication formats. We compare unrestricted chat to the structured announcements (of 'acceptable' prices or price ranges) the literature has, so far, largely focused on. We find that the implementation of private damage claims decreases cartel formation but makes cartels more stable. The overall impact of private damage claims is positive: cartel prevalence declines.

The second part of this thesis deals with regulatory affairs in a broader sense and refrains from antitrust policy. In markets with essential facilities markets fail due to a lack of competition and a monopoly sometimes may even be the most cost-efficient market form. In other markets, however, there are other failures which makes a regulation necessary. One example of such a market is the health care sector. One major problem in health care markets are information asymmetries. A doctor has an information advantage compared to the patient. For example, a physician might prescribe a drug that is not necessary and may even have negative long-term effects, e.g., an unnecessary treatment with antibiotics (McGuire, 2000). A doctor usually has no incentive to overprescribe, because he has no financial gain in doing so. Fundamental to this is a separation of the prescribing and dispensing process of drugs. In most countries, this separation prevents patients from possible unnecessary treatment. However, patients living in rural areas who have only limited access to pharmacies close by, might have a need for drugs dispensed by their doctor. This question was raised in 2018 by the chairman of the German General Practitioners' Association who says that drug dispensing by physicians "would allow better use of resources especially in rural areas" (F.A.Z., 8.10.2018).

In chapter 4 and 5, we study the behaviour of dispensing doctors. We would like to study this behaviour in Germany, however, we lack the counter-factual regulation, as well as the data. Therefore, we analyse the behaviour of dispensing doctors in England where this regulation is already in place since many years. In chapter 4, Hugh Gravelle, Nils Gutacker, Annika Herr and I evaluate drug dispensing by general practitioners in the English National Health Service (NHS) between 2011 and 2018. In the NHS, physicians may dispense drugs to patients who live more than a mile away from the next pharmacy where the number and share of eligible patients vary across practices (Department of Health, 2012) (Regulation 48(3)(a)). We identify causal effects of physician dispensing on sales, substitution, and proxies for quality (e.g., the use of antibiotics or opioids). We show that drug dispensing increases expenditures by 4.3% and the number of prescribed items by 2.4% per patient per year. Simultaneously, dispensing physicians prescribe smaller packages and more over-the-counter (OTC) drugs. This also holds true for opioids for which overprescribing is potentially harmful. In a second step, we restrict the analysis to all dispensing practices and exploit that the number of dispensing patients varies within a practice over time which confirms most of the results. Thus, dispensing physicians prescribe differently compared to non-dispensing physicians and they do this mostly in an inefficient way.

In chapter 4, the dispensing doctor market regulation is analyzed in a static, cross-sectional setting. Making dispensing and non-dispensing doctors directly comparable is a major challenge of these studies (Kaiser and Schmid, 2016; Burkhard et al., 2019; Ahammer and Zilic, 2017). We use a matching procedure to overcome this problem. In chapter 5, I use a policy change that gives a unique study setting for this market regulation because it makes it possible to directly compare dispensing doctors with each other. To study whether dispensing general practitioners (GPs) behave differently to non-dispensing GPs, I exploit the introduction of new prescribing guidelines regarding OTC drugs. The NHS introduces the OTC policy to decrease overall NHS spending (NHS Clinical Commissioners, 2018). However, the policy change has not been strictly enforced and, therefore, gives scope for an individual adaption decision for dispensing and non-dispensing GPs. The paper is the first to study the adaptation over time within practices, where a policy change was not strictly enforced. I compare pre- and post-reform OTC prescribing in England with non-affected prescribing behaviour in Wales. For the analysis, I use monthly prescription data from all general practitioner practices in both, the NHS England and the NHS Wales, from April 2015 to September 2019. The results show that the reduction in OTC volume of dispensing doctors is significantly smaller compared to non-dispensing doctors. However, the reduction in OTC expenses, which the NHS monitors, is not significantly different for dispensing and non-dispensing doctors. This is due to the result that dispensing doctors prescribe less expensive OTC drugs compared to non-dispensing doctors after the reform. The rewarding system for dispensing doctors with a fixed item component might be one possible reason for this finding.

The dissertation is organised as follows: In chapter 1, we present our study regarding monitoring market power in the electricity wholesale sector. Chapter 2 introduces a tool to identify upstream and downstream industries in large firm datasets. We then apply this approach to an European firm level dataset to discuss possible anti-competitive effects of minority shareholdings. In chapter 3, we analyse a new major tool in antitrust policy: private damage claims. We run an experiment to study the formation and stability of cartels under the new private damage claim regulation. Chapter 4 and 5 deal with regulatory affairs in the health care sector. We study the behaviour of dispensing physicians and analyse whether they behave differently compared to non-dispensing physicians.

Bibliography

Ahammer, A. and I. Zilic, "Do financial incentives alter physician prescription behavior? evidence from random patient-gp allocations.," *Working paper.*, 2017.

Bigoni, M., S.-O. Fridolfsson, C. Le Coq, and G. Spagnolo, "Fines, Leniency, and Rewards in Antitrust," *The RAND Journal of Economics*, 2012, *43* (2), 368–390.

Burkhard, D., C. P. R. Schmid, and K. Wüthrich, "Financial incentives and physician prescription behavior: Evidence from dispensing regulations," *Health economics*, 2019, *28* (9), 1114–1129.

Canenbley, C. and T. Steinvorth, "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*, 2011, (2), 315–326.

Cauffman, C. and N. J. Philipsen, "Who does what in competition law: Harmonizing the rules on damages for infringements of the EU competition rules?," *Maastricht European Private*, 2014, (2014/19).

Department of Health, "Regulations under the health act 2009: Market entry by means of pharmaceutical needs assessments information for primary care trusts executive summary and chapters 1-4.," 2012.

DG Competition, "Report on energy sector inquiry," 2007, (SEC(2006)1724).

European Commission, "Green paper - damages actions for breach of the EC antitrust rules," 2005.

_ , "Towards more Effective EU Merger Control.," *COM(2014) 449 final*, 2014a, *09/07/2014.*

_ , "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.*, 2014b.

_ , "Summary of replies to the public consultation on the White Paper Towards more effective EU merger control," 2015.

— , "Final Report of the Sector Inquiry on Capacity Mechanisms," *com(2016) 752 final*, 2016.

— , "Directive on antitrust damages actions: Transposition of the directive in member states.," 2018.

— , "Consolidated version of the Treaty on the Functioning of the European Union - Part three: Union Policies and internal actions- Title VII: Common rules on competition, taxation and approximation of laws."

Faruqui, A., H. Chao, V. Niemeyer, and K. Platt J.and Stahlkopf, "Analyzing California's Power Crisis," *The Energy Journal*, 2001, *22* (4).

F.A.Z., "Hausärzte stellen Monopol der Apotheker infrage," 8.10.2018.

Gans, J. S. and F. A. Wolak, "A Comparison of Ex Ante versus Ex Post Vertical Market Supply: Evidence from the Electricity Supply Industry," *SSRN Electronic Journal*, 2008.

German Federal Cartel Office, "Sector Enquiry Wholesale Energy Sector," *b10-9/09*, 2011.

— , "Wettbewerbsverhältnisse im Bereich der Erzeugung elektrischer Energie (Marktmachtbericht)," 2019.

Gilo, D. and Y. Spiegel, "Partial Vertical Integration in Telecommunication and Media Markets in Israel," *Israel Economic Review*, 2011, (9).

— , N. Levy, and Y. Spiegel, "Partial Vertical Integration, Ownership Structure, and Foreclosure," *American Economic Journal: Microeconomics*, 2018, *10* (1), 132–180.

Hunold, M. and K. Stahl, "Passive vertical integration and strategic delegation," *The RAND Journal of Economics*, 2016, *47* (4), 891–913.

Joskow, P. L. and E. Kohn, "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000," *The Energy Journal*, 2002, *23* (4).

Kaiser, B. and C. Schmid, "Does Physician Dispensing Increase Drug Expenditures? Empirical Evidence from Switzerland," *Health economics*, 2016, *25* (1), 71–90.

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

Kwoka, J. and V. Sabodash, "Price Spikes in Energy Markets: "Business by Usual Methods" or Strategic Withholding?," *Review of Industrial Organization*, 2011, *38* (3), 285–310.

McGuire, T. G., "Chapter 9 physician agency," 2000.

Migani, C., "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*, 2014, (7), 81–111.

NHS Clinical Commissioners, "Quick reference guide for healthcare professionals: Conditions for which over the counter items should not routinely be prescribed in primary care," 2018.

OECD, "Policy Roundtables. The Essential Facilities Concept," *OCDE/GD(96)113*, 1996.

Sheffrin, A., "Predicting market power using the residual supply index. Predicting market power using the residual supply index.," *FERC market monitoring workshop*, 2002.

U.S. Department of Justice, "Antitrust Enforcement and the Consumer," 2015.

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

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

1

Screening instruments for monitoring market power — The Return on Withholding Capacity Index (RWC)

Joint work with Marc Bataille, Alexander Steinmetz, Susanne Thorwarth

1.1 Introduction

Since the 1990s, energy markets have been liberalized in many countries all over the world. Wholesale electricity markets are a key part of the energy sector and many of these markets have reached a reasonable share of private competitors. Nevertheless, incumbents often still hold dominant positions due to their existing fleet of power stations. Thus, wholesale electricity markets are typically subject to market . In particular, specialized regulators, competition authorities and service operators are monitoring the market development to deduce substantial information about the degree of remaining market power.[1]

However, the most important development in electricity wholesale has been the increasing electricity production from renewable energies in recent years. In many countries, renewables have reached considerable market shares within a short time. In Germany, for instance, renewable energies accounted for already 40.36 percent of electricity production in 2018.[2] A special characteristic of renewable energies is that production is subsidized in many countries. In addition, the production volume of important technologies, such as wind and solar power, cannot be determined by the suppliers. With marginal costs close to zero, production is heavily dependent on meteorological factors. Therefore, cartel authorities generally regard renewable energies as a separate market, while conventional power plants compete for residual load (load minus generation from volatile renewables) (German Federal Cartel Office, 2011). Due to the rapidly increasing volume of renewable energies and the resulting overall increase in electricity generation capacity, market power problems on wholesale markets have recently been observed less frequently. Therefore only little research has been done in this field. However, due to the closure of power plants and increasing shortages a new focus is laid on this topic. In the German-Austrian bidding zone, the average day-ahead price for electricity rose from 30.23 euro/MWh to 45.62 euro/MWh between 2016 and 2018.[3] Most European countries

[1]Important examples of such a market monitoring include the European Commission (2016), Final Report of the Sector Inquiry on Capacity Mechanisms, COM(2016) 752 final, Nov 2016, `https://ec.europa.eu/energy/sites/ener/files/documents/com2016752.en_.pdf` and the German Federal Cartel Office (2011), Sector Inquiry into Electricity Generation and Wholesale Markets, Report in Accordance with Section 32e (3) of the German Act against Restraints of Competition-ARC, January 2011, Bonn.

[2]Own calculations based on production data of the ENTSO-E transparency platform.

[3]This is the load-weighted average price for these years, calculated with data from the ENTSO-E Transparency Platform. Since October 1, 2018, the German und Austrian markets are separated, hence only data for the German bidding zone was taken into account for 2018.

are planning or even have already implemented government mechanisms to secure capacity (European Commission, 2016).

As conventional electricity becomes scarce again, monitoring of market power on wholesale markets is gaining importance. The results of this kind of monitoring are important for political and institutional actors to further develop the market design, e.g. in terms of capacity mechanisms. Moreover, the knowledge gained and the methods used in market monitoring to assess market power are important in quasijudical investigations conducted by competition authorities. For all these institutions, reliable screening tools are important to monitor markets with foresight. As screening of market power in wholesale electricity markets is subject to particular challenges, we are proposing a new instrument, the "Return on Witholding Capacity Index" (RWC) and show its application on data for the German-Austrian electricity wholesale market in 2016. The RWC index is a measure which indicates the strategic incentive of capacity withholding by a supplier. Therefore, we consider the perspective of an electricity supplier: electricity markets are mostly organized via uniform price auctions. Hence, the market price can increase by a large amount in reaction to a small decrease in supply when demand is at a high level. Consequently, strategically withholding a fraction of running capacity leads to a net increase in profit if the earnings from the higher market price exceed the losses from the offset power plant.

The article is organized as follows: The subsequent section gives a literature review and describes the existing tools for monitoring market power. Section 1.3 presents the new instrument for market screening of market power, namely the RWC, whose benefits and practicability are shown by using data from the German-Austrian electricity wholesale market in section 1.4. We conclude in section 1.5.

1.2 Monitoring market power: Literature review

Economic research provides a large set of indices to measure market power. They can generally be used by monitoring units in energy markets. By contrast to other markets, wholesale electricity markets have some distinctive characteristics which have to be taken into account when measuring market power. It is generally assumed that these markets are characterised by a mean reversion of the price, sudden fluctuations in load and supply without strong opportunities to storage and low elasticity in demand, which is reflected in price spikes (Cartea and Figueroa, 2005). Consequently, it can be shown that typical market share indices, such as the Herfindahl-Hirschman Index (HHI), are not well suitable to investigate market power in electricity markets (Newberry, 2009)[4]. Thus, market monitoring as well as certain research has been focusing in particular on two methods: on the one hand emphasis is put on the Residual Supply Index (RSI) which is kind of a structural index and specialized for the needs of electricity markets. On the other hand more complex behavioral analysis is used. The latter is typically based on real cost data or cost estimation, such as the price-cost markup.

The RSI was initially introduced by Sheffrin (2002) who showed a strong relationship between the RSI and markups during the California Energy Crisis in 2001. The index was developed as a more differentiated extension of the Pivotal Supplier Index (PSI) which has been used for the first time by the US Federal Energy Regulator Commission (FERC) in 2000 as a measure called Supply Margin Assessment (SMA) to determine market power of electricity suppliers. Both, PSI and RSI measure on an hourly basis whether a supplier is pivotal in terms of its capacity being relevant for the market to serve total electricity demand. If this is the case, the supplier can determine the price if demand is inelastic. Since pivotal market power has a significant impact, the RSI has gained considerable importance as a market power indicator. The RSI is also suitable for market monitoring, because it can be calculated with a reasonable amount of load and market share data.[5] Since its first application, the RSI has become an important predictor for market power in electricity markets (e.g. Chang (2007), Lang (2007), Asgari and Monsef (2010), Kamiński (2012), Mulder and Schoonbeek (2013)). Even more relevant seems to be the use of the RSI by market monitoring units of US regional transmission organizations (RTO).[6] Additionally, the RSI has gained importance in market surveillance by European competition authorities. In Europe, the RSI was crucial for assessing European energy markets by DG Competition (2007).[7] Furthermore, the German

[4]Even though HHI and similar measures like the concentration rates are still used to determine market concentration on wholesale electricity markets, i.e. Frontier Economics (2010), European Commission (2012).

[5]The Residual Supply Index has reasonable requirements for the data which are necessary for calculation. However, load and other data should be available (at least) on an hourly basis. If data is used on a more aggregated level a lot of explanatory power is lost by this way of RSI calculation.

[6]In RTO the use of PSI/RSI or similar indicators can be intensive as they can be part of local market power mitigation mechanisms. E.g. the California Independent System Operator (CAISO) applies the "Three Pivotal Supplier Test". In order to prove the appropriateness of the test a surveillance report is published CAISO (2013).

[7]On behalf of DG Competition (2007) London Economics undertook a study in which the RSI

Federal Cartel Office applied an RSI calculation on an hourly basis for the years 2007 to 2008 in its sector enquiry (German Federal Cartel Office, 2011). More recently, the German Monopolies Commission conducted RSI analyses in their special reports on the German energy sector (Monopolies Commission, 2013, 2015, 2017).

As an alternative to the RSI calculation, the Lerner index as a well established measure of market power in economic research can be used. The Lerner index or likewise the very similar price-cost markup (PCMU) is specified as the proportional price-cost margin of a firm. Although these indices are usually considered as reliable to describe market power they do not serve as common screening instruments in market monitoring due to the limited availability of adequate cost data. Using the Lerner index in wholesale electricity markets requires hourly data – in particular regarding marginal costs – for each power generation unit. Only in quite extensive sector investigations it might be feasible for a competition authority to obtain this cost data directly from the suppliers. Despite the fact that the European Commission as well as the German Federal Cartel Office have retrieved this information from power generators once in their sector inquiries, this complex procedure seems to be unsuitable for continuous market monitoring.

Instead of gathering information on real marginal cost from suppliers one can estimate costs using a synthetic model of electricity dispatch. These kind of models simulate the market by combining behavioral assumptions with available information about input prices and power generation units. Regarding the measurement of market power only few studies make use of those models. For example Lang (2007) analyze the German wholesale electricity market using a simulation model. Additionally, Möst and Genoese (2009) investigate the exercise of market power with an agent-based simulation model, that uses detailed German wholesale power market data. In a more recent study Mulder (2015) tests whether the intensity of competition in the Dutch electricity wholesale market changed over the period 2006–2011. The marginal costs per firm are based on actual plant-level data, using engineering-cost estimates. However, the synthetic simulation of dispatch is usually not used by market monitoring units presumably due to missing confidence in appropriate estimation techniques and hence, the lack of empirical work.

While at least the RSI offers a prospective way of measuring market power in wholesale electricity markets there is still little empirical research on its appropri-

was calculated for several European countries. Their results showed substantial market power of huge electricity suppliers in the observed countries.

ate application. Since Sheffrin's initial idea, there have been very few attempts to provide evidence on the quality and appropriate quantification of the RSI. An exemption is the study of London Economics (2007) and subsequent research of the authors Swinand et al. (2010). They show, that market structure, as measured by the RSI, is a significant explanatory factor for markups, even when scarcity and other explanatory variables are included.

One major drawback of the RSI is that it does not fully consider suppliers' incentives to influence the market price. The RSI aims at the degree of pivotality of a supplier and thus, its ability to manipulate the price, but not its incentive to do so. As the RSI does not consider different technologies of electricity production and their related costs it cannot capture these incentives which are due to different technologies of each supplier. The consideration of incentives for abusive strategies constitutes an important advantage of behavioral indicators: these measures are more suitable to capture possible capacity withholding strategies. The dispatch model, for instance, reveals further details about the marginal cost structure of the analyzed coal and gas fired power plants. With the information on the ownership structure of each generating unit one can calculate markups for the total installed capacity of every supplier and then determine incentives to withhold specific generating units. A weakness of dealing with this information is, however, that the results are decisively depending on the accuracy of the estimated cost level. Therefore, a market power index, such as the Lerner index, which is sensitive to the absolute cost level is more vulnerable to possible impreciseness. In fact, this is a serious problem, because market monitoring and even more investigations by competition authorities need feasible and reliable monitoring techniques.

1.3 A method for monitoring capacity withholding incentives

Although a market can be distorted in several ways, generally two strategies are considered as important for abusive practices in power generation (Twomey et al., 2006; Helman, 2006; Biggar, 2011). Both strategies aim at rising prices by reducing supply, such that the cost of the marginal power plant increases: on the one hand this can be achieved by physical withholding of capacity (reducing output), e.g. a supplier temporarily reduces its capacity by claiming a unit is not operational.

On the other hand market power is exercised by financial withholding (raising the price of output), i.e. a supplier raises its bidding price above the marginal cost of a generation unit. Either strategy generates specific cost and price effects depending on the technology mix of the electricity supplier.

Physical withholding behavior restrains capacities from the market for strategic reasons, which results in a reduction of output. Electricity units are not provided, although they could be offered at a price that exceeds the marginal cost. The strategy can be implemented through planned outages or throttling (Twomey et al., 2006). In a normal competitive market situation any capacity that exceeds its short-term marginal costs is expected to be sold in order to achieve a positive contribution margin. In the market for electricity generation, however, this behavior is not always profit-maximizing due to special price generation on the electricity spot market.[8] Abusive practice can be identified if a supplier withholds capacity that he could sell above marginal cost in the short term. This is profitable as the supplier expects that the shortage of supply volume leads to a shift in the merit order. As a result of the price mechanism in electricity wholesales market, the price increases which leads to additional profits due to higher contribution margins for the remaining power plants in the portfolio. However, in order to maximize profits with this strategy, additional profits must exceed losses due to withheld capacity (German Federal Cartel Office, 2011). This mechanism is illustrated in Figure 1.1:

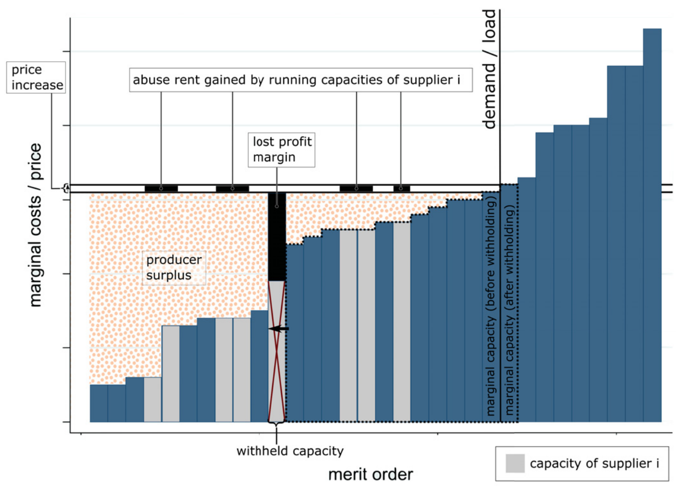


Figure 1.1: Rational calculation for capacity withholding

[8]For theoretical work see for example Crampes and Creti (2005), as well as LeCoq (2002).

Strategic withholding played an important role during the California Energy Crisis in 2001 and was a reason for the severity of this crisis (Kwoka and Sabodash, 2011). One of the best-known empirical studies in the field of capacity withholding incentives is that of Joskow and Kahn (2002), whose goal is to explain the sharp price increase in Californian energy wholesale in 2000 compared to previous years. They first examine whether the increase was caused by exogenous market changes, such as the rise in gas prices, demand or lower imports. In a further step, they investigate whether the increase in cost of emission rights led to higher prices in the wholesale energy sector. In their analysis, however, they conclude that a large part of the increased costs cannot be explained by changed market conditions and emission rights certificates. Finally, they analyze capacity withholding behavior as a reason of the price increase. They distinguish between actual scarcity caused by planned failures or transmission limitations and strategic withholding to increase profits. They limit their analysis to peak load periods in order to determine reasonable quantities by comparing the actually produced capacity with the maximum installed capacity. The maximum installed capacity serves as a proxy for the amount of quantity produced in a competitive benchmark during peak load periods. Joskow and Kahn (2002) show that the behaviour of capacity withholding in the Californian spot market during summer 2000 can at least partly explain the observed price increase. The California Public Utilities Commission also comes to the conclusion that five Californian electricity producers have withheld the capacity of their plants (Joskow and Kahn, 2002).

Prior to Joskow and Kahn (2002), Patrick and Wolak (2001) investigated possible strategies to raise prices in the wholesale energy sector. In their empirical analysis of the UK wholesale electricity market during April 1991 to March 1995, they found that National Power and PowerGen strategically withheld capacity in the market to achieve higher prices.

The pivotality of a supplier, as well as a diversified power plant portfolio increases the probability of capacity withholding. However, the strategy of physical capacity withholding is difficult to identify for regulators as it can be easily declared as necessary maintenance or technical failure. More recently, Fogelberg and Lazarczyk (2014) and Bergler et al. (2017) have found evidence that so-called unplanned power plant outages (defined as strategic failures by the authors) increase with rising prices in the Swedish or German-Austrian-wholesale market. In conclusion, Bergler et al. (2017) recommend extended monitoring by public authorities.

In fact, capacity withholding may be a serious problem in liberalized energy markets and reliable monitoring techniques are needed. Hence, we develop an indicator detecting abusive use of market power in terms of capacity withholding which is more suitable for practical use. It is quite important that the indicator addresses the incentives for capacity withholding on the one hand, but is not sensitive to the absolute level of costs.

1.3.1 The Return on Withholding Capacity Index (RWC)

To measure the incentive for capacity withholding, the costs and benefits of withholding for any supplier have to be examined: the lost profit margin on the one hand and the abuse rent due to a price increase on the other hand. The RWC takes into account that the supplier will induce a higher price if capacity is withheld due to the increasing scarcity (e.g. caused by the use of a more expensive power plant), without this price increase necessarily having to be particularly high. A corresponding unilateral effect can also occur if the provider does not have a pivotal position. Incentives for capacity withholding outside peak load hours also depend in particular on the power plant portfolio, which is considered by the RWC, contrary to other frequently used indicators, such as the RSI. Therefore, the RWC is suitable for assessing the risk of abusive capacity reduction.

Data on hourly produced quantities per plant type and the portfolio structure of each supplier enables us to derive the total capacity each power supplier provides per hour. Hence, multiplying these value with the induced increase of the market price, yields the profits any supplier can gain by withholding a capacity of one MWh. This value can be considered as the incentive for the abusive strategy of withholding physical capacity to increase prices. For interpretational reasons it is helpful to relate this rent to the actual market price. This yields an alternative measure of market power on firms' incentive for abusive behaviour which we label as the Return on Withholding Capacity Index (RWC). It is defined as follows:

$$RWC_{i,t} = \frac{\Delta p \times (runningcapacity_{i,t} - 1)}{marketprice_t}$$

with Δp as the estimated value for the price premium expected by the supplier i for withholding one MWh capacity at time t. For calculation of the RWC, information on suppliers' running capacity is essential though. Therefore, we will discuss the approximation of this figure in more detail in the following section.

The RWC works as a standardized indicator to quantify the incentive of a certain power supplier to withhold capacity. However, to interpret the results of a comprehensive RWC calculation, some aspects have to be considered: the calculated return on withheld capacity has to be compared with the lost profit margin due to reduced production (see Figure 1.1). An incentive for strategic withholding is given if the RWC is higher than the proportional profit margin for withheld capacity. At maximum the proportional profit margin equals one if the withheld capacity has marginal costs of zero.

Thus, the following rule can be applied:

RWC $\geq$ 1 the supplier has a strong incentive to withhold capacity since the lost profit margin is always smaller than the abuse rent gained if the supplier runs other capacities.

RWC $<$ 1 interpretation of this indicator is limited since it can solely provide information on the relative likelihood of strategic withholding (e.g. by intertemporal, inter-market or inter-firm comparison). For further interpretation of an RWC below one, extended in-depth data about the hourly profit margins of generation units would be necessary.

1.3.2 Data requirements for the index calculation

The purpose of our market power index is to provide an instrument for a wide range of users, such as monitoring units. The index is designed to provide information about the incentives of suppliers to withhold capacity with reasonable effort. Thus, it is important that data for the calculation of the RWC is publicly available.[9] Information to calculate the running capacity of a provider, as well as price and demand (load) data to estimate the price premium Δp are necessarily required.

For EU countries, the ENTSO-E transparency platform offers detailed information on production and load data.[10] To determine the incremental price increase Δp, data on the demand for electricity from price-setting power plants in the bidding zone under examination is required. Since we use day-ahead prices for the later estimation, it is further suitable to use day-ahead forecasts for the load and production

[9] While most of the data is available for free, exceptions are the commercial ORBIS database of Bureau van Dijk and information on price data.

[10] See `https://transparency.entsoe.eu/` for hourly load (including day-ahead-forecast) and production data.

values accordingly. Total demand is available as (forecasted) "total load" on the ENTSO-E platform. To calculate the incremental price increase, however, only demand for the production volume of power plants that produce electricity dependent on the market price is considered. This is the case if the output of the power plants can be controlled by the supplier and power plants are also working at marginal cost above zero. Power plants that do not meet these criteria are usually volatile renewable energy plants that produce electricity depending on meteorological factors. These plants feed into the grid independent of the current market price. As a result, they shift the residual demand for electricity from power plants that set prices.[11] While the production volume of ordinary power plants is weakly or strongly positively correlated with the price, inflexible power plants reduce the residual demand for price-setting plants and are negatively correlated with the price.[12] Hence, the appropriate residual demand value results from the difference of (forecasted) "total load" and (forecasted) production volume of inflexible power plants. In our case solar, wind power and run-of-the-river hydroelectricity plants in the German-Austrian price zone.[13]

There is often no precise data for the specific power plants of a supplier that produce at a certain point in time. However, the running capacity can be approximated quite well from data which is publicly available. For this purpose, production data from ENTSO-E differentiated by power plant type can be used. Within a power plant technology, marginal costs usually differ only slightly. By allocating data on hourly produced quantities per plant type to the corresponding supplier's market share of the existing capacities, it is possible to deduce the running capacity of each supplier.[14] Therefore, the market share of suppliers, differentiated by type of

[11]One issue in this calculation of the quantity in demand could be that the value for "total load" includes electricity trading (limited by the transmission volume) with suppliers and buyers from neighboring bidding zones. However, the non-price-setting power plants in these neighboring markets cannot simply be adjusted in value for the total load. The resulting distortion can be critical if smaller markets are examined in which supply by neighboring countries accounts for a considerable part of supply in the bidding zone under consideration. Thus, in these cases it can be an advantage to approximate the demand quantity instead of the adjusted "total load" by the accumulated supply of the price-setting power plants in the bidding zone under consideration.

[12]Significant negative price-generation correlation values for 2016 are: solar (-0.11), wind power (-0.41) and run-of-the-river hydroelectricity (-0.14).

[13]In consequence, subtracting non-controllable renewable energy plants leads to a substantially higher explanatory power on the effect on prices. This effect is shown by a R2 of 0.745 with adjusted load compared to a R2 of 0.325 in case of non-adjusted load.

[14]Capacity of production facilities should not be taken into account if they feed in at fixed subsidized prices and therefore do not (or only marginally) benefit from an increase in the market price. In Germany and Austria, this accounts for wind power and solar installations that receive

generation, has to be determined.

To do so, we use data for the installed capacity in Germany from the periodical power plant survey of the German Federal Network Agency.[15] The survey provides detailed information regarding e.g. the normal maximum operating capacity (MW), energy source (type of power plant), supplier (owner) company etc. for all German generation units with a net nominal output of at least 10 MW.[16] According to the Federal Network Agency the survey covers more than 95% of the total installed capacity produced by conventional power plants in Germany or rather in the German control areas.[17] In total, we observe 875 generation units with a total installed capacity of 110,346 MW. In order to determine the ownership structure for these generation units, or rather the total installed capacity, survey data on the respective owner companies was merged with Bureau van Dijk's ORBIS database. ORBIS provides information about the global ultimate owner for most of our sample firms. Firms which were not covered by ORBIS were manually researched by checking company websites and company reports. This enables us to identify which firms in our sample data belong to one of the four big generators in Germany. As a result, we obtain the market share in the production capacities of the individual generator companies, both for the overall market and for the individual types of power plant. The added production volumes of the individual power plant types from the ENTSO-E transparency platform multiplied by the respective market shares leads to the approximated running capacity of each supplier.

Moreover, energy carrier prices, such as gas, coal and carbon oxide were used as control variables for estimating the supply curve. Gas prices and carbon oxide as well as coal prices are provided by Energate. Spot market prices (day-ahead prices) for electricity can be observed on an hourly basis e.g. on the ENTSO-E transparency platform.

---

a feed-in tariff or a market premium.

[15] Version 31.03.2017. An updated version of the survey can be found here:`https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/Erzeugungskapazitaeten/Kraftwerksliste/Kraftwerksliste_2016.html`

[16] Note, that the survey also includes industrial generation units. However, they react differently on market signals, such as energy prices since they are operating as required to meet demand of the respective industrial company. Hence, all industrial power plants are discarded from our analysis.

[17] There are generation units not located in Germany, but in the border region of Austria, Switzerland, France and Luxembourg which feed power into the German grid. Hence, they are regarded as part of the German control area.

1.3.3 Estimation of the incremental price increase Δp

For the estimation of the price premium, i.e. the incremental price increase Δp, expected by the supplier for withholding one MWh capacity, we make use of the relationship between residual load and the price level. This relationship is used to estimate the supply function.

The general shape of the supply function depends on the market under consideration and has to be adjusted accordingly. In the German-Austrian bidding zone, the technological merit order and the fluctuations of residual load support a cubic shape. A cubic relationship is assumed as a result of the typical mixture of different power plant technologies, which differ in their fixed and marginal costs (Crew et al., 1995). Base load power plants (e.g. lignite, nuclear) have low marginal costs, but a limited flexibility. For this reason, shutting down base load power plants is only profitable in case of sharp price jumps to negative prices. Meanwhile, peak-load power plants, such as gas-fired ones, are only used during peak demand, have high marginal costs and have to finance these with short running times. This suggests that price jumps are stronger with fluctuations in base load and peak load ranges than with medium loads. The German-Austrian market is characterised by precisely these technological characteristics, which are reflected in price peaks and negative prices. The cubic shape of the supply curve is also confirmed by Figure 1.2 which shows the relationship between residual load and market price in 2016.

Since we intend to estimate the price increase expected by suppliers for the withdrawal of one unit of capacity, the appropriate investigation period has to be adjusted accordingly. A study of at least several months would be more appropriate if a period that is too short may give an incomplete picture of the different seasonal load situations. A period, that is too long, on the other hand, does not consider structural changes on the energy markets (e.g. construction or dismantling of power plants) and thus, attenuates the correlation of load and price. Therefore, for our study of the German-Austrian market, we consider the calendar year to be a suitable period of investigation. Nevertheless, in case of other markets or study years examined, the supply conditions may also fluctuate considerably during a calendar year, requiring the model to be adjusted in such cases accordingly.[18]

[18]Fluctuations, that can blur the relationship between price and residual load during a calendar year are, for example, the seasonal nature of certain price-setting power plant types (e.g. downtimes in the event of overheating in summer months), large-scale power plant shutdowns during the year or significant changes in fuel prices. Instead of dividing the investigation period into periods of less than one year, seasonal dummy variables (e.g. monthly) could also be considered in such cases.

We are able to show, that the calculated residual demand or load explains electricity prices to a great extent. We focus on the period of 2016, for which running capacity was measured likewise and which serves as the basis for determining the RWC in Section 1.4.

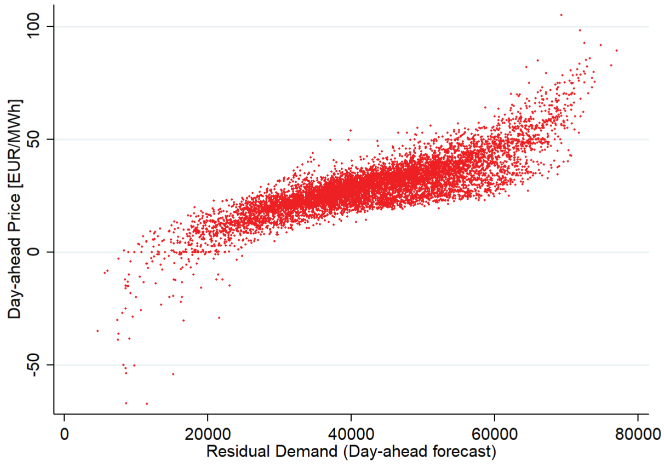


Figure 1.2: Relationship of residual load and market price in 2016

To investigate the incremental price increase Δp, we further quantify this effect as we estimate a regression model. A key point in our analysis is the estimation method, as we make use of wholesale electricity price as the dependent variable and (residual) load as a regressor. One could argue that such a model suffers from an endogeneity bias due to reverse causality since demand and price are usually simultaneously determined. However, this may not be the case here, as demand in wholesale energy markets is highly price inelastic. Our main focus lies on the real-time price elasticity of demand,[19] since the theory of capacity withholding is an unpredictable event for consumers. This means, that one can only react to the withholding price effect in real time. But a large proportion of electricity consumers are generally unaware of the real-time fluctuations in wholesale electricity prices (Knaut and Paulus, 2016; German Federal Cartel Office, 2011; Gelabert et al., 2011). Hence, they have no incentive to adjust their consumption in real-time, but may only react on price changes in the long run. However, elasticity of demand plays a decisive

However, it should be pointed out that using such dummy variables may lead to the disappearance of seasonal effects from the price/load relationship. The price premium expected by providers during capacity holdbacks may therefore deviate from the estimate in these periods.

[19]Real-time elasticity is defined as the price elasticity of demand on an hourly basis.

role, particularly with regard to the increased amount of volatile renewable energy which poses new challenges to electricity grids.

Due to the huge amount of data required, there are only a few studies that deal with the analyses of real-time elasticities of overall demand. Nevertheless, they come to similar conclusions. Lijesen (2007) measures the real-time elasticity of demand for electricity in the Netherlands for 2003. To address possible endogenity due to the simultaneity of demand and supply, lagged prices are used as instrument variables. Lijesen (2007) shows a price elasticity of -0.0014 in the linear model specification and -0.0043 in the log-linear model specification. He concludes that according to his study, consumers are not prepared to react to real-time prices because the elasticity of demand is low. Knaut and Paulus (2016) estimate a similar model for the German energy market in 2015 and come to similar results. They show that demand reacts only slightly to price changes. Elasticity varies between -0.004 and -0.006. These results are in line with Lijesen (2007). Genc (2014) chooses a different approach. He uses detailed power generation and market data from 2007 and 2008 to investigate market power in the wholesale energy sector in Ontario, Canada. He estimates the responsiveness of demand only for wholesale customers. However, the hourly elasticities are rather small. They lie within an interval -0.144 and -0.013 for 2007 and within -0.083 and -0.019 for 2008.

Thus, we conclude that real-time elasticity of demand is rather small in all these cases. If, in a specific wholesale market for which the RWC is calculated, an (almost) completely inelastic demand can be assumed, there is no problem with endogeneity. In the following, we use data for the German-Austria-wholesale market and perform both, simple OLS as well as IV regressions.

1.3.4 OLS estimation approach

In a first step we use the wholesale electricity price (day-ahead price) as the dependent variable and residual demand (day-ahead forecast) as a regressor to perform OLS estimations.[20] As we have already emphasized in the previous chapters residual demand L_{resid} enters the equation as a cubic function. Furthermore, we use control variables for input prices, namely fuel prices for coal, gas and CO2, to control for supply shocks. Augmented Dickey-Fuller test reveal that neither the wholesale elec-

[20] We took the day-ahead price, as it is the most important spot market price and all hourly prices for a day can be determined at the same time. Accordingly, we also use the day-ahead forecast for the demand data.

tricity price nor residual demand and our control variables show a unit root. That is, they are stationary and consequently enter the regression model in levels. Thus, the following equation is estimated:

$$p_t = \beta_0 + \beta_1 * Lresid_t + \beta_2 * Lresid_t^2 + \beta_3 * Lresid_t^3 + \beta_4 * pcoal_t + \beta_5 * pgas_t + \beta_6 * pCO2_t + \epsilon_t \tag{1.1}$$

By deriving the estimated model with respect to $Lresid_t$, it is possible to calculate the price increase Δp that occurs when load is changed by one unit.

Our results in Table 1.1 show that in the full sample residual load (i.e. residual demand), as well as changes in gas and coal prices have a significant effect on the wholesale electricity price, while changes in CO2 price show no significant effect. This result is also confirmed if we split our sample into peak and off-peak hours. From a theoretical point of view, commodity prices can have an impact on electricity prices as they influence the marginal costs of power plants. Due to the technology mix in a merit order, the influence on the price depends on how much a certain technology is used at a certain load phase and how much the marginal power plant provides. Therefore, the influence of certain fuel prices on the market price is not linear, but depends on load.[21] Consideration of fuel prices may distort the estimate, as the linear influence in the model will have an effect even if this is not actually the case due to the load situation.[22] Hence, in a further specification we simplify our model by omitting these control variables. As Table 1.1 shows omitting all control variables only led to a small decline in the goodness of fit measure R2 and - even more important - coefficient values of residual load. Since our goal is to develop a very simple measure of market power, that can be calculated and applied easily, this simplified model provides a very sound foundation. Although using the coefficient value of residual load of the simplified model may come at the cost of a small bias, it has the huge advantage that considerable less data is needed. Besides, omitting control variables allows for defining a standardized procedure for calculating the RWC. Thus, we propose using the result of the simplified regression model as the foundation of the RWC and build on this within the following analysis. Furthermpore, the simplified regression model yields to a rather high adjusted R2

[21]The correlation between the residual load and the coal and gas prices amounts to 0.23 and is statistically significant in each case.

[22]If there is a huge differential of fuel-prices during the calender year, it could be preferable to catch the effect by monthly dummy variables.

of 0.75.[23]

	Full sample	Peak hours	Off-Peak hours	Full sample
	I	II	III	IV
L_resid	0.0056***	0.0057***	0.0055***	0.0055***
	(0.0004)	(0.0001)	(0.0001)	(0.0004)
L_resid2	<-0.0001***	<-0.0001***	<-0.0001***	<-0.0001***
	(<-0.0001)	(<-0.0001)	(<-0.0001)	(<-0.0001)
L_resid3	<0.0001***	<0.0001***	<0.0001***	<0.0001***
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
pcoal	0.1670***	0.1740***	0.1700***	
	(0.0074)	(0.0130)	(0.0100)	
pgas	0.4330***	0.6830***	0.2150***	
	(0.0482)	(0.0800)	(0.0617)	
pCO2	-0.0863	0.1140	-0.1880	
	(0.0831)	(0.1310)	(0.1000)	
_cons	-80.18***	-85.43***	-76.47***	-63.21***
	(4.8530)	(1.9310)	(1.7050)	(4.7730)
Observations	8,687	4,345	4,342	8,687
R^2	0.800	0.804	0.785	0.745

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.1: Regression result of spot price on load

All our results show a highly significant relationship between residual load and the spot market price. If we take the first derivative of equation 1.1 with respect to $Lresid_t$ we receive

$$\frac{\delta p_t}{\delta Lresid_t} = \beta_1 + 2\beta_2 Lresid_t + 3\beta_3 Lresid_t^2 = \Delta p \tag{1.2}$$

Hence, by inserting the residual load we are able to calculate Δp, namely the price increase if residual load is raised by one MWh or rather the price increase if one MWh of capacity is withheld by the supplier, for each each hour t of the period

[23] As an robustness check, we show that this strong relationship is also present in 2017 (adjusted R2 of 0.87) and in 2015 (adjusted R2 of 0.78). This is also shown in the graphs presented in the appendix section 1.6.3.

	Mean	Std.Dev.	Minimum	Maximum	Observations
I	0.0839	0.0510	0.0457	0.4708	8,687
II	0.0872	0.0557	0.0431	0.4791	4,345
III	0.0752	0.0443	0.0437	0.3906	4,342
IV	0.0899	0.0517	0.0511	0.4637	8,687

in eurocent

Table 1.2: Summary statistics Δp

examined. Table 1.2 shows the summary statistic of Δp.

It can be seen that in the full sample withholding of one MWh unit leads to a price increase of 0.084 eurocent, on average. At first glance, this value may seem rather small, but it in fact, it is of great importance if one recalls, that, for example, RWE has a plant portfolio of 46,411 MW in Germany.[24]

The results of the subsamples reveal that another unit of residual load leads to an average price increase of 0.087 eurocent during peak hours and to 0.075 eurocent during off-peak hours. However, a Wald test on the difference of the coefficients of residual load during peak and off-peak hours shows that the coefficients are not significantly different from each other making the full sample model the preferred specification. Tabel 1.2 also shows that, if we use the simplified model, withholding one unit of capacity leads to an average price increase of 0.090 eurocent.

1.3.5 Instrument variable approach

As mentioned before, depending on the specific situation in a market, the OLS model may suffer from an endogeneity bias. To address this issue, we estimate an instrument variable approach in the following. We use lagged residual load as one instrumental variable. To ensure that the instrument captures the same time of the day as in the OLS model, residual load should be lagged by 24 hours. However, since demand for electricity may differ between weekdays and weekends, residual load is not lagged by 24, but by (7*24) 168 hours. Intuitively, the dynamics of the energy market ensure the exogenity of our instrument, as it seems reasonable that today's EEX spot market price is not directly dependent on demand decisions from the previous week.[25]

[24]Own calculations based on the periodical power plant survey of the German Federal Network Agency (Version 31.03.2017) linked with ORBIS ownership Data.

[25]This instrument variable approach with lagged instruments is also in line with the estimation of Lijesen (2007) who uses lagged prices to estimate the real-time elasticity of demand.

Since the amount of electricity produced by conventional power plants is highly dependent on the amount of renewable electricity, we use the production values of renewable electricity by wind, solar and run-of-the-river hydroelectricity as another instrument. One may argue, that these renewable technologies are determinants for the prices on the supply side rather than on the demand side, but this channel is rather indirect due to an intersection of lower remaining demand for conventional electricity and the merit order. In fact, wind and solar power plants operate at zero marginal cost. Their electricity is fed into the grid, regardless of the current market price. Their production quantity is determined based on exogenous weather conditions. In addition, wind and solar power are subsidized by fixed prices in many countries.[26] Therefore, the production of wind and solar power can be considered as an exogenous parameter, that influences the residual demand for electricity conventional power plants. Performing a Stock and Yogo test leads to partial F-values of the instrumental variables of 1564.10, 2946.3, 2156.6.[27] According to Staiger and Stock (1997) a partial F-value of the instrumental variable in the first stage regression should exceed the value of ten. Hence, we can conclude that the instruments are relevant and the IV-regressions will not suffer from a possible weak instrument bias. Furthermore, we test whether our instruments are exogenous employing the Sargan test. The statistic does not reject the null hypothesis of validity of instruments (Sargan $p = 0.8753$).

Hence, the model to be estimated can be written as follows:

$$p_t^{iv} = \beta_0+\beta_1*\widehat{Lresid}_t*\beta_2*\widehat{Lresid}_t^2+\beta_3*\widehat{Lresid}_t^3+\beta_4*pcoal_t+\beta_5*pgas_t+\beta_6*pCO2_t+\epsilon \quad (1.3)$$

As in the OLS-model this econometric framework enables us to determine Δp. We also conducted IV-regression without the control variables on input prices in order to define a standardized procedure to calculate the RWC. Results of the estimations are displayed in Table 1.3.

[26]Since the German Renewable Energy Act (EEG) electricity from renewable sources has dispatch priority leading to a decreasing demand for conventional power plants.

[27]Detailed results can be found in the Appendix in section 1.6.1.

	Full sample	Full sample
	I	II
L_resid	0.0056***	0.0063***
	(0.0004)	(0.0004)
L_resid2	<-0.0001***	<-0.0001***
	(<0.0001)	(<0.0001)
L_resid3	<0.0001***	<0.0001***
	(<0.0001)	(<0.0001)
pcoal	0.1740***	
	(0.0086)	
pgas	0.4600***	
	(0.0525)	
pCO2	-0.0465	
	(0.0864)	
_cons	-79.56***	-70.92***
	(4.400)	(4.927)
Observations	8,687	8,687

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.3: Regression results IV-estimation

All our results of the IV-regressions show a highly significant relationship between residual load and the spot market price. This is in line with our results obtained by the OLS-regressions in the previous chapter. Table 1.4 shows the summary statistics of Δp, i.e. the price increase if residual load is raised by one MWh. It can be seen, that withholding of one MWh unit leads to an average price increase of 0.079 eurocent. In the model without the input price control variables Δp amounts to 0.090 eurocent, on average and thus, confirms ours results obtained with the OLS model.[28]

[28]The similarity of the estimated relationship between OLS and IV estimation is also clearly shown by the graphical illustration in Section 1.6.3. Furthermore, as a robustness check we show that our model is robust by estimating the proposed model for the years 2015 and 2017. Results can be found in section 1.6.3

	Mean	Std.Dev.	Minimum	Maximum	Observations
I	0.0786	0.0513	0.0401	0.4690	8,687
II	0.0891	0.0629	0.0418	0.5266	8,687

in eurocent

Table 1.4: Summary statistics first derivative - IV estimates

Finally we perform a Hausman-Wu-test ($pvalue = 0.0501$) which supports our assumption that we do not face an endogeneity problem in this case. Hence, the Hausmann's test null hypotheses that OLS and IV lead to the same estimates cannot be rejected. Furthermore, we have conducted several robustness checks with alternative instruments, which lead to the same conclusion (see section 1.6.2). Intuitively, this can be explained by the low real-time elasticity of demand as also shown by Graf and Wozabal (2013), as well as Lijesen (2007). However, higher elasticity is generally observed in peak load periods (Patrick and Wolak (2001)) which is quite essentiell for the estimation of the RWC since withholding incentives may be even higher at the steep part of the merit order. It is also not clear that real-time elasticity of demand will remain low in the future. For example, with increasing digitization, consumer devices may automatically control their load according to the price level. Thus, demand elasticity would further increase which would bias OLS estimates. It is therefore particular important, that the RWC index can be calculated for high demand elasticity as well. Hence, we suggest to use the IV estimator rather than the OLS-estimator for market monitoring, as the results are more reliable for the whole range of hours and the volume of demand.

1.4 Results and application proposal for market monitoring units

By using estimation techniques as described in the previous chapter we are able to calculate RWC values for the German-Austrian market for 2016. In total, we receive 8,687 RWC values (i.e. for (almost) every hour of the year 2016) for the four largest providers in Germany. The values indicate how high the incentives of the individual providers were to withhold capacities on the spot market in the respective trading hour.

In order to interpret the values with regard to each suppliers' market power, it is necessary to adjust the calculated RWC values for off-peak periods. If load is low, the market price is low and the incremental price increase (Δp) is high. Hence, the RWC may show relatively strong incentives for withholding capacity in this situation. However, abusive capacity withholding is rather unlikely at very low market prices. In fact, baseload power plants are operated in low load phases at (even negative) prices, that do not cover short-term variable costs, because technical reasons prevent these baseload power plants from being completely shut down at short notice. Since the ability to control output represents a technological limitation, that can hardly be verified, the market price can be used as an indicator. We therefore clear the measured values for the RWC by those hours in the lower load range in which the price is very low and the possible price jumps are at the same time particularly high. To do this, we calculate the average incremental price increase Δp for the load values below the turning point of our function. For 2016 this gradient value corresponds to an estimated market price of 21.67 euros. We presume, that in cases where the market price is actually lower, electricity is potentially produced although the short-term variable costs of production are not covered. Although the hours below this threshold are taken into account for the subsequent calculation of the percentiles, the RWC value is set to zero in these cases. This procedure ensures that the number of hours with the highest RWC values, whose limit is indicated by the percentile value, does not fluctuate significantly. Therefore, these values remain comparable for different tests. Finally, there remain 6.858 out of 8.784 hours for which the RWC can be interpreted.

In order to assess market power of individual providers for a specific year, we suggest to consider the upper 90 or 95 percentile values of the RWC. These values indicate how high the incentives to withhold capacity are for the 5 or 10 percent of hours with the highest values. Using these thresholds is in line with the monopolist test (SSNIP test) from European antitrust law. The SSNIP test asks the question whether a hypothetical monopolist would be able to profitably increase prices by 5 or 10 percent.[29] Table 1.5 shows the calculation of the 90 and 95 percentile values as well as the mean for the OLS and IV estimates. Moreover, Figure 1.3 shows seasonal variations of the RWC values received by IV-estimations.

[29]Note, however, that this is quite a general comparison since the RWC only measures the incentive for capacity withholding and does not make a statement about a price increase.

	Mean		90% Percentil		95% Percentil		Hours> 1	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Company 1	0.3103	0.2968	0.4551	0 .4719	0.5455	0.5851	4	12
Company 2	0.1457	0.1396	0.2154	0.2240	0.2590	0.2775	0	0
Company 3	0.1946	0.1860	0.2868	0.2958	0.3429	0.3661	0	0
Company 4	0.1466	0.1403	0.2176	.2228	0.2609	0.2785	0	0

Table 1.5: Mean and fringe values for the RWC in 2016

The results of the OLS and IV estimates for the incremental price increase are similar on average as well as for the percentile values. We conclude from this that the elasticity of demand in the market under examination is low. This fact as well as our tests conducted in the previous section lead us to the conclusion that the endogeneity problem in the market under consideration is not that serious. Nevertheless, we still suggest to use the IV-estimator for market monitoring, as the results are more reliable if demand elasticity is elastic. One advantage of the presented IV-estimator is that the required data (lagged residual load and renewable infeed) are usually quickly accessible through public sources, such as ENTSO-E. Thus, RWC values can be very well determined and standardized using the simplified form of our model (with the residual load as the only regressor) by market monitoring units. The standardization of the model also simplifies intertemporal comparisons of the calculated RWC values making changes of withholding incentives visible in market monitoring.

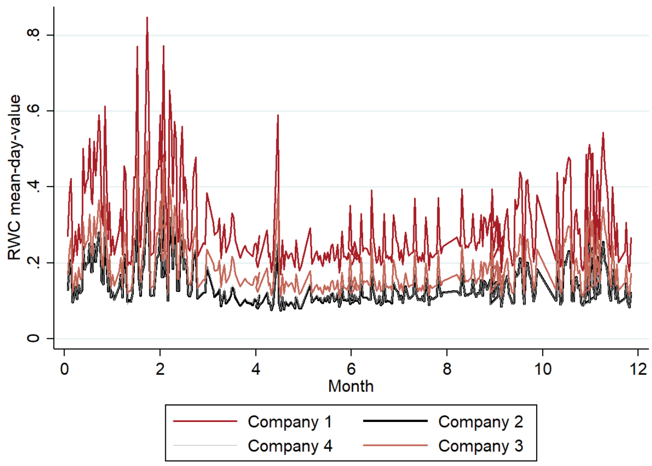


Figure 1.3: RWC results for 2016 monthly pattern

1.5 Conclusion

Withholding production capacity is a widely discussed economic problem of wholesale electricity markets. Antitrust authorities and market monitoring units are engaged in this issue since the liberalisation of these markets. We have developed and tested the Return on Withholding Capacity Index to analyse the risk of abusing market power by capacity withholding. The index should be available to market monitoring units, but also to antitrust authorities and academics as a new screening instrument. The RWC is also suitable as a complement to the Residual Supply Index, which, unlike the RWC, only reflects a certain (but important) type of market power by individual suppliers.

The proposed RWC index is based on the assessment of a power plant operator prior to its decision to hold back capacity. The power supplier maximises its profit by withholding capacity if the lost contribution margins of a power plant (or a unit of capacity) held back is at least offset by the triggered price increase and the resulting contribution margins at other power plants. An RWC value of one or higher indicates that the capacity withholding might be profitable for a supplier, regardless of which power plant is held back. To interpret values below one in terms of absolute market power, further analyzes of individual costs are necessary. However, these results show the relative incentives for capacity withholding and are suitable, for example, for intertemporal comparisons of market situations.

An important requirement for the construction of the RWC is the calculation with reasonable effort from available data. Using an application example with data for the day-ahead market in the German-Austrian bidding zone in 2016, it is shown how the calculation can be optimized and standardized. An important component of the index calculation is the estimation of the price increase in the case of holding back one unit of capacity. In particular, market data on constant day-ahead prices on the one hand and hourly load forecasts on the other hand were used. The price increase was determined by using both, OLS and IV estimation models. In order to address the possible issue of endogeneity in case of an elastic demand function, we recommend that monitoring units determine the incremental price increase preferably with the IV-estimator. We also suggest to evaluate the data on the basis of a 90 and 95 percentile, on which the assumption of market power of individual suppliers could be based.

Bibliography

Angrist, J.D. and A.B. Krueger, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 2001, *15* (4), 69–85.

Asgari, M.H. and H. Monsef, "Market power analysis for the Iranian electricity market," *Energy Policy*, 2010, *38* (10), 5582–5599.

Bergler, J., S. Heim, and K. Hüschelrath, "Strategic capacity withholding through failures in the German-Austrian electricity market," *Energy Policy*, 2017, *102*, 210–221.

Biggar, D., "The theory and practice of the exercise of market power in the Australian NEM, Australian Competition and Consumer Commission, Melbourne.," 2011.

CAISO, "Report on the Appropriateness of the Three Pivotal Supplier Test and Alternative Competitive Screens.," 2013.

Cartea, Á. and M.G. Figueroa, "Pricing in Electricity Markets: A Mean Reverting Jump Diffusion Model with Seasonality," *Applied Mathematical Finance*, 2005, *12* (4), 313–335.

Chang, Y., "The New Electricity Market of Singapore: Regulatory framework, market power and competition," *Energy Policy*, 2007, *35* (1), 403–412.

Crampes, C. and A. Creti, "Capacity Competition in Electricity Markets. Economia delle fonti di energia e dell'ambiente 2,," 2005, pp. 59–83.

Crew, Michael A., C.S. Fernando, and P.R. Kleindorfer, "The theory of peak-load pricing: A survey," *Journal of Regulatory Economics*, 1995, *8* (3), 215–248.

DG Competition, "Report on energy sector inquiry, SEC(2006)1724.," 2007.

European Commission, "Energy Markets in den European Union in 2011, Publications Office of the European Union.," 2012.

_ , "Final Report of the Sector Inquiry on Capacity Mechanisms, COM(2016) 752 final, Nov," 2016.

Fogelberg, S. and E. Lazarczyk, "Strategic Withholding through Production Failures," *SSRN Electronic Journal*, 2014.

Frontier Economics, "Market concentration in the German engery wholesale market.," 2010.

Gelabert, L., X. Labandeira, and P. Linares, "An ex-post analysis of the effect of renewables and cogeneration on Spanish electricity prices," *Energy Economics*, 2011, *33*, S59–S65.

Genc, T., "Market Power Indices and Wholesale Price Elasticity of Electricity Demand.," *Technical report.*, 2014.

German Federal Cartel Office, "Sector enquiry wholesale energy sector, B10-9/09.," 2011.

Graf, C. and D. Wozabal, "Measuring competitiveness of the EPEX spot market for electricity," *Energy Policy*, 2013, *62*, 948–958.

Helman, U., "Market power monitoring and mitigation in the US wholesale power markets," *Energy, Vol. 31, issue 6, 877-904*, 2006.

Joskow, P. and E. Kahn, "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000.," *The Energy Journal, Vol 23, Nr. 4.*, 2002, pp. 1–35.

Kamiński, J., "The development of market power in the Polish power generation sector: A 10-year perspective," *Energy Policy*, 2012, *42*, 136–147.

Knaut, A. and S. Paulus, "Hourly Price Elasticity Pattern of Electricity Demand in the German Day-ahead Market," *. EWI Working Paper, Nr. 16/07.*, 2016.

Kwoka, J. and V. Sabodash, "Price Spikes in Energy Markets: "Business by Usual Methods" or Strategic Withholding?," *Review of Industrial Organization*, 2011, *38* (3), 285–310.

Lang, C., "Marktmacht und Marktmachtmessung im deutschen Großhandelsmarkt für Strom, Wiesbaden.," 2007.

LeCoq, C., "Strategic use of available capacity in the electricity spot market. SSE/EFI Working Paper Series in Economics and Finance 496.," 2002.

Lijesen, M.G., "The real-time price elasticity of electricity," *Energy Economics*, 2007, *29* (2), 249–258.

London Economics, "Structure and Performance of Six European Wholesale Electricity Markets in 2003, 2004 and 2005, http://www.londecon.co.uk/le/publications/recent_reports.shtml," 2007.

Monopolies Commission, "Special Report 65: Energy 2013," 2013.

_ , "Special Report 71: Energy 2015: A competitive market design for the Energiewende," 2015.

_ , "Special Report 77: Energy: Avoid patchwork, targeted action required," 2017.

Möst, D. and M. Genoese, "Market power in the German wholesale electricity market," *Journal of Energy Markets 2*, 2009, pp. 47–74.

Mulder, M., "Competition in the Dutch Electricity Market: An Empirical Analysis over 2006-2011," *The Energy Journal*, 2015, *36* (2).

_ **and L. Schoonbeek**, "Decomposing changes in competition in the Dutch electricity market through the residual supply index," *Energy Economics*, 2013, *39*, 100–107.

Newberry, D., "Predicting Market Power in Wholesale Electricity Markets. EUI Working Papers, RSCAS 2009/03.," 2009.

Patrick, R. and F. Wolak, *Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices*, Cambridge, MA: National Bureau of Economic Research, 2001.

Sheffrin, A., "Predicting Market Power Using the Residual Supply Index. FERC Market Monitoring Workshop, http://www.caiso.com/docs/2002/12/05/2002120508555221628.pdf.," 2002.

Swinand, G., D. Scully, S. Ffoulkes, and B. Kessler, "Modeling EU Electricity Market Competition Using the Residual Supply Index," *The Electricity Journal*, 2010, *23* (9), 41–50.

Twomey, P., R.J. Green, K. Neuhoff, and D. Newbery, "A Review of the Monitoring of Market Power The Possible Roles of TSOs in Monitoring for Market Power Issues in Congested Transmission Systems," 2006.

1.6 Appendix

1.6.1 Robustness check IV results

Due to the cubic shape of the residual demand ($Lresid$) the first stage of the IV-model can be written as:

$$Lresid_t = \gamma_0 + \gamma_1 * Lresid_{t-168} + \gamma_2 * Lresid^2_{t-168} + \gamma_3 * Lresid^3_{t-168} + \gamma_4 * unflex_t + \gamma_5 * z + \mu \tag{1.4}$$

$$Lresid^2_t = \gamma_0 + \gamma_1 * Lresid_{t-168} + \gamma_2 * Lresid^2_{t-168} + \gamma_3 * Lresid^3_{t-168} + \gamma_4 * unflex_t + \gamma_5 * z + \mu \tag{1.5}$$

$$Lresid^3_t = \gamma_0 + \gamma_1 * Lresid_{t-168} + \gamma_2 * Lresid^2_{t-168} + \gamma_3 * Lresid^3_{t-168} + \gamma_4 * unflex_t + \gamma_5 * z + \mu \tag{1.6}$$

where $Lresid_{t-168}$ is the residual load lagged by 168 hour and $unflex$ is the forecasted load of unflexible power plants e.g. solar and wind generation.

	(1)	(2)	(3)
	L_resid	L_resid2	L_resid3
da_gen_sum_unflexibel	-0.6450***	-52924.8***	-3.54495e+09***
	(0.0100)	(803.6)	(59473659.3)
lag168L_resid	-0.8850***	-117181.3***	-9.41889e+09***
	(0.1400)	(10511.3)	(775585064.9)
lag168L_resid2	<0.0001***	3.5140***	253491.3***
	(<0.0001)	(0.2690)	(20532.8)
lag168L_resid3	<-0.0001***	<-0.0001***	-1.490***
	(<0.0001)	(<0.0001)	(0.1700)
_cons	50153.8***	3.24295e+09***	2.10003e+14***
	(1856.2)	(133734979.3)	(9.53591e+12)
Observations	8,687	8,687	8,687
F	1564.10	2946.3	2156.6
R^2	0.581	0.576	0.551

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A1: Stock and Yogo weak instruments test

1.6.2 Robustness check with other instruments

The use of lagged values as an instrument is not always seen as unproblematic in the economic literature. It is argued that in terms of prices and quantities there may be a strong correlation of the realizations of the variables over time (Angrist and Krueger, 2001). This is often the case in markets where long-term negotiations take place, as certain delivery conditions, such as in food retailing, are clearly determined on the basis of the previous period. By contrast, futures trading at the dynamic spot market for energy quantity takes place on an hourly basis. On this basis, a lag of 168 hours no longer appears to be directly dependent on the offer and an exogeneity seems plausible. In order to meet the criticism of Angrist and Krueger (2001), however, a modified specification is also tested to exclude a potential misinterpretation of the test results due to possible endogenous instruments. The specification is defined as follows:

$$Lresid_t = \gamma_0 + \gamma_1 * unflex_t + \gamma_2 * publichol_t + \gamma_3 * precipitation_t + \gamma_4 * Monday + \gamma_5 * Friday + \gamma_6 * z + \mu$$
$$Lresid_t^2 = \gamma_0 + \gamma_1 * unflex_t + \gamma_2 * publichol_t + \gamma_3 * precipitation_t + \gamma_4 * Monday + \gamma_5 * Friday + \gamma_6 * z + \mu$$
$$Lresid_t^3 = \gamma_0 + \gamma_1 * unflex_t + \gamma_2 * publichol_t + \gamma_3 * precipitation_t + \gamma_4 * Monday + \gamma_5 * Friday + \gamma_6 * z + \mu$$

and second stage as

$$p_t = \beta_0 + \beta_1 * \widehat{Lresid_t} * \beta_2 * \widehat{Lresid_t^2} + \beta_3 * \widehat{Lresid_t^3} + \beta_4 * pcoal_t + \beta_5 * pgas_t + \beta_6 * pCO2_t + \epsilon$$

where $unflex$ is the forecasted produced load of unflexible power plants e.g. solar and wind generation, *publichol* is a dummy variable which takes the value of one on German national public holidays and *precipitation* is the weighted daily amount of precipitation. [30]

Instruments are relevant (F-value: 586.74, 499.68, 423.73) and exogenouson a 10% level (Sargan test: $p = 0.2894$). By applying the econometric framework we are able to estimate the elasticity of supply and perform a Durbin-Wu-Hausman Test (Model I $p = 0.6831$ and model II $p = 0.4640$)[31]. Results of the estimation are displayed in Table A2. Summary statistics of the first derivative can be found in A3.

[30] To capture precipitation for whole Germany, we weight precipitation with the number of inhabitants of all large German cities with more than 100,000 inhabitants which are assigned to the closest weather station.

[31] We also conducted the IV regression without taking supply inputs into account. In order for defining a standard procedure, as explained in the previous section.

	Full sample	Full sample
	I	II
L_resid	0.0076*	0.0082*
	(0.0030)	(0.0038)
L_resid2	$<$-0.0001*	$<$-0.0001
	($<$0.0001)	($<$0.0001)
L_resid3	$<$0.0001	$<$0.0001
	($<$0.0001)	($<$0.0001)
pcoal	0.1700***	
	(0.0219)	
pgas	0.5100***	
	(0.0977)	
pCO2	-0.0758	
	(0.2530)	
_cons	-104.1***	-92.14*
	(31.11)	(39.06)
Observations	8,687	8,687

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Robustness check other IVs

	Mean	Std.Dev.	Minimum	Maximum	Observations
I	0.0790	0.0737	0.0238	0.6355	8,687
II	0.0887	0.0874	0.0228	0.6837	8,687

in eurocent

Table A3: Summary statistics first derivative - IV robustness check estimates

1.6.3 Relationship of residual load and market price OLS and IV for 2015, 2016 and 2017

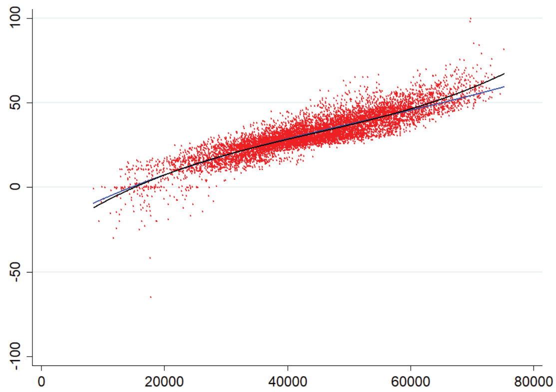


Figure A1: Relationship of residual load and market price OLS (black) and IV (blue) in 2015

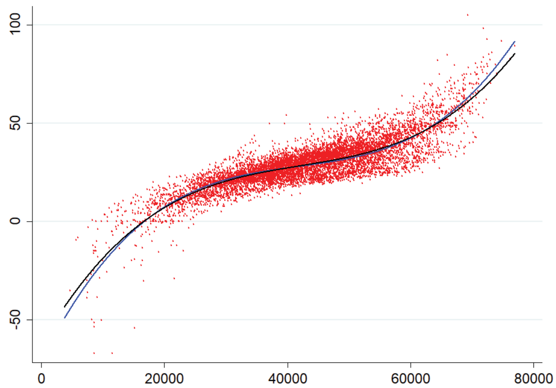


Figure A2: Relationship of residual load and market price OLS (black) and IV (blue) in 2016

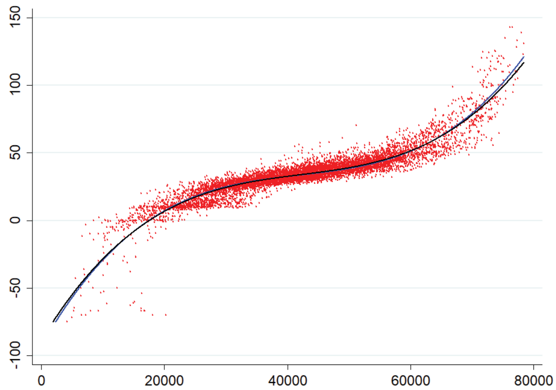


Figure A3: Relationship of residual load and market price OLS (black) and IV (blue) in 2017

2

The Identification of Up- and Downstream Industries using Input–Output Tables and a Firm-level Application to Minority Shareholdings

Joint work with Achim Buchwald and John Weche

2.1 Introduction

In its 21st main report, the Monopolies Commission, an independent expert committee that advises the German government and legislature in the areas of competition policy, empirically investigated the competitive effects through minority shareholdings between vertically linked firms (Monopolkommission, 2016). For the identification of vertically linked industries of EU-28 economies, trade flows from input–output tables were used. This approach has been applied in empirical research before, but mostly for US data. We take this as motivation to discuss the available strategies to identify up- and downstream industries via input–output tables in general and to present detailed results for European industries. Furthermore, we test the time-consistency of vertical trade flows between industries as well as the sensitivity of the industry assignments to varying definitions of the thresholds. Finally, the method is used to scientifically replicate and extend the analyses of the Monopolies Commission and test theories on foreclosure effects of non-controlling minority shareholdings.[1] The main contribution of this article is thus twofold: Firstly, it offers an in-depth tutorial for the identification of up- and downstream industries using input–output tables.[2] Secondly, it provides insights into the competitive effects of non-controlling minority shareholdings that are currently under debate within the framework of the European merger control.

There is a large theoretical literature on the specific relationship between upstream and downstream industries. For example, in the field of industrial organization, Ordover et al. (1990), Salinger (1988) and Hart and Tirole (1990) deal with the relation between mergers and acquisitions and market foreclosure. Other authors analyze vertical relations in the context of theories of the firm and strategic management (Williamson, 1971, 1979; Grossman and Hart, 1986; Hart and Moore, 1990; Pfeffer, 1992) or with regard to minority shareholdings (Gilo et al., 2006; Reynolds and Snapp., 1986). Thus, the empirical identification of up- and downstream industries to test various theoretical predictions is of relevance in many areas.

However, empirical evidence for theoretical predictions that are based on a differentiation between up- and downstream industries is scarce, due to the lack of an

[1] For the necessity and discussion of pure and scientific replication in economics, see Hamermesh (2007).

[2] We provide both the dataset of up- and downstream links based on European input–output information for the years 2008–2011 derived from Eurostat as well as the Stata code for reproducing the assignment of up- and downstream industries in the online appendix.

appropriate classification of industries with regard to their position within a value chain. This is especially true when it comes to firm-level studies with large datasets which cover a broad range of industries. The empirics face a dilemma here: the larger the population covered in a study, the less precise will be the classification into up- and downstream industries. This is because it is not possible to manually identify such relations with proportionate effort compared to case studies, which in turn cannot be statistically representative for a broader population.

Previous empirical firm-level studies have used a makeshift differentiation between so called intra-industry links on the one hand and inter-industry, intersectoral or cross-industry links on the other hand. While it seems plausible to describe firms to be horizontally connected when the involved firms are classified with the same industry code it is questionable to classify two industries as vertically connected when the codes differ. This approach is neither suitable to capture whether a vertical link is of an up- or downstream nature, nor does it consider that industries are not necessarily connected via a value chain in the case of no substantial trade relations and that links therefore can also be conglomerate. The terms inter-industry and vertical relations are occasionally used as a synonym in empirical studies e.g. Monopolkommission (2014) or Buchwald (2014).

We discuss more advanced approaches for identifying up- and downstream industries in inter-industry datasets in the following Section 2.2. Although the surveyed approaches have been predominantly applied to US data, a recent exception for European data has recently been conducted by the German Monopolies Commission in its 21st main report. We describe the procedure to identify up- and downstream industries, which is based on inter-industry trade flows, in detail in Section 2.3. We apply this approach to aggregated European input–output data and present detailed results on the number of identified industry links and their sensitivity to the definitions of the thresholds. We furthermore test the time-consistency assumption of up- and downstream classifications based on input–output tables, which has been taken for granted in previous applications, in Section 2.4. In Section 2.5, we describe and scientifically replicate the application concerning foreclosure effects through minority shareholdings and extend it by performing several robustness checks related to the identification of up- and downstream assignments. Finally, we discuss the limitations of the proposed approach in general and give an outlook in Section 2.6.

2.2 Approaches to identify up- and downstream industries

As stated above, it is essential to identify the vertical links and their direction in inter-industry firm-level datasets, for example, to measure the effects of vertical integration. There is only a limited number of previous empirical studies incorporating large firm-level datasets, because the firms in a sample first have to be identified as horizontally or vertically related. Firm links are generally defined as horizontal if the involved firms are classified with the same industry code (Fan and Goyal, 2006; Fee and Thomas, 2004). Firms with different industry codes can either be vertically related or conglomerate. In the case of vertical relations, it is important to identify up- and downstream firms along the supply chain. There is very little empirical literature that uses measures of the direction of vertical industry linkages. The most important exceptions are summarized in the following:

Atalay et al. (2014) use a measure to identify vertical integration of firms to assess how production in vertically integrated firms differs from that of unlinked producers in the same industries. The authors use microdata from two sources: the first is the US Economic Census and the second is the US Commodity Flow Survey. They corroborate their investigation with the classical theories of the firm, which focus on the determinants and effects of vertical integration and assume that vertical integration is often about transfers of intangible inputs rather than physical ones. They assess which businesses are vertically integrated by identifying the industry affiliation of every establishment using Economic Census data. These affiliations are based on the 1992 input–output Industry Classification System, the taxonomy used by the Bureau of Economic Analysis (BEA). Input–output (IO) tables are part of the system of national accounts and describe the sale and purchase relations between producers and consumers within an economy (cf. Section 2.3.1). They define an existing substantial link between one industry and another based on the relative volume of trade flows between the two industries. The industries are identified by 4-digit Standard Industrial Classification (SIC) codes. According to their definition, a substantial link exists between industry A and another industry B when one percent of industry A's sales are sent to establishments in industry B. The authors admit that the one percent cutoff is chosen arbitrarily, however they make several sensitivity checks and find only few differences. They assume the IO structure of the economy to be stable over time and use data for the year 1992 to identify links in a sample of establishments from the 1977, 1982, 1987, 1992 and 1997 censuses.[3]

Matsusaka (1996) uses a similar approach to find empirical evidence for the antitrust hypothesis, postulating that industries diversified in the 1960s because antitrust authorities prevented them from expanding in their home industries. Later, when antitrust policy became less stringent in the 1980s, firms were again allowed to expand horizontally, which could have led to de-diversification and a refocussing on their core businesses. The sample consists of 549 mergers which were identified from listing statements of the New York Stock Exchange and took place in 1968. Matsusaka (1996) also added observations listed in Ravenscraft and Scherer (1987). The author uses IO tables of the US Census Bureau for the year 1972 and defines two industries as vertically related if they buy at least five percent of their input or sell at least five percent of their output to each other.

The approach of Matsusaka (1996) is in line with the definitions of McGuckin et al. (1991) and Shahrur (2005). McGuckin et al. (1991) use a sample of 94 takeovers between 1977 and 1982. The analysis relies on data from the Census Bureau's Longitudinal Research Database, a database on the activities of manufacturing establishments in the US which is also used by Atalay et al. (2014). Shahrur (2005) examines the wealth effects of horizontal takeovers on rivals of the merging firms as well as on firms in the takeover supplier and customer industries. They use 1987, 1992 and 1997 benchmark IO tables in order to identify firms that supply inputs to the takeover industry and firms that utilize the takeover industry's output. They only consider customer industries with a "customer input coefficient", which measures the importance of the takeover industries output in the customers production, greater than one percent. Later they repeat their analysis for three and five percent cutoffs. They use the Worldwide M&A Section of the Securities Data Company (SDC) database to obtain their takeover sample for the period 1987–1999. Shahrur (2005) finds positive abnormal returns to rivals, suppliers and corporate customers in his subsample of takeovers, suggesting that takeovers are driven by efficiency considerations.[4]

In alignment with the studies mentioned above, Fan and Goyal (2006) use US data of 2,162 mergers from the Centre of Research in Security Prices in the period 1962–1996. They find that vertical merger activity is more intensive in the 1980s

[3] More specifically, they use the establishments in the Longitudinal Business Database, which includes all US business establishments with paid employees.

[4] See also Williamson (1971, 1979); Klein et al. (1978); Grossman and Hart (1986); Hart and Moore (1990) on this.

and 1990s. Another important finding is the positive wealth effect of vertical mergers. Industry commodity flow information in IO tables is utilized to define vertical relations in mergers.

Kedia et al. (2011)and Shenoy (2012) follow Fan and Goyal (2006) and use the industry commodity flow information in the use-table of benchmark IO tables in order to identify vertical links. Consistent with the studies above they classify mergers as vertically related if the corresponding vertical coefficient is larger than a certain cutoff. They consider alternative cutoffs (one, five and ten percent) to test for robustness. After defining the vertical links, Kedia et al. (2011) study the market reaction to vertical mergers. In addition, they explore reasons for vertical integration which is analyzed in the industrial organization literature. They confirm the findings of Fan and Goyal (2006) that returns for vertical merger announcements were positive until the late 1990s. Another finding, consistent with theory, is that vertical deals in non-competitive environments are associated with higher returns relative to other vertical deals.

Consistent with the literature, Shenoy (2012) uses the value of commodity flows between industries in order to identify customer industries for each vertical takeover in the period 1981–2004. The main finding is an average positive wealth effect for the merging firm, which is consistent with Fan and Goyal (2006).[5]

A relatively new paper by Ahern and Harford (2014) describes the economy as a network of industries connected through customer and supplier trade flows. They investigate if industries may be affected by shocks. Consistent with the studies above, they use IO tables from the BEA. In their main analysis, they present results on industry links using the 1997 IO table. In addition, they run several robustness checks with IO tables for the years 1982, 1987, 1992 and 2002. Ahern and Harford (2014) find that cross-industry mergers are highly clustered in a small number of industry pairs. Another finding is the fact that industry merger activity takes place in a wave-like pattern through customer and supplier links. They also outline the importance of this network approach beyond the merger context (Acemoglu et al., 2012; Ahern, 2013).

Fee and Thomas (2004) examine the reaction of customers and suppliers to merger announcements in a large cross-sectional setting. Different to the methods above they identify individual companies with actual product-market relationships.

[5] These results are also consistent with the efficiency hypothesis, but inconsistent with foreclosure theories, cf. Salinger (1988); Hart and Tirole (1990); Ordover et al. (1990).

As already mentioned at the beginning, one can see the two approaches as complementary. The IO approach is able to identify a large pool of potential customers and suppliers. However, the approach used by Fee and Thomas (2004) identifies those firms that have significant product-market relationships with the merging firms. A sample of horizontal mergers from the SDC M&A database is used, that were proposed between 1980 and 1997. Firms are required to report information such as any customer representing more than 10 percent of the firm's total sales. This information is included in the industry segment files.

Similar to Fee and Thomas (2004), another possible measure of vertical relatedness is the 10-K-text-based measure to identify the extent to which a firm's products span vertically related product markets. Fresard et al. (2013) use this method by focussing on how firm product vocabularies relate to commodity descriptions from the BEA IO tables. They extend the work of Hoberg and Phillips (2010), who focus on horizontal links only. This method requires multiple data sources: 10-K business descriptions from annual reports, the database of financial statistical and market information on active and inactive global companies (Compustat), the SEC Electronic Data Gathering, Analysis and Retrieval system (Edgar) database, the BEA IO tables, and data on merger cases. They start with the Compustat sample of firm-years from 1996 to 2008, and proceed by using the Edgar database to extract the text in the business description section of annual firm 10-Ks. They use web crawling and text parsing algorithms to construct a database of business descriptions. Further, to define vertical relatedness, they also rely on IO tables from the BEA. Ahern (2012) uses a similar approach based on 10-K-text product market descriptions to control for product similarity using data on mergers from the US SDC.

All approaches surveyed so far have been exclusively applied to US data. An exception is the work by Acemoglu et al. (2010) who use UK firm-level data of the Annual Respondents Database, over the period 1996 to 2001, to identify vertical relations between firms and to analyze the determinants of vertical integration. They try to find empirical evidence that mergers help to solve the hold-up problem based on the theoretical work of Grossman and Hart (1986).

Although the use of IO tables to identify up- and downstream relations is well established in empirical economic research, the existing studies predominantly focus on US data. To the best of the authors' knowledge, the first application in a European context was recently conducted by the Monopolkommission (2016) and will be discussed in more detail in Section 2.5. Another insight obtained from our literature survey is that previous studies using firm-level panel data for several years only extract cross-sectional information on whether an industry is up- or downstream. It is assumed that the commodity trade flows in the IO tables are rather constant over time, and hence also an industry's classification as up- or downstream.

2.3 Identification of up- and downstream industries based on European IO tables

2.3.1 European IO tables

To measure up- and downstream industry relations for European economies, we use aggregate trade data from Eurostat. IO tables are part of the system of national accounts and describe the sale and purchase relations between producers and consumers within an economy. Eurostat provides consolidated IO tables for the Euro area and the EU with a breakdown into 64 products and industries respectively.[6] Information on commodity trade flows is available in the Statistical Classification of Economic Activities in the European Community (NACE) Rev. 2 format for the years 2008 to 2011. Consolidated tables for the years 2000 to 2007 were disseminated on the basis of the NACE Rev. 1 industry classification. Information on the European level is available only in terms of a two-digit industry classification scheme.

There are three types of available IO tables (Eurostat, 2008): i) The supply table gives the flow of goods and services at basic prices. It represents the value of produced goods in the domestic market, structured into product divisions. ii) The use table documents the flows of goods at purchasers' prices. It shows the usage of products for example as an input for other products, exports or consumption which were produced in the domestic market. The supply and the use table can be compiled by converting the supply and use tables at basic prices into one symmetric product-by-product IO table. iii) A symmetric product-by-product IO table can be compiled by converting the supply and use tables at basic prices. This procedure involves a change in format, from two asymmetric tables to one symmetric table. We use the symmetric EU aggregates product×product matrix at current prices, with 64 product categories. It is classified according to the Statistical Classification of

[6] For a comprehensive description see Eurostat (2008).

Products by Activity in the European Economic Community (CPA), which generally matches the NACE Rev. 2. However, a more detailed industry classification at the European level is not available. Some industry codes are pooled. For a better fit with an industry dataset it is useful to split the pooled industry codes. As a result, we obtain the trade flows between 79 different industry divisions.[7] For each year 6,162 industry pairs exist which results in 24,644 industry pairs over the years 2008 to 2011.

2.3.2 Identification of up- and downstream industries

In line with previous work, discussed in Section 2.2, we identify the up- and downstream industries using the IO tables provided by Eurostat. Figure 2.1 illustrates the flows of goods and services between three industries for a better understanding of the IO tables: ϵ denotes the flow of goods and services and i denotes any further related industry. For example ϵ_{AB} stands for the value of the flows of goods and services which are delivered from industry A to industry B. On the one hand, this shows how much of A's output is delivered to industry B. On the other hand, this denotes how much input industry B purchased from industry A.

Atalay et al. (2014) and Matsusaka (1996) assume two industries to be vertically related if industry A receives at least five percent of its input from industry B or industry A sells at least five percent of its output to industry B. According to this definition, there are two possible ways to identify the vertical links between industries:

The first approach focuses on the output of an industry and thus the resulting input supply for another industry (output-related approach). The flow of goods ϵ_{AB} from industry A to industry B is compared to the sum of industry A's total output. This can be written formally as

$$\vartheta_{AB} \quad = \quad \frac{\epsilon_{AB}}{\epsilon_{AB} + \epsilon_{AC} + \epsilon_{Ai}}. \tag{2.1}$$

The value ϑ_{AB} shows what proportion of the total outcome of industry A is delivered to industry B. It therefore describes whether industry B is an important buyer of

[7] Contrary to the Monopolies Commission, we split the pooled industries beforehand in order to identify every possible vertical link in the industry dataset. To do so we assume the pooled industries to be equal linked to other industries. This assumption is due to the lack of appropriate information about the accurate flow size of every single pooled industry.

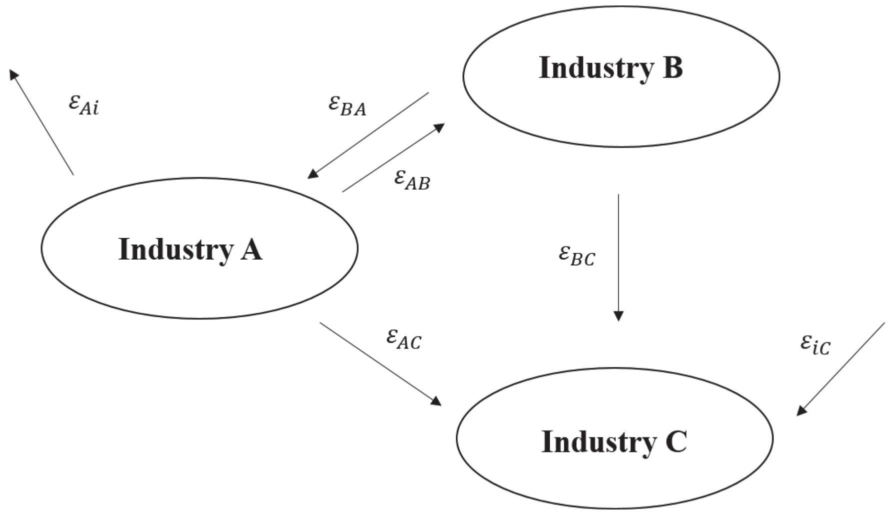


Figure 2.1: Structure of IO tables

industry A's output. It depends on the selection of the threshold θ if there exists a vertical link between the two industries. If $\vartheta_{AB} \geq \theta$, a vertical link can be identified. Depending on which threshold exceeds the cutoff, industry A can be identified as upstream and industry B as downstream (if $\vartheta_{AB} > \theta$ & $\vartheta_{BA} < \theta$) or vice versa, because of the symmetry of the IO table. This approach is also used by Atalay et al. (2014).

The second approach focuses on the input of an industry and the resulting relation with a supplier (input-related approach). This approach has been used also by the Monopolkommission (2016), as described in Section 2.5. The flow of goods and services ϵ_{AC} from industry A to C is compared to the sum of industry C's total input, which can be written formally as

$$\mu_{AC} \quad = \quad \frac{\epsilon_{AC}}{\epsilon_{AC} + \epsilon_{BC} + \epsilon_{iC}}. \tag{2.2}$$

The value μ_{AC} shows what ratio of the total input of industry C is purchased from industry A. It therefore reflects whether industry A is an important seller for industry C. As in the output-oriented approach, the threshold θ determines whether a vertical link between two industries is identified. For instance, this identification method is used by Matsusaka (1996).

Both approaches can be used to identify vertical links and thus to decide whether

Cutoff	No. of vertical links (Percentage of all possible links in parentheses)	No. of vertically linked industries (Percentage of all industries in parentheses)			Average no. of vertical links per industry (standard deviation in parentheses)	
		Upstream	Downstream	Both	Upstream	Downstream
3%	575	22	0	57	7.14	6.95
	(9.33)	(27.85)	(0.00)	(72.15)	(3.54)	(9.29)
5%	243	30	2	44	3.09	3.05
	(3.94)	(37.97)	(2.53)	(55.70)	(2.09)	(4.84)
10%	92	23	8	23	1.16	1.16
	(1.49)	(29.11)	(10.13)	(29.11)	(1.38)	(2.01)

Note: The sample covers 6,162 industry pairs of 79 different industries.

Table 2.1: Summary statistics of up- and downstream assignments

an industry is up- or downstream. Which approach should be used depends on whether the focus is on the up- or on the downstream industry. For instance, if one is interested in the importance of upstream links for a downstream industry it can make a huge difference whether the output-oriented approach is used or whether the input-oriented one is used. Say, for instance, that one upstream industry U delivers 90 percent of its output to the downstream industry D. In this case we would expect that D is an important buyer of U, but one cannot conclude conversely that U is also the most important input industry for D. The 90 percent of U's output might only make up a negligible proportion of D's total inputs. Thus U would not be the most important input supplier for D.

Table 2.1 presents a descriptive analysis of the first approach. It was applied to the symmetric IO table for aggregated European data. It shows the number of vertically linked industries and also the average number of vertical links per industry in Europe. We compute the mean value over the time period 2008 to 2011 to straighten out the variation in the proportion of output delivered from one industry to another. In general, we find relatively small differences, but it is possible that the proportion drops slightly under the chosen cutoff in one year.[8]

We apply several cutoffs to control for the sensitivity of the results. In addition to the 5 percent cutoff, which is often used in the literature (Matsusaka, 1996; Atalay et al., 2014; Monopolkommission, 2016), we report descriptive findings for the 3 and

[8] The assumption of time consistency of these assignments is tested explicitly in Section 2.4.

10 percent cutoff. As expected, the overall level of integration is higher for the 3 percent cutoff (575 links) compared to the more stringent case of a 5 percent cutoff (243 links) or the 10 percent cutoff (92 links). We additionally report the number of one-way linked industries. For instance, we find 30 purely upstream and only 2 purely downstream industries for a 5 percent cutoff. More than half of all industries in the sample have both up- and downstream connections.

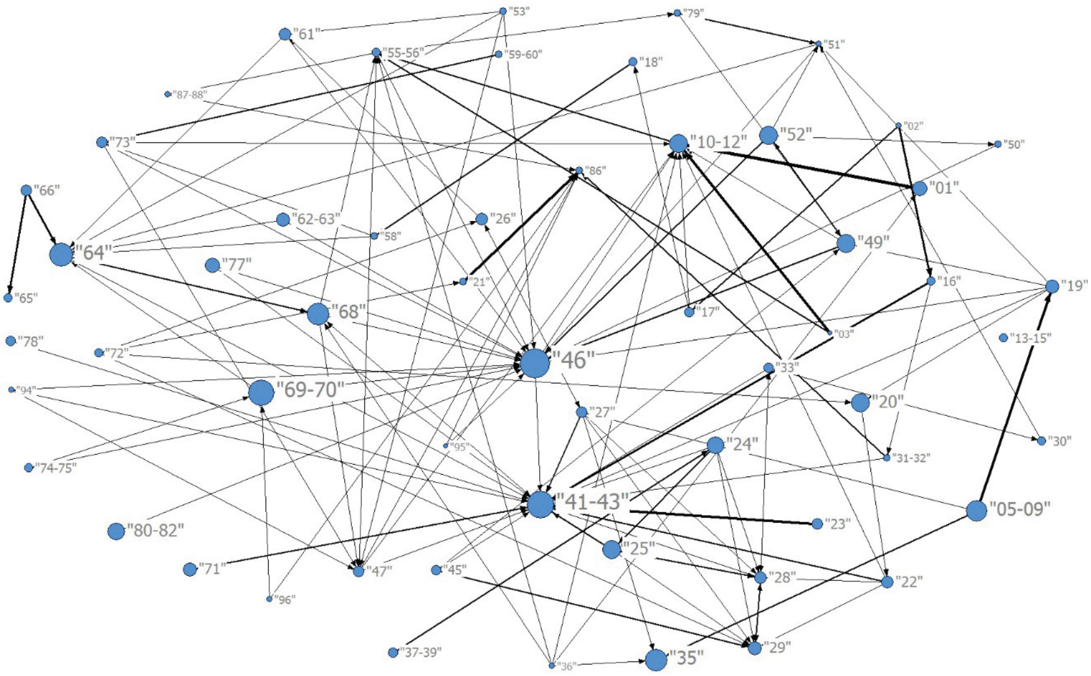
Alternatively, we examined indirect vertical industry relations. These second degree links occur if industry A delivers goods and services to industry C via industry B. Analogous to Figure 2.1, the flow of goods ϵ_{AB} from industry A to industry B and ϵ_{BC} from industry B to industry C compared to the sum of the respective industry's total output has to exceed the threshold. Following this approach, we receive a total number of 1,565 indirect vertical connections based on a 3 percent cutoff. For the 5 percent and 10 percent cutoff we find 552 and 72 links, respectively.[9]

As a further robustness check, we contrasted the output- and input-related approach, but found relatively few differences in the identified vertical industry links. Figures 2.2 and 2.3 illustrate the relevance of vertical links between industries using the arithmetic mean over the period 2008 to 2011 and a threshold of 5 percent. The arrows represent the commodity flows between two industries, shown as circles.[10] The strength of an arrow is a function of the proportion of the total sales of firms in an upstream industry to customers in the downstream industry (represented by ϑ in Equation 2.1). In Figure 2.3, the strength of an arrow symbolizes the percentage of the total inputs of a downstream industry which is purchased from an upstream industry (described by the value of μ in Equation 2.2). The size of a circle is calculated based on the value of the industry's total output in Figure 2.2 and on the value of the industry's total input in Figure 2.3.

Not surprisingly, we find that the wholesale trade sector (NACE Rev. 2 division 46) is central in both approaches. Other important industries in the network in Figure 2.2 are the manufacture of food products, beverages and tobacco products (divisions 10–12), the manufacture of basic metals (division 24), construction activities (divisions 41–43), and financial service activities (division 64). In Figure 2.3 it is real estate activities (division 68) as well as legal, accounting and management

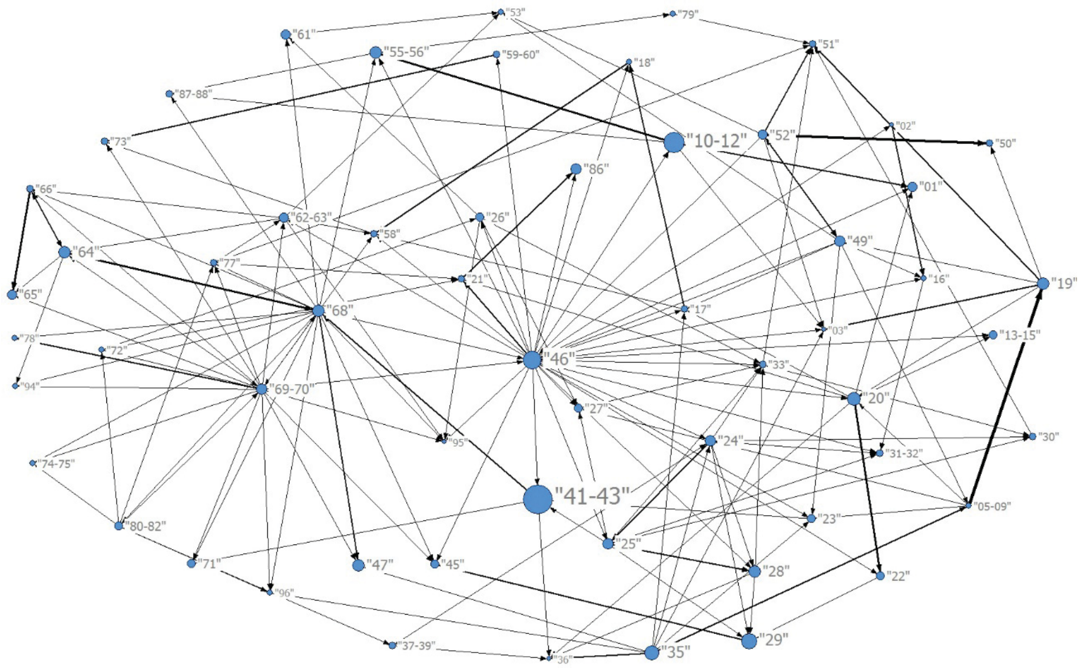
[9] The descriptive statistics of this approach are displayed in the appendix in Table B2. The higher number of identified indirect links compared to Table 2.1 results from the much larger variety when combining three industries (resp. triangles) out of the 79 industry divisions in the sample.

[10] A detailed NACE Rev. 2 industry description is offered in the appendix in Section 2.7.1.



Note: A detailed industry description can be found in the appendix.

Figure 2.2: Network analysis of the output-related approach



Note: A detailed industry description can be found in the appendix.

Figure 2.3: Network analysis of the input-related approach

consulting activities (divisions 69–70). It has to be noted that the size of the figures does not allow any conclusions regarding the absolute economic importance of an industry in global markets, since the figures exclusively refer to intra-European trade flows, while imports from and exports to non-European markets are neglected.

2.4 Time consistency of industry assignments

The existing studies that use IO tables for the identification of up- and downstream industries in panel data rely on the assumption that the trade flows in these tables are more or less constant over time (cf. Section 2.2). Therefore, it is common to use IO data from only one cross section for classifying industries in the previous and following years. We test this hypothesis so as to evaluate whether this procedure is appropriate or whether the classification of industries varies considerably over time.

We are not interested in the detailed variation of inter-industry trade flows but rather in whether this variation leads to changes in the up- and downstream classification of industries according to particular threshold criteria. Hence, we classified industries according to the input-oriented approach based on IO tables for the years 2008 to 2011 separately and compared the steadiness of these assignments. The results in Table 2.2 show that the number of industry pairs that fulfill the three, five, and ten percent thresholds within the observed period is higher than using the time averages as shown in Table 2.1. This points to considerable variation and that relatively high ratios of ϑ in some years cause industry pairs to fulfill the threshold criteria in other years when mean values are focused. However, more than 70 percent of the identified up- and downstream relations persist over the entire period observed. In the case of the three percent threshold, this is even almost 80 percent. Therefore, we note that most industry assignments appear to be stable over time, irrespective of the particular threshold, but that there is also a considerable share of assignments that do show volatility. Whether these ambiguous industry links affect the results of firm-level studies depends crucially on the weight of the observations of these industries.

Threshold	N	No. of years in which threshold applies			
		1	2	3	4
3%	640	39	55	38	508
	(1.00)	(0.06)	(0.09)	(0.06)	(0.79)
5%	281	25	15	34	207
	(1.00)	(0.09)	(0.05)	(0.12)	(0.74)
10%	105	12	5	10	78
	(1.00)	(0.11)	(0.05)	(0.10)	(0.74)

Note: Numbers of industry pairs and percentages in parentheses; N refers to the total number of industry pairs that fulfill the threshold criteria in at least one period.

Table 2.2: Threshold consistency over time (2008–2011)

2.5 Application in the context of minority shareholdings and EU merger control

The German Monopolies Commission recently used an IO table based approach in its 21st main report for the identification of up- and downstream industries in a broad European firm-level dataset (Monopolkommission, 2016).[11] The objectives of the Monopolies Commission's analysis are mainly i) to assess the overall frequency of up- and downstream minority shareholdings in the European economies, ii) to determine the yearly number of firms' minority share purchases in up- or downstream targets that would additionally fall within the scope of the European Merger Regulation in case of a modification, and iii) to estimate the effect of minority shareholdings on the intensity of competition in the affected industries.

With regard to the European Commission (EC) white book *Towards a More Effective EU Merger Control*, the German advisory body investigated potential anti-competitive effects of foreclosure strategies of firms that hold minority shares in up- or downstream industries (European Commission, 2013). A differentiation between up- and downstream ownership links is of crucial importance in this context, as theory predicts different outcomes for different levels of the value chain. For example, if a downstream firm holds shares in an upstream firm, the downstream firm can

[11] The German Monopolies Commission is an independent expert committee that advises the German government and legislature in the areas of competition policy-making, competition law, and regulation (see www.monopolkommission.de for further information).

have incentives to foreclose its downstream competitors. This is the case if the partially integrated upstream firm stops supplying downstream rivals or discriminates them by charging relatively higher prices. For some minority shareholdings, input-foreclosure may be even more likely than in the case of full vertical integration, because the acquirer of the minority share only internalizes a part of the target's losses in the upstream market, but internalizes all the profits in the downstream market. According to the same logic, upstream minority shareholders could benefit from preferential treatment through customers at the downstream level. Such customer-foreclosure could again be more likely in the case of minority shareholdings, because the shareholder firm could profit disproportionately from higher sales prices or quantities in relation to the suffering from losses of the target firm due to worse possibilities of changing the input supplier.

2.5.1 Data and variables

We build on the Monopolies Commission's analysis to offer insights into the competitive effects of minority shareholdings and to extend it by conducting several sensitivity checks related to the applied up- and downstream classification. The identification of vertical links and their separation into up- and downstream vertical links is crucial here to evaluate the specific effects through input- and customer-foreclosure. Eurostat IO tables can be used here for the years 2008 to 2011 using a five percent threshold for the share of average inflows (over time) from industry A to industry B of industry B's total inflows, to classify industry B as a downstream industry of A (cf. Section 2.3). Applying this method, we are able to identify up- and downstream shareholdings in an EU-28 subsample of the Bureau van Dijk's Orbis firm-level database combined with historic ownership data.[12]

The subsequent analysis refers to non-controlling minority shareholdings (NCMS), which would not fall within the scope of the current European Merger Regulation and are thus of priority interest from a regulators perspective. Unlike other (controlling) minority shareholdings, NCMS do not allow the shareholder a significant strategic influence over the target. Whether or not the strategic influence is sufficient for an application of the EC Merger Regulation is ideally subject to a case-specific assessment, which is not feasible in a large dataset. Therefore, we are dependent on a rather makeshift definition of non-controlling minority shares in our data and

[12] For a detailed description of the data and the data preparation, see the data appendix in Section 2.7.2.

assume NCMS when other shareholders of the target firm exist that hold equal or larger shares.

Table 2.3 shows the number of identified NCMS in 2013 by identified direction in terms of a value chain. In order to account for the fact that many firms in our sample are multi-product firms and are active in more than one industry, we consider not only the primary industry of a firm, but also the secondary, as a robustness check.[13] This appears to be important, since it significantly raises the number of links between firms either operating in the same industry or producing in the same value chain. It is noteworthy that the number of NCMS between firms in the same 4-digit industry increases more than fivefold. Furthermore, many links exist between industries that have a mutual or twoway trade relation and thus cannot be straightforwardly classified as up- or downstream. The category *conglomerate* is a residual category that covers all links that cannot be identified otherwise. Therefore, it may not only contain truly conglomerate links, but also i) truly vertical inks that cannot be identified due to the 2-digit IO data and ii) links which cannot be identified as vertical or horizontal due to missing industry information. In the subsequent analysis we focus on firm links for which anti-competitive effects are most likely, namely up- and downstream as well as horizontal NCMS.

	Using primary industries	Using primary and secondary industries
Upstream	5,067	6,942
Downstream	4,335	6,365
Twoway	2,742	3,933
Horizontal (4-digit)	11,110	58,057
Conglomerate*	64,349	20,995

Note: Identified NCMS exclude intra-UCI links (between entities with the same ultimate controlling institutional unit) and shareholdings of institutional investors or natural persons; *Additional to truly conglomerate links, this category may also contain i) truly vertical inks that cannot be identified due to the 2-digit input-output data and ii) links which cannot be identified as vertical or horizontal due to missing industry information.

Table 2.3: Non-controlling minority shareholdings (2013)

To measure competitive effects of NCMS, we calculate an empirical Lerner index

[13] The problem of multi-product firms is however alleviated in our analysis, because we use unconsolidated balance-sheet data.

as an indicator for market power. The theoretical idea behind the Lerner index L is that firm i exercises market power, if it is able to raise prices P over marginal costs MC ($L_i = 1 - \frac{P_i - MC_i}{P_i}$) (Lerner, 1934; Elzinga and Mills, 2011). In a situation of perfect competition, where prices equal marginal costs, the index is one, and it indicates market power, whenever $L_i < 1$. A great advantage of this indicator is that it is much less dependent on a proper market definition as structural measures, such as the Herfindahl-Hirschmann index or concentration ratios. In line with Aghion et al. (2005) we approximate the Lerner index using operating revenue minus capital costs over total revenue, due to a lack of information on prices and marginal costs.[14]

2.5.2 Results

Figure 2.4 illustrates the development of the market power of a shareholder firm and a target firm, measured via each of their Lerner indexes, in the years prior to a NCMS transaction and in subsequent years.[15] For upstream transactions, there is only a slight convergence of the market power indicator between shareholder firms and target firms following a transaction. In the case of an input foreclosure, a decreasing market power of target firms would be expected together with a market power increase of shareholder firms. However, the market power of target firms decreases and shareholders do not appear to gain market power. In the case of downstream transactions, the level of the shareholders' market power again looks relatively constant, while the market power of target firms increases. Although the directions, in which the market power develops for the shareholder one year after the transaction, fit the picture of customer foreclosure, the evidence from Figure 2.4 must regarded weak as the changes in market power appear to be rather small. In

[14] In contrast to information on revenues and the value of tangible and intangible assets, information on the economic capital costs is not available in balance-sheet data. Therefore, Aghion et al. (2005) assume a constant cost factor of 8.5 across all industries and time periods. To approximate a firm-specific cost factor, we follow an approach of Nickell (1996), which was also adopted by the Monopolkommission (2016), and calculate the capital cost factor as the sum of the country-specific long-term interest rate, the industry-specific depreciation rate, and a firm-specific risk premium. Information on interest and depreciation rates are taken from the EC's Ameco database and the OECD Stan database. For the risk premium, an interest rate is derived from the balance-sheet information on the paid interest on borrowed capital and the overall debt, which is adjusted by the short-run interest, again taken from the Ameco database.

[15] A transaction is assumed, when a link is reported in year$_t$, but not in year$_{t-1}$. To ensure that the observed transactions are in fact real transactions and not merely represent new information about the shareholder structure of a firm, NCMS are only considered when there is general information on shareholders available in year$_{t-1}$. For further details on the panel preparation, see the appendix.

the case of horizontal transactions, the changes are again very small and rather point to a strengthening of competition, instead of anti-competitive effects, for which a parallel downward movement of the graphs would be expected.

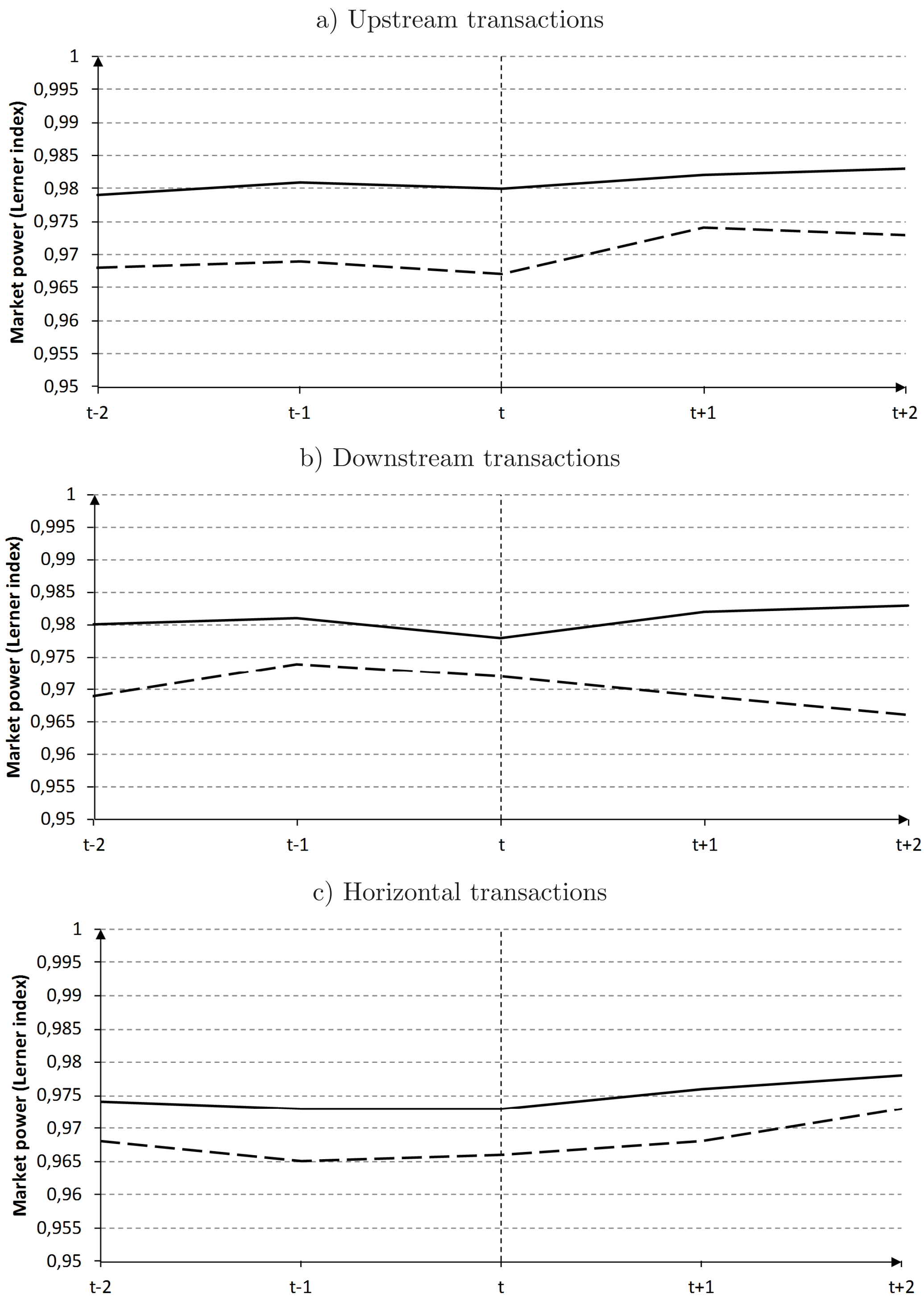
In a next step, we hold other factors beyond NCMS constant, which may affect our profit indicator for market power, via estimating a dynamic fixed-effects regression model. In particular, we estimate the percentage changes of the firm-specific Lerner index and include a set of dummy variables that indicate the year of a transaction as well as the two previous and two follow-up years. The dynamic nature of our revenue-based measure for market power is considered by including an autoregressive term. The first years after a firm entry and the years before an exit are considered to capture a likely explanatory power for market power variation. Moreover, the turnover, firm age, and the membership in a corporate group are considered as control variables.[16] Unobserved time-invariant firm-level characteristics are considered by the within fixed-effects estimator and general cyclical fluctuations are captured by year dummies. The standard errors are adjusted for firm clusters.

Table 2.4 presents the estimation results for shareholder firms and target firms. Unsurprisingly, the autoregressive term and whether or not a firm is about to exit the market are both significantly correlated with changes in market power. However, the results for the transaction dummies do not indicate changes in market power due to NCMS at all. Neither for shareholder firms nor for their targets.

As discussed above, the existence of multi-product (and therefore multi-industry) firms in the sample may bias the results if only their primary industry is considered in the assignment procedure of their NCMS transactions. Therefore, we have also conducted estimates that consider both the primary and secondary industries (cf. Table 2.3). Alternatively, we have performed the estimates solely for single-product firms which do not report operations in multiple industries. However, the results do not change significantly and also do not give reason to assume anti-competitive effects due to NCMS.[17]

Finally, we can conclude that there is no evidence for anti-competitive effects due to NCMS from a broad European inter-industry perspective. These results are in line with findings of the Monopolies Commission in its 21st main report. However, in our analysis we focus on average effects that do not say anything about individual cases and also not about the general potential for NCMS to unfold anti-competitive effects.

[16] Summary statistics for all variables are reported in Table B3 in the Appendix.
[17] The results are not reported for brevity.



Note: The averages of individual Lerner index values for shareholder firms (solid line) and target firms (dashed line) are shown. The Lerner index ranges from 0 (monopoly) to 1 (perfect competition). t denotes the year in which a transaction was identified.

Figure 2.4: Market power and non-controlling minority transactions

Moreover, it needs to be stressed that our econometric design aims primarily on the identification of correlations rather than causal effects and that it cannot be ruled out that there may be other (time-variant) factors determining the market power in our model or that the results are biased through reverse causality. Another caveat is that we use a revenue-based measure for market power that may not capture other effects of foreclosure or collusion, for example, in terms of entry barriers, although a correlation with the wedge between prices and marginal costs should be assumed also in these cases.

	Shareholder-level		Target-level	
	Coefficient	|t-value|	Coefficient	|t-value|
Market power$_{t-1}$	-3.59***	4.15	-3.59***	4.15
Upstream transaction$_{t-2}$	0.03**	2.01	0.03	1.46
Upstream transaction$_{t-1}$	-0.003	0.27	0.01	0.56
Upstream transaction$_{t}$	0.15	1.34	0.09	0.84
Upstream transaction$_{t+1}$	0.003	0.11	-0.16	0.99
Upstream transaction$_{t+2}$	-0.001	0.08	-0.04	0.65
Downstream transaction$_{t-2}$	0.02	1.34	0.02	1.19
Downstream transaction$_{t-1}$	-0.02	1.20	0.003	0.23
Downstream transaction$_{t}$	-0.003	0.22	0.004	0.24
Downstream transaction$_{t+1}$	0.01	0.45	-0.001	0.04
Downstream transaction$_{t+2}$	-0.02	1.26	-0.003	0.25
Horizontal transaction$_{t-2}$	0.05	1.26	-0.01	0.65
Horizontal transaction$_{t-1}$	0.09	1.01	-0.02	1.39
Horizontal transaction$_{t}$	-0.39	0.97	-0.03	1.64
Horizontal transaction$_{t+1}$	0.03	0.45	0.06	1.20
Horizontal transaction$_{t+2}$	0.004	0.13	0.04	1.14
Entry$_{t+1}$	0.02	0.60	0.02	0.61
Entry$_{t+2}$	-0.02	1.23	-0.02	1.21
Exit$_{t-2}$	-0.32	1.17	-0.32	1.18
Exit$_{t-1}$	-0.15	0.88	-0.15	0.88
Exit$_{t}$	-0.15**	2.43	-0.15**	2.42
Turnover$_{t}$ (mill. EUR)	-0.0001	1.08	-0.0001	1.09
Age$_{t}$	0.01	1.51	0.01	1.53
Group member$_{t}$	0.06	1.45	0.06	1.44
N	4,368,225		4,368,225	
n	797,300		797,300	

Note: *** p<0.01, ** p<0.05, * p<0.10; Estimated are percentage changes of an empirical Lerner index; Standard errors are adjusted for firm clusters; All estimates include year dummies.

Table 2.4: Dynamic fixed-effects estimates of market power

2.6 Limitations and outlook

This article has discussed input–output based approaches to identify up- and downstream industry relations and has presented results for the European member states during the period 2008 to 2011. The results complement previous work that has focused on identifying vertical industry relations based on UK and US data. Finally, the proposed method was used to identify up- and downstream minority shareholdings in a broad European firm-level database and enabled the estimation of anti-competitive effects due to foreclosure and/or collaborative strategies. The results do not suggest anti-competitive effects on average.

Although the proposed method can be potentially useful for empirical research in the fields of empirical industrial organization, international trade, and strategic management, some noteworthy limitations have to be mentioned.

First, as compared to US datasets, which are available at the 4-digit SIC Code level, consolidated IO tables provided by Eurostat are limited to 64 two-digit NACE Rev. 2 product or industry codes. Therefore, the identified up- and downstream relations are comparatively more aggregated, leading to a significant loss of information. As a consequence, the extent of vertical inter-industry links is systematically underestimated. Analogously, the degree of horizontal intra-industry relations is overestimated. An example will illustrate these two opposing effects: The NACE Rev. 2 division 29 includes the three subgroups "manufacture of motor vehicles" (29.1), "manufacture of bodies (coachwork) for motor vehicles; manufacture of trailers and semi-trailers" (29.2) and "manufacture of parts and accessories for motor vehicles" (29.3). Firms operating in these subgroups are allocated to division 29 in the Eurostat IO tables, resulting in horizontal links. Vertical relations between motor vehicle manufacturers and their various suppliers are therefore neglected by definition. The structural problem summarized above provides motivation for future work. In particular, an important contribution requiring further attention is to track up- and downstream relations on a more detailed level. Based on the data at hand, it is also not possible to distinguish between the respective subgroups of the divisions wholesale (46) and retail trade (47). As a consequence, most of the industries under investigation are connected to the trade sector.

Second, it would also be interesting to examine the assumption of the time consistency of the vertical relations between industries over a longer period of time, to disclose possible developments, such as regional or technological changes. The present paper has focused on the time period 2008 to 2011. Future research should therefore extend the time horizon.

Third, it has to be noted that the selected thresholds are somewhat arbitrary, putting emphasis on sensitivity checks.

Finally, previous studies and also the application in this paper have focused solely on direct up- and downstream relations, and have neglected indirect and more complex network relations within the entire value chain. While in the case of non-controlling minority shareholdings this does not seem to be crucial, for example, in the case of full mergers, indirect value chain relations may be of importance and should be explicitly considered.

Bibliography

Acemoglu, D., P. Aghion, R. Griffith, and F. Zilibotti, "Vertical Integration and Technology: Theory and Evidence," *Journal of the European Economic Association*, 2010, *8* (3), 989–1033.

_ **, V. Carvalho, A. Ozdaglar, and Tahbaz-Salehi; A.**, "The network origins of aggregate fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt, "Competition and Innovation: an Inverted-U Relationship," *Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Ahern, K., "Bargaining power and industry dependence in mergers," *Journal of Financial Economics*, 2012, *103* (3), 530–550.

_ , "Network Centrality and the cross section of stock returns," *SSRN working paper*, 2013, (December).

_ **and J. Harford**, "The importance of Industry Links in Merger Waves," *Journal of Finance*, 2014, *69* (2), 527–576.

Atalay, E., A. Hortacsu, and C. Syverson, "Vertical Integration and Input flows," *The American Economic Review*, 2014, *104* (4), 1120–1148.

Buchwald, A., "European Firm Networks and Competition: Theoretical Background and Empirical Evidence," *International Journal of Networking and Virtual Organisations*, 2014, *14* (4), 355–376.

Elzinga, K.G. and D.E. Mills, "The Lerner Index of Monopoly Power: Origins and Uses," *American Economic Review: Papers and Proceedings*, 2011, *101* (3), 558–564.

European Commission, "Towards More Effective Merger Control, Commission Staff Working Document," *Commission Staff Working Document*, 2013, *SWD 239* (June 25th), 558–564.

Eurostat, "Eurostat Manual of Supply, Use and Input-Output Tables," *Eurostat methodologies and working papers*, 2008.

Fan, J. and V. Goyal, "On the Patterns and Wealth Effects of Vertical Mergers," *Journal of Business*, 2006, *79* (2), 877–902.

Fee, E. and S. Thomas, "Sources of Gains in Horizontal Mergers: Evidence from Customer, Supplier and Rival Firms," *Journal of Financial Economics*, 2004, *74* (3), 423–460.

Fresard, L., G. Hoberg, and G.M. Phillips, "The Incentives for Vertical Mergers and Vertical Integration," *Working paper, University of Maryland, College Park*, 2013.

Gilo, D., Y. Moshe, and Y. Spiegel, "Partial Cross Ownership and Tacit Collusion," *The RAND Journal of Economics*, 2006, *37* (1), 81–99.

Grossman, S. and O. Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 1986, *94* (4), 691–719.

Hamermesh, D., "Viewpoint: Replication in economics," *Canadian Journal of Economics*, 2007, *40* (3), 715–733.

Hart, O. and J. Moore, "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 1990, *98* (6), 1119–1158.

_ **and J. Tirole**, "Vertical Integration and Market Foreclosure," *Booking papers on Economic Activity*, 1990, pp. 205–76.

Hoberg, G. and G.M Phillips, "New dynamic product based industry classifications and endogenous product differentiation," *Working Paper, University of Maryland, College Park*, 2010.

Kedia, S., R. Abraham, and V. Pons, "When do vertical mergers create value?," *Financial Management*, 2011, *40* (4), 845–877.

Klein, B., R.G. Crawford, and A.A. Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 1978, *21* (2), 297–326.

Lerner, A., "The concept of monopoly and the measurement of monopoly power," *Review of Economic Studies*, 1934, *1* (3), 157–175.

Matsusaka, J., "Did Tough Antitrust Enforcement Cause the Diversification of American Corporations?," *Journal of Financial and Qualitative Analysis*, 1996, *31* (2), 283–294.

McGuckin, R.H., S.V. Nguyen, and S.H Andrews, "The Relationships Among Acquiring and Acquired Firms' Product Lines," *Journal of Law and Economics*, 1991, *34*, 477–502.

Monopolkommission, "XX. Main Report: Eine Wettbewerbsordnung für die Finanzmärkt," *Baden-Baden*, 2014.

_ , "XXI. Main Report: Wettbewerb," *Baden-Baden*, 2016.

Nickell, S.J., "Competition and Corporate Performance," *Journal of Political Economy*, 1996, *104* (4), 724–746.

Ordover, J., G. Saloner, and S. Salop, "Equilibrium Vertical Foreclosure," *The American Economic Review*, 1990, *80* (1), 127–142.

Pfeffer, J., "A Resource Dependence Perspective on Intercorporate Relations," *Mark S. Mizruchi and Michael Schwartz eds., Intercorporate Relations: The Structural Analysis of Business. Cambridge: Cambridge Univ. Press*, 1992, pp. 25–55.

Ravenscraft, D. and F. Scherer, "Mergers, Sell-offs and Economic Efficiency," *Washington D.C. Brookings Institutions*, 1987.

Reynolds, R. and Snapp., "The competitive effects of partial equity interests and joint ventures," *International Journal of Industrial Organization*, 1986, *4* (2), 141–153.

Salinger, M., "Vertical Mergers and Market Foreclosure," *The Quarterly Journal of Economics*, 1988, *103* (2), 345–356.

Shahrur, H., "Industry structure and horizontal takeovers: Analysis of wealth effects on rivals, suppliers, and corporate customers," *Journal of Financial Economics*, 2005, *76* (1), 61–98.

Shenoy, J., "An examination of the efficiency, foreclosure, and collusion rationales for vertical takeovers," *Management Science*, 2012, *59*, 1482–501.

Williamson, O., "The Vertical Integration of Production: Market Failure Considerations," *The American Economic Review*, 1971, *61* (2), 112–123.

_ , "Transaction-Cost Economics: The Governance of Contractual Relations," *The Journal of Law and Economics*, 1979, *22* (2), 233–261.

2.7 Appendix

2.7.1 Industry classification and indirect links

Table B1: NACE Rev. 2 Industry Classification

2-digit Code	Description
01	Crop and animal production, hunting and related service activities
02	Forestry and logging
03	Fishing and aquaculture
05	Mining of coal and lignite
06	Extraction of crude petroleum and natural gas
07	Mining of metal ores
08	Other mining and quarrying
09	Mining support service activities
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
22	Manufacture of rubber and plastic products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals
25	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment n.e.c.

Continued on next page

Table B1 – *Continued from previous page*

29	Manufacture of motor vehicles, trailers and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment
35	Electricity, gas, steam and air conditioning supply
36	Water collection, treatment and supply
37	Sewerage
38	Waste collection, treatment and disposal activities; materials recovery
39	Remediation activities and other waste management services
41	Construction of buildings
42	Civil engineering
43	Specialised construction activities
45	Wholesale and retail trade and repair of motor vehicles and motorcycles
46	Wholesale trade, except of motor vehicles and motorcycles
47	Retail trade, except of motor vehicles and motorcycles
49	Land transport and transport via pipelines
50	Water transport
51	Air transport
52	Warehousing and support activities for transportation
53	Postal and courier activities
55	Accommodation
56	Food and beverage service activities
58	Publishing activities
59	Motion picture, video and television programme production, sound recording and music publishing activities
60	Programming and broadcasting activities
61	Telecommunications
62	Computer programming, consultancy and related activities
63	Information service activities
64	Financial service activities, except insurance and pension funding
65	Insurance, reinsurance and pension funding, except compulsory social security
66	Activities auxiliary to financial services and insurance activities

Continued on next page

Table B1 – *Continued from previous page*

68	Real estate activities
69	Legal and accounting activities
70	Activities of head offices; management consultancy activities
71	Architectural and engineering activities; technical testing and analysis
72	Scientific research and development
73	Advertising and market research
74	Other professional, scientific and technical activities
75	Veterinary activities
77	Rental and leasing activities
78	Employment activities
79	Travel agency, tour operator and other reservation service and related activities
80	Security and investigation activities
81	Services to buildings and landscape activities
82	Office administrative, office support and other business support activities
86	Human health activities
87	Residential care activities
88	Social work activities without accommodation
94	Activities of membership organisations
95	Repair of computers and personal and household goods
96	Other personal service activities

2.7.2 Data preparation

For the financial data, the Orbis version as of June 2015 was used, and all European firms with a turnover of at least 2 million EUR in at least one year during the observation period and an available unconsolidated account were extracted for 2006–2013. The sample covers firms from NACE 2-digit industries 01 to 82 (without financial services) and EU-28 countries without Cyprus and Malta. In contrast to the Monopolies Commission, we conduct our analysis for a sample including firms from Germany, Austria, and the UK.

Historic shareholder information stems from the Orbis company ownership module and was merged to firm IDs both on shareholder and on target level. Therefore,

Cutoff	Number of vertical links (Percentage of all possible links in parentheses)	Number of vertically linked industries (Percentage of all industries in parentheses)		
		Upstream	Downstream	Both
3%	1,565	27	0	52
	(0.32)	(34.18)	(0.00)	(65.82)
5%	552	37	1	37
	(0.11)	(46.84)	(1.27)	(46.84)
10%	72	29	12	8
	(0.01)	(36.71)	(15.19)	(10.13)

Note: The sample covers 493,039 triangular relationships between 79 different industries.

Table B2: Summary statistics of indirect up- and downstream assignments

at the target firm level only capital links can be considered whose shareholders are located within the EU-28 countries. Shareholdings of institutional investors and natural persons were not considered. A detected minority shareholding in in year$_t$ was also assumed to be existent in year$_{t+n}$, when shareholder information is missing in these years, but the firm was active in terms of reported turnover.

The Lerner index variable was restricted to values between zero and one, whereas all values below zero were treated as zero values. Particularly with regard to panel analyses, a careful preparation of the Orbis database is needed. Therefore, observations with asymmetric observations—which is observations with non-missing information for the outcome variable, but missing information for control variables or vice versa—were excluded. Furthermore, all firms with inconsistent missing values were dropped from the unbalanced sample. Inconsistency here refers to cases in which missing values occur in between reported figures and thus are unlikely to be caused by a firm's inactivity. Missing information can also occur in years in which firms had not been established or already had exited the market: these observations remain in our unbalanced sample. Moreover, only firms with at least three consecutive observations were kept in the sample.

Transaction variables				
	Shareholder-level		Target-level	
	No. of obs.	%	No. of obs.	%
Upstream transaction$_t$	1,945	0.04	1,159	0.03
Downstream transaction$_t$	1,387	0.03	1,175	0.03
Horizontal transaction$_t$	4,681	0.10	2,748	0.06
Vertical transaction$_t$	12,021	0.27	10,658	0.24
Considering secondary industries				
Upstream transaction$_t$	2,703	0.06	1,710	0.04
Downstream transaction$_t$	2,207	0.05	1,899	0.04
Horizontal transaction$_t$	13,302	0.30	14,839	0.33
Other variables				
	Mean	Std.dev.	Min.	Max.
Lerner index	0.96	0.08	0.00002	1
Turnover (mill. EUR)	20.77	348.47	0.001	163,000
Age (years)	19.39	15.68	1	896
Group member	0.35	0.48	0	1

Note: The number of firm-year observations for all variables is 4,467,488.

Table B3: Summary statistics for the firm dataset (2006–2013)

3

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

Joint work with Melinda Fremerey, Hans-Theo Normann, Jannika Schad

3.1 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 the US, private damage claims have existed since the early 20$^{\text{th}}$ 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

[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, Rn.(124), 2016, Rn.(31), 2017b, Rn.(28)).

[3] Private damage claims account for 90 to 95 percent of US cartel cases (Knight and Ste. Claire

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 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

(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.

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 3.1 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 3.1 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.

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.

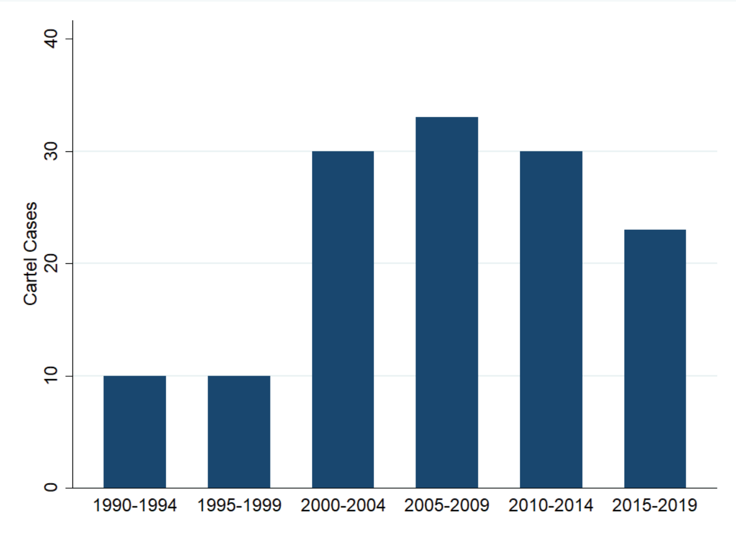


Figure 3.1: Cartel cases decided by the European Commission 1990–2019. Source European Commission (2019, section 1.10).

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 experi-

[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 (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.

ments 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 and Gonzalez (2016)), it seems promising to conduct this comparison with the inclusion of leniency. Likewise, Apesteguia et al. (2007) and Dijkstra et al. (2018) 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

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

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 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 3.3 presents our hypotheses which are the basis for our further analyses in section 3.4. Section 3.5 provides insights of an additional treatment that protects the leniency applicant from damage suits. We conclude in section 3.6.

3.2 The experiment

3.2.1 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:

[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.

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

3.2.2 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 3.2.3).

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:

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

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 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

---

[12]The revenue is defined as the quotient of the market price and the number of firms that sell at market price, see 3.7.1

[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)).

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 3.1 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 3.1: Fines and private damage claims of one firm.

3.2.3 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]

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 3.2 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 3.2). 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]

[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.

[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.

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 3.2: 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.

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. 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 and Gonzalez, 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 and Gonzalez (2016) have used similar chat devices in oligopoly experiments. Brosig et al. (2003) generally investigate the issue of the communication format on cooperation.

Table 3.3 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 3.5.

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 3.3: Overview of treatments.

3.2.4 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 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.

3.3 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 3.7.1). 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 3.7.1.

Our first hypothesis is about cartel formation, that is, the number of newly

[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.

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 3.7.1 for details) from the cartel's *participation constraint* which must necessarily be met, see also Bigoni et al. (2015) or Chowdhury and Wandschneider (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

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

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

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

where $E(D_i^c)$ is the expected, dicsounted, 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 3.7.1). 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 paid in either case, but they differ in magnitude (see Appendix 3.7.1 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.*

3.4 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 3.4. The exact definition of each variable can be found in Table C2 in the Appendix. The exact values underlying Figures 3.2 to 3.9 can be obtained from Table 3.4.

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

3.4.1 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 C2 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 3.2 and Table 3.4. 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 3.5 are also consistent with hypothesis 1. We see that the dummy variable PDC is highly significant and economically substantial.

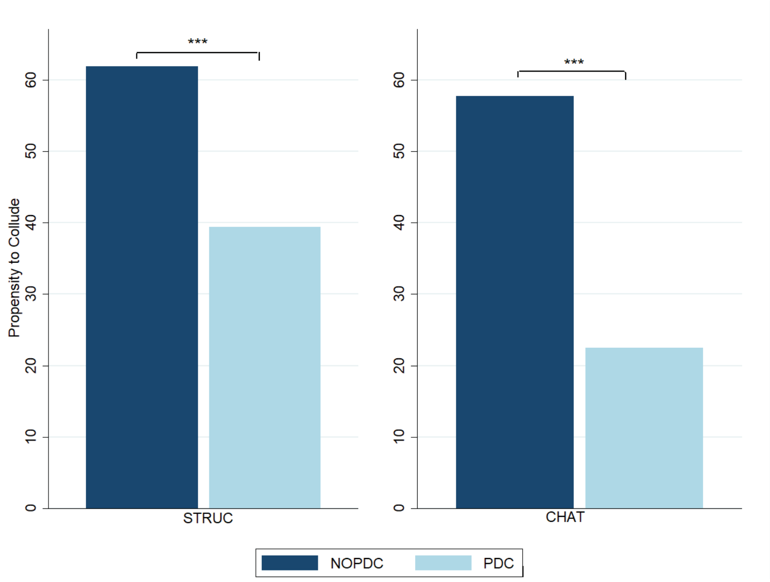


Figure 3.2: The impact of PDC on the individual propensity to collude in STRUC (left) and CHAT.

	(1) Collude	(2) Collude	(3) Collude	(4) 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 3.5: 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 C2 in the Appendix). Figure 3.3 and

Table 3.4 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 3.6 confirm that the effect is significant.

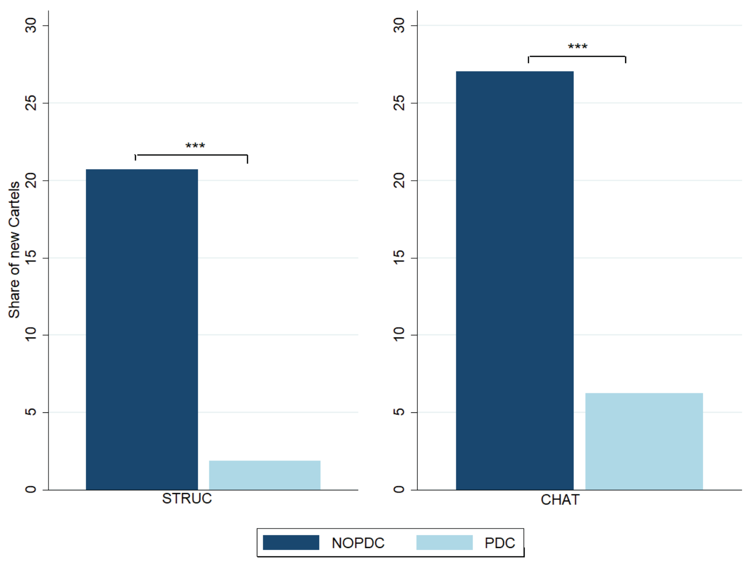


Figure 3.3: The impact of PDC on the number of cartels in STRUC (left) and CHAT.

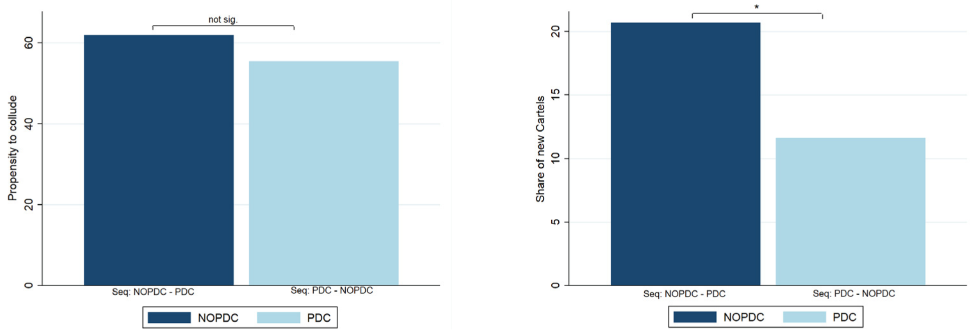
	(1) Collusion	(2) Collusion	(3) Collusion	(4) Collusion
PDC	-0.181***	-0.0813***	-0.125+	-0.375***
	(0.0311)	(0.0130)	(0.0817)	(0.116)
constant	0.194***	0.0875***	0.125+	0.375***
	(0.0344)	(0.0172)	(0.0817)	(0.116)
Time FE	No	No	Yes	Yes
Sample STRUC	Yes	No	Yes	No
Sample CHAT	No	Yes	No	Yes
N	320	320	320	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 3.6: 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 3.7.6. Figure 3.4 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]

[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.



(a) Propensity to collude (b) Share of cartelized markets

Figure 3.4: 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).

3.4.2 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 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 C2 in the Appendix for a detailed explanation of the variable *share report*).

Figure 3.5 and Table 3.4 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.

[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.

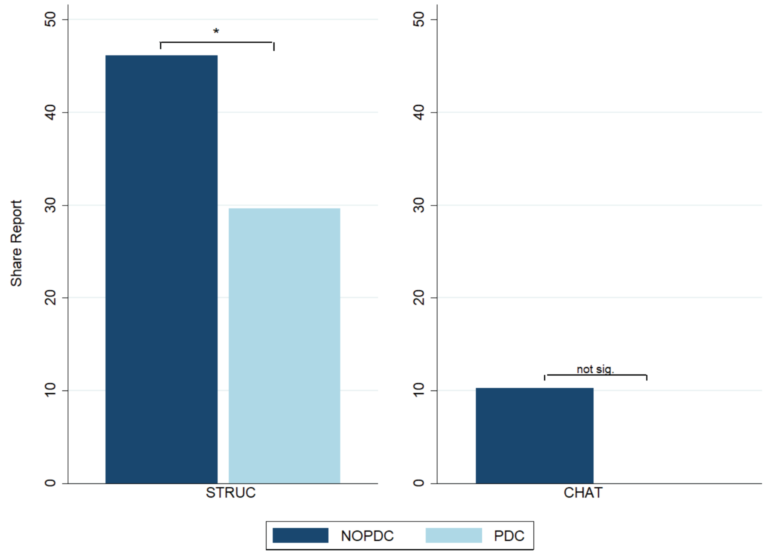


Figure 3.5: The impact of PDC on the individual reporting decision in STRUC (left) and CHAT.

Table 3.7 reports a linear regression of PDC on the individual decision to report 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 3.7: 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 C2 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 3.4). 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

[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.

enforcing effect on stability over time because damages cumulate. Cartels should, accordingly, be more strongly discouraged from reporting the longer they exist.

3.4.3 Cartel prevalence

We finally look at cartel prevalence, defined as the percentage of periods where a stable cartel existed (Table C2 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 3.3 is not a directed hypothesis about prevalence.) What is the overall balance?

Figure 3.6 and Table 3.4 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: $p-value = 0.0051$ and CHAT: WMP, $p-value = 0.0139$). The linear regressions in Table 3.8 confirm this.

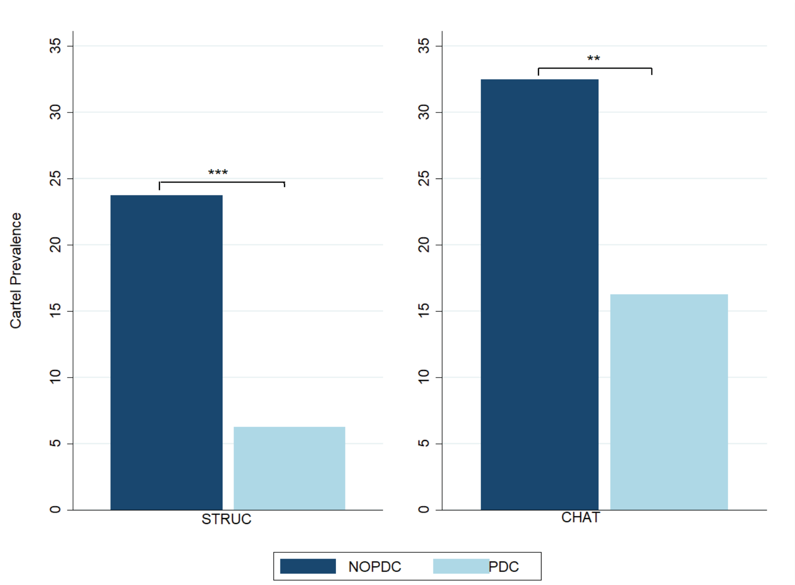


Figure 3.6: The impact of PDC on cartel prevalence in STRUC (left) and CHAT.

	(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 3.8: 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, $p - value = 0.0842$).

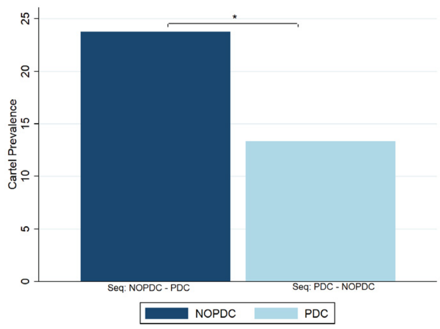


Figure 3.7: Cartel prevalence in STRUC: between-subjects comparison with PDC data from treatment with reverse order (PDC-NOPDC).

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

3.4.4 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 3.3).

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

We compare the average market price with and without private damage claims across the CHAT and STRUC treatments as shown in Table 3.9 and Figure 3.8. 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 3.9 verifies the lower overall market prices in PDC with STRUC communication (WMU, $p - value = 0.0511$).

[24]For an analysis of individual ask prices see Appendix 3.7.5.

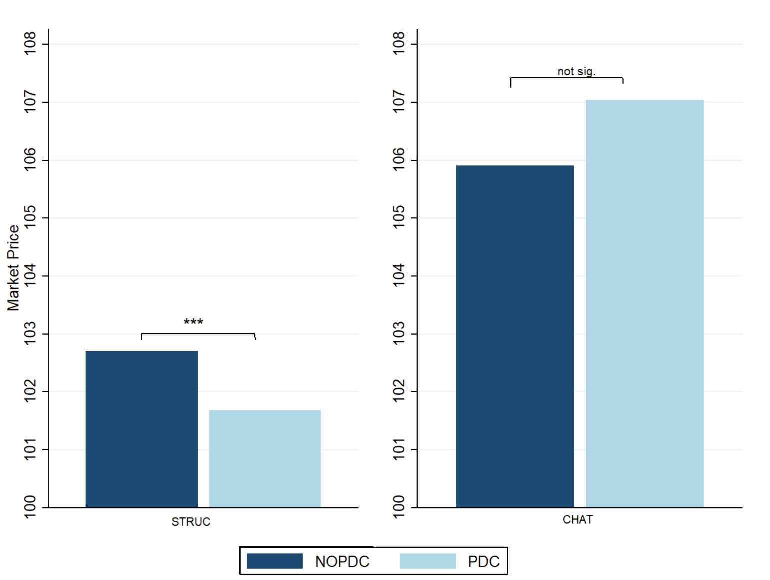


Figure 3.8: The impact of PDC on market prices in STRUC and CHAT.

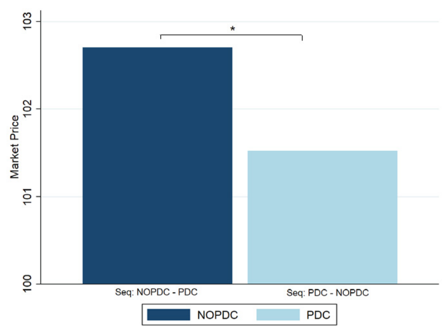


Figure 3.9: Market price in STRUC: between-subjects comparison with PDC data from the treatment with reverse order (PDC-NOPDC).

Table 3.10 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 3.10, column 1). They significantly increase in CHAT.

	(1) MarketPrice	(2) MarketPrice	(3) MarketPrice	(4) 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 3.10: 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 3.3 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 an hysteresis effect (see also 3.7.4 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.

3.4.5 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 3.4). This is also emphasized by the result that infringers apply less often for leniency ($p - value = 0.0011$) (see Figure 3.5). 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 3.2). 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 3.9). As already mentioned, higher prices in CHAT can be explained by a 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 3.9).

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.

3.5 Protection from damages for leniency applicants

Although in an overall assessment of PDC we find a decreasing cartel prevalence in PDC, the results of the preceding section 3.4 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 3.2. 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

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

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 3.7.1 for details.

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

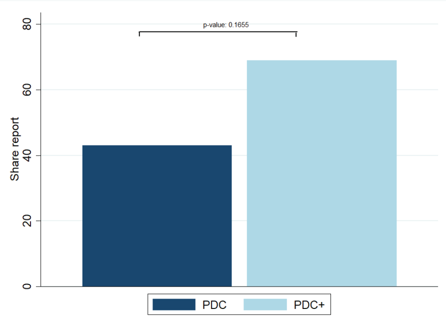


Figure 3.10: Share report in STRUC: within-subjects comparison from the treatment PDC-PDC+.

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 3.10. 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 3.11) ($p - value = 0.0929$).

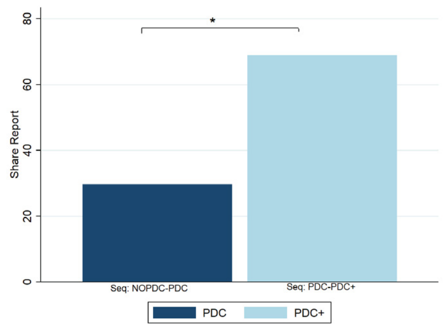


Figure 3.11: Share report in STRUC: between-subjects comparison with PDC data from NOPDC-PDC and PDC+ data from PDC-PDC+.

3.6 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, at least they not completely, reduced or they are capped only 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.

Bibliography

Andres, M., L. Bruttel, and J. Friedrichsen, "The effect of a leniency rule on cartel formation and stability: Experiments with open communication," 2019. *Preliminary draft.*

Apesteguia, J., M. Dufwenberg, and R. Selten, "Blowing the Whistle," *Economic Theory*, 2007, *31* (1), 143–166.

Becker, G., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217.

Bigoni, M., J. Potters, and G. Spagnolo, "Frequency of interaction, communication and collusion: An experiment," *Economic Theory*, 2019, *68* (4), 827–844.

_ , S.-O. Fridolfsson, C. Le Coq, and G. Spagnolo, "Fines, Leniency, and Rewards in Antitrust," *The RAND Journal of Economics*, 2012, *43* (2), 368–390.

_ , _ , _ , and _ , "Trust, Leniency, and Deterrence," *Journal of Law, Economics, and Organization*, 2015, *31* (4), 663–689.

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

Bourjade, S., P. Rey, and P. Seabright, "Private Antitrust Enforcement in the Presence of Pre-Trial Bargaining," *The Journal of Industrial Econommics*, 2009, *57* (3), 372–409.

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

Brosig, J., J. Weimann, and A. Ockenfels, "The Effect of Communication Media on Cooperation," *German Economic Review*, 2003, *4* (2), 217–241.

Brown Kruse, J. and D. Schenk, "Location, cooperation and communication: An experimental examination," *International Journal of Industrial Organization*, 2000, *18* (1), 59–80.

Buccirossi, P., C. Marvao, and G. Spagnolo, "Leniency and Damages," *CEPR Discussion Paper No. DP10682*, 2015.

Camera, G., M. Casari, and M. Bigoni, "Communication, commitment, and deception in social dilemmas: Experimental evidence," *Quaderni - Working Paper DSE No. 751*, 2011.

Canenbley, C. and T. Steinvorth, "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*, 2011, *2* (4), 315–326.

Cauffman, C. and N. Philipsen, "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*, 2014.

Chowdhury, S. and F. Wandschneider, "Antitrust and the Beckerian Proposition: The effects of investigation and fines on cartels," in "Handbook of Behavioral Industrial Organization," Cheltenham: Edward Elgar Publishing, 2018.

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

Cooper, D. and K.-U. Kühn, "Communication, Renegotiation, and the Scope for Collusion," *American Economic Journal: Microeconomics*, 2014, *6* (2), 247–278.

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

Department of Justice, "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," 2007. `https://www.justice.gov/archive/atr/public/press_releases/2007/224928.htm`. Accessed: 2019-05-19.

Dijkstra, P., M. Haan, and L. Schoonbeek, "Leniency Programs and the Design of Antitrust: Experimental Evidence with Free-Form Communication," *Working Paper*, 2018.

Dufwenberg, M. and U. Gneezy, "Measuring Beliefs in an Experimental Lost Wallet Game," *Games and Economic Behavior*, 2000, *30* (2).

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

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

_ , "Green Paper - Damages Actions for Breach of the EC Antitrust Rules," 2005. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52005DC0672&from=en`. Accessed: 2019-06-19.

_ , "Commission Notice on Immunity from Fines and Reduction of Fines in Cartel Cases (2006/C 298/11)," *Official Journal of the European Communities*, 2006.

_ , "Fines for breaking EU Competition Law," 2011. `http://ec.europa.eu/competition/cartels/overview/factsheet_fines_en.pdf`. Accessed: 2019-06-19.

_ , "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*, 2014.

_ , "The Damages Directive - Towards More Effective Enforcement of the EU Competition Rules," *Competition policy brief*, 2015.

_ , "CASE AT.39824 - Trucks," 2016. `http://ec.europa.eu/competition/antitrust/cases/dec_docs/39824/39824_6567_14.pdf`. Accessed: 2019-06-19.

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

_ , "Case AT.39258 — Airfreight," *Official Journal of the European Union*, 2017b.

_ , "Directive on Antitrust Damages Actions: Transposition of the Directive in Member States," 2018. `http://ec.europa.eu/competition/antitrust/actionsdamages/directive_en.html`. Accessed: 2019-03-25.

_ , "Cartel Statistics," 2019. `http://ec.europa.eu/competition/cartels/statistics/statistics.pdf`. Accessed: 2019-04-01.

Feltovich, N. and Y. Hamaguchi, "The Effect of Whistle-Blowing Incentives on Collusion: An Experimental Study of Leniency Programs," *Southern Economic Journal*, 2018, *84* (4), 1024–1049.

Fischbacher, U., "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 2007, *10*, 171–178.

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

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

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

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

Harrington, J. and M.-H. Chang, "Modeling the Birth and Death of Cartels with an Application to Evaluating Competition Policy," *Journal of the European Economic Association*, 2009, *7* (6), 1400–1435.

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

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

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

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

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

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

Jones, A., "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, 2016.

Kersting, C., "Removing the Tension Between Public and Private Enforcement: Disclosure and Privileges for Successful Leniency Applicants," *Journal of European Competition Law & Practice*, 2014, *5* (1), 2–5.

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

Kirst, P. and R. van den Bergh, "The European Directive on Damages Actions: A Missed Opportunity to Reconcile Compensation of Victims and Leniency Incentives," *Journal of Competition Law and Economics*, 2016, *12* (1), 1–30.

Knight, T. and C. Ste. Claire, "Reconciling the Conflict: Antitrust Leniency Programs and Private Enforcement," *Working Paper*, 2019.

_ **and W. de Weert**, "On Implementing Private Damages in European Competition Cases," *Working Paper*, 2015.

Land, R. and J. Davis, "Comparative Deterrence from Private Enforcement and Criminal Enforcement of the U.S. Antitrust Laws," *BYU Law Review 315*, 2011.

Landeo, C. and K. Spier, "Naked Exclusion: An Experimental Study of Contracts With Externalities," *American Economic Review*, 2009, *99*, 1850–1877.

Mechtenberg, L., Muehlheusser G., and A. Roider, "Whistle-Blower Protection: Theory and Experimental Evidence," *IZA DP No. 10607*, 2017.

Michaels, D., "Deutsche Bahn to Claim Damages of More Than USD 3 Billion Over Air-Cargo Cartel," *The Wall Street Journal from 2014-11-30*, 2014.

Migani, C., "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*, 2014, (7), 81–111.

Miller, N., "Strategic Leniency and Cartel Enforcement," *American Economic Review*, 2009, *99* (3), 750–768.

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

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

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

Spagnolo, G., "Divide et Impera Optimal Deterrence Mechanisms Against Cartels and Organized Crime," *mimeo, Mannheim*, 2003.

_ , "Leniency and whistleblowers in antitrust," in "Handbook of antitrust economics," Cambridge, Massachusetts and London: MIT, 2008.

Spagnolo, G. Luz R.D.and, "LENIENCY, COLLUSION, CORRUPTION, AND WHISTLEBLOWING," *Journal of Competition Law and Economics*, 2017, *13* (4), 729–766.

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

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

_ , "The Relationship between Public Antitrust Enforcement and Private Actions for Damage," *World Competition*, 2009, *32* (1), 3–26.

3.7 Appendix

3.7.1 Variables 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 C1. 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 C1: 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)

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

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

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

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

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

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

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

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 $E(D_{it}^j) = 1.3916 D_{it}^j$.

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

$$12.443 \quad \geq \quad 9.333.$$

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

3.7.2 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 C2: Definition of the main variables.

3.7.3 Group dynamics over time

Figures C1 and C2 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.

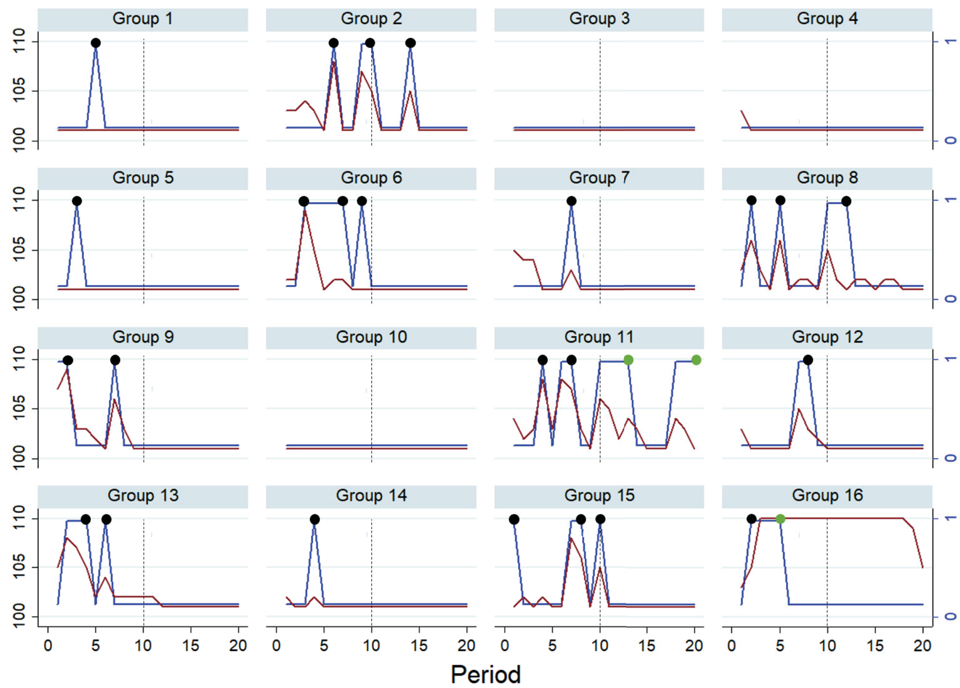


Figure C1: Collusive activity and market price by group for the treatment in STRUC.

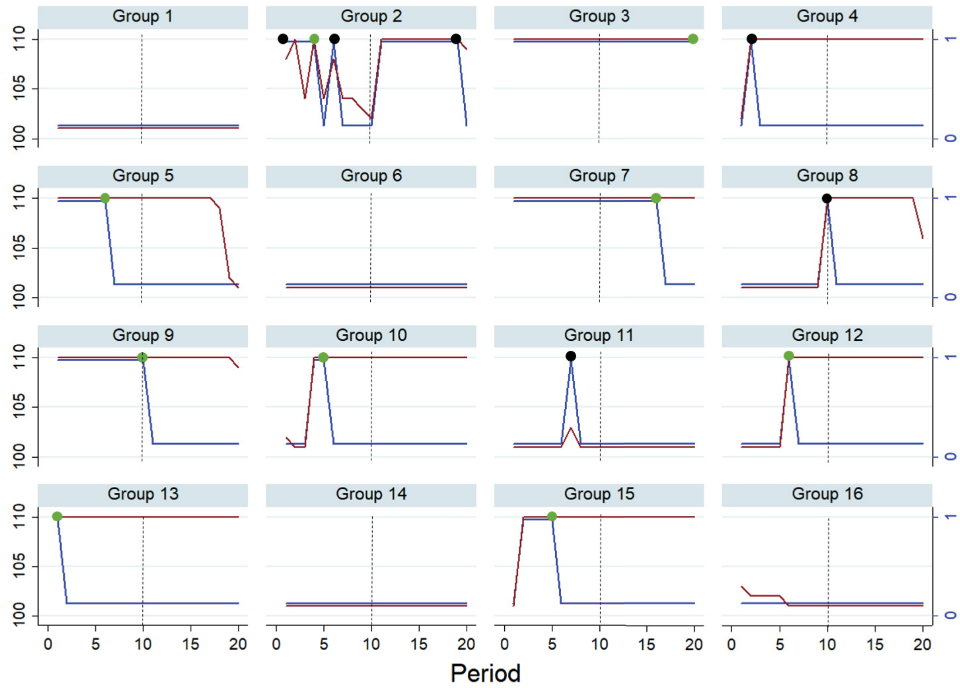


Figure C2: Collusive activity and market price by group in CHAT.

3.7.4 Deviations from agreed price

Figures C3 and C4 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 C3 indicates this by the upper and lower bound of agreed prices (see e.g., group 9).
In Figure C4, we can observe a more stable price setting following the agreed price even in periods without a cartelized market in CHAT. Figure C3, 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 C4 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.

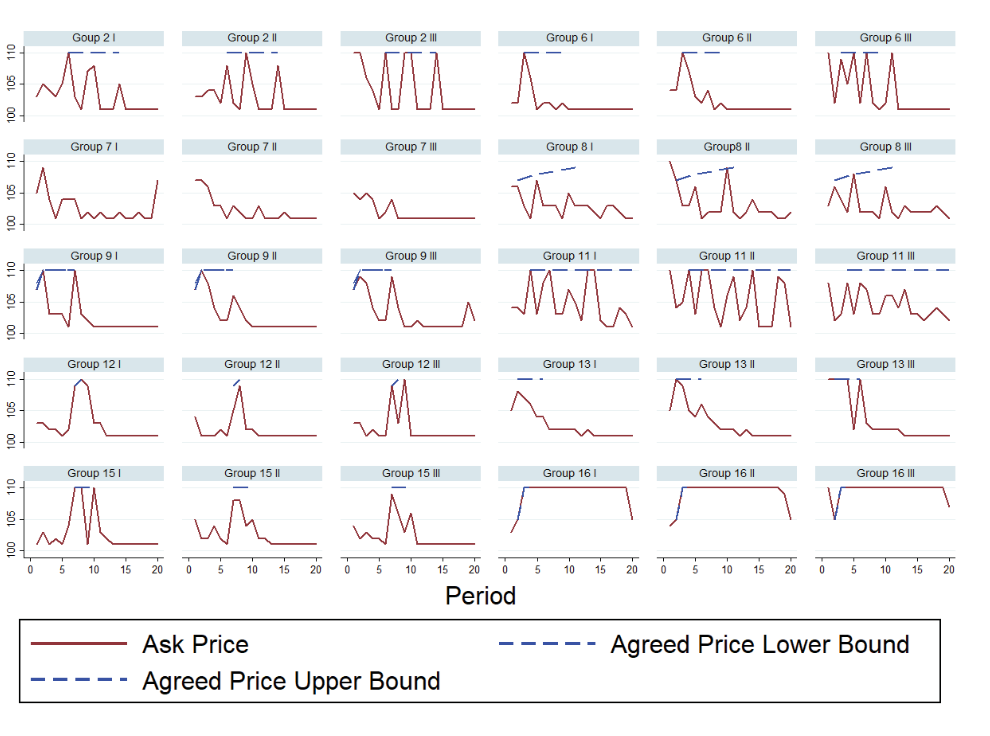


Figure C3: Agreed price and set price by subject in STRUC.
Note: Groups that do not discuss prices or could not agree on an interval other than 101 to 110 are excluded.

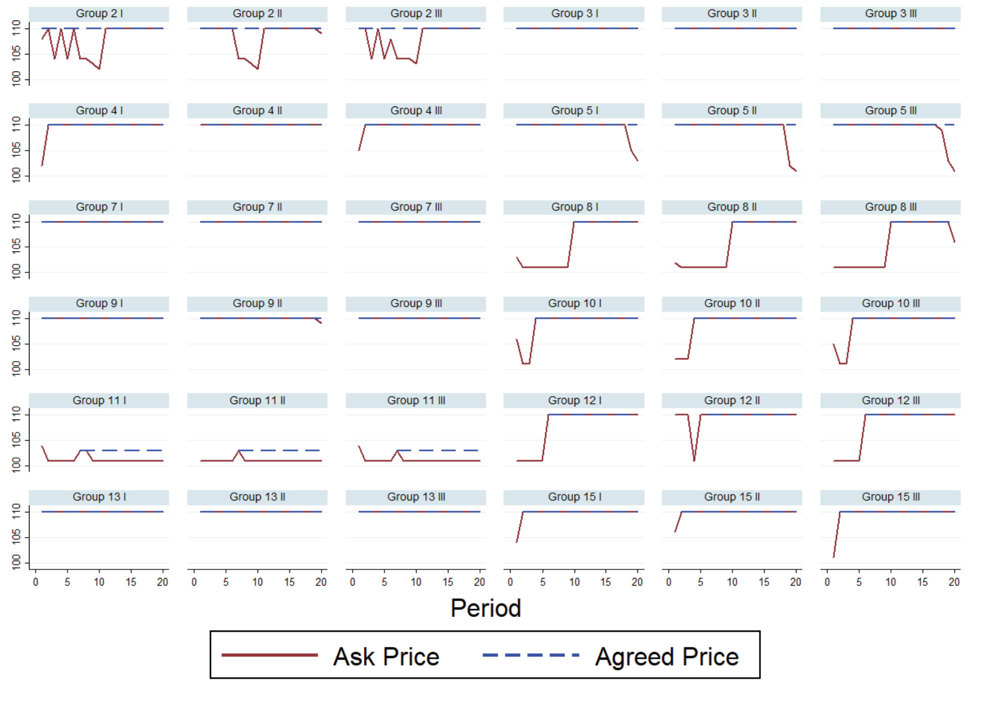


Figure C4: Agreed price and set price by subject in CHAT.
Note: Groups that do not discuss prices or could not agree on an interval other than 101 to 110 are excluded.

3.7.5 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 C5 (and the bottom line in Table C3) 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 C3: Ask price – averages per treatment (standard deviations in parenthesis).

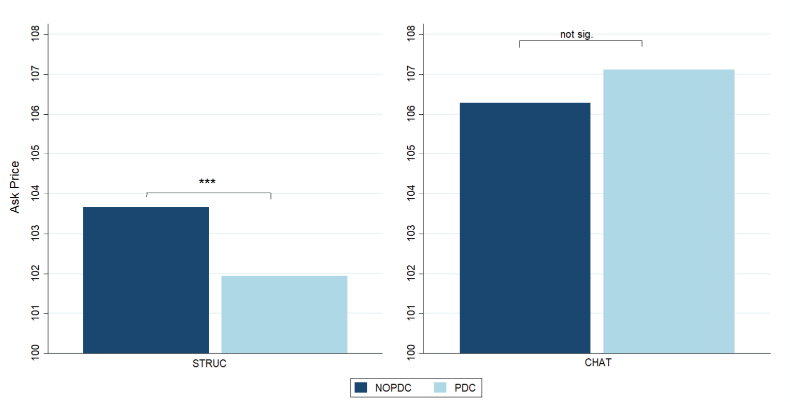


Figure C5: The impact of PDC on ask prices in STRUC (left) and CHAT.

In Table C4 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 C4, column 1), whereas PDC have a positive impact on ask prices in CHAT on a 15% level (Table C4, 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 C4: Ask price – linear regression (standard errors in parentheses).

Figure C6 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, $p - value = 0.0785$).

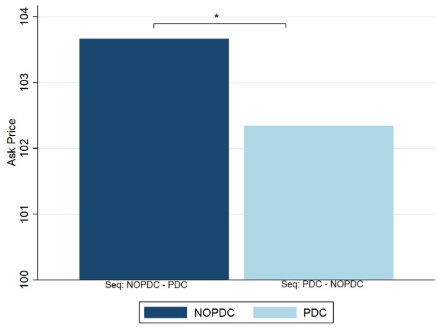


Figure C6: Ask price in STRUC: between-subjects comparison with PDC data from treatment with reverse order (PDC-NOPDC).

3.7.6 Within-subjects results of reverse-order treatment

For the robustness check of our main analysis we only use the PDC data from the session PDC-NOPDC (see chapter 3.4). 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 C5 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 C5: Summary statistics of the results in treatments PDC–NOPDC (STRUC); average results per treatment (standard deviations in parentheses).

3.7.7 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,

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 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 about prices in step 1 has taken place. If not all group members have agreed to

you can discuss the price with your group members by entering a text in the communication field and 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.

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 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.

4

Financial incentives and prescribing behaviour in primary care

Joint work with Hugh Gravelle, Nils Gutacker, Annika Herr

4.1 Introduction

In the English National Health Service (NHS), family doctors (known as general practitioners (GPs)) may dispense drugs to patients who live more than a mile away from their nearest community pharmacy. There are several reasons why prescribing behaviour in dispensing practices may differ from non-dispensing practices. First, by design, dispensing practices are predominantly located in rural areas and so may have patients who differ in their health conditions and their demand for prescriptions. Second, GPs may be imperfect agents because of asymmetric information (McGuire, 2000) and dispensing GPs may respond to financial incentives which reward practices for dispensing more and more expensive prescriptions. GPs with eligible patients can choose whether to dispense or not but if they dispense to one eligible patient they must dispense to all eligible patients who request it. We use instrumental variables to solve the potential endogeneity arising if GP decisions on whether to dispense are affected by unobservable factors affecting their prescribing. Physicians dispense drugs to, on average, 6% of their patients. We analyse whether physician dispensing leads to more prescriptions, pack size substitution and consequential, higher drug costs. We also look at the prescription of potentially harmful drugs, i.e. antibiotics and opioids. We identify the causal effect of physician dispensing by using an instrument based on patient distances from local pharmacies and their practice: we calculate the total miles saved by patients who live more than one mile away from their nearest pharmacy if their practice chooses to dispense.

Our work contributes to the literature in several ways. First, the specific type of market regulation, where the dispensing rights depend on patients' residence and not on GP practice location, has not been studied before in the economic literature. Second, we estimate causal effects using spatial and intertemporal variation and account for non-random allocation of physicians into the dispensing regime. Third, in addition to examining the effect on drug expenditures we provide evidence on the effects of dispensing by physicians on prescribing quantity, quality and pack size substitution. Fourth, we do not only study the extensive margin (the effects of a practice decision to open an on-site dispensary) but also the "intensive margin": the effects dispensing status on practices with varying numbers of patients to who they dispense.

Using quarterly data from all general practices in England for the calendar years 2011 to 2018 we find that practices which dispense have on average 4.3 percent higher expenditures on pharmaceuticals per patient. This effect is mainly driven by an increase in the number of prescriptions and a substitution to smaller pack sizes, which can be explained by the fixed-fee payment for each dispensed prescription. Dispensing physicians also prescribe more potentially harmful drugs, such as opioids, per patient. They also prescribe more Over the Counter (OTC) drugs which ought not to be prescribed routinely in primary care.

Our results fit into a broad literature on physician behaviour (McGuire, 2000) and, in particular, on supplier-induced demand (Iversen and Luras, 2000; Iversen, 2004; Clemens and Gottlieb, 2014). There is empirical literature that focuses on the impact of physician dispensing (for a systematic review see Lim et al. (2009)). The first stream studies the effect of the dispensing status on drug expenditures and reports mixed results. Chou et al. (2003); Kaiser and Schmid (2016); Burkhard et al. (2019) show that imperfect agency results in higher drug costs per patient. More recently, Goldacre et al. (2019) find that dispensing practices are more likely to prescribe high-cost drugs, especially when having a high share of dispensing patients. In contrast, Trottmann et al. (2016); Ahammer and Zilic (2017) find that physician dispensing is associated with lower drug expenditures per patient in the canton of Zurich and in Austria. The second stream of the literature focuses on the effect of drug mark-ups on physician prescribing behaviour. Using a dynamic probit model, Iizuka (2007, 2012) find that mark-ups influence the choice of anti-hypertensive drugs by Japanese physicians. Liu et al. (2009); Rischatsch et al. (2013) also show that higher mark-ups available for generic drugs than for brand-name drugs increase the prescription of generic drugs in dispensing practices. The third stream of the literature studies the effect of dispensing practice on prescribing quality, which is proxied by the use of antibiotics to treat bacterial infections. Park et al. (2005); Trap (2002); Filippini et al. (2014) find that dispensing increases antibiotic prescribing in South Korea, Switzerland and Zimbabwe.

The study is organised as follows. Section 4.2 provides the institutional background and the dispensing regulation in the NHS. Section 4.3 describes the data. Section 4.4 outlines the empirical methods. Section 4.5 shows the results. In section 4.6 we conclude.

4.2 Institutional setting

There is a list system in primary care in the English NHS. Patients register with a single general practice that acts as the gatekeeper to most other NHS services including non-emergency hospital care. Almost all general practices are small businesses owned and run by partnerships of GPs who share profits and losses. In 2018, there were 7,148 general practices in England with an average list size of 8,279 patients and 3.37 full time equivalent GPs.

The English NHS is funded almost entirely by taxation. There is a small patient charge (£8.80 in 2018/19) when a primary care prescription is dispensed, whether by a pharmacy or an on-site GP dispensary. Around half of the population are exempt from this charge on grounds of age (under 16, in full time education and under 18, or over 60), current or recent pregnancy, or specified medical conditions (House of Commons Library, 2020). As a result, approximately 90% of prescriptions are dispensed without charge.

General practices have contracts with the NHS under which they are paid via a mixture of capitation payments, quality incentives, and items of service such as vaccinations. They are reimbursed for the costs of their premises and information technology but meet all other costs, such as hiring practice staff, including salaried non-partner GPs, from their revenue. There are two main type of contracts. The General Medical Services (GMS) contract is negotiated centrally by the Department of Health & Social Care and the British Medical Association, the doctors' trade union. The most common non-GMS contract is the Primary Medical Services (PMS) contract, which is negotiated between individual practices and local healthcare purchasers, known as clinical commissioning groups (CCGs).[1] Under PMS contracts, the practice receives a lump sum for providing a set of services similar to those required by the GMS contract plus additional services for particular groups of patients.

4.2.1 General practice dispensing

Most patients who receive a drug prescription from their GP must take it to a pharmacy in contract with the NHS to have it dispensed. However, patients who would have serious difficulty in accessing a pharmacy or who live in a controlled

[1]Prior to 2012/13, healthcare was purchased by Primary Care Trusts. We will use the term CCG for both types of purchasers.

Figure 4.1: Small areas in England with at least one dispensing GP practice

area which has been designated as rural in character and who are more than 1 mile (1.6km) away from a pharmacy, can ask their general practice to dispense drugs to them.[2] The practice decision on whether to dispense is all or nothing: if it agrees to dispense for one eligible patient the practice must dispense to any eligible patient who requests it.

In 2018, 2.9m (5.5%) of 53.8m general practice patients were in the 916 dispensing practices (16.3% of all practices) which had agreed to dispense to their eligible patients. As Figure 4.1 shows, most dispensing practices are in rural areas, though there are some in urban areas where a small number of patients have claimed eligibility on grounds of serious difficulty in accessing community pharmacies.

Like community pharmacies, practices receive two main types of payments for dispensing:

(a) a fee per prescription they dispense which is independent of the type or quantity

[2]There are regulations restricting the entry of new pharmacies into rural areas (Department of Health, 2012) and attempts to enter are strongly resisted by local dispensing general practices.

of the drug.[3] Thus, a practice is paid more for dispensing two separate prescriptions on separate occasions each for a month's supply of a drug, than for a single prescription for two months' supply of the same drug. The dispensing fee declines with the total number of prescriptions dispensed by the practice in the financial year. The maximum dispensing fee per prescription is currently £1.99 and the minimum is £1.76.

(b) reimbursement for the dispensed drugs bought by the practice. The reimbursement is based on the manufacturers' list price of drugs, known as the Net Ingredient Cost (NIC), minus an adjustment for purchasing discounts. Dispensing practices can usually buy drugs from wholesalers at a discount, which depends on a number of factors such as volume purchased, temporary promotions, etc. In recognition of such discounts, the NHS reduces the reimbursement by a fraction of the NIC, known as the "clawback". The clawback increases with the total NIC of all drugs dispensed and ranges from around 3% to 11%, with most dispensing practices facing the full clawback of 11%. Even the maximum clawback is less than the discount that practices receive on the NIC.

In addition, dispensing practices can earn up to £2.58 per dispensing patient if they meet various process requirements of the Dispensary Services Quality Scheme. The scheme aims to ensure minimum competency standards for dispensing staff by mandating training requirements and standard operating procedures for dispensaries. It also requires practices to review medicine prescriptions for 10% of their dispensing patients each year.

Dispensing is profitable for dispensing practices. In 2017/18, partner GPs in dispensing practices had a mean pre-tax income of £121,000 compared with £111,700 for partner GPs in non-dispensing practices (NHS Digital, 2019).

4.2.2 Physician incentives

We assume that GPs in dispensing practices, like those in non-dispensing practices, are partially altruistic (McGuire, 2000) and care about the effect of their decisions on their income and on the well-being of their patients. Practices with patients who are eligible for dispensing (e.g. those living more than a mile from the nearest pharmacy) must decide whether to dispense to all or none of them. GPs also decide

[3] We define a prescription as a specified amount of a specified drug. The prescription *form* given to the patient by a GP could contain more than one prescription. A prescription form with prescriptions for two different drugs counts as two separate prescriptions.

what they prescribe and dispense to dispensing patients and what they prescribe to non-dispensing patients.[4] The resulting patient utilities for patient h in practice i in period t are b^d_{hit} if patient h is on the dispensing list and b^{nd}_{hit} if h is not on the dispensing list (either because the practice has chosen not to dispense, or, if the practice dispenses, h is not eligible or has not asked the practice to be put on the dispensing list). Patient utility depends not just on the prescriptions that they are given but also on the time, travel, and other costs of having their prescriptions dispensed in the practice or in a pharmacy. If their practice has eligible patients and decides to dispense, the change in utility for an eligible patient who requests that the practice dispense to them is $g^d_{hit} = b^d_{hit} - b^{nd}_{hit} \geq 0$.

Given the prescribing decision, the utility gain G to a practice i that has eligible patients and decides to dispense is

$$G_{it}(L^d_{it}) = \pi_{it}(L^d_{it}) - F_{it} + \alpha \sum\nolimits_{h \in DL_{it}} g^d_{hit} \tag{4.1}$$

where $\pi_{it}(L^d_{it})$ is the gross profit from dispensing to the L^d_{it} patients in the set of DL_{it} patients on the dispensing list, F_{it} is the fixed cost of running a dispensary, and $\alpha > 0$ is a parameter reflecting GP altruism. A dispensing practice cannot control the size (L^d_{it}) of its dispensing list. Once it has decided to dispense, it must dispense to any eligible patient who requests it.

A practice with eligible patients will decide to dispense if and only if $G_{it}(L^d_{it}) \geq 0$. The total utility gain to dispensing patients will increase with L^d_{it}. There are also likely to be economies of scale affecting gross profit since practices can achieve bigger discounts on larger drug purchases and use their specialist dispensary staff more productively. Thus, it is plausible that practices with more eligible patients are more likely to choose to dispense.

Gross profit from dispensing is

$$\pi_{it} = \sum\nolimits_k \sum\nolimits_\ell p_{\ell k} n_{it\ell k} + r(n_{it}) + mL^d_{it} - c(n_{it}) \tag{4.2}$$

[4]We assume that all patients who are given a prescription have it dispensed, either by their practice or by a pharmacy. Thus, we do not distinguish between prescribed and dispensed drugs. A proportion of prescriptions are never dispensed: the patient may subsequently decide that they have recovered and do not need the drug or that it is not worth paying the prescription charge if they are not exempt (Beardon et al., 1993). We assume that GPs can predict the probability that the prescription is not dispensed and allow for this in their decisions on prescribing and whether to open a dispensary.

where $p_{\ell k}$ is reimbursement from the NHS, net of the clawback, for drug k dispensed in pack size ℓ minus the wholesale price and $n_{it\ell k}$ is the number of prescriptions of drug type k and pack size ℓ. $r(n_{it})$ is dispensing fee income from dispensing $n_{it} = \sum_k \sum_\ell n_{it\ell k}$ prescriptions. m is payment per dispensing patient if the practice meets the quality standards of the Dispensing Services Quality Scheme. $c(n_{it})$ is the cost of the additional staff required to dispense n_{it} prescriptions.

Practices can increase revenue from dispensing in a number of ways. They can prescribe more expensive prescriptions with a higher NIC and hence a higher payment net of the clawback ($p_{\ell k}$). For example, they can prescribe proprietary drugs rather than cheaper equivalent generics. Second, they can increase the number of prescriptions ($n_{it\ell k}$) of a given drug of a given pack size, for example by prescribing antibiotics for upper respiratory tract infections. This will increase their reimbursement $p_{\ell k} n_{it\ell k}$ and also their dispensing fee income $r(n_{it})$. Third, since dispensing fee income varies with the number of prescriptions, not the quantities of drugs prescribed, practices will increase their dispensing fee income if they prescribe several smaller packages rather than a single larger package with the same total quantity of the drug. In some of these decisions GPs will be trading off patient utility against greater income. For example, prescribing smaller pack sizes will require patients to make more trips to the practice to receive the prescriptions and to have them dispensed. Other responses may increase patient utility as well as practice profit. For example, patients may regard receiving a prescription as a validation of their decision to consult the GP and so will be more satisfied (Zgierska et al., 2012; Ashworth et al., 2016).

If they respond to these incentives dispensing practices will have different prescribing patterns compared to non-dispensing practices: they will prescribe more expensive drugs, they will prescribe a greater total quantity of drugs, and, on average, each prescription will be for a smaller dose. In the next section, we describe how we measure practice prescribing patterns and our data on other practice characteristics, which we use to control for other factors that may affect prescribing.

4.3 Data

We link administrative data using GP practice identifiers to construct a quarterly panel of 7,979 practices for 2011Q1 to 2018Q4. The panel covers all practices in England, but is unbalanced due to market entries, exits and mergers. We exclude practice-quarter observations on very small practices (less than 1,000 patients) since these are likely to be in the process of reorganisation, which may affect their prescribing behaviour.[5] The final sample has 229,178 practice-quarter observations.[6] For each GP practice, we have data on their organisational structure, the characteristics of their patient population, and detailed information on their prescribing behaviour. Appendix Table D1 lists the data sources and reporting frequencies. Practices are clustered geographically in 195 Clinical Commissioning Groups (CCGs) with, on average, around 40 practices within each CCG.

4.3.1 Prescribing measures

We construct quarterly measures of practice prescribing from monthly prescribing data. The data covers all medicines, dressings and appliances prescribed by English practices and dispensed to patients anywhere in the United Kingdom. We will use the shorthand "drug" to cover all types of prescription (medicines, dressings and appliances). Drugs are labelled with a 15 digit British National Formulary (BNF) code which identifies the name of the drug, whether it is generic or branded, its formulation (e.g. capsule, tablet, liquid), its strength, and the quantity (number of pills, volume of liquid..). For each drug type (ie each BNF code) the data reports, for each month for each practice, the total number of prescriptions dispensed, the total quantity, and the total NIC. The latter is based on the the list price for the drug excluding VAT, and does not take account of discounts, dispensing costs, or prescription charge income.

The monthly prescribing data do not differentiate between prescriptions issued to dispensing and non-dispensing patients, nor by whether the prescription was dispensed in a community pharmacy or practice on-site dispensary. Our prescribing variables are therefore measured at practice level and are a weighted average of prescribing for dispensing and non-dispensing patients. Since we do observe the number of dispensing and non-dispensing patients in each practice we use this to test our hypotheses about the effects of having dispensing patients (see the discussion of Methods in Section 4.4).

Our discussion in Section 4.2.2 suggests that practices with on-site dispensaries are likely to have greater drug costs per patient because they gain financially by

[5] We also exclude two practices that report unusually high prevalence rates for chronic diseases. These practices appear to serve very unusual patient populations.

[6] This is 94.1% of the initial sample prior to applying exclusion criteria.

prescribing more expensive drugs and by prescribing a greater amount of drugs. Because they also receive a fee per prescription dispensed they have an incentive to write more prescriptions with smaller drug quantities. We construct nine measures of practice prescribing to test these hypotheses. (Details are in Appendix Table D2)

Four measures are based on overall practice prescribing: total NIC per patient, total NIC per prescription, the number of prescription per patient, and relative pack size. The latter compares the quantity of drugs per prescription of a given type with the modal pack size across all prescriptions of this drug across all dispensing and non-dispensing practices in England. We expect this to be smaller in dispensing practices because the fee per prescription dispensed creates an incentive to prescribe a given quantity of a given quantity of a drug in several small packs rather than in a single large pack. To allow for this when examining the effect of dispensing total NIC per prescription and on the number of prescriptions, we adjust the number of prescriptions by the relative pack size so that for a practice with small relative pack sizes the adjusted number of prescriptions is smaller than the number observed.

We also test whether dispensing status affects prescribing of particular types of drugs. Generic drugs are usually cheaper than equivalent patented proprietary versions and GPs are encouraged by the NHS to prescribe generic drugs whenever possible. But prescribing cheaper drugs is likely to reduce dispensing practice income and so we compare the percentage of generic prescriptions in dispensing and non-dispensing practices. Over the counter (OTC) drugs can be bought by patients without the need for a prescription from a GP and usually treat minor or short-term conditions. They are widely available in supermarkets, petrol stations and community pharmacies. To reduce costs, the NHS discourages GPs from prescribing OTCs. Dispensing practices, however, will lose financially if they reduce the number of OTC prescriptions that are dispensed in their on-site pharmacies. We compare the number of OTC prescriptions per patient in dispensing and non-dispensing practices. In both prescribing measures we again adjust the number of prescriptions by relative pack sizes to allow for the possibility that dispensing practices may prescribe more but smaller packages.

Overuse of some drugs, such as opioids, antidepressants and antibiotics, can be harmful to patients and to wider population health (e.g. through increased antimicrobial drug resistance). The NHS, like many other healthcare systems, discourages GPs from prescribing these drugs if there is no clear medical reason to do so. We measure the number of prescriptions per patient for each of these three potentially harmful drugs to test dispensing practices also take financial incentives into account when prescribing them. We again adjust the number of prescriptions by relative pack size for each type of drug.

4.3.2 Practices and their registered patient populations

We have quarterly data on practice list size and the number of their dispensing patients. We measure practice organisational structure using annual data on the number of full time equivalent GPs, the proportion of them who are partners (and so residual claimants on practice profit), rather than salaried, their age, gender and whether they qualified in the UK or elsewhere. We also know the type of practice contract with the NHS (GMS vs all other types).
To control for differences in patient case-mix across practices, we use data on patient demographics (fourteen age by gender categories). We attribute a measure of average patient deprivation to practice by using the proportions of practice patients resident in each Lower Super Output Area (LSOA).[7] and the Index of Multiple Deprivation (IMD) for each LSOA. Finally, we use annual prevalence data for 12 major diseases treated in primary care, which are reported for each practice as part a national pay for performance programme - the Quality and Outcomes Framework (QOF).

4.3.3 Eligible population for dispensing services

The observed number of dispensing patients in a practice will be less than the true eligible population if some eligible patients do not request dispensing services or if the practice decides not to dispense. We construct two variables to investigate practice decisions on whether to provide a dispensing service to the eligible patients who request it. The first is the number of people in the surrounding area who are eligible to request dispensing on the grounds of distance from their nearest pharmacy. We calculate the straight-line distance from the centroid of every Output Area (OA)[8] in England to the nearest community pharmacy and compare it to the 1-mile threshold for eligibility. The median area covered by OAs in England is 0.03 square miles (7 hectares), which suggests that OA centroids are a reasonable approximation to the location of individual residents. We then compute the eligible population for each

[7] There are 34.753 LSOAs in England with a mean population of 1500 (ONS, 2012).
[8] There are 181,408 in England and Wales with an average population of 309 (ONS, 2012).

practice as the total population of OAs whose centroids are within 3 miles (4.8km)[9] of the practice and more than 1km from the nearest community pharmacy.
The second measure is the total travel distance per prescription which would be saved by the eligible population within 3km of the practice if it had a dispensary. For each eligible OA (with centroid more than 1 mile from the nearest pharmacy) within 3km of a practice we calculate the difference between twice the straight line distance from the centroid of the OA to the pracice (i.e. the minimum distance a patient would have to travel if the practice dispenses) and the sum of the distances from the OA centroid to the practice, from the practice to the pharmacy nearest to the patient, and from the nearest pharmacy to the OA centroid (i.e. the minimum distance a patient would have to travel to obtain a prescription and to have it dispensed in a community pharmacy). We then multiply this quantity by the population of the OA and sum over all OAs within 3 miles (4.8km) of the practice to give an estimate of the potential miles saved by the local population if the practice had an on-site dispensary. This provides us with a measure of the total potential demand for dispensing services in the area around the practice.

4.4 Methods

We study the effect of GP dispensing on our prescribing measures both at the extensive margin (i.e. whether a practice has any patients to whom it dispenses) and at the intensive margin (i.e. the proportion of a practice's patient list to whom it has agreed to dispense).

4.4.1 Extensive margin

To study the effect at the extensive margin, we define GP practices as dispensing practices if they have a positive number of patients on their list for which they have registered dispensing rights with the NHS. Our baseline regression model takes the form

$$y_{ijt} = \alpha + \beta x_{it} + \delta D_i + \omega_t + \gamma_j + \varepsilon_{ijt} \tag{4.3}$$

[9]Santos et al. (2017) report that patients in England are registered with general practices that are, on average, 1.2 miles (1.9km) away from their home. We use a slightly larger radius to better capture the wider catchment area of practices in rural areas.

where y_{ijt} is the prescribing measure for practice i in CCG j in quarter t, x_{it} is a row vector of practice and patient characteristics, D_i is an indicator for practice dispensing status, ω_t are quarter t fixed effects, γ_j are CCG fixed effects, and ε_{ijt} is a zero mean error. The coefficient of interest is δ, which measures the difference in prescribing behaviour between dispensing and non-dispensing practices.

A practice's dispensing status may vary over time due to changes in their patient population or local market entry of community pharmacies. However, during our study period, dispensing status was essentially time-invariant, with only 95 out of 7,850 (1.2%) practices in our sample changing status over time. We therefore model dispensing status as a time-invariant practice characteristic and exclude practices that switch dispensing status throughout the study period. This also precludes the use of practice fixed effects, which would allow us to control for time-invariant unobserved practice factors not picked up in x_{it}. In contrast, a model with CCG fixed effects can be estimated since dispensing status varies sufficiently across, on average, around 40 practices within each CCG. These CCG fixed effects are expected to absorb a large part of unobserved heterogeneity. CCGs can influence the prescribing of its practices via its clinical governance procedures and local prescribing incentive schemes. Local hospital provision and policies can also affect practice prescribing. For example practices in areas where patients wait longer for hip replacements may prescribe more pain killers. The CCG effect will also allow for local area characteristics, such as air quality, availability of outside spaces, housing stock, etc. which may affect patient morbidity and, hence, prescribing. Finally, it will control for the overall rurality of an area which will influence patient eligibility for dispensing and thus practice decisions about dispensing status.

Equation 4.3 includes a rich set of GP and patient characteristics in x_{it} which helps to reduce concerns over differences in patient or GP characteristics that affect prescribing and the practices' decisions to dispense. However, there may still be residual differences in observed and unobserved characteristics between dispensing and non-dispensing practices that would bias our estimates. We therefore expand our analytical strategies in two directions.

First, we pre-process our data using Entropy Balancing (EB) techniques as proposed by Hainmueller (2012) to reduce imbalance in observable characteristics that may determine selection into treatment.[10] EB is a re-weighting approach in which

[10] We use the user-written Stata command `ebalance` (Hainmueller and Xu, 2013) to estimate EB weights.

the weights for each observation are chosen to approximately equalise pre-specified moments of the covariate distributions between the treatment and control groups. The pre-processed data are then analysed using conventional weighted least squares controlling for observable characteristics. The resulting estimates have been shown to have doubly-robust properties (Zhao and Percival, 2016), which reduces model dependence and mis-specification bias (Robins, 1994). On the assumption that observable and unobservable practice characteristics are highly correlated, EB yields unbiased estimates of the effect of treatment (here: dispensing status). In our case, we require the weighted data to exhibit balance in both the means and the variances of all control variables including CCG membership. By definition, no set of weights can satisfy these requirements for practices in CCGs without any (or with only) dispensing practices, and practices in these CCGs are therefore dropped from the sample. The subsequent analysis of the re-weighted data recovers the average treatment effect on the treated (ATT), which is informative about the likely consequences of a hypothetical policy that disallows dispensing by English GP practices.

Second, we adopt an instrumental variable (IV) strategy to control for potential selection into dispensing status based on unobservable patient or practice characteristics. Note that the regulatory framework of the NHS prohibits practices who do not treat eligible patients (e.g. those located in urban areas with close proximity to community pharmacies) from opting into dispensing services. Hence, we analyse a situation with potential one-sided non-compliance, which results in a downward biased estimate of δ. To address endogenous selection, we estimate two-stage least squares (2SLS) models on our unweighted data using the potential local demand for dispensing services that each GP practice faces as an instrument for observed dispensing status. Specifically, our IV Z_i is defined as the total miles saved by the local resident population eligible for dispensing services who live within a 3 mile radius of practice i (see Section 4.3.3). Under the assumption that GPs are at least partially altruistic, we expect practices who can achieve a more significant reduction in the travel burden of their patient population to be more likely to select into dispensing, all else equal. Figure 4.2 demonstrates a strong positive, monotonic relationship between miles saved by the local eligible population Z_i and dispensing status D_i. The resulting estimates are local average treatment effects (LATE) and, therefore, are informative for the policy question of what would happen if all English GP practices (that respond to the IV) were allowed to dispense.

The validity of our IV may be compromised if the potential number of eligible pa-

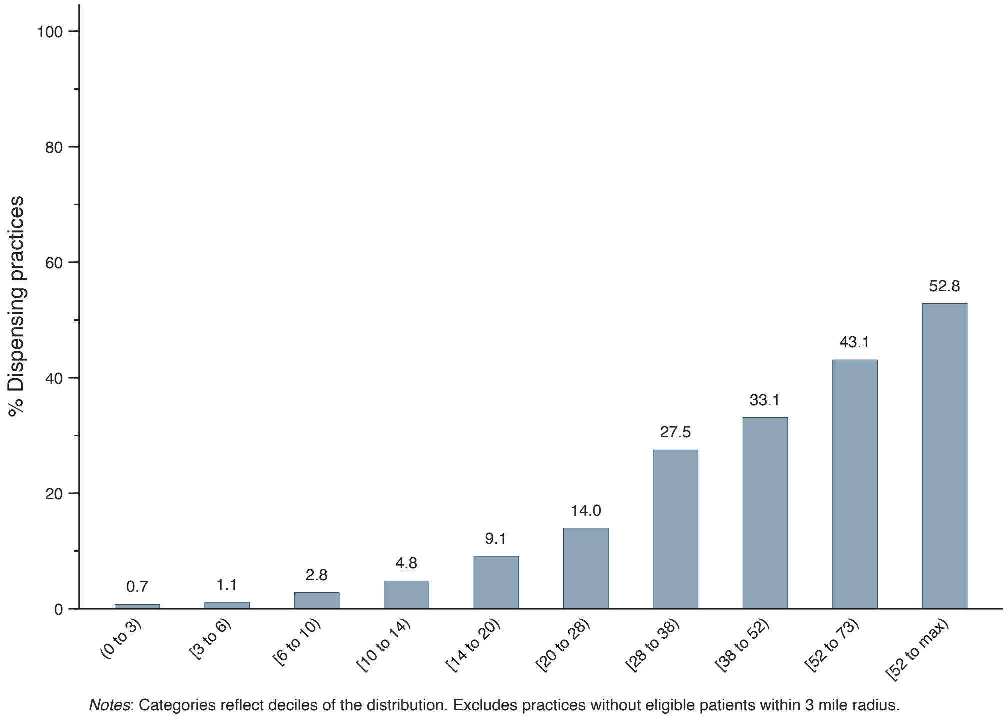


Figure 4.2: Probability of dispensing by potential miles saved (3 mile radius around the practice)

Potential miles saved is the difference between the straight line distance from the patient's OA centroid to the practice and back and the sum of the distances via the nearest pharmacy, which is then multiplied with the eligible population in this OA and summed over all OAs within a 3 mile radius.

tients in the local vicinity of the practice is in itself correlated with its prescribing behaviour. This may be the case if practices compete for patients and seek to attract demand by prescribing over-generously. However, GP practices located in rural areas, where dispensing status is potentially endogenous, face little competition for patients and, therefore, have less need to attract patients than their urban counterparts. This is consistent with Schaumans (2015), who finds that Belgian GPs write fewer prescriptions per patient contact in areas where there was less competition. Furthermore, prescribing behaviour is difficult to ascertain for patients at the stage of practice selection since these data are not reported publicly, e.g. on websites such as nhs.net. This reduces the scope for quality competition based on prescribing. We therefore argue that our IV is both strong and valid.

All standard errors are clustered at practice level to account for serial correlation.

In the context of the EB analysis, inference is conditional on the estimated weights.

4.4.2 Intensive margin

Our weighted prescribing measures are averages at practice level where the weights are the proportions of dispensing and non-dispensing patients. Thus,

$$y_{ijt} \equiv s_{ijt}y^{d}_{ijt} + (1 - s_{ijt})y^{nd}_{ijt} = y^{nd}_{ijt} + s_{ijt}(y^{d}_{ijt} - y^{nd}_{ijt}) = y^{nd}_{ijt} + s_{ijt}\Delta_{ijt} \tag{4.4}$$

where y^{d}_{ijt} and y^{nd}_{ijt} are the prescribing measures for dispensing and non-dispensing patients, Δ_{ijt} denotes the difference in prescribing measures between both patient groups, and $s_{ijt} = L^{d}_{ijt}/L_{ijt}$ is the proportion of patients to whom the practice dispenses.

The analysis at the extensive margin recovers Δ_{ijt} at $\bar{s}|s > 0$, i.e. at the average value of s_{ijt} over all dispensing practices. We now seek to establish whether Δ_{ijt} is in itself a function of s_{ijt} and thus varies across the range $s_{ijt} \in (0, 1]$. This would indicate that GP practices change their relative prescribing behaviour as the proportion of dispensing patients in the practice changes, for example because dispensing income becomes a more important source of overall practice income or, conversely, because the fixed costs of operating a dispensary are distributed over more patients. For simplicity, we assume that Δ_{ijt} is approximately linear in s_{ijt} so that

$$\Delta_{ijt} = \phi_1 + \phi_2 s_{ijt} \tag{4.5}$$

where ϕ_1 is the constant difference in prescribing measures between dispensing and non-dispensing patients at any level of s_{ijt}, and ϕ_2 is a behavioural parameter that reflects the responsiveness of GPs' prescribing behaviours to the share of dispensing patients in their practices.

We estimate an intensive margin model for the sub-sample of practice-quarter observations with a positive number of dispensing patients to recover ϕ_1 and ϕ_2. We exploit variation within practices in terms of the number of patients to whom they can dispense drugs assuming that this varies exogenously within and between practices once a dispensary has been established. This is a reasonable assumption since dispensing practices cannot exclude eligible patients from drug dispensing services without terminating dispensing entirely. GP practice dispensing status is largely time-invariant and so any inter-temporal variation in the number of patients receiving dispensing services is solely the effect of variation in local demand and not an effect of selection.

We use similar covariates to the extensive margin model in Eq. (4.3) but replace the binary dispensing status variable with the share of dispensing patients s_{ijt} and its squared value so that

$$y_{ijt} = \alpha_i + \beta x_{it} + \phi_1 s_{ijt} + \theta s_{ijt}^2 + \omega_t + \varepsilon_{ijt} \tag{4.6}$$

where α_i is a practice fixed effect and $\theta = \frac{1}{2} * \phi_2$. Taking the first derivative of Eq (4.6) with respect to s_{ijt} yields the expression in Eq. (4.5). Note that if $\theta = 0$ then Δ_{ijt} is constant in s_{ijt} and the model collapses to the extensive margin model in Eq. (4.3). Since we have limited within GP practice variation in s_{ijt}, we also estimate a model with CCG fixed effects instead of practice fixed effects.

4.5 Results

4.5.1 Descriptive statistics

Table 4.1 presents the descriptive statistics, by practice dispensing status, for the prescription measures, practice and patient characteristics as well as the distance measure. Figure D1 in the Appendix presents kernel density plots for the prescribing indicators. Thirteen percent of practices (N=1,023) in our sample have dispensing patients. These practices prescribe, on average, more and smaller packages than non-dispensing practices, and they have higher net ingredient costs (NIC) per patient. Dispensing practices also appear to be more homogeneous in their prescribing behaviour as indicated by the lower variance of the prescribing indicators. The mean number of GPs educated in the UK is slightly higher for dispensing practices and the share of female patients between 20 and 44 is slightly lower. Most other practice characteristics, such as their patient populations' age structure, the number of patients per GP, as well as their list sizes are similar.

The average dispensing practice has dispensing rights for approximately 3,100 patients, or 49% of their list (Figure 4.3). There is substantial variation in the number of dispensing patients across practices, with some practices being allowed to dispense for up to 10,000 patients. The share of dispensing patients in dispensing practices ranges from nearly 0% to 100% and has a bimodal distribution.

	Dispensing practices		Non-dispensing practices	
	Mean	SD	Mean	SD
Prescribing measures				
Cost per patient	41.34	7.78	37.01	10.21
Cost per prescription	6.73	0.70	6.74	0.80
Prescriptions per patient	6.19	1.32	5.55	1.60
OTC prescriptions per patient	0.92	0.25	0.90	0.30
Antibiotic prescriptions per patient	0.18	0.04	0.17	0.08
Opioid prescriptions per patient	0.23	0.08	0.21	0.11
Antidepressant prescriptions per patient	0.36	0.09	0.31	0.13
Relative pack size	1.11	0.16	1.28	0.23
% generic prescriptions	0.80	0.06	0.79	0.06
Organisational structure of practice				
List size	7873.76	4707.96	7377.11	4466.22
log(list size)	8.80	0.59	8.73	0.61
Full-time equivalent GPs per 1,000 patients	0.52	0.30	0.46	0.27
GP partners (%)	0.69	0.24	0.67	0.30
UK-trained GPs (%)	0.65	0.36	0.51	0.38
Contract type				
GMS	0.74	0.44	0.60	0.49
other (incl. PMS)	0.26	0.44	0.40	0.49
Age structure of GPs (headcount)				
Age <40	0.27	0.22	0.29	0.26
Age 40 to 59	0.67	0.24	0.56	0.31
Age 60+	0.07	0.16	0.16	0.27
Patient population				
Demographic composition by age-sex band (%)				
Male - 0 to 4	0.02	0.01	0.03	0.01
Male - 5 to 19	0.08	0.01	0.09	0.02
Male - 20 to 44	0.13	0.02	0.18	0.05
Male - 45 to 59	0.11	0.01	0.10	0.02
Male - 60 to 74	0.10	0.02	0.07	0.02
Male - 75 to 84	0.03	0.01	0.02	0.01
Male - 85+	0.01	0.00	0.01	0.00
Female - 0 to 4	0.02	0.01	0.03	0.01
Female - 5 to 19	0.08	0.01	0.08	0.02
Female - 20 to 44	0.13	0.02	0.18	0.04
Female - 45 to 59	0.11	0.01	0.10	0.02
Female - 60 to 74	0.10	0.02	0.07	0.02
Female - 75 to 84	0.04	0.01	0.03	0.01
Female - 85+	0.02	0.01	0.01	0.01
Prevalence of chronic conditions or major health shocks (per 1000)				
Coronary heart disease	3.80	0.91	3.21	1.16
Stroke	2.11	0.55	1.63	0.65
Hypertension	16.05	2.83	13.62	3.63
Chronic obstructive pulmonary disease	1.82	0.63	1.86	0.94
Cancer	2.93	0.78	2.02	0.85
Mental health problems	0.65	0.22	0.94	0.48
Asthma	6.47	0.98	5.84	1.34
Heart failure	0.85	0.34	0.73	0.35
Palliative care	0.35	0.37	0.28	0.31
Dementia	0.75	0.37	0.63	0.43
Atrial fibrillation	2.21	0.60	1.49	0.71
Cardiovascular disease (aged 30-74)	1.92	1.11	1.75	1.02
Index of Multiple Deprivation (2015)	0.09	0.03	0.17	0.08
Potential miles saved if dispensing (in 1000)				
3 mile radius	57.17	32.08	16.27	23.73
Practices	6,857		1,023	
Practice-quarter observations	197,915		31,263	

Note: Data based on quarter-practice level 2011Q1 to 2018Q4. Data sources can be found in the Appendix in Table D1.

Table 4.1: Descriptive statistics by practice dispensing status

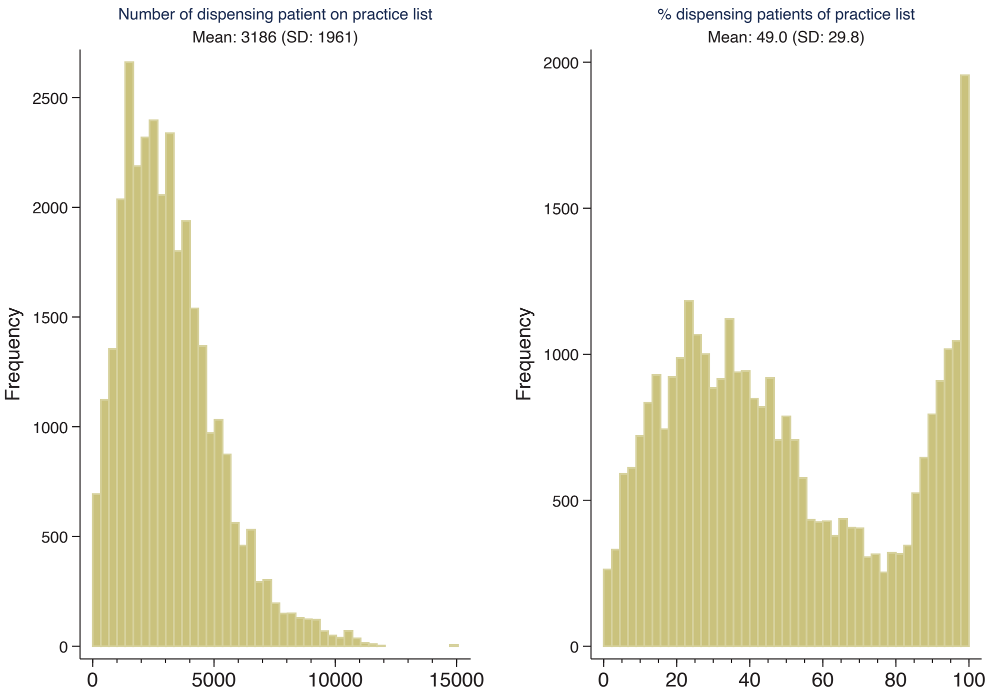


Figure 4.3: Histogram of (i) number and (ii) share of dispensing patients in dispensing practices

4.5.2 Extensive margin

Table 4.2 presents the effects of dispensing status on our prescribing measures. The OLS results are consistent with the predictions set out in Section 4.2.1. Dispensing practices maximise dispensing fee income with 0.07 more prescriptions per quarter and with, on average, 22.7% smaller packages. The average NIC per prescription is £0.09 higher than in non-dispensing practices, which is in part explained by the lower use of cheaper generic drugs (-0.03%). Dispensing practices prescribe more OTC drugs, which is a pre-requisite for these to be reimbursed, as well as more antibiotics and opioids, but fewer antidepressants.[11]. Overall, these differences in prescribing behaviour are associated with an additional expenditure of £1.04 per patient on the list and calendar quarter, or 2.8% of the mean quarterly expenditure

[11]While the effects for antibiotics, antidepressants and opioids appear to be small, it is worth noting that there are usually only few patients in a practice in need of these specific types of medication. For example, the Adult Psychiatric Morbidity Survey 2014 estimates that approximately 3.3% of people aged 16 or over experience depression in any given week; only a subset of which will receive pharmacological treatment.

per patient for non-dispensing practices. For a dispensing practice of average list size, this amounts to £32,587 of additional prescribing expenditure per year.

	Pooled OLS		EB + WLS		2SLS	
	Est	SE	Est	SE	Est	SE
Cost per patient	1.038***	0.153	1.605***	0.189	0.824	0.619
Cost per prescription	0.091***	0.016	0.108***	0.021	0.078	0.068
Prescriptions per patient	0.066***	0.023	0.138***	0.030	0.135	0.097
OTCs prescriptions per patient	0.064***	0.005	0.049***	0.006	0.042**	0.021
Antibiotic prescriptions per patient	0.004***	0.002	0.002	0.002	0.021**	0.009
Opioid prescriptions per patient	0.004**	0.002	0.009***	0.003	0.014**	0.007
Antidepressant prescriptions per patient	-0.005**	0.002	0.010***	0.003	0.020**	0.009
Relative pack size	-0.227***	0.007	-0.207***	0.011	-0.262***	0.028
% generic prescriptions	-0.003***	0.001	-0.005***	0.002	0.0002	0.006
Practice-quarter observations	229,178		130,113		229,178	
Partial F-test of excluded instrument					195.7	
Test of endogeneity (p-value)					0.497	

*** p<0.01, ** p<0.05, * p<0.1
Est = Coefficient estimate; SE = Standard error; OTC = Over-the-counter; OLS = Ordinary Least Squares; EB = Entropy balancing; WLS = Weighted Least Squares; 2SLS = Two-stage Least Squares.
Note: All models control for a full set of control variables for characteristics of the patient population and the organisational structure of the practice. Quarterly data 2011Q1 to 2018Q4 (Sources see Table D1).

Table 4.2: Effect of dispensing status at the extensive margin

There are considerable differences in observable characteristics of dispensing and non-dispensing practices, which may not be fully accounted for by the OLS regression adjustment. EB pre-processing successfully equalises the first two moments of the covariate distributions (see Appendix Table D3 for descriptive statistics). The resulting WLS estimates are generally in line with the OLS estimates, although there are two noteworthy differences. First, the effect of dispensing status on both the prescription costs and the number of prescriptions issued is approximately 50% larger than under OLS. The increase in the latter appears to be driven nearly entirely by increased prescribing of non-OTC medications, which now account for 65% (=1 - 0.049/0.138) of the additional prescribing volume associated with dispensing status. Second, we now find dispensing status to be linked to increased prescribing of antidepressants, whereas there is no longer a measurable effect on antibiotic prescribing.

Finally, we turn to the 2SLS estimates which allow for selection into treatment due to unobservable practice characteristics and identify LATEs. We use as an IV the total potential travel distance saved for the resident population within a 3-mile radius around the practice if the practice operated an on-site dispensary and all residents were registered with this practice. (Semi-)altruistic practice owners are expected to consider possible reductions in travel burden for their patient population when

deciding whether to opt into dispensing (see Section 4.2.2). However, differences in travel burden between both states of the world are unlikely to affect prescribing behaviour directly, i.e. they are uncorrelated with unobserved patient morbidity or provider characteristics not already adjusted for. We find our IV to be a strong predictor of dispensing status with a partial F-statistic of 195.7 (see Appendix Table D4 for first stage results). The resulting point estimates are broadly consistent with the OLS and EB+WLS estimates, but the lower efficiency of the 2SLS estimator means that some estimates are no longer statistically different from zero. As before, we find evidence that dispensing practices prescribing statistically significantly more OTCs as well as specific medications (i.e. opioids, antidepressants, antibiotics). They also choose, on average, smaller pack sizes than non-dispensing practices. We do not find statistically significant difference in the overall number of prescriptions, the cost per prescription, or the overall cost per patient (all $p > 0.10$). However, since the robust score test of endogeneity (Wooldridge, 1995) fails to reject the null hypothesis of conditional exogeneity (p=0.497) we prefer the more efficient OLS estimates to judge the ATE of dispensing status.
Table 4.3 shows that our results are robust to alternative adjustments for selection on observables, namely inverse probability weighting with regression adjustment (IPWRA). The method allows calculating ATT and ATEs under the assumption that practices do not select into dispensing based on unobservable characteristics.

	IPWRA (ATT)		IPWRA (ATE)	
	Est	SE	Est	SE
Cost per patient	1.615***	0.067	1.819***	0.109
Cost per prescription	0.109***	0.006	0.016	0.012
Prescriptions per patient	0.132***	0.013	0.291***	0.026
OTC prescriptions per patient	0.051***	0.002	0.063***	0.005
Antibiotic prescriptions per patient	0.003***	0.000	0.005***	0.001
Opioid prescriptions per patient	0.010***	0.001	0.028***	0.001
Antidepressant prescriptions per patient	0.011***	0.001	0.017***	0.001
Relative pack size	-0.218***	0.003	-0.103***	0.004
% generic prescriptions	-0.006***	0.001	-0.007***	0.001
Practice-quarter observations	229,178		229,178	

*** p<0.01, ** p<0.05, * p<0.1
Est = Coefficient estimate; SE = Standard error; OTC = Over-the-counter; IPWRA = Inverse probability weighting with regression adjustment. ATE = Average treatment effect; ATT = Average treatment effect on the treated.
Note: All models control for a full set of control variables for characteristics of the patient population and the organisational structure of the practice. Quarterly data 2011Q1 to 2018Q4 (Sources see Table D1). All estimations were performed using Stata's `teffects` routine.

Table 4.3: Extensive margin - alternative adjustment for selection on observables

4.5.3 Intensive margin

Table 4.4 presents the results of our analysis of the intensive margin. For each prescribing measure we present two coefficient estimates: the difference in prescribing between dispensing and non-dispensing patients that is independent of the actual share of dispensing patients in the practice (ϕ_1), and the estimated marginal effect of increasing the share of dispensing patients in the practice by 100 percentage points (ϕ_2). For ease of interpretation, the share variable s_{ijt} is mean centered so that ϕ_2 denotes deviations away from the sample mean of approximately 49%.

Most (97.5%) of the variation in the share of dispensing patients in our sample is between GP practices so that point estimates are poorly identified once GP fixed effects are introduced. We therefore prefer to focus on the model with CCG fixed effects. In line with the pooled OLS and EB+WLS results at the extensive margin (Table 4.2) we find that dispensing status increases prescribing costs, prescriptions per patient (both OTC and non-OTC), and the cost per prescription but reduces pack size and % generic prescribing. In addition, our estimates of θ (and thus ϕ_2) suggest a diminishing marginal effect of dispensing share on most prescribing measures as evidenced by the opposing signs of ϕ_1 and θ. Put differently, GPs in dispensing practices prescribe more similarly to their peers in non-dispensing practices when the share of patients for which they can dispense is high. However, this diminishing effect is often small in magnitude.[12] Figure 4.4 plots the predicted values of parametric regression models assuming either a linear or parabolic relationship between s_{ijt} and the prescribing measures of interest. In addition, a spline model with 20 equally spaced knots serves as a non-parametric approximation of these relationships.

4.6 Conclusion

The English NHS is one of few healthcare systems in high-income countries that permits GPs to prescribe *and* dispense medicines under specific circumstances. GP practices are allowed to dispense prescriptions to patients who live more than 1 mile away from the nearest community pharmacy or who otherwise struggle to access a pharmacy. We present evidence that English GPs respond to the financial incentives created by geographic, exogenous variation in dispensing rights. Practices

[12]Table D5 in the Appendix reports the regression coefficients of a model assuming that prescribing behaviour is constant in s_{ijt}.

	Pooled OLS		CCG fixed effects		GP fixed effects	
	Est	SE	Est	SE	Est	SE
Cost per patient						
ϕ_1	2.630***	0.593	2.175***	0.517	1.276	1.148
θ	-5.449***	1.682	-3.355**	1.392	2.739	2.582
Cost per prescription						
ϕ_1	0.229***	0.067	0.142**	0.058	0.282*	0.154
θ	-0.655***	0.186	-0.325**	0.146	0.374	0.362
Prescriptions per patient						
ϕ_1	0.257***	0.090	0.272***	0.082	0.083	0.212
θ	-0.028	0.217	-0.078	0.207	-0.233	0.557
OTC prescriptions per patient						
ϕ_1	0.049**	0.023	0.072***	0.018	0.028	0.031
θ	-0.075	0.048	-0.124***	0.043	-0.011	0.103
Antibiotic prescriptions per patient						
ϕ_1	0.008*	0.004	0.009**	0.004	-0.001	0.008
θ	-0.025**	0.012	-0.020*	0.012	0.001	0.019
Opioid prescriptions per patient						
ϕ_1	0.006	0.006	0.007	0.006	0.009	0.009
θ	-0.016	0.017	-0.005	0.016	0.023	0.022
Antidepressant prescriptions per patient						
ϕ_1	0.011	0.008	0.001	0.007	-0.002	0.013
θ	-0.026	0.023	-0.007	0.020	0.035	0.032
Relative pack size						
ϕ_1	-0.163***	0.020	-0.161***	0.018	-0.036	0.035
θ	0.359***	0.051	0.292***	0.044	0.129	0.103
% generic prescriptions						
ϕ_1	-0.014***	0.005	-0.012**	0.005	-0.007	0.019
θ	-0.049***	0.014	-0.036***	0.013	-0.074	0.045
Practice-quarter observations	31,263		31,263		31,263	

*** p<0.01, ** p<0.05, * p<0.1
Est = Coefficient estimate; SE = Standard error.
All models control for a full set of control variables for characteristics of the patient population and the organisational structure of the practice.
Note: ϕ_1 and $\theta = \frac{1}{2}\phi_2$ denote the regression coefficients on the share of dispensing patients s_{ijt} and s^2_{ijt}, respectively. See Section 4.4.2 for further details.

Table 4.4: Effect at intensive margin

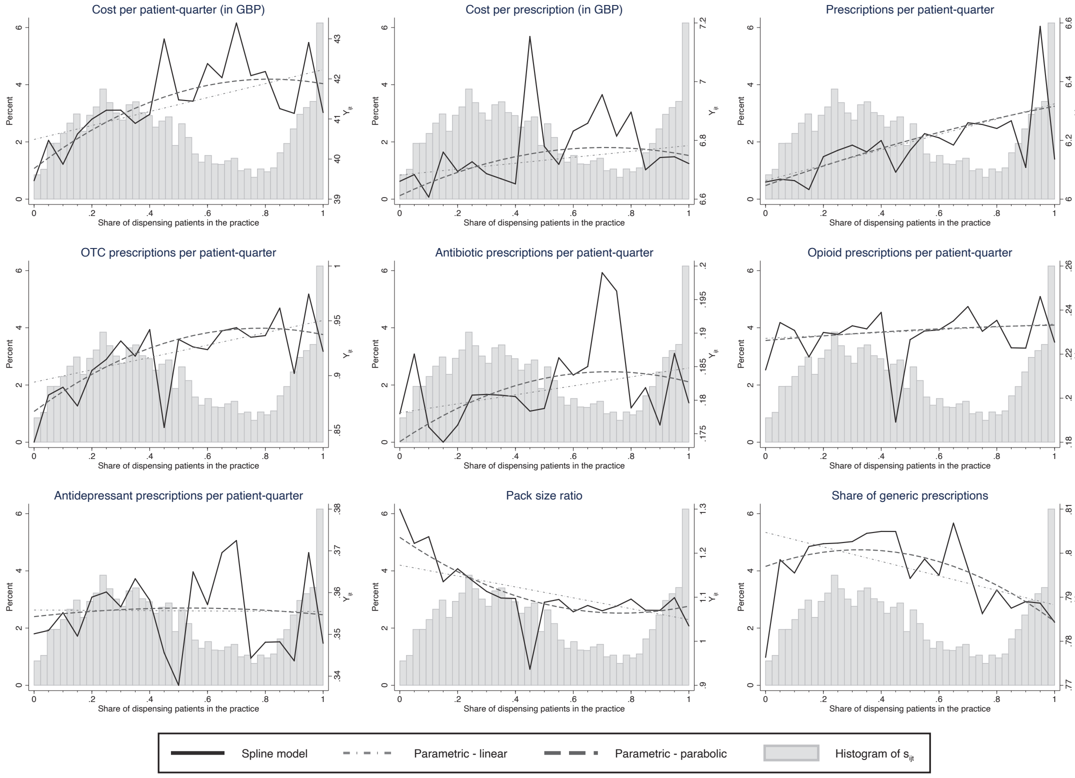


Figure 4.4: Effect of dispensing status on prescribing measures over different shares of dispensing patients in the practice

with a on-site dispensary prescribe differently compared to non-dispensing practices. Specifically, they prescribe, on average, more often and in smaller pack sizes. We first present the average treatment effect on the treated applying entropy balancing with weighted least squares. This presents effects of a policy that disallows dispensing by English GP practices. Second, we estimate a local average treatment effect (LATE) using a 2SLS framework to understand what would happen if all English GP practices (that respond to the instrumental variable (IV)) were allowed to dispense. The IV exploits the idea that physicians behave altruistically and open dispensaries to reduce the travel distance for their patients. As expected, the 2SLS results exceed the OLS estimates of dispensing status (where significant), but are less precisely measured. They also point in the same direction as the ATT results.
Third, we observe an exogenously varying share of dispensing patients within and across practices with dispensaries. We exploit this variation to estimate effects at the intensive margin. We find that marginal effects are largest for small to medium

shares of dispensing patients.

Our estimates present average effects across all patients, where we cannot differentiate prescriptions by eligibility. One could argue that doctors treat all patients equally and prescribe more to everybody when running a dispensary. However, it is also possible that doctors distinguish patient types in their practice and the average effects underestimate the true effect on GPs' prescribing behaviours for eligible patients. The analysis at the intensive margin shows that marginal effects decrease with a higher share of eligible patients where there is less need to gain more per eligible patient. Thus, the effect is most pronounced when practices have relative few eligible patients. While we do not observe the reasons for this behaviour, we note that his behaviour is consistent with an attempt to recover the fixed costs of running a dispensary and, potentially, to compensate for lower wholesale margins.

Our analysis implies three policy conclusions: First, when dispensaries are opened, the number of eligible patients should be sufficiently high to cover the running cost. However, this does not imply that everybody should be eligible in a practice with dispensary since the number and costs of prescribing increase across all patients. Second, the fee per dispensed item should be sufficiently small to not incentivise practices to prescribe inefficiently more drugs. Not only producing more cost to the NHS, this may also harm patients' health. Prescribing smaller packages may increase the number of visits to get a new prescription and thus may crowd-out other patient visits. Third, it may be useful to monitor OTC prescribing to reduce this effect.

Bibliography

Ahammer, A. and I. Zilic, "Do Financial Incentives Alter Physician Prescription Behavior? Evidence from Random Patient-GP Allocations," *Working paper*, 2017, *2017.*

Ashworth, Mark, Patrick White, Hannah Jongsma, Peter Schofield, and David Armstrong, "Antibiotic prescribing and patient satisfaction in primary care in England: cross-sectional analysis of national patient survey data and prescribing data," *British Journal of General Practice*, 2016, *66* (642), e40–e46.

Beardon, P.H., M.M. McGilchrist, A.D. McKendrick, D.G. McDevitt, and T.M. MacDonald, "Primary non-compliance with prescribed medication in primary care.," *British Medical Journal*, 1993, *307* (6908), 846–848.

Burkhard, D., C. Schmid, and K. Wüthrich, "Financial incentives and physician prescription behavior: Evidence from dispensing regulations," *Health Economics*, 2019.

Chou, YJ, W.C Yip, C.-H. Lee, N. Huang, Y.-P. Sun, and H.-J. Chang, "Impact of separating drug prescribing and dispensing on provider behaviour: Taiwan's experience," *Health Policy and Planning*, 2003, *18* (3), 316–329.

Clemens, J. and J. Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

Department of Health, "Regulations under the Health Act 2009: Market entry by means of Pharmaceutical Needs Assessments: Information for Primary Care Trusts Executive Summary and Chapters 1-4," 2012. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/212861/Chapters-1-4-introduction-background.pdf`. Accessed: 24.06.2019.

Filippini, M., F. Heimsch, and G. Masiero, "Antibiotic consumption and the role of dispensing physicians," *Regional Science and Urban Economics*, 2014, *49*, 242–251.

Goldacre, B., C. Reynolds, A. Powell-Smith, A.J. Walker, T.A. Yates, R. Croker, and L. Smeeth, "Do doctors in dispensing practices with a financial conflict of interest prescribe more expensive drugs? A cross-sectional analysis of English primary care prescribing data," *BMJ Open*, 2019, *9* (2), e026886.

Hainmueller, J., "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies," *Political analysis*, 2012, pp. 25–46.

_ **and Y. Xu**, "Ebalance: A Stata package for entropy balancing," *Journal of Statistical Software*, 2013, *54* (7).

House of Commons Library, "House of Commons Library briefing on NHS charges.," 2020. `https://commonslibrary.parliament.uk/research-briefings/cbp-7227/`. Accessed: 13.07.2020.

Iizuka, T., "Experts' agency problems: evidence from the prescription drug market in Japan," *The RAND Journal of Economics*, 2007, *38* (3), 844–862.

_ , "Physician Agency and Adoption of Generic Pharmaceuticals," *American Economic Review*, 2012, *102* (6), 2826–2858.

Iversen, T., "The effects of a patient shortage on general practitioners' future income and list of patients," *Journal of Health Economics*, 2004, *23* (4), 673–694.

_ **and H. Luras**, "The effect of capitation on GPs' referral decisions," *Health Economics*, 2000, *9* (3), 199–210.

Kaiser, B. and C. Schmid, "Does Physician Dispensing Increase Drug Expenditures? Empirical Evidence from Switzerland," *Health Economics*, 2016, *25* (1), 71–90.

Lim, D., J. Emery, J. Lewis, and V.B. Sunderland, "A systematic review of the literature comparing the practices of dispensing and non-dispensing doctors," *Health Policy*, 2009, *92* (1), 1–9.

Liu, Y.-M., Y.-H. Yang, and C.-R. Hsieh, "Financial incentives and physicians' prescription decisions on the choice between brand-name and generic drugs: Evidence from Taiwan," *Journal of Health Economics*, 2009, *28* (2), 341–349.

McGuire, T.G., "Chapter 9 Physician agency," in "Handbook of Health Economics," Vol. 1 2000, pp. 461–536.

NHS Digital, "GP Earnings and Expenses Estimates 2017/18," 2019. https://digital.nhs.uk/data-and-information/publications/statistical/gp-earnings-and-expenses-estimates/2017-18. Accessed: 14.07.2020.

ONS, "2011 Census: Population and Household Estimates for Small Areas in England and Wales, March 2011," 2012. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouse//holdestimatesforsmallareasinenglandandwales/2012-11-23. Accessed: 13.07.2020.

Park, S., S.B. Soumerai, A.S. Adams, J.A. Finkelstein, S. Jang, and D. Ross-Degnan, "Antibiotic use following a Korean national policy to prohibit medication dispensing by physicians," *Health Policy and Planning*, 2005, *20* (5), 302–309.

Rischatsch, M., M. Trottmann, and P. Zweifel, "Generic substitution, financial interests, and imperfect agency," *International Journal of Health Care Finance and Economics*, 2013, *13* (2), 115–138.

Robins, J.M., "Correcting for non-compliance in randomized trials using structural nested mean models," *Communications in Statistics-Theory and methods*, 1994, *23* (8), 2379–2412.

Santos, R., H. Gravelle, and C. Propper, "Does Quality Affect Patients' Choice of Doctor? Evidence from England," *Economic journal (London, England)*, 2017, *127* (600), 445–494.

Schaumans, C., "Prescribing behavior of general practitioners: competition matters," *Health Policy*, 2015, *119* (4), 456–463.

Trap, B., "Prescription habits of dispensing and non-dispensing doctors in Zimbabwe," *Health Policy and Planning*, 2002, *17* (3), 288–295.

Trottmann, M., M. Frueh, H. Telser, and O. Reich, "Physician drug dispensing in Switzerland: association on health care expenditures and utilization," *BMC Health Services Research*, 2016, *16* (1), 1976.

Wooldridge, J.M., "Score diagnostics for linear models estimated by two stage least squares," *Advances in econometrics and quantitative economics: Essays in honor of Professor CR Rao*, 1995, pp. 66–87.

Zgierska, A., M. Miller, and D. Rabago, "Patient satisfaction, prescription drug abuse, and potential unintended consequences," *Jama*, 2012, *307* (13), 1377–1378.

Zhao, Q. and D. Percival, "Entropy balancing is doubly robust," *Journal of Causal Inference*, 2016, *5* (1).

4.7 Appendix

4.7.1 Data sources

Data	Frequency	Source
Prescriptions per GP practice	Monthly	NHSBSA
Number of dispensing patients per practice	Quarterly	NHSBSA
Patient demographics	Annual	NHS Digital
GP characteristics	Annual	NHS Digital
QOF disease prevalences	Annual	NHS Digital
Population density	Annual	ONS

Abbreviations: NHSBSA = NHS Business Services Authority; ONS = Office for National Statistics; QOF = Quality & Outcomes Framework.

Table D1: Data sources and reporting frequencies

4.7.2 Practice prescribing measures

Prescribing measure	Definition
Cost per patient	$(\sum_k \sum_\ell N_{k\ell it} C_{k\ell t}) / / L_{it} = C_{it} / L_{it}$
Cost per prescription	C_{it} / N_{it}^A
Relative pack size	$\sum_k \left(\frac{N_{kit}}{N_{it}}\right) \left(\frac{1}{M_k} \frac{\sum_\ell N_{k\ell it} Q_{k\ell t}}{\sum_\ell N_{kit}}\right)$ $= \sum_k \left(\frac{N_{kit}}{N_{it}}\right) RPS_{itk}$
Prescriptions per patient	$(\sum_k N_{itk} RPS_{itk}) / L_{it}$ $= \left(\sum_k N_{ikt}^A\right) / L_{it} = N_{it}^A / L_{it}$
% generic prescriptions	$\sum_{k \in generic} N_{kit}^A \Big/ N_{it}^A$
Variable	**Definition**
k	drug name and formulation
ℓ	pack size category
i	practice
t	period (quarter)
L_{it}	list size of practice i in period t
$N_{k\ell it}$	number of prescriptions (items) of drug k pack size ℓ
N_{it}	total prescriptions
$C_{k\ell t}$	net ingredient cost drug k in pack size ℓ
C_{it}	total net ingredient cost
L_{it}	list size of practice i in period t
$Q_{k\ell t}$	quantity drug k supplied in pack size ℓ
M_k	Modal pack size drug k (over all periods, practices)
$RPS_{itk} = \left(\frac{1}{M_k} \frac{\sum_\ell N_{k\ell it} Q_{k\ell t}}{\sum_\ell N_{kit}}\right)$	Relative pack size drug k, practice i, period t
$N_{itk}^A = N_{itk} RPS_{itk}$	Quantity adjusted number of items drug k
$N_{it}^A = \sum_k N_{itk} RPS_{itk}$	Total quantity adjusted items

Note: Prescribing volume of specific medications (i.e. antibiotics, opioids, and antidepressants) and OTCs are defined in the same way as prescriptions per patient.

Table D2: Definition of prescribing measures

4.7.3 Density outcomes

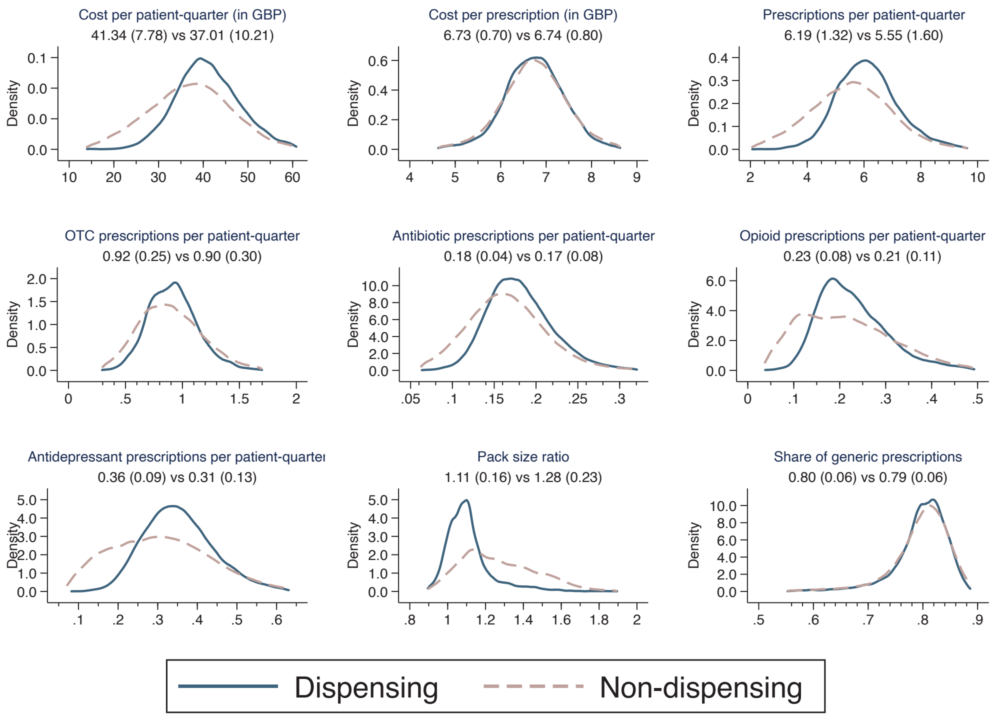


Figure D1: Kernel densities for prescribing measures

Note: Shown densities exclude top and bottom percentiles of sample values. Mean and SD are reported for the full sample.

4.7.4 Mean and SD after EB+WLS

	Unweighted				Weighted	
	Treated		Control		Control	
	Mean	SD	Mean	SD	Mean	SD
Organisational structure of practice						
log(list size)	8.80	0.59	8.85	0.61	8.80	0.59
Full-time equivalent GPs per 1,000 patients	0.53	0.30	0.46	0.27	0.53	0.30
GP partners (%)	0.69	0.24	0.68	0.29	0.69	0.24
UK-trained GPs (%)	0.65	0.36	0.53	0.38	0.65	0.36
Contract type						
GMS	0.75	0.43	0.62	0.49	0.75	0.43
other (incl. PMS)	0.25	0.43	0.38	0.49	0.25	0.43
Age structure of GPs (headcount)						
Age <40	0.27	0.22	0.28	0.24	0.27	0.22
Age 40 to 59	0.66	0.25	0.60	0.28	0.66	0.25
Age 60+	0.07	0.16	0.12	0.24	0.07	0.16
Patient population						
Demographic composition by age-sex band (%)						
Male - 0 to 4	0.02	0.01	0.03	0.01	0.02	0.01
Male - 5 to 19	0.08	0.01	0.09	0.01	0.08	0.01
Male - 20 to 44	0.13	0.02	0.17	0.04	0.13	0.02
Male - 45 to 59	0.11	0.01	0.10	0.01	0.11	0.01
Male - 60 to 74	0.10	0.02	0.08	0.02	0.10	0.02
Male - 75 to 84	0.03	0.01	0.03	0.01	0.03	0.01
Male - 85+	0.01	0.00	0.01	0.00	0.01	0.00
Female - 0 to 4	0.02	0.01	0.03	0.01	0.02	0.01
Female - 5 to 19	0.08	0.01	0.08	0.01	0.08	0.01
Female - 20 to 44	0.13	0.02	0.16	0.03	0.13	0.02
Female - 45 to 599	0.11	0.01	0.10	0.02	0.11	0.01
Female - 60 to 74	0.10	0.02	0.08	0.02	0.10	0.02
Female - 75 to 84	0.04	0.01	0.03	0.01	0.04	0.01
Female - 85+	0.02	0.01	0.01	0.01	0.02	0.01
Prevalence of chronic conditions or major health shocks (per 1000)						
Coronary heart disease	3.78	0.91	3.46	1.06	3.78	0.91
Stroke	2.10	0.55	1.85	0.62	2.10	0.55
Hypertension	15.99	2.82	14.32	3.42	15.99	2.82
Chronic obstructive pulmonary disease	1.81	0.62	1.96	0.84	1.81	0.62
Cancer	2.92	0.77	2.28	0.83	2.92	0.77
Mental health problems	0.64	0.22	0.85	0.35	0.64	0.22
Asthma	6.45	0.96	6.15	1.15	6.45	0.96
Heart failure	0.85	0.34	0.78	0.35	0.85	0.34
Palliative caree	0.35	0.37	0.30	0.31	0.35	0.37
Dementia	0.74	0.36	0.71	0.42	0.74	0.36
Atrial fibrillation	2.20	0.59	1.73	0.67	2.20	0.59
Cardiovascular disease (aged 30-74)	1.92	1.10	1.77	1.03	1.92	1.10
Index of Multiple Deprivation (2015)	0.09	0.03	0.14	0.07	0.09	0.03
Practice-quarter observations	30,203		99,910		99,910	

Note: CCG membership is also balanced after weighting but is not reported here. Data based on quarter-practice level 2011Q1 to 2018Q4. Data sources can be found in the Appendix in Table D1.

Table D3: Descriptive statistics after EB+WLS

4.7.5 2SLS first-stage results

	Est	SE
Potential miles saved if dispensing (in 1000)	0.003***	0.000
log(list size)	-0.021***	0.006
Full-time equivalent GPs per 1,000 patient	0.058***	0.009
GP partners (%)	-0.033***	0.008
UK-trained GPs (%)	0.022***	0.007
PMS contract	-0.010	0.006
Age structure of GPs (%)		
Age 40 to 59	0.010	0.009
Age 60+	-0.030***	0.011
Demographic composition by age-sex band (%)		
Male - 0 to 4	-1.014*	0.527
Male - 5 to 19	0.459	0.390
Male - 45 to 59	-1.133***	0.297
Male - 60 to 74	6.905***	0.486
Male - 75 to 84	7.272***	0.876
Male - 85+	4.368***	1.690
Female - 0 to 4	-0.949*	0.527
Female - 5 to 19	1.408***	0.391
Female - 20 to 44	-0.133	0.176
Female - 45 to 59	0.931***	0.278
Female - 60 to 74	-0.986**	0.485
Female - 75 to 84	-8.096***	0.775
Female - 85+	-8.632***	0.918
Prevalence of chronic conditions or major health shocks (per 1000)		
Coronary heart disease	-0.039***	0.006
Stroke	0.002	0.011
Hypertension	-0.004**	0.002
Chronic obstructive pulmonary disease	-0.028***	0.005
Cancer	0.014*	0.007
Mental health problems	-0.020***	0.007
Asthma	0.008***	0.003
Heart failure	-0.032**	0.013
Palliative care	0.019**	0.008
Dementia	0.028***	0.009
Atrial fibrillation	0.062***	0.011
Cardiovascular disease (aged 30-74)	0.006**	0.003
Index of Multiple Deprivation (2015)	-0.077	0.070
Practice-quarter observations	229,178	

Table D4: Predictors of observed dispensing status

4.7.6 Alternative specification - Intensive margin

	Pooled OLS		GP fixed effects		CCG fixed effects	
	Est	SE	Est	SE	Est	SE
Cost per patient	1.989***	0.585	1.508	1.113	1.742***	0.504
Cost per prescription	0.152**	0.065	0.313**	0.150	0.101*	0.057
Prescriptions per patient	0.254***	0.092	0.063	0.192	0.261***	0.085
OTC prescriptions per patient	0.040*	0.021	0.027	0.033	0.056***	0.017
Antibiotic prescriptions per patient	0.005	0.004	-0.001	0.008	0.007	0.004
Opioid prescriptions per patient	0.004	0.006	0.011	0.008	0.006	0.006
Antidepressant prescriptions per patient	0.008	0.008	0.001	0.013	0.000	0.007
Relative pack size	-0.120***	0.018	-0.025	0.033	-0.124***	0.017
% generic prescriptions	-0.019***	0.005	-0.013	0.018	-0.016***	0.005
Practice-quarter observations	31,263		31,263		31,263	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Est = Coefficient estimate; SE = Standard error.
All models control for a full set of control variables for characteristics of the patient population and the organisational structure of the practice.

Table D5: Effect at intensive margin - assumed linear relationship

5

Incentives not to adopt NHS prescribing guidelines

5.1 Introduction

NHS spending on prescribed medicines that patients can purchase privately over the counter (OTC drugs) sums up to £569 million in 2016 (NHS England, 2018). To reduce these spending on low-value care, NHS England released guidelines on OTC prescribing in primary care in December 2017. England is one country that allows medicines to be distributed through two types of dispensaries. A regular community pharmacy and an on-site general practitioner (GP) dispensary where a GP can dispense drugs to patients who live more than a mile away from the next pharmacy. For some of the practices, the market for pharmaceuticals diagnosis and service (in terms of drug redemption) is therefore combined. This may lead to own financial incentives for dispensing doctors compared to non-dispensing doctors and might be in contradiction to a long-run overall cost-efficient prescribing aim of a policymaker. To study whether dispensing GPs behave differently to non-dispensing GPs, we exploit the introduction of over-the-counter (OTC) prescribing guidelines as an exogenous policy change. Using this variation as a 'natural experiment', we analyse if dispensing doctors follow prescribing guidelines implemented by the policymaker in the same way as non-dispensing doctors.

This paper compares pre- and post-reform OTC prescribing in England with non-affected prescribing behaviour in Wales and identify guideline adoption in general and across dispensing and non-dispensing doctors. Key to our estimation is the exogenous policy change in the NHS England for all general practitioners. Our work contributes to the literature in several ways. First, the specific market regulation of dispensing doctors has not been studied over time in a setting where a policy change was not strictly enforced before. This gives scope for an individual adaption decision and allows a within practice comparison. Second, we estimate causal effects using a difference-in-difference approach where we can directly compare dispensing doctors' behaviour with non-dispensing doctors' behaviour using non-affected Welsh practices. Third, we not only make a statement about one specific aspect but also provide a broad overview of different prescribing indicators and show underlying substitution mechanisms. Fourth, it is the first economic study evaluating the OTC policy in general.

We observe monthly prescription data from all general practitioners in the NHS England and Wales from April 2015 to September 2019. We control for a mix of patient characteristics, such as major disease categories and general practice characteristics. To avoid unobservable heterogeneity in our model, we use practice fixed effects. We find that doctors adopt the prescribing guideline in general even without strict enforcement. Our results show that the reduction in OTC volume of dispensing doctors is significantly smaller compared to non-dispensing doctors. However, the reduction in OTC expenses is not significantly different for dispensing and non-dispensing doctors. This is due to the result that dispensing doctors prescribe less expensive OTC drugs compared to non-dispensing doctors after the reform. Since NHS only monitors OTC expenses and not volume, there seems to be no adoption difference between dispensing and non-dispensing doctors for the policymaker. The rewarding system for dispensing doctors with a fixed item component might be one possible reason for this finding. Package size substitution does not only matter for one doctor type. Both doctor types substitute to smaller package sizes. This effect is not surprising since the NHS monitors OTC expenses. If a doctor decides to still satisfy a patient request for an OTC prescription it is efficient to prescribe a smaller package to decrease expenses. Substitution to non-OTC drugs does not seem to matter. Moreover, results indicate a possible spillover effect in terms of lower non-OTC expenses after the OTC policy introduction. Our results fit into a broad literature on guideline adoption. Rashidian et al. (2008) find that policymakers should not use strict enforcement regularly. Strict enforcement should be a tool for interventions where a softer policy does not work because otherwise, it can increase barriers to adopt guidelines (Raisch, 1990). However, there are important factors in the implementation process that policymakers need to consider to see an uptake of guidelines[1]. The NHS considered these factors in the implementation of the OTC guideline, such as credibility of the source, GP involvement, and explicitly named cost-containment motives.

In particular, my results contribute to the literature on the dispensing practice of doctors in primary care. The first stream studies the effect of the dispensing status on drug expenditures where evidence is mixed. Chou et al. (2003); Kaiser and Schmid (2016); Burkhard et al. (2019) find an imperfect agency that results

[1]The credibility of the source is important. Well-known national bodies such as the National Institute for Clinical Excellence (NICE) should plan guidelines (Fairhurst and Huby, 1998; Rashidian et al., 2008). It is also important to publish the guidelines in credible sources (Carthy et al., 2000). General practitioners do not perceive guidelines from the pharmaceutical industry as a credible source (Choudhry et al., 2002). The credibility of the content is another important factor for uptake. The involvement of GPs in the development of the guidelines makes adoption easier (Rashidian et al., 2008). If there is a cost-containment motive behind the guideline, it should be explicitly defined in the guidelines (Mayer and Piterman, 1999).

in higher drug costs per patient due to induced demand. More recently, Goldacre et al. (2019) find that dispensing practices are more likely to prescribe high-cost drugs, especially when having a high share of dispensing patients. In contrast to the results, Trottmann et al. (2016) and Ahammer and Zilic (2017) find that physician dispensing is associated with lower drug expenditures per patient in the canton of Zurich and Austria. The second stream of the literature focuses on the effect of markups on physician prescribing behaviour. Using a dynamic probit model, Iizuka (2007, 2012) finds that mark-ups influence the choice of antihypertensive drugs by Japanese physicians. Liu et al. (2009) and Rischatsch et al. (2013) also show that the mark-up differences between branded drugs and generic drugs influence prescribing behaviour in favour of generic drugs. The third stream of the literature studies the effect of dispensing on prescribing quality, with antibiotic drugs as a proxy. Park et al. (2005); Trap et al. (2002); as well as Filippini et al. (2014) find that dispensing increases antibiotic prescribing in South Korea, Switzerland, and Zimbabwe[2]. Recently Bodnar et al. (mimeo) find that dispensing doctors in England prescribe more volume compared to non-dispensing doctors and especially substitute to smaller package sizes due to a fixed item fee.

The article is organised as follows. Section 5.2 describes the English institutional background, with a focus on the dispensing regulation, as well as the policy implementation. Section 5.3 presents the data with the main variables of interest and Section 5.4 presents the econometric framework. Section 5.5 shows the main results of the paper where we (1), evaluate the policy adoption in general and (2), analyse the behaviour of dispensing doctors compared to non-dispensing doctors. Section 5.6 concludes.

5.2 General institutional background

To receive primary care, individuals must register with a specific general practitioner (GP) in the NHS England. In most cases, GPs own the practice themselves and contract with the NHS to provide primary care to their patients (Health and Social Care Information Center, 2015). Practices are assigned to a Clinical Commissioning Group (CCG), which is responsible for the planning and commissioning of health care services for their specific local area. In total, there are 191 CCGs in England (NHS Clinical Commissioners, n.d.).

[2]Parts of this section are borrowed from Bodnar et al. (mimeo).

The NHS is financed almost entirely from general taxation and almost free for patients. In primary care in England, patients need to pay a fee for dispensed drugs, which is £9 per dispensed item. According to the government, 50% of people in England are exempt from any charges, which results in 87% of all items prescribed without any charges (House of Commons Health Committee, 2006).
NHS England allows some of its physicians to distribute drugs through on-site dispensaries to some of their patients. Operating an on-site dispensary is only possible in a few other countries such as Switzerland, Austria, Japan and lately discussed in Germany. In England, this fact is rooted in historic regulations from at least the 1920s. To operate an on-site dispensary, a GP needs a request of a patient whose residence is more than one mile (1.6 km) away from the next pharmacy (Regulation 48(3)(a)). This implies that a doctor can only dispense drugs to some of his patients. However, the GP is not obliged to operate an on-site pharmacy even if he receives a request from a patient. He can decide whether he wants to meet this demand. However, if there is no eligible patient, the GP is not allowed to operate an on-site pharmacy, even if he wants to do so (Department of Health, 2012). Around 1,000 out of 8,000 practices have dispensing rights. Figure 5.1 shows the Lower Layer Super Output Areas (LSOAs) in which at least one doctor has dispensing rights. The figure clearly shows that practices are scattered across the country.

5.2.1 Reform in low-value care prescribing

In general, all GPs enjoy autonomy over the prescribing decision and there is only a very limited scope for pharmacies to substitute. However, Clinical Commissioning Groups (CCGs) demanded a nationally co-ordinated approach to provide some commissioning guidance to decrease expenditure growth in primary care. This approach restricts inefficient prescribing. The main reason is to allocate the NHS budget to more important areas. At the beginning of 2017, the NHS England and the NHS Clinical Commissioners (NHSCC) formed a joint working group to create guidelines about prescribing recommendations. This group consists of various national stakeholders, for example, the Royal College of GPs, Royal Pharmaceutical Society, the National Institute for Health and Care Excellence (NICE) and the Department of Health and Social Care. From July to December 2017, the NHS England discussed prescribing guidelines for prescription and OTC drugs. This discussion resulted in two consultations. First, prescribing guidelines for prescription drugs and second,

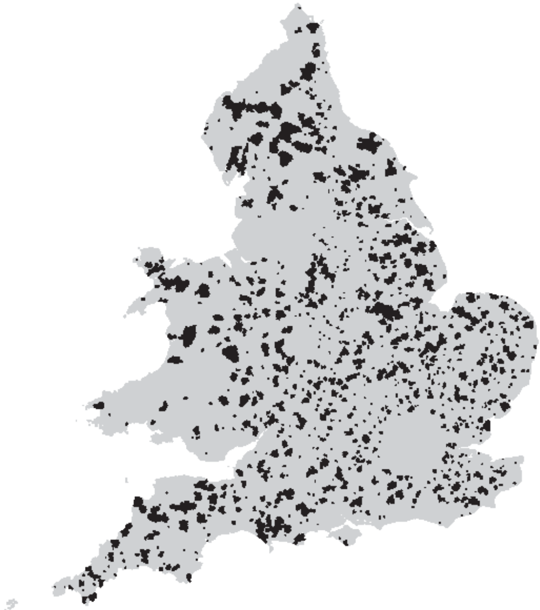


Figure 5.1: Small areas in England and Wales with at least one dispensing GP practice

prescribing guidelines for OTC drugs. In this paper, we will only focus on the latter OTC guideline.
The OTC guideline restricts general prescription of medicines which are 'readily' available over the counter[3]. The idea behind the guideline, published by NHS England and NHS Clinical Commissioners, is a restriction of prescribing OTC medicines that belong to one of the three following areas: (1) a self-limiting condition (2) a condition that is suitable for self-care and (3) items of limited clinical effectiveness. Based on several proposed approaches the clinical working group decided to map OTC products to conditions for which they are usually prescribed[4]. This approach makes the guideline easier to understand and follow for patients and doctors. From December 20, 2017, to March 14, 2018, the OTC guideline was publicly consulted on. CCGs provided information material for practices and patients. To analyse the effectiveness, NHS England announced that it would monitor prescribing data (NHS Clinical Commissioners, 2018).

[3]However, there are some exceptions to the guideline, for example, when the OTC drug is needed for the treatment of a chronic illness, a more complex form of a minor illness (e.g., migraines) or a patient who cannot treat himself. For a detailed overview of exceptions see NHS Clinical Commissioners (2018).

[4]An overview of these conditions can be found in NHS Clinical Commissioners (2018).

Besides, Professor Stephen Powis, the National Medical Director of the NHS England, sent a letter to GPs in January 2019. The letter states:

> 'The OTC guidance includes specific reference to prescribers, and requires prescribers to reflect local policies in prescribing practice. In NHS England's view, this guidance is 'relevant guidance' under Regulation 94 and other relevant regulatory references. Contractors are therefore required to have regard to this guidance and are able to follow the guidance [..] without any risk that they will be in breach of their contract.' (NHS England, 2019)

5.3 Data

We link several administrative datasets to construct a monthly panel of GP practices in England and Wales from April 2015 to September 2019. The constructed panel covers all GP practices in England and Wales. It is not balanced due to entries and exits of some practices during the observation period. We follow Bodnar et al. (mimeo) and exclude practices with less than 1,000 patients because they seem to be in a re-organization process and might behave differently. The final sample includes 355,068 practice-month observations. For each GP practice in England and Wales, we use data on their prescribing facts (e.g. volume, expenses), the organizational structure, and patient information. Detailed source information can be found in the Appendix in Section 5.7.1.

5.3.1 Prescribing data

The NHS Business Services Authority (NHSBSA) provides a monthly list of all medicines, dressings, and appliances that are prescribed by English practices in the NHS and are dispensed in a community pharmacy or an on-site GP dispensary. Prescriptions that are issued in England but are dispensed outside the country are also included in this data source. NHS Wales also provides the same information for Welsh practices. Practice level prescribing data records information about the total volume prescribed within one practice, as well as information about the total net ingredient cost[5] and total quantity (e.g., total number of pills, applications, ml).

[5] According to NHS Digital (2018) 'Total net ingredient costs (NIC) is the basic cost of a drug in pounds as used in primary care, which is the list price excluding VAT. It does not account for

All prescriptions are uniquely identified by their British National Formula (BNF) code, which is a 15-digit code that allocates the medical product to the categories in the BNF. This makes it possible to identify drugs as over-the-counter drugs using a list provided by NHSBSA[6].

The volume refers to a single item of a medicine, dressing or appliance prescribed on a prescription form. This means that, if a prescription form includes two medicines, these count as two items prescribed. However, the British National Formula (BNF) does not define a package size. To receive this information, we calculate the adjusted volume dispensed based on a 'standard pack size', which we estimate from the data[7].

We calculate three indicators to analyse guideline adoption and three indicators to analyse a possible substitution pattern behind the adoption process. We aggregate all indicators to the practice-month level and standardise it by the list size of the practice. Table 5.1 shows an overview of all indicators. OTC expenditures per patient are the total net ingredient costs of all OTC drugs prescribed by a practice in one month standardised by the list size of the practice. NHS monitors these expenses. OTC volume per patient is the single supply of a medicine, dressing or appliance prescribed on a prescription form not adjusted by the quantity. For example, the variable 'volume' counts a prescription with 20 pills of ibuprofen, as well as 40 pills as one prescription. Meaning the volume is one in both cases. The dispensing fee depends on the volume prescribed independently on the content (e.g., therefore counting a 20 or 40 pill package as one in that particular case). However, to get information about the length of treatment, we also use a quantity adjusted volume per patient. In the example before, this means the quantity adjusted volume is one in the case of 20 pills, but two in the case of 40 pills. The first substitution indicator, package size directly indicates whether a practice prescribes on average smaller or larger packages compared to the standardised package. The second sub-

---

any discounts, dispensing costs or prescription charge income, which makes it the standardised cost throughout prescribing nationally, and allows comparisons of data from different sources.'

[6]As explained in Section 2.1. the clinical working group mapped OTC products to conditions to make adoption easier. However, a drawback of the analysis is that we cannot map the drug to the actual condition. Nevertheless, the specified areas (1) a self-limiting condition (2) a condition that is suitable for self-care and (3) items of limited clinical effectiveness, apply to almost all OTC drugs.

[7]Adjusted volume is calculated as follows based on Bodnar et al. (mimeo): First, we divide the total quantity (e.g., pills, ml) by the total number of prescriptions, which provides the average pack size for drug k in month t for each practice i. In the next step, we calculate the mode package size over the whole dataset over all practices, months and years. The standard package size is used to calculate the quantity adjusted volume. To make a statement about package size substitution of each general practice, the quantity-adjusted volume is divided by the total volume.

Indicator	Category
OTC expenditure per patient	Adoption
OTC volume per patient	Adoption
OTC quantity adjusted volume per patient	Adoption
Package size	Substitution
Average cost per unit	Substitution
Non-OTC expenditure per patient	Substitution

Table 5.1: Prescription indicators

stitution indicator average cost per unit shows the price per pill or ml. We calculate non-OTC expenditures per patient based on the net ingredient costs, but only for prescription drugs.

We choose the adoption indicators to analyse a change in prescribing behaviour after the policy introduction. We expect OTC expenditures per patient, OTC volume per patient and quantity adjusted volume per patient to decrease after the guideline introduction for English GP practices. One might expect that the uptake is rather small in magnitude because the OTC guideline is not strictly enforced. However, as already discusses in the literature part in Section 5.1, NHS considered specific factors in the implementation of the OTC guideline. These factors are the credibility of the source, GP involvement, and explicitly named cost-containment motives, which increase the adoption of guidelines (Fairhurst and Huby, 1998; Rischatsch et al., 2013; Carthy et al., 2000; Mayer and Piterman, 1999). Additionally, NHS monitors OTC expenses. However, the adoption process may need some time. We, therefore, expect a larger decrease at the end of the observation period.
If we find a decrease in OTC prescribing of English GP practices after the guideline introduction, it is interesting to know whether there is an underlying mechanism behind the adoption. We expect English GPs to substitute to smaller packages. If a doctor still decides to satisfy a patient's request for an OTC prescription, it is better to prescribe a smaller package. If a GP reduces the number of pills he prescribes this decreases OTC expenses, which are monitored. A second possibility for English GPs to lower OTC expenses, but still prescribe OTC drugs, is to decrease the average cost per unit. That means GPs substitute to less expensive OTC drugs. A third possibility to still satisfy a patient request for a prescription is a substitution to

non-OTC drugs. We, therefore, expect non-OTC expenditures to increase after the OTC policy introduction[8].

As already explained in Section 5.2, NHS England allows medicines to be distributed through an on-site GP dispensary if a patient lives more than one mile away from the next pharmacy. This rule results in two different types of GPs in England, as well as in Wales. Since non-dispensing GPs have no own financial incentives to prescribe OTC drugs, we expect them to follow the guidelines accordingly. Compared to non-dispensing GPs, GPs operating an on-site dispensary have their own financial incentives. The dispensary is one income source if items are constantly prescribed and afterward dispensed. Previous literature shows that dispensing doctors follow these incentives and prescribe more items Chou et al. (2003); Kaiser and Schmid (2016); Burkhard et al. (2019); Bodnar et al. (mimeo) show that the rewarding system with a fixed items fee component in England especially leads to substitution to smaller package sizes. We, therefore, expect dispensing GPs to decrease OTC volume prescribed significantly less compared to non-dispensing doctors. Since NHS monitors OTC expenses, we expect no difference in the reduction between dispensing and non-dispensing GPs. However, it is not possible to decrease expenses in the same way but still prescribe more volume without any other adjustment. The first possibility is a substitution to smaller package sizes, meaning a reduction in quantity adjusted volume, but not in volume. The second possibility is a substitution to less expensive OTC drugs.

5.3.2 Practice characteristics and patient population

The NHS Business Authority (NHSBSA) provides quarterly information for NHS England on the number of total patients, patient demographics and the number of patients a practice can dispense drugs to. NHS Wales provides yearly information on the number of total patients and the practice dispensing status for Welsh practices. 15.6% of practices (1,100) in the final dataset have dispensing rights. Patient age data on practice level is not available for Welsh practices. However, there is only little within practice age variation for English practices in the observation period. We also use annually prevalence data for 11 major diseases treated in primary care to control for a patient disease-mix. The Quality of Outcomes Framework (QOF) reports this data under the pay-for-performance program. This is especially important because

[8]However, one needs to interpret the results of non-OTC expenditures with caution, because there had also been an introduction of guidelines regarding prescription drugs in June 2017.

OTC prescribing is still encouraged for the treatment of a chronic illness.

5.4 Methods

Our empirical analysis seeks to establish whether dispensing doctors behave differently to their non-dispensing colleagues due to own financial incentives. In order to do so, we first test the adoption of prescribing guidelines across all GP's to analyse whether doctors have a general incentive to change their prescribing behaviour according to new rules. In the second step, we then focus on our main hypothesis and study heterogeneous effects comparing the adoption of guidelines between dispensing and non-dispensing doctors.

5.4.1 Baseline method

We first test the adoption of prescribing guidelines across all GP's to analyse whether doctors have a general incentive to change their prescribing behaviour according to new rules. To identify the causal treatment effect on OTC prescribing, we apply a difference-in-differences approach and exploit the introduction of new low-value care prescribing guidelines. Our baseline regression model for practice i in month t takes the form:

$$ln(y_{it}) = \alpha + g_i\gamma + tg_i\beta + X_{it}^{\prime}\pi + \mu_t + \lambda_i + \epsilon_{it} \tag{5.1}$$

where y_{it} is one of the six prescribing indicators in Table 5.1, for GP practice i in month t. g is a fixed group effect that accounts for differences between the treatment group and the control group. It takes the value 1 for English practices and the value 0 for Welsh practices. t is a dummy variable for the post guideline introduction period. It takes the value 1 for prescriptions in January 2018 or later, and 0 otherwise. The interaction term tg_t captures the impact of the guideline introduction because those in the treatment group are exposed to the guideline introduction, whereas there is no rule change in the control group. X_{it} is a vector of observed practice characteristics, including a large set of variables, such as practice list size information and disease prevalence data[9]. μ_t are time fixed effects, λ_i are practice fixed effects, which capture unobserved individual and regional differences and ϵ_{it} is the error term. Since we

[9]A detailed list of variables can be found in the Appendix in Table E2.

observe the same practices over time, their unobservable characteristics will likely be correlated. Therefore, we cluster all standard errors at the practice level.
In other words, the empirical strategy in our analysis aims at identifying the causal effect of the guideline introduction by comparing the OTC prescribing behaviour of English GPs before and after the guideline introduction against non-affected Welsh GPs. The setup also allows to analyse heterogeneous treatment effects between different doctors types in a second step.
English and Welsh GPs are similar concerning their observable characteristics. They are exposed to a comparable regulation framework, patient groups, as well as seasonal trends. This is also true for the dispensing regulation. Since NHS England and NHS Wales are both part of the NHS in general and practice rules are very similar in both countries the common trend assumption is likely to hold. In other words, it is reasonable to assume that, in the absence of the reform, English GPs would follow the same trend as the prescribing behaviour of their Welsh colleagues. Overall, seasonal pre-trends do not show any differences between OTC prescribing in England and Wales. We conduct an event study to provide suggestive evidence in favour of the common trend assumption in Section 5.5.2.
The major difference between Welsh and English GPs are the prescription charges a patient needs to pay if he gets his drug dispensed: only patients in England need to pay the fee. However, according to the government, 50% of people in England are exempt from any charges, which results into 87% of all items prescribed without any charges (House of Commons Health Committee, 2006). Patients' co-payments should also not directly affect the doctor's adoption of a guideline. To avoid any bias due to guideline creation and consultation, we exclude the interim period from June to December 2017 from the main analysis. However, results also hold if the period is included[10].

5.4.2 Heterogeneous treatment effects

To study whether dispensing doctors behave differently to their non-dispensing colleagues, we exploit the exogenous policy change as a natural experiment, which occurred to both types of doctors in England. If dispensing and non-dispensing doctors behave differently, one would expect to see a difference in the adoption process between the two groups. To test this underlying hypothesis, we estimate the

[10]Results of this robustness check can be made available upon request.

following equation:

$$ln(y_{it}) = \alpha + I_i \cdot g_i\gamma + I_i \cdot tg_i\beta + X'_{it}\pi + \mu_t + \lambda_i + \epsilon_{it} \tag{5.2}$$

where y_{it} is one of the six prescribing indicators of Table 5.1, for GP practice i in month t. In addition to equation 5.1, we interact the fixed group effect g, as well as the dummy variable for the post-introduction period t, as well as the interaction term tg_t with a dummy variable I indicating the dispensing status (yes/no) of a GP practice[11]. The interaction between the dispensing status, the guideline introduction indicator, and group variable $I_{it} \cdot tg_{it}$ will indicate whether there are any underlying differences in the adoption process between doctors who might have financial incentives from prescribing OTC drugs (dispensing GPs) and those who do not (non-dispensing GPs).

In other words, the empirical strategy in our main analysis aims at identifying the causal effect of the guideline across different doctor types. We do so by comparing the OTC prescribing behaviour of English dispensing GPs before and after the guideline introduction against Welsh Dispensing GPs with OTC prescribing behaviour of English Non-dispensing GPs prior to and subsequent to the guideline introduction against Welsh Non-dispensing GPs.

As already discussed in Section 5.4.1, we assume the common trend assumption to hold. We conduct an event study by dispensing status to provide suggestive evidence in favour of the common trend assumption in Section 5.5.3. Section 5.5.1 also presents pre-trend graphs by doctor type. As before, we cluster all standard errors at the practice level and exclude the interim period from June to December 2017 from the main analysis.

In order to study adoption over time and to analyse whether effects between dispensing and non-dispensing doctors diverge, we also conduct an event study. We estimate the following equation:

$$ln(y_{it}) = \alpha + I_{iq} \cdot g\gamma + \sum_{q=1}^{18} \mu_q \cdot I_{iq} \cdot qg_{iq}\beta + X'_{iq}\pi + \lambda_i + \epsilon_{iq} \tag{5.3}$$

where v is one of the six prescribing indicators of Table 5.1, for GP practice i in

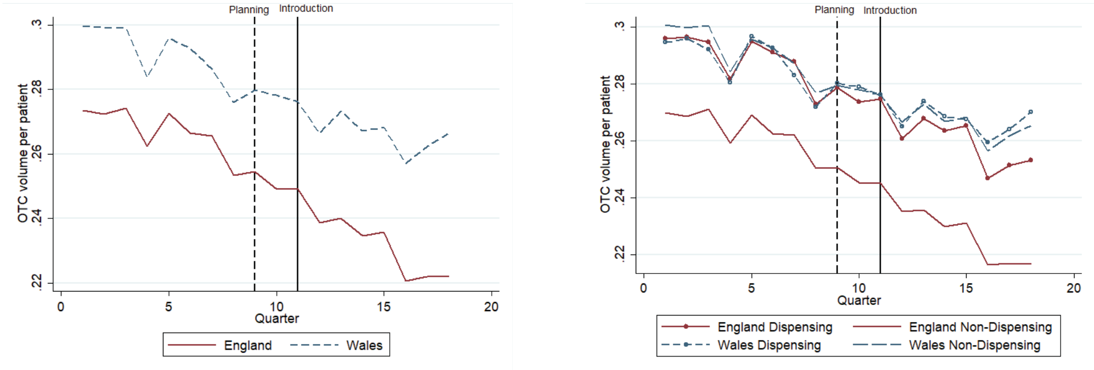
[11]Practices included in the dataset do not switch the dispensing status over the observation period.

quarter q. In order to smooth seasonal trends, we conduct the event study on a quarterly basis q. We interact the dispensing status, the guideline introduction indicator and group variable $I_{iq} \cdot g\gamma$ with every quarter. We take the pre-reform quarter as the baseline quarter to avoid any potential bias of anticipation behaviour due to planning and consulting.

5.5 Results

5.5.1 Descriptive statistics

Table 5.2 presents descriptive statistics for the prescribing indicators for the control and the treatment group before and after the guideline introduction by dispensing status. After the introduction of the prescribing guidelines, the volume and expenses of OTC drugs prescribed per patient decrease in England. However, the decrease is only moderate. OTC volume is higher for dispensing GPs compared to non-dispensing GPs. The same holds for non-OTC expenses per patient. The descriptive statistics also indicate that dispensing doctors prescribe smaller package sizes. Descriptive statistics for the practice characteristics can be found in Table E2 in the Appendix. The total number of patients within a practice increases in England and Wales for dispensing and non-dispensing doctors after the guideline introduction. This makes it necessary to weight all the indicators by the list size. Overall, major disease categories do not change a lot after the guideline introduction.
Figure 5.2 shows in panel (a) the pre- and post-policy trends for the prescribing indicators in general and panel (b) differentiates between dispensing and non-dispensing GPs in England and Wales. The y-axis presents one of the outcome variables: OTC volume per patient. In order to smooth seasonal trends, the x-axis shows quarters instead of months. The vertical dashed line shows the start of the planning period in June 2017. The pre-policy OTC volume per patient follows a parallel trend in England and Wales before the planning period.
Figure 5.2 panel (b) presents the pre- and post-policy trends by dispensing status. It shows that English dispensing doctors have a higher OTC volume per patient compared to non-dispensing English GPs. Pre-policy trends by dispensing status also follow a parallel pre-trend. Trend graphs for the other five prescribing indicators by dispensing status can be found in the Appendix in Section 5.7.3. The event studies in Sections 5.5.2, 5.5.3 and Appendix 5.7.4, 5.7.5 provide suggestive evidence



(a) OTC volume per patient

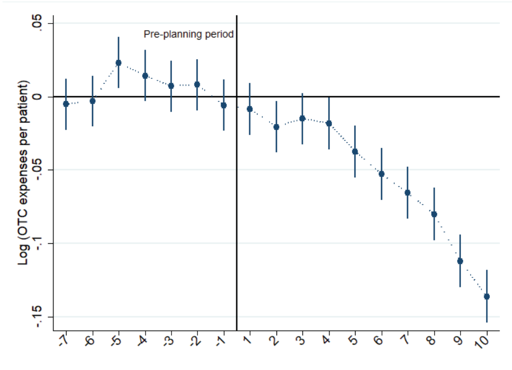
(b) OTC volume per patient by dispensing status

Figure 5.2: Time Trend OTC volume per patient Q2 2015-Q3 2019

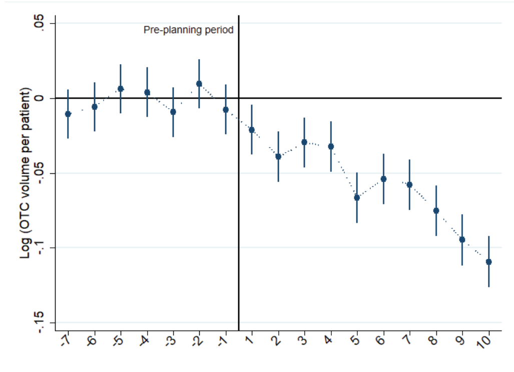
in favour of the common trend assumption overall and by dispensing status for the six prescribing indicators.

	Treatment group - England				Control group - Wales			
	Non-Dispensing		Dispensing		Non-Dispensing		Dispensing	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Prescribing indicators								
OTC expenses per pat	1.007	0.804	0.996	0.810	0.928	0.790	0.889	0.775
	(0.384)	(0.310)	(0.262)	(0.216)	(0.218)	(0.192)	(0.194)	(0.164)
OTC volume per pat	0.263	0.226	0.288	0.259	0.291	0.265	0.288	0.267
	(0.106)	(0.0933)	(0.0777)	(0.0744)	(0.0713)	(0.0661)	(0.0610)	(0.0536)
OTC adj. volume per pat	0.298	0.256	0.309	0.275	0.308	0.287	0.301	0.284
	(0.111)	(0.0972)	(0.0780)	(0.0739)	(0.0697)	(0.0672)	(0.0595)	(0.0563)
Package size	1.161	1.160	1.083	1.079	1.071	1.092	1.055	1.069
	(0.190)	(0.211)	(0.133)	(0.175)	(0.122)	(0.126)	(0.100)	(0.109)
Average cost per unit	0.0346	0.0337	0.0350	0.0337	0.0339	0.0313	0.0351	0.0329
	(0.0101)	(0.00576)	(0.00448)	(0.00479)	(0.00427)	(0.00428)	(0.00409)	(0.00430)
Non-OTC expenses per pat	11.56	11.01	13.15	12.74	14.09	13.65	13.94	13.55
	(3.626)	(3.291)	(2.706)	(2.603)	(2.816)	(2.718)	(2.360)	(2.106)
N	165,988	123,243	26,021	20,003	9,246	6,897	2,053	1,617

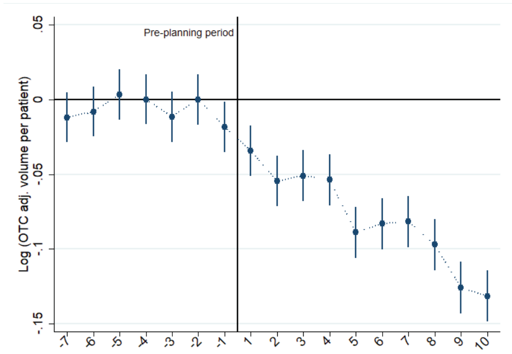
Table 5.2: Descriptive Statistics outcome variables - Control and Treatment group - Q2 2025-Q3 2019



(a) OTC expenses per patient



(b) OTC volume per patient



(c) OTC quantity adjusted volume

Adoption results England vs. Wales on 3 outcomes. Controls: models are adjust for average practice characteristics (incl. disease prevalence, patient list size), time fixed effects and practice fixed effects. Observation period from Q2 2015-Q3 2019. Standard errors are clustered at GP practice level. Figures show 95% percentiles.

Figure 5.3: Event study adoption indicators

5.5.2 General adoption of the policy across all English GPs

Table 5.3 presents the effect of the OTC policy on six prescribing indicators. Columns (1) to (3) report the results for the adoption indicators and columns (4) to (6) report the results for the substitution indicators. Overall, we see a significant adoption of the policy, since all adoption indicators in column (1) to (3) decrease after the guideline introduction. After the policy, OTC expenses per patient decrease by 7.4%. For a pre-treatment average English list size of 7,740, this results in £562.45 per practice or up to around £4,332,552 over all English GPs. Volume per patient and quantity-adjusted volume per patient also decrease by 6.5% and 8.5%, respectively. We also find significant results for all substitution indicators in column

(4) to (5), indicating a change in the prescribing mechanism behind the decrease in expenses. After the policy, OTC package size decreases by 1.9%, which indicates that English GPs substitute to smaller packages after the guideline introduction. This effect is not surprising since the NHS monitors OTC expenses. If a doctor decides to satisfy a patient request for an OTC prescription, it is better to prescribe a smaller package to decrease his prescribing-costs. Furthermore, if an English GP prescribes an OTC drug the average cost per unit increases by 4.8% after the policy. One would expect doctors to substitute to non-OTC drugs to satisfy patients' demand for treatment. Results indicate that non-OTC expenses per patient decrease by 2.1%. The OTC prescribing guideline may have spill-over effects on other prescribing areas. However, one should interpret this result with caution because there had also been an introduction of prescribing guidelines in the area of prescription drugs in the observation period.

As already discussed before, the adoption process may need some time. We, therefore, expect a larger decrease at the end of the observation period. Thus, considering a dynamic treatment effect gives a better understanding of how GPs adopt OTC prescription guidelines over time. Figure 5.3 presents the adoption process over time and indicates a successive increase in guideline adoption. In Q3 2019, OTC expenses and adjusted volume per patient decreased by almost 15%. In the beginning, adoption was only around 5%. Event studies for the substitution indicators can be found in the Appendix in Section E6. They show a similar pattern over time. The graphs give also suggestive evidence in favour of the common pre-trend assumption since the pre-planning period coefficients are statistically insignificant[12].

In general, English GPs seem to change their prescribing behaviour according to the new guidelines which gives a readily available testbed for a comparison between dispensing and non-dispensing doctors.

[12]There is only one period in Figure 3 (a) where this assumption is not fulfilled.

	Adoption indicators			Substitution indicators		
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(otcexpenpat)	ln(volumepat)	ln(adjvolumepat)	ln(packsize)	ln(avcost)	ln(nonotcpat)
did [tg_{it}]	-0.074***	-0.065***	-0.085***	-0.019***	0.048***	-0.021***
	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
GP FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	355,068	355,068	355,068	355,068	355,068	355,068
Mean pre-treatment	0.982	0.263	0.295	1.150	0.034	11.819

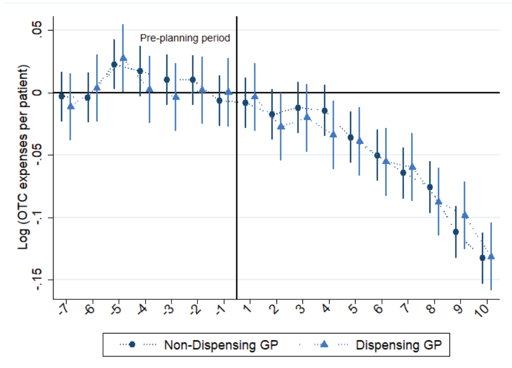
Adoption results England vs. Wales on 6 outcomes. Controls: models are adjust for average practice characteristics (incl. disease prevalence, patient list size), time fixed effects and practice fixed effects. Observation period from Q2 2015-Q3 2019. Standard errors (in parentheses) are clustered at GP practice level.
*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table 5.3: Adoption results across all English GPs

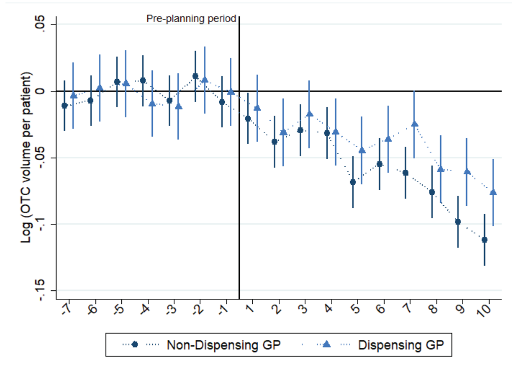
5.5.3 Behaviour of dispensing doctors compared to non-dispensing doctors

Self-dispensing doctors have a greater financial disincentive to adopt the policy since one of their substantial income components is driven by the dispensing service. Recall that dispensing income consists of two main components, (i) the variable margin and (ii) the fixed items fee. The latter might result in more prescriptions in general.

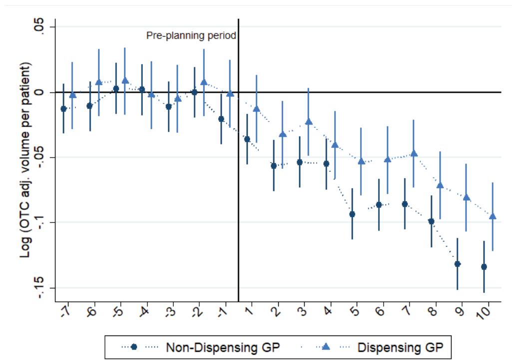
First, we look at the adoption of the guidelines between dispensing and non-dispensing GPs in columns (1) to (3) in Table 5.4. The reduction in OTC volume of dispensing doctors is significantly smaller compared to non-dispensing doctors. Column (2) of Table 5.4 shows that non-dispensing GPs OTC volume per patient decreases by 6.7% after the OTC guideline introduction, whereby dispensing doctors volume per patient decreases only by 4.7%. However, OTC expenses per patient, which the NHS monitors, are not significantly different for dispensing and non-dispensing doctors (column (1)). At first glance, this result might seem counterintuitive, but two main underlying mechanisms can explain this effect. The underlying mechanism can be found in Table 5.4 in column (4) to (6). First, dispensing doctors could decrease the number of pills, but hold total, not quantity adjusted volume constant by only vary the package size. Therefore they would still receive the same amount from the fixed items fee. However, column (4) shows that both doctor types substitute in a similar way to smaller package sizes after the policy introduction. This effect is also supported by column (3). Column (3) shows



(a) OTC expenses per patient



(b) OTC volume per patient



(c) OTC quantity adjusted volume

Adoption results England vs. Wales by dispensing status on 3 outcomes. Controls: models are adjust for average practice characteristics (incl. disease prevalence, patient list size), time fixed effects and practice fixed effects. Observation period from Q2 2015-Q3 2019. Standard errors are clustered at GP practice level. 95% percentiles.

Figure 5.4: Event study adoption indicators by dispensing status

that not only the volume decrease is significantly different, but also the decrease in the adjusted volume of total pills prescribed. The adjusted volume is 2.1% lower for dispensing doctors. A second possible mechanism is price substitution. Column (5) shows the average cost per unit increases by 5.3% for non-dispensing doctors but only by 2.6% for dispensing doctors. That means dispensing doctors prescribe less expensive OTC drugs compared to non-dispensing GPs. Another possibility to keep the earnings from dispensing with the fixed items fee constant is a substitution to non-OTC drugs. However, results indicate no difference between dispensing and non-dispensing GPs.

	Adoption indicators			Substitution indicators		
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(otcexpenpat)	ln(volumepat)	ln(adjvolumepat)	ln(packsize)	ln(avcost)	ln(nonotcpat)
did [tg_{it}]	-0.072***	-0.067***	-0.087***	-0.019***	0.053***	-0.022***
	(0.006)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
did*d_disp [$I_{it} \cdot tg_{it}$]	-0.004	0.020*	0.021*	0.001	-0.027***	0.010
	(0.013)	(0.010)	(0.009)	(0.006)	(0.008)	(0.008)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
GP FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	355,068	355,068	355,068	355,068	355,068	355,068

Adoption results England vs. Wales by dispensing status on 6 outcomes. Controls: models are adjust for average practice characteristics (incl. disease prevalence, patient list size), time fixed effects and practice fixed effects. Observation period from Q2 2015-Q3 2019. Standard errors (in parentheses) are clustered at GP practice level.
*** p<0.001, ** p<0.01, * p<0.05, + p<0.10

Table 5.4: Adoption results by dispensing status

Another interesting point is the evolution of the adoption of the OTC guideline for both doctor types. Figure 5.4 presents the results of the event study and shows that (a) the adoption of OTC expenses per patient follows a very similar trend over time for both doctor groups. However, panels (b) and (c) reveal that volume per patient and quantity-adjusted volume per patient differ between doctor types. Adoption effects diverge over time and the gap becomes larger. The same is true for the average cost per unit, which can be found in the Appendix in Section 5.7.5.

5.6 Conclusion

The paper analyses whether dispensing doctors follow own financial incentives and behave differently to non-dispensing doctors. To answer this question, we exploit an OTC prescribing guideline introduced in December 2017 in the NHS England. The guideline has been introduced to decrease expenditure growth in primary care and restrict inefficient prescribing. The idea behind the new guidance is not to prescribe OTC medicines that are from one of the three following areas (1) a self-limiting condition (2) a condition that is suitable for self-care and (3) items of limited clinical effectiveness. The introduction of the prescribing guidelines was an exogenous decision made by CCGs in close contact with the NHS England.
Using several prescribing indicators, we find that doctors adopt the prescribing guideline in general even without strict enforcement. Self-dispensing doctors, who have a greater financial disincentive to adopt the policy, adopt it less. The reduction in OTC volume of dispensing doctors is significantly smaller compared to

non-dispensing doctors. However, reduction in OTC expenses, which NHS England monitors, do not differ significantly between dispensing and non-dispensing doctors. At first glance, this result might seem counterintuitive, but one main underlying mechanism can explain this effect. Results indicate that dispensing doctors prescribe less expensive OTC drugs to prescribe more volume compared to non-dispensing doctors. The rewarding system for dispensing doctors with a fixed item component might be one possible reason for this finding. Package size substitution does not only matter for one doctor type. Both doctor types substitute to smaller package sizes. This effect is not surprising since the NHS monitors OTC expenses. If a doctor decides to satisfy a patient request for an OTC prescription it is better to prescribe a smaller package to decrease costs. Substitution to non-OTC drugs does not seem to matter.

Previous studies analysed dispensing GPs behaviour in static cross-sectional settings. Making both doctor types directly comparable is a major challenge of these studies. Most papers use a matching procedure to overcome this problem. Few papers analysed the specific market regulation of dispensing doctors in a dynamic setting. However, they analyse a strict dispensing ban. The OTC policy gives scope for an individual adoption process, which offers a unique study setting for this market regulation.

The paper shows that dispensing doctors behave differently compared to non-dispensing doctors resulting in higher costs for the NHS. These financial incentives might be in contrast to a long-run overall cost-efficient prescribing aim of a policy-maker. Previous research, as well as our results, help to classify costs from a dispensing regulation. However, in rural areas, physician dispensing might be important for a good health infrastructure. These positive welfare effects are though complicated to measure. Up to this point, it is therefore not clear whether positive effects, for example, better healthcare provision, can predominate the negative effects, we find.

Bibliography

Ahammer, A. and I. Zilic, "Do financial incentives alter physician prescription behavior? Evidence from random patient-gp allocations.," *Working paper*, 2017.

Bodnar, O., H. Gravelle, N. Gutacker, and A. Herr, "Financial incentives and precribing behaviour in primary care," mimeo.

Burkhard, D., C.P.R. Schmid, and K. Wüthrich, "Financial incentives and physician prescription behavior: Evidence from dispensing regulations," *Health Economics*, 2019, *28* (9), 1114–1129.

Carthy, P., I. Harvey, R. Brawn, and C. Watkins, "A study of factors associated with cost and variation in prescribing among GPs," *Family practice*, 2000, *17* (1), 36–41.

Chou, Y.J., W.C. Yip, C.H. Lee, N. Huang, Y.P. Sun, and H.J. Chang, "Impact of separating drug prescribing and dispensing on provider behaviour: Taiwan's experience," *Health policy and planning*, 2003, *18* (3), 316–329.

Choudhry, N.K., H.T. Stelfox, and A.S. Detsky, "Relationships between authors of clinical practice guidelines and the pharmaceutical industry," *JAMA*, 2002, *287* (5), 612–617.

Department of Health, "Regulations under the Health Act 2009: Market entry by means of Pharmaceutical Needs Assessments Information for Primary Care Trusts Executive Summary and Chapters 1-4," 2012. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/212861/Chapters-1-4-introduction-background.pdf` Accessed: 23.01.2020.

Fairhurst, K. and G. Huby, "From trial data to practical knowledge: qualitative study of how general practitioners have accessed and used evidence about statin drugs in their management of hypercholesterolaemia," *BMJ (Clinical research ed.)*, 1998, *317* (7166), 1130–1134.

Filippini, M., F. Heimsch, and G. Masiero, "Antibiotic consumption and the role of dispensing physicians," *Regional Science and Urban Economics*, 2014, *49*, 242–251.

Goldacre, B., C. Reynolds, A. Powell-Smith, A.J. Walker, T.A. Yates, R. Croker, and L. Smeeth, "Do doctors in dispensing practices with a financial conflict of interest prescribe more expensive drugs? A cross-sectional analysis of English primary care prescribing data," *BMJ open*, 2019, *9* (2), e026886.

House of Commons Health Committee, "Nhs charges," 2006. `https://publications.parliament.uk/pa/cm200506/cmselect/cmhealth/815/815-i.pdf` Accessed: 23.01.2020.

Iizuka, T., "Experts' agency problems: evidence from the prescription drug market in Japan," *The Rand journal of economics*, 2007, *38* (3), 844–862.

_ , "Physician Agency and Adoption of Generic Pharmaceuticals," *The American economic review*, 2012, *102* (6), 2826–2858.

Kaiser, B. and C. Schmid, "Does Physician Dispensing Increase Drug Expenditures? Empirical Evidence from Switzerland," *Health Economics*, 2016, *25* (1), 71–90.

Liu, Y.M., Y.H.K. Yang, and C.R. Hsieh, "Financial incentives and physicians' prescription decisions on the choice between brand-name and generic drugs: evidence from Taiwan," *Journal of health economics*, 2009, *28* (2), 341–349.

Mayer, J. and L. Piterman, "The attitudes of Australian GPs to evidence-based medicine: a focus group study," *Family practice*, 1999, *16* (6), 627–632.

NHS Clinical Commissioners, "Quick Reference Guide for Healthcare Professionals: Conditions for which over the counter items should not routinely be prescribed in primary care.," 2018. `https://www.england.nhs.uk/wp-content/uploads/2018/05/over-the-counter-quick-reference-guide.pdf` Accessed: 13.02.2020.

_ , "About CCGs," n.d. `https://www.nhscc.org/ccgs/` Accessed: 23.01.2020.

NHS England, "Publications clearance: 00095 from Professor Stephen Powis National Medical Director NHS England," 2019. `https://www.england.nhs.uk/wp-content/uploads/2019/01/otc-gms-gp-practice-letter.pdf`.Accessed: 23.01.2020.

Park, S., S.B. Soumerai, A.S. Adams, J.A. Finkelstein, S. Jang, and D. Ross-Degnan, "Antibiotic use following a Korean national policy to prohibit medication dispensing by physicians," *Health policy and planning*, 2005, *20* (5), 302–309.

Raisch, D. W., "A model of methods for influencing prescribing: Part II. A review of educational methods, theories of human inference, and delineation of the model," *DICP : the annals of pharmacotherapy*, 1990, *24* (5), 537–542.

Rashidian, A., M.P. Eccles, and I. Russell, "Falling on stony ground? A qualitative study of implementation of clinical guidelines' prescribing recommendations in primary care," *Health policy (Amsterdam, Netherlands)*, 2008, *85* (2), 148–161.

Rischatsch, M., M. Trottmann, and P. Zweifel, "Generic substitution, financial interests, and imperfect agency," *International journal of health care finance and economics*, 2013, *13* (2), 115–138.

Trap, B., E.H. Hansen, and H.V. Hogerzeil, "Prescription habits of dispensing and non-dispensing doctors in Zimbabwe," *Health policy and planning*, 2002, *17* (3), 288–295.

Trottmann, M., M. Frueh, H. Telser, and O. Reich, "Physician drug dispensing in Switzerland: association on health care expenditures and utilization," *BMC health services research*, 2016, *16*, 238.

5.7 Appendix

5.7.1 Data sources

Data	Frequency	Source
Prescriptions per GP practice	Monthly	NHSBSA; NHS Wales (Welsh: GIG Cymru)
Listsize and status of practice	Annual	NHSBSA ; NHS Wales (Welsh: GIG Cymru)
Patient demographics	Annual	NHS Digital; Gov. Wales
OTC list	Once	NHSBSA

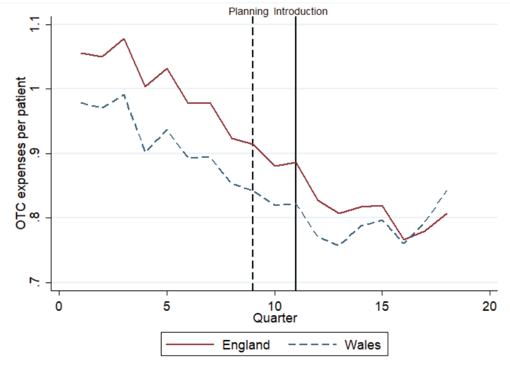
Table E1: Data Sources and frequencies

5.7.2 Descriptive statistics

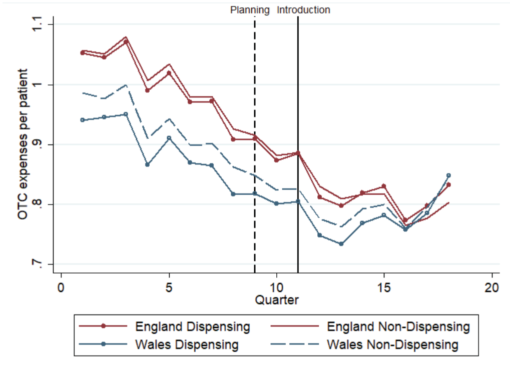
	Treatment group - England				Control group - Wales			
	Non-Dispensing		Dispensing		Non-Dispensing		Dispensing	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Characteristics								
Conorary Heart Disease	3.145	3.076	3.776	3.747	3.729	3.605	3.967	3.876
	(1.155)	(1.124)	(0.905)	(0.876)	(0.876)	(0.857)	(0.752)	(0.722)
STIA	1.642	1.674	2.139	2.221	1.995	2.021	2.231	2.281
	(0.696)	(0.686)	(0.555)	(0.563)	(0.531)	(0.535)	(0.476)	(0.495)
Hypertension	13.65	13.78	16.26	16.58	15.52	15.45	17.44	17.55
	(3.652)	(3.667)	(2.824)	(2.818)	(3.306)	(3.251)	(2.798)	(2.764)
COPD	1.905	1.964	1.888	1.965	2.246	2.319	2.373	2.440
	(0.961)	(0.977)	(0.626)	(0.622)	(0.801)	(0.806)	(0.896)	(0.841)
Cancer	2.195	2.517	3.202	3.688	2.452	2.796	3.029	3.496
	(0.872)	(0.994)	(0.680)	(0.738)	(0.772)	(0.848)	(0.741)	(0.806)
Mental Health	0.968	1.012	0.667	0.699	0.934	0.980	0.827	0.865
	(0.498)	(0.524)	(0.222)	(0.224)	(0.312)	(0.313)	(0.336)	(0.326)
Asthma	5.823	5.819	6.499	6.598	6.913	6.989	7.165	7.172
	(1.330)	(1.343)	(0.984)	(0.981)	(1.180)	(1.150)	(1.219)	(1.161)
Heart Failure	0.735	0.804	0.861	0.982	0.964	0.999	1.191	1.231
	(0.365)	(0.400)	(0.344)	(0.417)	(0.398)	(0.407)	(0.424)	(0.409)
Palliative Care	0.332	0.383	0.408	0.470	0.295	0.300	0.395	0.444
	(0.592)	(0.543)	(0.409)	(0.461)	(0.302)	(0.248)	(0.311)	(0.356)
Dementia	0.723	0.731	0.862	0.884	0.601	0.648	0.684	0.723
	(0.880)	(0.753)	(0.355)	(0.351)	(0.315)	(0.325)	(0.342)	(0.374)
Astrial Fibrillation	1.546	1.740	2.320	2.650	1.929	2.128	2.340	2.610
	(0.743)	(0.816)	(0.576)	(0.601)	(0.610)	(0.656)	(0.560)	(0.598)
Number of patients	7586.1	8422.8	8076.8	8700.9	7472.2	8097.1	6354.9	6615.0
	(4568.5)	(5281.9)	(4904.2)	(5596.9)	(3670.0)	(3894.3)	(3714.3)	(3723.3)
N	165,988	123,243	26,021	20,003	9,246	6,897	2,053	1,617

Table E2: Descriptive Statistics - Control and Treatment group - Q2 2025-Q3 2019

5.7.3 Time trends

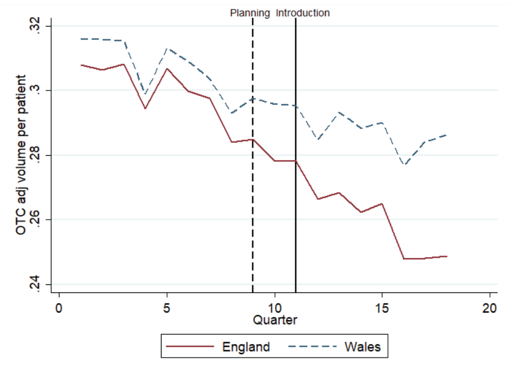


(a) OTC expenses per patient

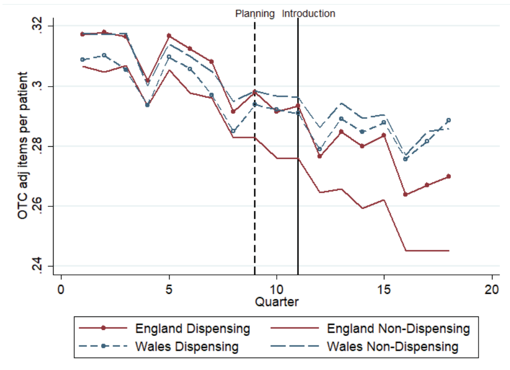


(b) OTC expenses per patient by dispensing stautus

Figure E1: Time Trend OTC expenses per patient Q2 2015-Q3 2019

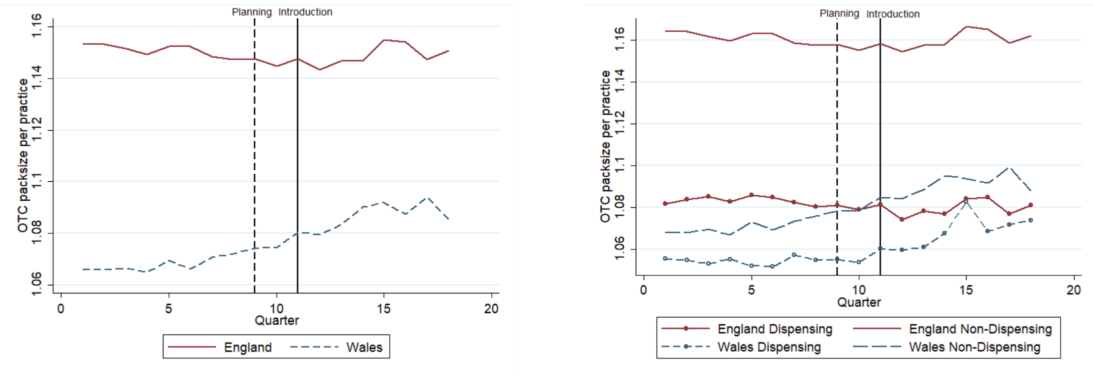


(a) OTC quantity adjusted volume per patient



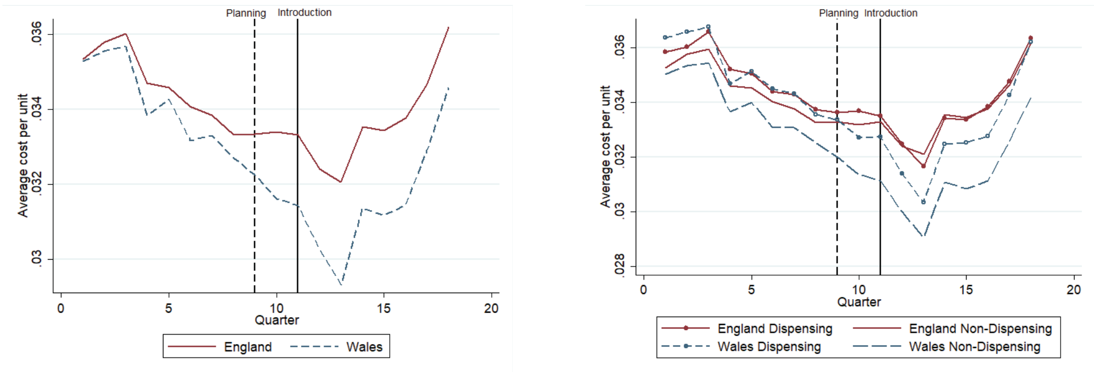
(b) OTC quantity adjusted volume per patient by dispensing status

Figure E2: Time Trend OTC quantity adjusted volume per patient Q2 2015-Q3 2019



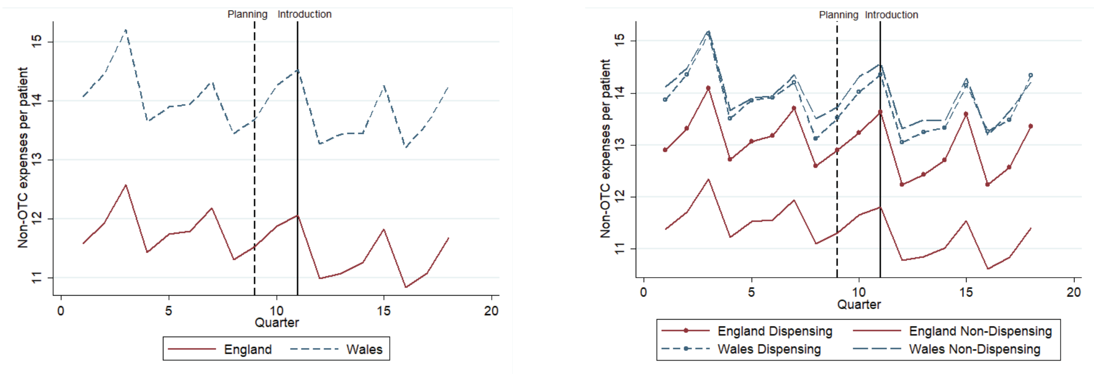
(a) OTC package size (b) OTC packagesize by dispensing status

Figure E3: Time Trend OTC package size per practice Q2 2015-Q3 2019



(a) OTC cost per unit (b) OTC cost per unit by dispensing status

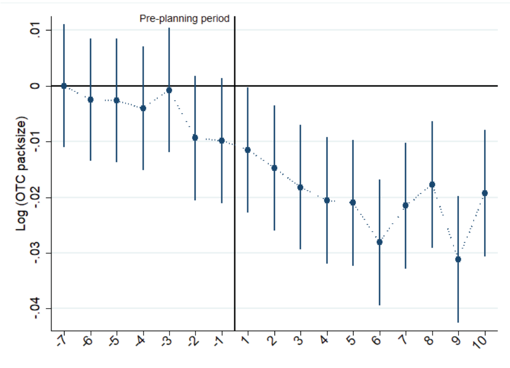
Figure E4: Time Trend OTC average cost per unit Q2 2015-Q3 2019



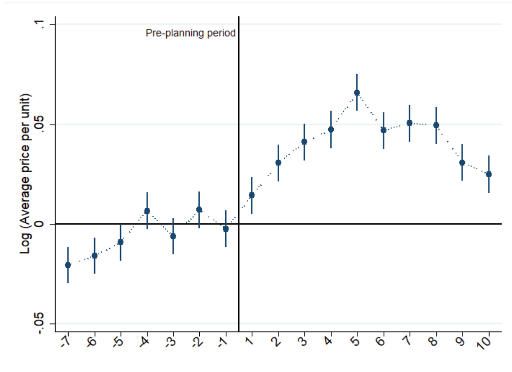
(a) Non-OTC expenses per patient (b) Non-OTC expenses per patient

Figure E5: Time Trend non-OTC expenses per patient Q2 2015-Q3 2019

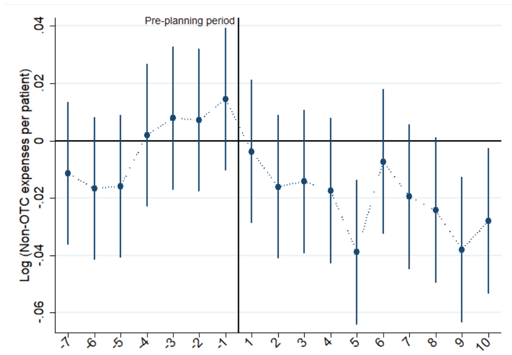
5.7.4 Event studies general adoption



(a) OTC package size



(b) OTC average cost per unit

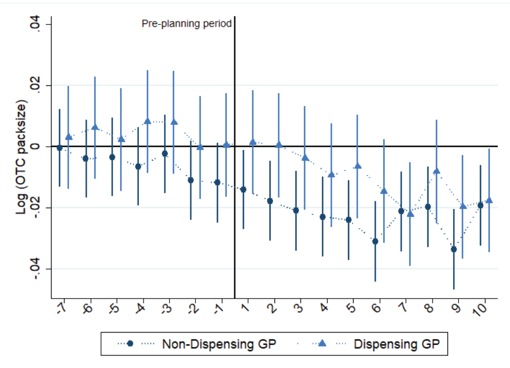


(c) Non-OTC expenses per patient

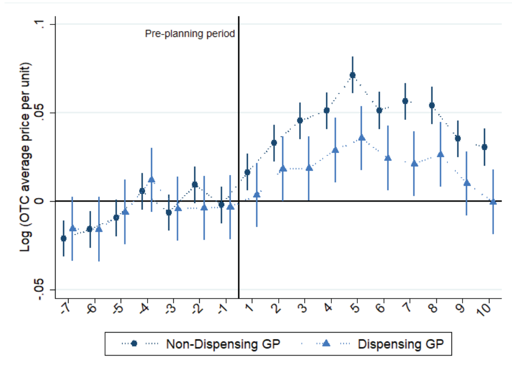
Substitution results England vs. Wales on 3 outcomes. Controls: models are adjust for average practice characteristics (incl. disease prevalence, patient list size), time fixed effects and practice fixed effects. Observation period from Q2 2015-Q3 2019. Standard errors are clustered at GP practice level. 95% percentiles.

Figure E6: Event study substitution indicators

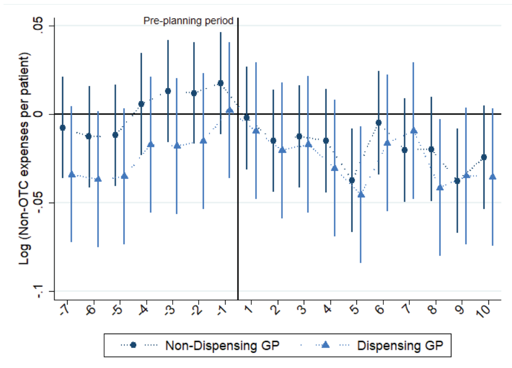
5.7.5 Event studies by dispensing



(a) OTC package size



(b) OTC average cost per unit



(c) Non-OTC expenses per patient

Substitution results England vs. Wales by dsiepnsing status on 3 outcomes. Controls: models are adjust for average practice characteristics (incl. disease prevalence, patient list size), time fixed effects and practice fixed effects. Observation period from Q2 2015-Q3 2019. Standard errors are clustered at GP practice level. 95% percentiles.

Figure E7: Event study substitution indicators by dispensing status

Eidesstattliche Versicherung

Ich, Frau Olivia Alexandra Bodnar, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.
Düsseldorf, der 20. August 2020