FOUR ESSAYS ON THE DESIGN AND IMPLEMENTATION OF EFFECTIVE ANTITRUST ENFORCEMENT

A dissertation

presented to the School of Economics of

the University of East Anglia

in candidacy for the degree of

Doctor of Philosophy

by

Carsten Jürgen Crede

University of East Anglia

School of Economics

Presented: 25 September 2017

Passed: 08 December 2017

Abstract

This thesis consists of four essays that relate to the design and implementation of effective competition policy in the fields of cartels and merger control. The first essay contains a new empirical cartel screen that can be used to detect cartels in market data to increase deterrence. The second essay uses an experimental analysis to study the determinants of tacit collusion after the end of cartels and develops policy suggestions to prevent such outcomes. The third essay provides an experimental study on the use of endogenous fines for cartels to reduce the harm caused by them by lowering the optimal cartel price. In the fourth essay a new empirical framework for the *ex post* evaluation of the effects of mergers and acquisitions in clustered data is developed and used to study the occurrence of spillover effects in multi-market merger control cases.

Contents

List of Tables

List of Figures

Acknowledgements

In the course of conducting the research contained in this thesis, I have benefited from comments and support from many researchers. My biggest thanks go to my supervisors Stephen Davies and Subhasish Modak Chowdhury, who have provided me with outstanding support throughout my research career at the University of East Anglia (UEA). Whenever I hear the phrase "Standing on the shoulders of giants", my supervisors come to my mind as role models both as academics and human beings. In addition, I would like to thank the staff of the School of Economics at the UEA, the Centre for Competition Policy (CCP), and the Centre for Behavioural and Experimental Social Science (CBESS) for valuable comments and support throughout the course of conducting research at the UEA. Further, I would like to thank my numerous fellow PhD students of the School of Economics at the UEA. Additional thanks go to Sebastian Otten and Julia Bredtmann for their support during my early ventures into research.

Chapter 2 has profited from comments from Rosa Abrantes-Metz, Giuliana Battisti, Farasat Bokhari, Steve Davies, Andreas Gerster, Nils Gutacker, Franco Mariuzzo, Peter Ormosi, George Papadopoulos, Daniel Rubinfeld, Maarten Pieter Schinkel, Lawrence White, Achim Zeileis, two anonymous referees as well as participants at the Workshop of the Law & Economics of Antitrust 2016 in Zurich, the RGS Doctoral Conference 2016 in Bochum, the 2015 Network of Industrial Economists Doctoral Student Colloquium in Nottingham, CCP seminar participants in 2014 and 2015 in Norwich, the 2014 Competition Law and Economics European Network Workshop in Norwich, and the 2014 Competition Law and Scholars Forum in London.

For Chapter 3, I would like to thank my co-author Subhasish Modak Chowdhury for valuable discussions and ideas, the UEA for funding the research, and the CBESS for proving access to its lab. I would like to thank David Angenendt, Maria Bigoni, Steve Davies, Luke Garrod, Joris Gillet, Morten Hviid, Kai-Uwe Kühn, Claudia Möllers, Matt Olczak, Sander Onderstal, Sigrid Suetens, Fred Wandschneider, Alexandra Zaby, participants at the 2015 London Experimental Workshop, the 2015 Spring Meeting of Young Economists in Gent, the 2015 Mannheim Centre for Competition and Innovation Conference, the 2014 London Experimental Economics PhD Workshop, as well as seminar participants at the UEA for helpful comments.

Internal examiner: Professor Bruce Lyons

External examiner: Professor Maarten Pieter Schinkel

Date of oral defence: 08/12/2017

Chapter 1

Introduction

The four essays in this thesis in broad terms are concerned with the design and implementation of effective antitrust enforcement in the fields of cartels and merger control. In principle, each chapter is self-contained and can be read independently. Yet, at a higher level, they are connected by two overarching themes. The first is the use of empirical methods to improve the effectiveness of antitrust enforcement. First, this idea motivates a new screen for the detection of cartels to increase the deterrence properties of anti-cartel laws. Second, a method is developed that allows to uncover ineffectiveness in merger control regimes relating to spillover effects. The second purpose is the use of experimental methods to contribute to solutions to current problems of competition policy relating to the deterrence of harmful cartel behaviour.

A new empirical screen for detecting cartels is developed in Chapter 2. Cartels screens are subject to an increased interest both in academia and in the field and complement other methods of cartel detection with the prospect to enhance the deterrence of cartels. In particular, they might help to target stable cartels not threatened by leniency programmes and high fines and improve deterrence by increasing the exogenous risk of detection. For this purpose, they aim at detecting cartels by discovering suspicious patterns left by collusion in market data. The new empirical cartel screen detects collusion by scanning for structural breaks that are induced by cartels in the data generating process of industry prices. Technically, it tests reduced form price equations modelling industry pricing dynamics for structural instability. It is applied to three European markets for pasta products, in which it successfully reports the cartels that were present in the Italian and Spanish markets, but finds no suspicious patterns in the French market, which was not cartelised. The screen can also be used to date the beginning of known conspiracies, which is often difficult in practice.

Even if cartels are detected – for example by a cartel screen as the one above – they often continue to induce welfare damages after their dissolution. These welfare damages can persist if the industry does not immediately revert to competition but continues to adhere to the collusive prices and strategies. As such a behaviour rests on a tacit understanding of the market participants, it is referred to as post-cartel tacit collusion (PCTC). Despite

the fact that PCTC can have important consequences on the welfare properties of antitrust enforcement and has been observed repeatedly in the field, little is known about it. Chapter 3 attempts to tackle this gap and contains an experimental investigation of the determinants of PCTC, its effects on market outcomes, and potential policy measures aimed at its prevention. PCTC is found to be determined both by collusive price hysteresis and learning about cartel partners' characteristics and strategies. It induces a downward bias in the estimation of cartel overcharges, which plays an important role in private damage litigation and contributes to the deterrence of cartels. The results show that this bias increases with preceding cartel stability such that the most stable cartels might be deterred the least by fines imposed for collusion. In the experiment, rematching colluding subjects with strangers within a session prevents PCTC by disrupting the learning process that contributes to the tacit understanding of the market participants. The findings imply that manager disqualification orders, also known as debarment, might hold the promise to reduce the welfare damages created by PCTC.

An effective design of competition law might not only prevent the damages caused by cartels after their demise, but also reduces their harm to welfare while they are active. As one cannot hope to prevent cartel formation altogether or detect all cartels, a strategic design of fines to reduce the optimal cartel price can help to reduce their harm too. Such a design of cartel fines is the subject of Chapter 4, in which experiments are used to study the impact of antitrust enforcement on cartel price decisions when fines and detection probabilities depend on them. Expected punishment is imposed that creates two payoff-equivalent collusive price equilibria, of which one features a lower riskiness of collusion. Subjects are found to behave strategically in that they choose the equilibrium with a lower riskiness of collusion. This suggests that competition authorities can exploit the effects of such endogenous enforcement on strategic uncertainty between cartelists, i.e. the *a priori* uncertainty about the actions of the other cartel members, to lower cartel prices. However, frequency deterrence might be reduced as well such that the overall welfare effects of endogenous fines may be ambiguous.

Ambiguity in effects can also be a problem in merger control. In Chapter 5, the overall effects of spillovers in multi-market mergers and acquisitions (M&As) are studied to assess the role that they should play in merger control. For this purpose, a new empirical framework for the *ex post* evaluation of multi-market M&As is proposed. Approaches to correct for self-selection bias and for the estimation of the effects of M&As on clustered and interrelated product markets are suggested. The framework is applied to estimate the price effects of multi-market pharmaceutical M&As in the United Kingdom that were cleared by the European Commission (EC). Results suggest that the case practice developed by the EC might have limited success only at preventing price increases triggered by M&As. This follows from the fact that spillover effects can foster market power in markets, in which the market shares and (non-)overlapping presence in isolation appear unproblematic.

A summary of the results in this thesis can be found in Chapter 6.

Chapter 2

A structural break cartel screen for dating and detecting collusion

2.1 Introduction

[1] Recent success of empirical methods used by competition authorities in the Netherlands, Brazil, and Mexico leading to the detection of several cartels as well as the spectacular case of the LIBOR market manipulation have increased the interest in cartel screens (Abrantes-Metz, 2014). The purpose of these empirical methods is to detect a cartel by identifying patterns in market outcomes that suggest collusion. They are meant "*not to deliver the final evidence based on which colluders will be convicted, but instead to identify markets where empirical red flags are raised and which are worth further investigations.*" (Abrantes-Metz, 2014, p.7). Increasing the detection probability might be necessary to prevent cartel recidivism and to increase deterrence with respect to potentially more successful and stable cartels (Harrington, 2007).

There are two categories of cartel screens. Structural screens identify markets which are likely to be subject to cartelisation due to industry characteristics. Behavioural screens detect cartels by detecting patterns in market outcomes treated as signs of collusion (Abrantes-Metz, 2014; Harrington, 2007). The literature on behavioural cartel screens has grown significantly in the last decade. Most notable are the contributions of Abrantes-Metz *et al.* (2006) and Bolotova *et al.* (2008), who propose cartel screens based on the analysis of price variance in an industry.

In this essay, a novel behavioural cartel screen is proposed that is based on identifying structural breaks in the data generating process (DGP) of industry prices that are induced by cartel activity. The idea to trace structural breaks originating from changes in pricing dynamics

---

[1]This chapter is partly based on Crede (2015).

to detect cartels has first been suggested by Harrington (2008). A DGP characterising competition is established with a reduced form price equation based on periods characterised by competition and then used to test suspicious periods for collusion. In the absence of other explanations for the structural breaks, their existence raises an empirical red flag suggesting that there might be a cartel and that the market requires further investigation.

In addition to cartel screening, the structural break cartel screen is suitable to date the start of a conspiracy that has already been detected. This is often an issue in antitrust litigation, when it is uncertain whether the earliest written evidence that has been found of a cartel represents the start of the conspiracy.

Application of the new screen is discussed based on three European pasta markets, two of which were cartelised. Each market has different features, which allows to test the screen under varying circumstances: the market in Italy featured a cartel and altered industry pricing dynamics consistent with tacit collusion after cartel breakdown. The cartel in the Spanish market lasted only 3 months, which renders it difficult to detect its impact on market prices. The French market did not feature a cartel, but saw prices rise significantly during 2007 due to a strong input cost shock. Unlike variance-based cartel screens, the proposed structural breaks screen successfully detects the cartels in Italy and Spain, but does not report a false positive of a cartel in France. Robustness checks show that the power of the test increases with the precision, with which the DGP of competition in an industry can be modelled.

2.2 Literature review

Most behavioural screens so far have been suggested for bid-rigging conspiracies, which are now regularly used in auctions (Porter, 2005). For example, in a study of a bidding ring for highway-paving construction tasks in New York Porter and Zona (1993) find that while the lowest bid of a conspirator was most likely to be related to the fact that this firm had the lowest cost, such a correlation did not exist for the higher bids of other ring members.

In the last decade, researchers began to develop behavioural screens for cartels outside auctions. Abrantes-Metz and Bajari (2010) and Blair and Sokal (2013) provide an overview of the different applications of screens for detecting collusion, and Crede (2016) describes the intuition behind several behavioural cartel screens. In particular one class of behavioural screens has recently received much attention: variance-based price screens rely on the idea that the reduced price variance of firms across time or within geographical clusters is an indicator of collusion.[2]

Abrantes-Metz *et al.* (2006) find a significant reduction in the price variance for a frozen perch cartel in Philadelphia between 1984 and 1989 based on a comparison of the price coefficient of variation (standard deviation divided by the mean) between periods marked by collusion

---

[2]This is not to be confused with price dispersion between firms at a point in time (see, e.g. Connor, 2005).

and competition. Given the absence of changes in costs around significant price reductions, they are also able to track the collapse of the conspiracy. They further test for geographical clusters of gasoline retail stores in Louisville, USA, based on the concept of a reduced price variance indicating collusion. In a similar study, Esposito and Ferrero (2006) find reduced price variances for two Italian cartels fixing prices for motor fuel and products sold in pharmacies. A similar but more sophisticated approach is suggested by Heijnen *et al.* (2015), who test the Netherlands' gasoline market for suspicious local clusters with a reduced price variance.

Bolotova *et al.* (2008) use ARCH and GARCH models to analyse price and price variance changes of the citric acid and lysine cartels. Finding strong support for the former cartel, only mixed evidence is found for the latter. The authors provide several reasons that might explain the lack of robust findings regarding the price variance. Abstracting from the methodological explanations for this result, they stress that variance-based screens can fail when cartels are not all-inclusive and do not have full control over the price or abnormal supply or cost shocks affect market outcomes.

Blanckenburg *et al.* (2012) test whether the other moments of price change distributions mean, kurtosis, and skewness of price changes can be used as collusive markers for 11 cartels. They conclude that only the variance is a robust indicator for collusion. A different conclusion is reached by Hüschelrath and Veith (2013), who use sequential t-tests to test for changes in the mean of prices in the German cement cartel to show that with this approach the cartel could have been detected before it was uncovered by the German competition authority.

2.3 Methodology

2.3.1 The screen and its identification strategy

Abrantes-Metz (2014) points out two important rules with respect to designing and applying a cartel screen. First, the screen must be fitted to the industry under investigation to ensure proper identification. Second, identification further depends on the quality of the data that is used for the screen. This knowledge must be used to develop an idea how collusion or cheating might alter market outcomes to use statistical methods to test for significant changes in the considered market outcomes. These changes are signs of collusion. Harrington (2007) provides an extensive discussion of these changes, which are sometimes labelled *collusive markers*.

The existing behavioural screens in the literature exploit the existence of these collusive schemes to empirically test for specific changes in market outcomes that are assumed to result from the specific operation of a cartel. For example, the price variance-based screens assume that cartel activity reduces the price variance potentially indicating collusion. This is a shortcoming of the existing screens, as they only work in situations in which the cartel shows the very specific behavioural pattern assumed by the screen. Therefore, their functionality critically depends on choosing the correct screen for the market under investigation.

The structural break cartel screen suggested in this essay overcomes the problem of sensitivity of behavioural cartel screens to the underlying assumptions about cartel conduct by relying on a more general approach. It merely assumes that a cartel induces a structural break in the DGP determining industry pricing over time – whether there are reductions in the price variance, price wars, changes in cost pass-through, more cost increases than reductions, sudden price increases after cartel meetings, and so on does not matter. The approach of the screen is to estimate a reduced form equation of industry pricing for the industry under consideration and control for structural breaks induced by potential cartel activity. This idea is similar in spirit to the seminal work of Porter (1983), who shows how cartel activity induces regime switching measurable in market outcomes.

One way to determine the existence of breaks would be to run simple multiple change point models that test for significant changes in the mean or the variance as before, which is a more sophisticated way to implement existing approaches in the literature. However, such an approach has high risks of both type I and type II errors, as it cannot evaluate whether (no) changes in the price are unproblematic outcomes under competition or changes of the DGP arising from cartel activity.[3] Therefore, a regression-based approach to detect structural breaks is proposed that controls for all factors that should govern price changes under competition in a reduced form price regression, such that unexplainable price shifts can be attributed to cartel activity. Thus, the regression-based approach relies on fewer assumptions to identify cartel behaviour than the established methods.

Reduced form price equations are an established and the most common approach to estimate cartel overcharges in antitrust litigation (Brander and Ross, 2006; Nieberding, 2006; Baker and Rubinfeld, 1999). Advantages of this approach are its simplicity of use and its limited demands towards data compared to full demand and supply systems. Like these systems, reduced form equations explain price changes with supply and demand shifters as well as variables capturing changes in the market structure. Yet, unlike full structural systems, no instruments are needed for the identification of separate demand and supply equations (see, e.g. Davis and Garcés, 2009). The coefficient of an indicator variable flagging the cartel periods measures the average overcharge generated by a cartel in the dummy variable approach. In case of a cartel screen, whether and when a cartel exists is unknown, such that cartel dummies or forecasting cannot be used to identify suspicious pricing.

In case the reduced form price equation captures all demand, supply, and market characteristic factors that affect pricing in the industry, it can be used for cartel detection. Equation 2.1 shows a reduced form price equation that can be used to detect cartels

$$\Delta P_t = \alpha_1 \Delta C_t + \alpha_2 \Delta D_t + \alpha_3 \Delta S_t + \epsilon_t \ , \tag{2.1}$$

---

[3]See, e.g. 2.4: without controlling for significant input cost changes faced by the French pasta industry, price increases in the industry under competition would wrongly be detected as structural breaks.

where ΔP_t denotes a price change of the industry under consideration in period t, C_t and D_t are vectors of exogenous supply and demand shifters, S_t is a vector of market characteristics, and ϵ_t denotes the model error.[4] All variables are included in the form of the first difference to ensure stationarity and prevent spurious results.[5] As a result, the model describes the short-run relationship between the price and its shifters. First-differencing is a transformation of the data that incurs some loss of information because it removes undesired (unrelated) deterministic trends and unit roots. As such, it is not equivalent to reduced form equations of the price based on its absolute or demeaned value. Yet, in terms of the identification strategy of the screen this makes little difference provided that the DGP is specified correctly in each case. The screen detects a cartel from a significant relative change in the DGP – irrespective how exactly the DGP looks like – but not from specific absolute parameter values that describe it.

The choice of the specification should be driven by priors derived from an understanding of the industry under investigation. These priors relate to the selection of essential price shifters and, e.g. the inclusion of reasonable lags and leads that mirror reported dynamics of cost pass-through in the industry. These priors dictate the boundaries in which the specification of the DGP can be modelled based on a reasonable fit of the data. If available, priors relating to the signs of the coefficients or the functional form of the specification derived from an underlying structural model of the industry provide further guidance (see, e.g. Baker and Rubinfeld, 1999). While sensible tests of the robustness of the specification vary by application, testing the screen's robustness to the inclusion of price shifters with limited theoretical importance or statistical significance strengthens the credibility of the results.

Before a cartel forms, the price changes are determined by a DGP of competition. In this situation, the coefficients show the relationship between the regressors and the price changes in the industry when it is characterised by competition. A cartel that affects prices induces a structural break in the DGP of competition resulting in a bad model fit and sudden instability of coefficients during the cartel periods. Thus, the structural break cartel screen tests whether there are significant fluctuations of model parameters in the DGP of competition as identified above over time. The hypothesis H_0 and alternative hypothesis H_1 that are tested are

$$H_0 : \alpha_t = \overline{\alpha} \quad \text{and} \quad H_1 : \alpha_t \neq \overline{\alpha} \text{ for } t = 1, \dots, T, \tag{2.2}$$

where α_t denotes the full vector of all coefficients at any point of time and $\overline{\alpha}$ is a vector containing the corresponding whole-sample averages of the coefficients. Suitable methods to test the hypothesis in Equation 2.2 are introduced below. These tests provide only a single p-value for the stability of coefficients in the whole sample.

The screen incorporates testing for many different collusive technologies. Any successful

---

[4]This implies that, e.g. the cartel does not strategically alter its input costs (see, e.g. Mueller and Parker, 1992) leading to endogeneity of some of the variables. I am thankful to Daniel Rubinfeld for pointing this out.

[5]Further differencing of the variables needs to be conducted in case a first difference is nonstationary.

attempt of a cartel to affect prices creates a structural break that theoretically can be picked up by the screen. For example, sudden price increases increase residuals significantly and might lead to arbitrary fluctuations of coefficients provided that the price changes cannot be explained by the factors that usually induce price changes. Similarly, a cartel that links prices to a certain input good creates a significant change of the coefficient of the corresponding regressor in the regression. A reduced price variance due to collusion in turn should point to a structural break because price changes are predicted based on the DGP of competition leading to increased residuals.

2.3.2 Selection of suitable structural change tests

Dozens of structural break tests can be found in the literature, because "[...] *there are infinitely many conceivable ways of deviation from the null hypothesis of structural stability.*" (Zeileis *et al.*, 2005, p.100). Most of these structural change tests are designed against a specific H_1 hypothesis and feature the highest power against these specific alternatives rendering them too restrictive to be used for the cartel screen. Examples include tests for parameter constancy against the alternative hypotheses of a single shift (Andrews and Ploberger, 1994; Andrews, 1993), random walks (Nyblom, 1989), and unit roots (Zivot and Andrews, 1992).

For brevity, the discussion of the relevant literature on structural breaks is limited to seminal articles and recent fluctuation-based test introduced below that have been developed for linear models, which can be applied to reduced form equations and are flexible in application with favourable properties for the tasks at hand. As such, this selective literature review does not cover the enormous amount of work that the field of structural instability generates every year. First, in favour of a flexible approach specialised methods for specific applications such as cointegrated models, long memory processes, trending variables, unit roots, or heteroskedasticity are not discussed. Second, this focus excludes, e.g. non-linear and threshold models, Bayesian, bootstrap or non-parametric methods, and methods relating to the GMM framework.[6]

A suitable methodology for a cartel screen that has to deal with unknown forms of structural instability is the *generalized fluctuation test framework* that "*includes formal significance tests but [...] the techniques are designed to bring out departures from constancy in a graphic way instead of parameterizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives.*" (Brown *et al.*, 1975, pp.149-150). Kuan and Hornik (1995) show that the different fluctuation tests can be combined in the generalized fluctuation test framework. Fluctuation tests test for parameter consistency against the alternative of non-constancy, i.e. they do not rely on any assumption about the type of structural break – may it be triggered by one or several structural breaks, or different sources of departure from constancy. Further, they do not require any assumption with respect to the date(s) of the structural break(s).

---

[6]The interested reader is referred to Aue and Horváth (2013), Andreou and Ghysels (2009), and Perron (2006) for more general recent overviews over the literature.

To assess whether coefficient fluctuations are significant, the fluctuation tests calculate a functional called an empirical fluctuation process (EFP), which captures the parameter fluctuations in the data window. The EFP is then compared to a benchmark to assess the stability of the coefficients. In case of parameter constancy, the EFP converges to a functional central limit theorem. Yet if fluctuations are too large, no such convergence exists and the null hypothesis of no structural instability is rejected. Fluctuation tests provide both formal significance tests for the H_0, i.e. a single hypothesis test of Equation 2.2 for the existence of fluctuations in the sample, as well as allow for a graphical inspection of the EFPs. While the former provides an easy-to-interpret and familiar procedure to determine whether there are structural breaks, the latter enables the econometrician to gain information on its dates and lengths.

Different fluctuation tests exist, which vary with respect to the construction of the EFPs and feature different approaches to test the H_0 of constancy of parameters. While no test is dominated by any other test under all circumstances, the OLS-based cumulative sum (*OLS-CUSUM*) and the moving sum (*OLS-MOSUM*) residual-based tests as well as the Moving Estimates (*ME*) parameters-based test introduced below in many circumstances have high statistical power in the presence of either one or several structural breaks (Kim, 2011). For a more in-depth discussion of the approaches, the interested reader is referred to Kuan and Hornik (1995).

The first test used is the residual-based OLS-CUSUM test by Ploberger and Krämer (1992) with the improved alternative boundaries suggested by Zeileis (2004). It outperforms the OLS-MOSUM test when there is a single break, features good finite sample properties, and is suited well to detect relatively short-lasting structural instability (Chu *et al.*, 1995a). Starting from an initial data window at the beginning of the sample, the OLS-CUSUM test sequentially tests the data for structural breaks based on the cumulative sum of residuals. In period t, the model is calibrated for periods 1 until $t-1$ that serve as the benchmark and an expected value of the dependent variable is predicted for period t. If the regression parameters are constant across time, the residuals should fluctuate around zero. If, however, there is structural instability, the residuals systematically increase. The OLS-CUSUM test statistic is based on the accumulation of these residuals, such that the structural instability results in the test statistic to systematically drift away from zero. If this drift crosses a boundary determining whether fluctuations are significant, the test finds evidence for structural instability.

Instead of relying on a growing window of observations like the OLS-CUSUM test, the OLS-MOSUM test is based on a window of fixed width that is sequentially "moved" through the whole sample to calculate the moving sums of OLS residuals. It reports structural instability when the sum of the residuals in the data window exceeds a boundary. As it is only based on a window of the sample that changes across time, it detects multiple structural breaks faster and with a higher power than tests based on a growing window such as the OLS-CUSUM test (Chu *et al.*, 1995a, p.603). For example, Chu *et al.* (1995a, pp.610-611) show that the OLS-MOSUM test has a higher power than F-based tests or CUSUM tests when there are two

structural breaks. Further, it outperforms F-tests in many other circumstances, in particular if they rely on wrong alternative assumptions (Chu *et al.*, 1995a, p.612).

In contrast to the OLS-CUSUM and OLS-MOSUM tests, which are based on the fluctuation of the residuals, the ME test of Chu *et al.* (1995b) extends the analysis to the fluctuations within all regression parameters. As such, an EFP can be observed for each coefficient. This makes the ME test an attractive choice, because it allows to gain inference about the source of a structural break, i.e. which coefficients in the model are subject to structural instability. Further, unlike the residual-based tests, the ME test is sensitive to orthogonal shifts of the mean regressor (Zeileis, 2005; Chu *et al.*, 1995a; Ploberger and Krämer, 1992).[7] The ME test measures the fluctuations of coefficients by comparing the moving-window estimates to those based on the whole sample to pick up structural breaks. As such, it has approximately constant detection probabilities across time and can detect several structural changes.

While the fluctuation tests above are useful for determining the existence of structural breaks in the data, the approach suggested by Bai and Perron (1998, 2003) is better suited to determine the number of breaks as well as their dates.[8] This follows from the fact that information about the number and dates of structural breaks determined with the generalized fluctuation tests are obtained from a graphical inspection of the EFPs. This does not always allow to identify a precise date of a structural break.

Bai and Perron (1998, 2003) develop a procedure to determine the number of breaks as well as the optimal breakdates within a regression. They propose an algorithm based on dynamic programming to determine the number and dates of structural breaks in the estimated regression. The structural breaks are simultaneously obtained based on the global minimisation of the sum of squared residuals (RSS). Selection of the number of breaks is carried out by sequential checks of optimal single break partitions (for details, see Bai and Perron, 2003). The approach requires assumptions similar to those required by the generalized fluctuation test framework, i.e. detrended and stationary variables are required, lagged (dependent) variables can be included, and heteroskedasticity and autocorrelation (between segments) is allowed provided that consistent estimators are used. Thus, we can comfortably apply the approach of Bai and Perron (1998, 2003) for dating structural breaks to the same specification used for the generalized fluctuation tests.

Their approach computes the RSS for models based on different numbers of segments in the data. Increasing the number of segments towards their true number leads to a significant reduction in the overall RSS. Yet, assuming more than the true number of segments does not induce a further significant reduction in the RSS. Similar to the bandwidth in the OLS-

---

[7]Another estimates-based test is the Recursive Estimates (*RE*) test of Sen (1980) and Ploberger *et al.* (1989). Here, the ME test is preferred, as unlike the RE test it provides non-parametric estimates of the corresponding mean functions (Kuan and Hornik, 1995, p.136). Further, it usually has higher power than the RE test when there are multiple structural breaks (Chu *et al.*, 1995b, pp.713-714).

[8]Carlton (2004) and Boswijk *et al.* (2016) apply the methodology of Bai and Perron (1998, 2003) to date cartels as well, albeit based on different econometric approaches and to answer different research questions – I am indebted to Maarten Pieter Schinkel for pointing this out.

MOSUM and ME tests, the econometrician needs so specify a minimal segment width in this test. Different information criteria are considered to assess how many segments (and therefore structural breaks) are in the data. Simulations of Bai and Perron (2003) show that the different information criteria to different degrees might struggle to establish the existence of structural instability. For the purpose of the screen, the Bayesian Information Criterion (BIC) is suited best.[9]

Due to the potential problems in determining the existence of any structural instability in the dating approach of Bai and Perron (1998, 2003), for screening it can be used in conjunction with the generalized fluctuation tests. The existence of structural instability should be determined with the generalized fluctuation framework. If one or several structural breaks have been found with different tests from the framework, the approach of Bai and Perron (1998, 2003) can be used to establish the number and dates of the breaks based on the BIC criterion. This approach combines the strengths of both methodologies and reduces the risk of type I errors.

2.3.3   Estimation procedure and choice of specifications

In the following, the application of the cartel screen is discussed, i.e. how the reduced form price equation has to be combined with the structural break tests to ensure correct inference of the new approach. Chu *et al.* (1996) provide a useful list of aspects with respect to factors that determine the power of fluctuation tests. First, their precision depends on how accurately the DGP is estimated: the better the fit of the model, the higher is the chance to pick up a structural break. Second, chances of detecting a structural break increase with the magnitude of the structural break (i.e. the change of the DGP) and the induced parameter change in the model. Third, shorter windows in the OLS-MOSUM and ME tests, i.e. a smaller bandwidth h, lead to a faster detection of breaks but longer windows have better finite sample properties. It is not known which bandwidth performs best. However, it can be small if the data set is large, but should be increased in small data sets.

The specification should be determined based on competitive periods only to identify a valid DGP characterising competition in the industry. For this purpose, the suggested fluctuation tests can be used by regressing the dependent variable on a constant to assess when sufficient changes in the DGP might occur. This check can be supplemented by the structural break dating methodology suggested by Bai and Perron (1998, 2003) to determine the breakdate. Further, if data on unaffected industries in other regions/countries is available, these regional benchmarks can be used to determine the point of time from which the market under consideration was subject to different pricing dynamics.

To improve the finite sample properties of the estimates given the limited number of observations available in growing or moving windows, the inclusion of regressors should be

---

[9]The BIC does not perform well when there are no structural breaks present by overstating the true number of breaks in the data, but performs well in the presence of structural breaks.

parsimonious. This also prevents the risk of overfitting in the model from affecting inference in the structural break tests. The rule here is as many regressors as necessary, but as few as possible. Lagged regressors as well as lagged dependent variables are allowed and a compromise between goodness of fit and number of included lags has to be made. Priors do not only guide the inclusion of variables into the specification, but also the decisions which lags to use. The parsimonious lag structure of the variables should further be adapted to achieve a good fit of the data. The better the DGP can be modelled, the higher the power of the structural break tests. As such, the specification is not either correct or wrong. Lag specifications that differ from the true specification but capture some of the correlations that determine the DGP can result in the screen to work reasonably well albeit with less power.

The regression model must not feature any endogenous regressors, such that appropriate corrections need to be conducted if deemed necessary. Further, all variables need to be detrended if necessary and then checked for stationarity.[10] These steps are necessary to prevent spurious relationships from biasing the estimates and "exploding" variances to affect the statistical inference.

Seasonality needs to be removed from the data to ensure correct inference. While different methodologies exist to remove seasonality, the seasonal-trend decomposition procedure based on non-parametric loess smoothers (STL) by Cleveland *et al.* (1990) is used here. A strength of this approach is that it allows for a robust estimation of seasonality that is not biased by aberrant behaviour of the data, as can follow from the effect of cartels on prices. Removing rather than modelling seasonality is necessary for the screen, as the growing or moving data windows do not allow for a robust distinction between seasonality and cartel price changes.

To reduce the risk of type I errors, the OLS-MOSUM and ME tests are used with multiple bandwidths and it is only concluded that breaks are present if the majority of tests point towards structural instability. First, this ensures that the bandwidth choice does not drive the results. Second, this tackles potential convergence issues of the structural break tests: given that the power of the tests critically depends on the size, number, and type of structural breaks, no clear guideline can be provided on their convergence properties, although the tests tend to perform reasonably well for a sample size of 100 observations (see, e.g. Kim, 2011; Kuan and Hornik, 1995). Further, such handling of the results in a pre-defined and planned way addresses potential multiple comparison problems without relying on potentially problematic corrections of p-values.[11]

---

[10]Tests for unit roots have to be chosen carefully. Structural breaks in the time series can be misinterpreted as nonstationarity by Augmented Dickey Fuller and Phillips-Perron tests. An often suitable unit root test is proposed by Zivot and Andrews (1992), which tests for unit roots against the alternative of a structural break.

[11]This approach is suitable because the p-value comparisons are few in numbers only and complementary. Further, as the tested p-values are strongly and positively correlated with each other, many approaches to adjust p-values to address multiple testing are likely to produce misleading results by being overly conservative. The fluctuation tests used here tend to be conservative reducing the risk of Type I errors (Kuan and Hornik, 1995).

Segment widths for the dynamic programming algorithm to date breaks by Bai and Perron (1998, 2003) should be smaller than window widths of the fluctuation tests. If the distance between two breaks in the data is smaller than the minimal segment width defined for the algorithm, correct inference on the number and exact dates of the breaks is not possible. This is discussed in greater detail in Section 2.4.4.

2.4 Application: European pasta industries

2.4.1 The industry

As an application, the structural break cartel screen is used on the pasta industries in Italy, Spain, and France based on monthly data between 2003 and 2012. While the former two were subject to cartels between October 2006 to March 2008 and July to October 2007, respectively, the latter did not feature any cartel (Notaro, 2014; Ordóñez-de Haro and Torres, 2014). As such, we would expect a properly working cartel screen to detect cartels in Italy and Spain, but not to report structural breaks in France. Cartel formation in Italy and Spain was triggered by significant price increases in the main input good in pasta production – durum wheat – during 2006 and 2007.

These markets represent useful cases to test the screen under difficult conditions: In Italy, potential post-cartel tacit collusion might have prevented a clear breakdown of the DGP of collusion. In Spain, the cartel lasted only for three months, providing few observations only to detect collusive behaviour. In France, market prices rose significantly due to a strong durum wheat cost shock in the absence of a cartel. Absent significant changes in supply, demand, and market characteristics across countries, differences in the reaction of markets to common supply shocks allow for a preliminary analysis of market prices to the cost shock.

Figure 2.1: Cumulated price changes in the pasta price indices
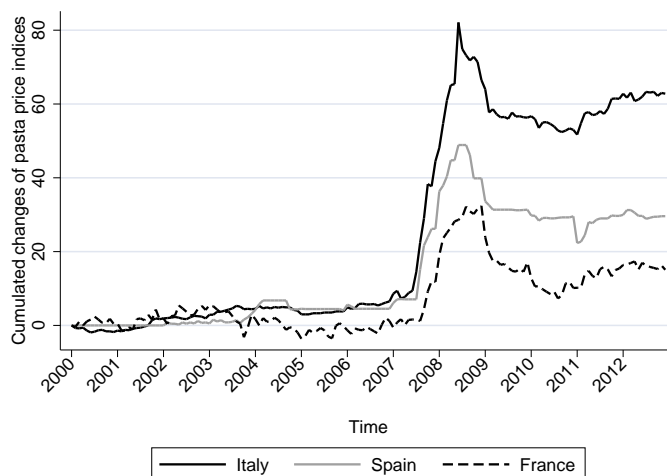
Notaro (2014) discusses the Italian pasta cartel in detail and estimates the average overcharge to be around 11 % of the price under competition. The cartel covered 76% of the market (90% if private labels produced by the cartelists are included). Market coverage of the Spanish cartel was less complete with a joint market share of 55-60% including private labels produced by the cartel members (Spanish Competition Authority, 2009). Between May 2006 and May 2008, the price charged by the producers to retailers in Italy increased by about 51.8% (Italian Competition Authority, 2009). Using a dynamic treatment effects analysis Notaro (2014) shows that the observed cost changes including the significant price increases for durum wheat could not explain the price increases for pasta. In addition, he argues that the market was characterised by a significant level of tacit collusion after breakdown of the cartel.

Figure 2.1 contains the cumulative price changes in % between 2000 and 2012 in the pasta price indices of Italy, Spain, and France (International Pasta Organisation, 2012).[12] While the prices of pasta develop similarly in the three countries before 2007, the indices rise significantly further in Italy and Spain than in France. While the pricing of the French industry suggests competitive price increases due to the common input cost shock of around 30%, prices at times rose by up to 82.1% and 48.9% in Italy and Spain. The pronounced price differences between the Italian and Spanish pasta indices compared to French prices after breakdown of the cartels are hard to explain by potential changes in the industry structures.

2.4.2   Specification of the DGPs of competition

In the first step in the application of the screen, models describing the markets' DGPs of competition are established based on periods assumed to be characterised by competition. The strong dependence of pasta prices on few input goods renders the industry suitable for an analysis with the proposed cartel screen: Notaro (2014) reports that domestic and imported durum as well as labour and energy costs represented 73% and 77% of the total direct costs and 54% and 60% of the total costs of the Italian pasta industry in 2006 and 2007.

In the following, five cost shifters are used together with a demand shifter for the screen. All variables are detrended and included as first differences to achieve stationarity. Augmented Dickey-Fuller and Phillips-Perron tests for unit roots confirm that all variables are stationary at a 1% significance level. A potential caveat of the analysis below is the absence of variables that control for market characteristics due to a lack of data. Thus, the assumption is imposed that market characteristics in the industries under consideration were roughly stable across time which should be likely given the short time window included in the analysis.

---

[12]Figure 2.1 presents cumulated price changes instead of the absolute price levels to simplify the graphical comparison of commonalities and differences in the trends. Otherwise, the time series would lie further apart in (absolute) values because of different base years and require a scaling of the vertical axis that makes it harder to visually derive information from the Figure. Further, this renders it easier to observe how long and pronounced certain price levels persisted compared to the price changes that are used in the empirical analysis and which are displayed for the Italian market in Figure 2.4.

Denote *Domestic durum* and *International durum* as the domestic and international durum price indices. The domestic durum wheat prices capture the majority of actual costs of the main input good in pasta production. The domestic prices are usually fixed in contracts between suppliers and customers for several months in advance. The international durum price measures both import prices and expected price changes of durum wheat. Further, let *Labour costs*, *Energy*, *Energy sq.*, and *Borrowing costs* denote the Italian price indices for industry labour, (squared) energy costs, and costs of borrowing capital for companies. *Expenditure* measures household expenditure on goods to approximate demand fluctuations. Tables 2.7 and 2.8 in the appendix contain descriptions, sources, and descriptive statistics of all variables used in the analysis.

The optimal specification for a DGP that characterises competition in the industry must be determined based on unsuspicious, non-collusive periods only. An inspection of the pasta industry prices in Figure 2.1 shows that prices increase significantly during 2007 suggesting that these periods should be excluded to establish the specification. Further, regressing price changes on a constant and testing for parameter instability with the structural break tests indicates changes in the average price changes in the mid of 2007 for all three countries. Therefore, only periods before 2007 are considered for the estimation of the DGPs of competition.[13]

The pasta industry is not characterised by contemporaneous cost pass-through, as input prices are often determined by fixed contracts several months in advance. The three national pasta industries are characterised by different market characteristics and are dominated by different firms. As such, their DGPs of competition are characterised by divergent lag structures. Parsimonious lag structures are used that in each case provide the best fit given the priors for the observed pricing dynamics.

While not all supply and demand shifters are always significant, they are included due to their theoretical importance for prices as well as the fact that they might become significant determinants of price changes outside the time frame used to determine the DGP of competition to prevent potential omitted variable bias. No lagged dependent variables are included in any of the regressions, as there is neither a theoretical reason to do so in the pasta industry nor are there observable patters of price changes that would suggest an inclusion. The specifications of the different DGPs of competition can be found in Table 2.1.

In line with theoretical expectations the coefficients of all significant (non-squared) cost and demand shifters are positive. The results are robust to different functional forms. The estimates in Table 2.1 are not used for the analysis. They are merely used to establish the lag structure that is later tested for stability with the structural break tests and for various specification tests.

---

[13]For Italy, only the periods prior to the cartelisation of the industry in October 2006 are used. Yet, this exclusion restriction has no effects on results, as the cartel did not influence market prices before June 2007.

Table 2.1: Specification of DGPs of competition

	Italy	Spain	France
Domestic durum	0.017**	0.047***	0.093***
	(0.008)	(0.012)	(0.032)
	4	3	4
International durum	0.007**	0.009**	0.013**
	(0.003)	(0.003)	(0.006)
	3	3	0
	-	-	0.021***
			(0.007)
			4
Labour costs	0.012	0.069	2.791*
	(0.011)	(0.053)	(1.425)
	0	2	5
Energy costs	0.175***	0.394***	0.331***
	(0.044)	(0.061)	(0.103)
	3	1	6
Energy costs sq.	-	-	-0.209***
			(0.045)
			6
Borrowing costs	-1.114	-0.911	3.967***
	(0.692)	(0.551)	(1.340)
	2	4	1
Household expenditure	0.000	0.140*	0.414
	(0.000)	(0.083)	(0.412)
	4	6	2
Adj. R^2	0.373	0.486	0.523
Observations	42	43	46

Notes: The dependent variables are the domestic producer prices for pasta products. All variables are included as first differences. The first row for each variable features the coefficient, with ***, **, and * showing significance of the coefficient at a 1%, 5%, and 10% significance level. The second row contains the (robust) standard errors in brackets. The third row indicates which lag of the variable is used with 0 denoting the contemporaneous value.

Heteroskedasticity is suspected for the Spanish market, such that heteroskedasticity and autocorrelation-robust standard errors are reported throughout the analysis for this market.[14] The Italian and French markets show no signs of either heteroskedasticity or autocorrelation as reported by appropriate tests, such that standard OLS residuals are used for these markets throughout the analysis. The careful use of heteroskedasticity and autocorrelation-consistent

---

[14]This does not affect results of the structural break tests in Section 2.4.3. Using unadjusted covariance matrices for the structural break tests in the Spanish test rather than HAC-consistent matrices provides p-values of 0.006 for the OLS-CUSUM test, 0.010, 0.013, 0.021, 0.021, and 0.023 for the OLS-MOSUM tests with window widths of 15%, 20%, 25%, and 30%, and p-values of 0.01 for all ME tests based on the same window widths.

covariance matrices is necessary, as samples in the growing or moving windows in the fluctuation tests typically are small. This ensures that poor asymptotics in small samples do not lead to significant biases in the covariance matrices leading to wrong inference (Wooldridge, 2012; Angrist and Pischke, 2008).

Seasonality is only detected in regressions for the French pasta industry, whereas no seasonality is present in those for the Italian and Spanish markets. Therefore, the seasonality in the French data is removed with the nonparametric STL deseasoning approach by Cleveland *et al.* (1990). Ramsey's RESET test results for all regressions indicate that the linear functional form is correct (Ramsey, 1969): in case of the French industry, energy costs must be quadratic such that the functional form of the specification is valid.

2.4.3 Determining the existence of structural breaks

In the second analytical step, the fluctuation tests are applied to the specifications proposed in Table 2.1 to test for structural breaks in the DGPs identified above. As such, the regressions are tested for stability throughout all available time periods between 2003 and 2012, which yields sample sizes of 115 (Spain), 117 (Italy), and 118 (France) observations. The sample size should be high enough to obtain reasonably reliable inference (in particular because the price changes are very pronounced between 2006 and 2008, see, e.g. Kuan and Hornik, 1995). All structural break estimations are carried out in the statistical software package R with the package *strucchange* by Zeileis *et al.* (2002).[15]

To address potential convergence issues arising from the sample size, all three tests are applied to all three countries and different bandwidths are specified for both OLS-MOSUM and ME tests. The choice of the bandwidth parameter h denoting the % of total observations to be included in the moving windows is no obvious task and requires the researcher to make a choice. As pointed out in Section 2.3.3, shorter windows increase the sensitivity of the tests to fluctuations of the coefficients and improve detection of short periods of instability in the data. Yet, overly small window widths also increase the risk of false positives. OLS-MOSUM and ME estimates based on smaller window widths might be subject to low power.

Therefore, four window widths $h = [0.15, 0.2, 0.25, 0.3]$ are tested and results across the different tests are compared with each other. To reduce the risk of Type I errors, we will only conclude that there is structural instability if evidence for it arises in the majority of the tests. This also helps to ensure that the manual choice of window widths does not determine the conclusions drawn from the results.

Table 2.2 reports the p-values for the tested H_0 hypotheses of stability of the regression models throughout time. The results show that there is strong evidence for structural breaks in the

---

[15]I am indebted to Achim Zeileis for providing access to a developer version of *strucchange* that allows to use HAC covariance matrix estimators for the estimation of empirical fluctuation processes.

Table 2.2: Break test p-values

	h	Italy	Spain	France
OLS-CUSUM	-	0.000	0.015	0.755
OLS-MOSUM	0.15	0.010	0.014	0.088
	0.20	0.010	0.036	0.043
	0.25	0.010	0.050	0.123
	0.30	0.010	0.052	0.151
ME	0.15	0.010	0.010	0.034
	0.20	0.010	0.010	0.276
	0.25	0.089	0.010	0.235
	0.30	0.160	0.010	0.212

DGP of pasta prices in Italy and Spain. The ME test merely fails to detect breaks in the Italian market for large window widths. This might occur if the industry is characterised by a distinct DGP for few periods only, such that the coefficients in the window insufficiently capture the changed DGP. Yet, given the strong and largely robust results, the structural break cartel screen suggests that DGPs changed significantly both in Italy and Spain, such that price changes were governed by distinct processes at different points in time in these markets.

A different conclusion arises for the French market, in which there was no cartel. Indeed, despite the strong and sudden price increases during 2007, most of the tests report that no structural break is present in the French market. Note that there are Type I errors for small window widths for the OLS-MOSUM and ME tests pointing towards false positives for the French market. Nevertheless, given that most results fail to reject the H_0 hypothesis of stability of the coefficients over time, based on the pre-defined rule to interpret the p-values it can be concluded that the DGP is stable across time in France. Put differently, price changes in the French pasta industry can be explained throughout time by established patterns of cost pass-through and the reaction to changes in demand.

The absence of reported instability in France shows the advantage of including price shifters in the estimated regressions. This relaxes the assumption about constancy of all supply and demand factors needed for the approach to provide valid inference. As is shown in a robustness check in Section 2.4.7, the tests would have reported structural breaks if the price changes are only regressed on a constant controlling for changes in the mean of prices changes over time.

A graphical inspection of the EFPs of the residuals in the different tests can be used to gain an impression of the structural instability in the models. The EFPs of the OLS-MOSUM and ME tests for the Italian and Spanish markets are plotted (solid lines) together with the boundaries of the corresponding 95% confidence intervals (horizontal dashed lines) in Figure 2.2.[16]

---

[16]Results of the OLS-CUSUM tests are plotted in Figure 2.6 in the appendix in Section 2.7.

Figure 2.2: OLS-MOSUM and ME EFPs for Italy and Spain

(a) OLS-MOSUM Italy

(b) OLS-MOSUM Spain

(c) ME Italy

(d) ME Spain
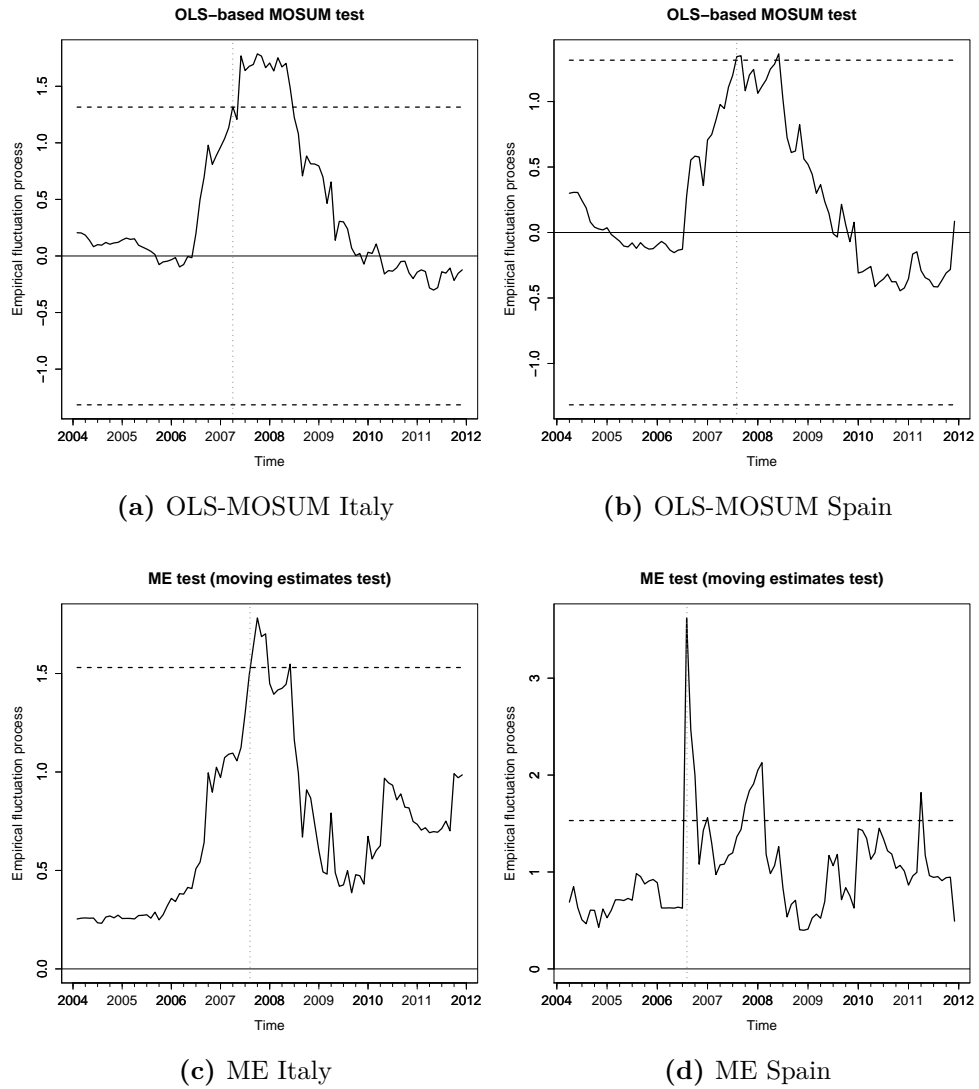
For the OLS-CUSUM and OLS-MOSUM tests, these EFPs are functions of residuals from a model fit on a growing or moving data window. An increasing EFP in a period represents a decline in the goodness of fit of the model. This results in larger residuals in periods with structural instability or change that can be spotted in the graphical representation of the EFP. For the ME test, the interpretation of EFP fluctuations is similar, although it is a function of the difference between coefficients in the models based on the moving window and on the whole sample.

The DGP of pasta prices in Italy is stable until the mid of 2006 when the EFP suddenly increases significantly and crosses the boundary in early 2007. Both OLS-MOSUM and ME tests point towards a significant instability of the DGP between early 2007 and the mid of 2008, after which the industry reverts to a stable DGP. Yet, as the higher EFP in the ME test after the mid of 2008 shows, the regression model fitted to the DGP of competition provides a worse and less stable fit after cartel breakdown, possibly due to post-cartel tacit collusion. A similar picture arises for the Spanish market.

As the ME test is a parameter-based test, we can also look at the EFPs of the model coefficients for, e.g. the Italian market, which are plotted in Figure 2.3. This provides insights on the sources of instability in the model. In each case the EFP is a function of the difference of the coefficients of the models fit on the moving data window and the whole period. The difference between the estimated coefficients gets larger and results in a more positive or negative EFP the more the coefficient diverge between the two models. This divergence implies a difference between the two models that is the result of structural instability of the DGP.

On the one hand, the cartel could engage in arbitrary price changes not linked to price developments of price shifters. As the estimated regression has no intercept, such a collusive behaviour likely results in instability of all coefficients, and (arbitrary) significant fluctuations in one or more coefficients. On the other hand, if the cartel links price changes to the development of specific price shifters, a strong and persistent change for that price shifter can be expected, while the other coefficients are not subject to strong and arbitrary fluctuations.

Figure 2.3: ME test results for coefficient stability in Italy
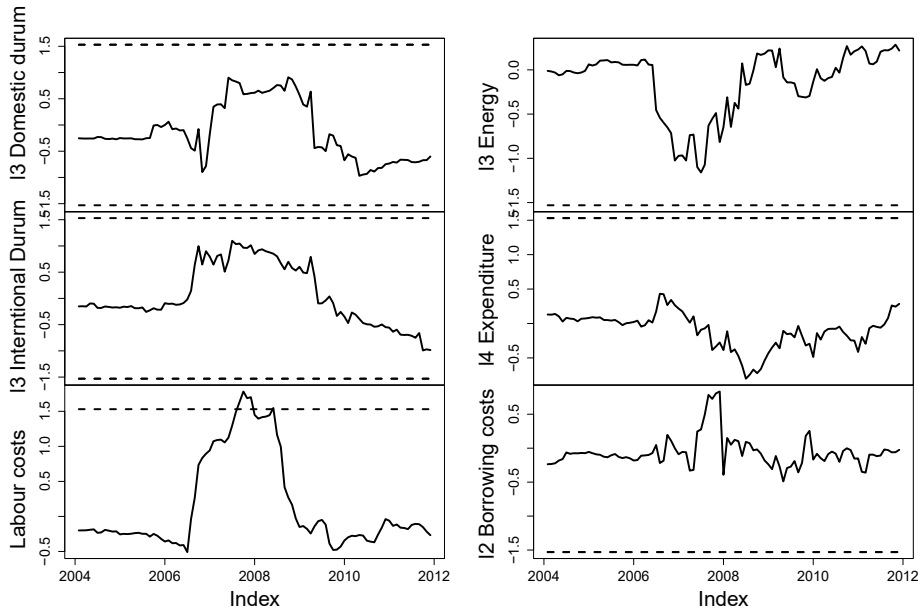
The Figure shows that all coefficients are subject to instability starting in early 2007, when the cartel first affected market prices. However, only the fluctuations in the Labour cost coefficient are significant, even though – unlike durum wheat – it was not subject to significant price changes. Taken together, this suggests that the structure of cost pass-through was not significantly altered by the cartel, but that it engaged in arbitrary price increases.

2.4.4   Dating the breaks

In the third step of the analysis, the dynamic programming algorithm of Bai and Perron (1998, 2003) is used to determine the number and dates of the structural breaks in the Italian and Spanish markets. For this purpose, the specifications established in Section 2.4.2 are

applied to the whole sample, and the optimal partition into different segments is carried out by the algorithm.

Similar to the choice of window widths for the OLS-MOSUM and OLS-CUSUM tests, a minimal segment width has to be chosen manually for the dynamic programming algorithm. Segment widths should not be chosen too large, as this could negatively affect identification: if the segment width is larger than the distance between two structural breaks, the breaks cannot be dated correctly. To reduce the chance that arbitrary choices affect the conclusions drawn from the results, again different values are tested within reasonable bounds. Therefore, segment widths of $s = [0.07, 0.08, 0.1]$ are used for dating the structural breaks. Results are reported in Table 2.3.

Table 2.3: Dates of structural breaks

s	Italy			Spain			
0.07	6/2007	5/2008	1/2009	4/2007	12/2007	10/2008	5/2011
0.08	6/2007	5/2008	2/2009	3/2007	12/2007	10/2008	5/2011
0.10	6/2007	5/2008	4/2009	7/2007	-	10/2008	5/2011

The results show that the dating algorithm detects three structural breaks for the Italian industry, and either three or four breaks for the Spanish industry. For Italy, the algorithm detects a first break induced by cartel formation in June 2007. A comparison of this result with the cumulated price changes as shown in Figure 2.1 suggests that this break coincides with the first period when the cartel that came into existence in October 2006 had a first visible impact on industry prices. Further, industry prices collapsed by about 20% after cartel detection in March 2008. The dating algorithm dates the breakdown of collusive pricing to June 2008. The third reported break varies by segment width and captures the point in time in which the industry entered a state of suspected tacit collusion.

A visualisation for the breaks identified for the Italian market for $s = 0.1$ can be found in Figure 2.4, in which the price changes over time are shown by the grey line, and the mean in each of the four distinct price phases in the industry are shown by the black line. The vertical dashed lines show the detected structural breaks, and the small horizontal black lines at the bottom of the break lines represent the 95% confidence intervals for the breaks. Thus, the cartel screen works well for the Italian market, in which it detects the cartel 9 months before the Italian Competition Authority learned about it.

The choice of the segment width parameter can affect the dating of a break when the distance between two breaks is smaller than the minimal segment width. This becomes evident for the Spanish market. The dating algorithm reports three or four structural breaks for the Spanish pasta prices. The short length of the Spanish pasta cartel – it lasted only between July and September 2007 – causes problems for the dating algorithm, as the structural instability is much shorter than the minimum segment width of the algorithm. Compared to the Italian market, this leads to less stable break dates across the different segment widths in the Spanish

Figure 2.4: Identified structural breaks in the Italian pasta industry
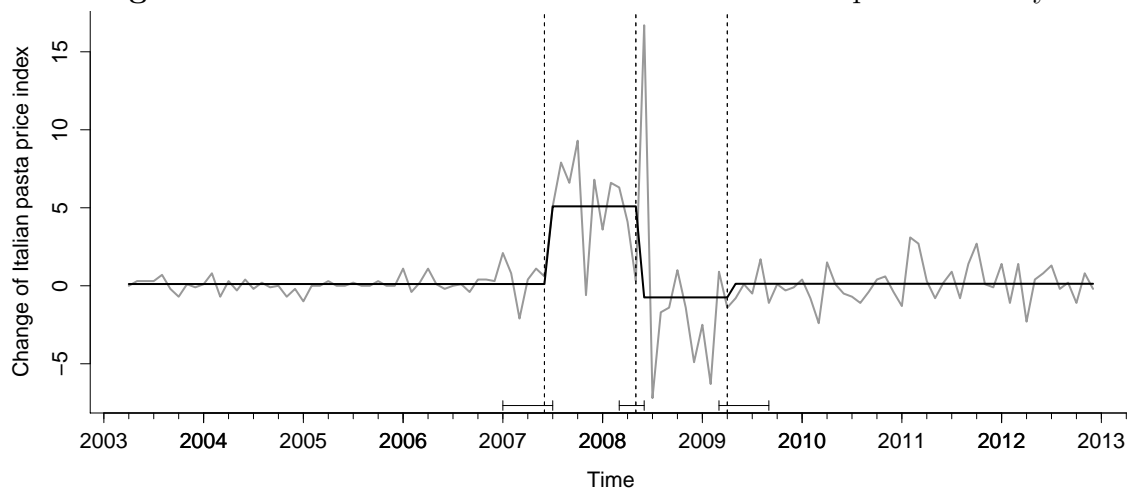
markets. Therefore, neither the establishment nor the breakdown of the Spanish pasta cartel after detection can be dated with precision for $s = 0.07$ and $s = 0.08$ leading to some arbitrary results for the exact break dates. These break dates that vary significantly between different segment widths point to a lack of robustness of the results and should not be considered when drawing conclusions on the dates of structural instability.

The above result highlights a caveat of the methodology used to date the structural breaks: cartels need to last sufficiently long enough to enable a proper dating. Only for $s = 0.1$ does the reported structural break with the start of the cartel, but no break is reported for its breakdown. However, two breaks that are further apart from each other can be dated consistently. The first is found for October 2008, when Spanish pasta prices stop to decline and converge towards the relative price level in the non-cartelised French market. Like Italy, this might indicate the start of post-cartel tacit collusion in the Spanish market. The second break is found for June 2011, when a noticeable price decline not existent in the other markets occurs, after which the price variance in the Spanish market increases. This suggests a collapse of the suspected tacit collusion in the market ending the resulting price hysteresis and bringing price variance more in line with the French market.

2.4.5   Comparison to other screens

To highlight the greater flexibility of the structural break screen compared to other commonly used screens, the coefficient of variation screen of Abrantes-Metz *et al.* (2006) and the GARCH variance screen of Bolotova *et al.* (2008) are subsequently applied to the three pasta markets. As these screens require the user to specify which periods are tested for collusion, in both cases it is assumed that in all three markets the observations from 2003 to the end of 2006 serve as the benchmark of competition (*Phase 1*). The potential periods of collusion are divided into two phases: *Phase 2* lasts from January 2007 to June 2009 and *Phase 3* from July 2007 to December 2012. This segmentation roughly follows the noticeable patterns in the prices in Figure 2.1.[17]

---

[17]The succeeding results are robust to reasonable alternative specifications of the tested periods.

Table 2.4: Variance test: coefficient of variation

		Time period			Diff. coeff. of var. (in %)	
		Phase 1	**Phase 2**	**Phase 3**	**Phase 1 to 2**	**Phase 1 to 3**
Italy	Std. dev.	0.920	24.722	3.661		
	Mean	101.149	143.100	154.517	1798.65%	145.94%
	Coeff. of var.	0.009	0.173	0.024		
Spain	Std. dev.	1.583	14.793	1.955		
	Mean	75.198	99.769	99.990	1272.27%	-14.96%
	Coeff. of var.	0.021	0.148	0.020		
France	Std. dev.	2.018	11.801	2.857		
	Mean	99.000	115.591	112.684	817.09%	49.67%
	Coeff. of var.	0.020	0.102	0.025		

Table 2.4 reports the mean, the standard deviation, and the coefficient of variation of the price indices for all three countries and phases. The last two columns report the percentage change of the coefficient of variation from Phase 1 to Phase 2 and 3, respectively. As variance-based screens see reduced price variances as signs of collusion, a reduction of the price variance from Phase 1 to Phase 2 and/or Phase 3 in these screens indicate a "detected" cartel. Yet, the results show that the variances during the cartel activity are significantly increased in Italy and Spain. The only sign of potential collusion is reported for post-cartel periods in Phase 3 for Spain.

Results of GARCH-based variance tests in the spirit of Bolotova *et al.* (2008) are reported in Table 2.9 in the appendix. They are conducted by applying the same variables and lag structures as reported in Table 2.1 to the whole sample from early 2003 to the end of 2012. However, dummies for Phases 2 and 3 are added to the specification to allow for significantly higher price changes during the suspected periods of collusion. To test for significantly increased variances, the dummies for Phases 2 and 3 are added to the variance equation as well to model a changed variance in the periods of collusion. The GARCH models do not report reduced but increased variances for all three markets both in Phase 2 and Phase 3. Thus, like the coefficient of variation approach, the GARCH tests do not detect any signs of collusion. This result is robust to various specifications of the GARCH models.

A potential reason for the variance-based screens' failure to detect the cartels in Italy and Spain is that they might require comparably long periods of stable cartel activity. In the markets under investigation, the cartels were short-lasting and featured steep price rises and declines, such that measurable reduced price fluctuations based on stable collusion could not set in.

2.4.6 Extension to monitoring

The cartel screen might not only be used to test past prices for signs of collusion, but also automatised and extended to monitor incoming data for structural breaks. Monitoring is in

particular relevant for competition authorities or compliance efforts by companies, which want to learn about potential collusion in markets so far marked by competition as soon as possible.

Extensions of the fluctuation tests to monitoring have been developed by Chu *et al.* (1996), Leisch *et al.* (2000), and Zeileis (2005). The monitoring variants of the OLS-CUSUM, OLS-MOSUM, and ME tests essentially work like their variants for "historical" data. Similar to the historical testing above, monitoring requires a DGP of competition to be defined for a collusion-free "historical" benchmark. This benchmark is then sequentially compared to data from succeeding periods as it becomes available. Monitoring tests monitor new data by comparing the estimates from growing or moving windows to the whole-sample estimates of the parameter(s) derived from the historical periods. Structural instability results in the difference between the two (sets of) estimates to exceed a boundary. Once the monitoring framework has been set up, screening can be conducted automatically and merely requires to re-run the monitoring test as new observations become available. Additional work or attention is not required until a break is reported.

Figure 2.5: Monitoring in the Italian pasta industry
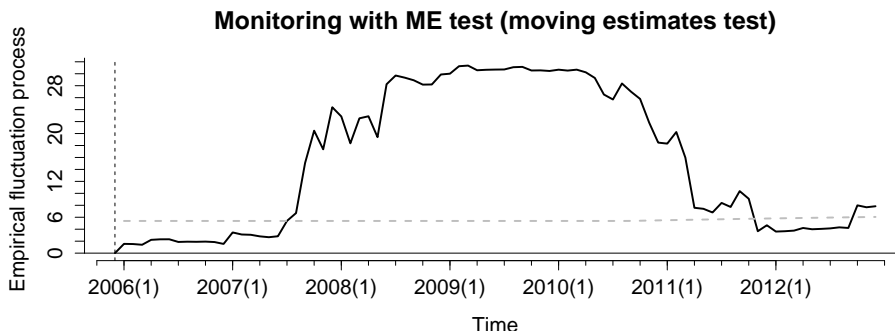
As an example, monitoring is conducted for the Italian pasta industry using the DGP specified in Table 2.1 and using observations between early 2003 and the end of 2005 as the historical benchmark. Figure 2.5 shows the EFP of the ME monitoring test with window width $h = 0.1$. In the ME monitoring test, the EFP is a function of the differences between the estimated coefficients in the moving window in the monitoring period and those from the historical period. An increase in the EFP indicates a growth in the difference of the estimated regression parameters that results from a change in the DGP in the monitoring period compared to the historical period.

Starting from the first period outside the historical data periods, structural stability is constantly monitored. In the example, the monitoring screen would have detected and reported collusion by July 2007 (i.e. 9 months before the Italian Competition Authority), which is shown by the EFP of the test (black line) crossing the boundary (horizontal grey dashed line) representing a 5% significance level.

2.4.7 Robustness checks

General statements about the robustness of the structural break cartel screen are hard to derive because of the almost infinite number of possible combinations of types, numbers, and magnitudes of structural breaks. Yet, one might try to address the robustness in a specific application by testing how changes in the specification of the DGP affect the results.

Two alternative DGPs are specified as robustness checks. In *Constant*, the regression consists of a constant only. The specification controls for changes in the mean of price changes, i.e. to what extent results might be driven by fluctuations in the dependent variable over time.[18] This represents the extreme case of having no price shifters available. In *Durum*, all variables other than the domestic and international durum wheat price changes are dropped. This specification tests how well the screen performs when only the primary price shifters are available/included but other main price shifters are missing. P-values of the structural break tests for both specifications are reported in Table 2.5.

Table 2.5: Alternative specifications of the DGPs – Break test p-values

Specification	h	Italy		Spain		France	
		Constant	Durum	Constant	Durum	Constant	Durum
OLS-CUSUM	-	0.012	0.000	0.130	0.004	0.491	0.395
OLS-MOSUM	0.15	0.000	0.000	0.010	0.010	0.041	0.075
	0.20	0.010	0.010	0.033	0.013	0.036	0.052
	0.25	0.010	0.010	0.077	0.021	0.099	0.139
	0.30	0.010	0.010	0.130	0.029	0.131	0.162
ME	0.15	0.010	0.256	0.010	0.113	0.041	0.211
	0.20	0.010	0.236	0.031	0.013	0.036	0.207
	0.25	0.010	0.160	0.074	0.010	0.099	0.148
	0.30	0.010	0.113	0.126	0.010	0.131	0.112

Qualitatively, the results of the specification *Durum* are mostly in line with those of the full specification in Table 2.1 indicating a certain level of robustness of the above results: evidence of a structural break in Italy is less pronounced but still present, whereas there is still strong evidence for structural instability in Spain and for no structural breaks in France. Therefore, in the present case, controlling for the main input costs is sufficient to ensure correct albeit less clear results. A different picture arises in the specification *Constant*, in which structural instability is consistently reported throughout all three countries. Therefore, failure to control for the relevant price shifters would wrongly raise a red flag with potential signs of collusion in the French market. This highlights the necessity to limit the application of the screen to cases in which at least the most relevant price shifters are available.

Results of the robustness check might not carry over to other applications. For example, for price time series that do not feature large price rises during periods of collusion, it is more likely that incomplete specifications of the DGP result in no detected structural breaks. As

---

[18]This approach is very similar to the price-change based cartel screen of Hüschelrath and Veith (2013).

pointed out before, the empirical power of the tests is higher the better the DGP can be modelled.

Table 2.6 reports results of the dating algorithm for the two alternative DGPs intended as robustness checks. For the specification *Constant*, the dating algorithm quite precisely detects the start of price increases induced by the cartel as well as the turning point towards declining prices after their collapses both in Italy and Spain. In addition, in both cases the end of price declines towards new market equilibria is dated as well. For the *Durum* specification, the algorithm reports a single break only for Italy for February 2009, but no break in Spain.

Table 2.6: Alternative specifications of the DGPs – Dates of structural breaks

| Specification | Constant | | | | | | Durum | |
s	Italy			Spain			Italy	Spain
0.07	6/2007	6/2008	4/2009	7/2007	6/2008	3/2009	2/2009	-
0.08	6/2007	6/2008	2/2009	7/2007	6/2008	3/2009	2/2009	-
0.10	6/2007	6/2008	5/2009	7/2007	6/2008	-	2/2009	-

Thus, at least in the present application, dating collusion appears more vulnerable to misspecification than detection, which naturally results from greater challenges related to counting and dating breaks instead of detecting instability *per se.* In such cases, a graphical inspection of the EFPs similar to that shown for the full specification of the DGPs in Figures 2.2 and 2.6 of the structural break tests can provide insights on the break dates albeit with less precision.

2.5 Discussion

In the following, advantages and limitations of the screen are discussed. With respect to limitations in scope, first and foremost – as all behavioural cartel screens – its application assumes that a cartel has significant control over the market price affecting the DGP. Thus, a cartel is only detected once it has a significant effect on prices. As a consequence, depending on the speed and magnitude of the cartel's impact on prices after its establishment or the time it takes participating firms to change their pricing strategy after its demise, the screen might only detect the start and end of the cartel with lags. Second, determining the number of breaks and the correct dates can be difficult if the break dates are not far apart. Third, in line with some of the other behavioural screens, it cannot distinguish between explicit and tacit collusion. Fourth, the screen assumes that at least some data is available about price shifters, and that unobserved important variables are not subject to significant changes across time.

The cartel screen suggested in this essay has several advantages compared to other behavioural cartel screens. The biggest advantage is that it is not based on a particular collusive marker for the identification of cartel activity – it only requires the cartel to affect industry prices. Thus, it can detect different collusive technologies which reduces the risk of type II errors

compared to other cartel screens. In addition, the approach works equally well for both stable and unstable/short-lasting cartels. Further, the identification strategy is not based on a manual (arbitrary) choice of time periods that are to be formally tested for collusion. Therefore, it avoids the problem faced, e.g. in the ARCH-based price variance test of Bolotova *et al.* (2008), in which the results of formal hypothesis tests depend on the exact definition of the time periods to be tested for collusion. Finally, it is not as prone to manipulation as the established cartel screens.

It is often argued by critics of cartel screens that they are useless because cartels will react to screening by developing mechanisms that trick the screens. For example, a cartel that knows that the market is screened with a price variance-based cartel screen could easily avoid detection by engaging in arbitrary price changes to keep the price variance stable over time. This problem does not exist for the proposed structural break cartel screen. Any artificial cartel price increase does alter the DGP. The only way for a cartel to avoid detection is to engage in many small price increases that would not be detected as significant alterations of the DGP but being hidden in the margin of error. However, this has very destabilising effects on collusion rendering such behaviour unlikely to occur, as it reduces the profits from collusion compared to the (fixed) penalties imposed for the formation of the cartel (see, e.g. Harrington, 2005, 2004a).

2.6   Conclusion

In this essay a new empirical methodology to detect previously unknown cartels and to date the beginning and end of cartels is proposed and tested. Based on the idea that cartels change the DGP of industry prices compared to periods of competition, an approach is presented that detects collusion by testing the stability of this DGP with structural break tests. Different structural break tests suitable for this task are discussed and those from the generalized fluctuation test framework are found to be the most useful for determining the existence structural breaks in an industry. The approach by Bai and Perron (1998, 2003) is suggested as a useful complement to provide exact estimates of both the number and dates of structural breaks. In addition to historical testing for breaks, an extension of the methodology to monitoring of markets is presented.

The new structural break cartel screen is applied to three European pasta industries: the markets in Italy and Spain were cartelised, while no cartel was present in France. The screen successfully detects the Italian pasta cartel 9 months before the Italian Competition Authority learned about the cartel. Despite its short length of three months only, the Spanish cartel is detected as well. However, the screen cannot precisely date the beginning and end of the cartel due to its short length. Although the pasta prices were characterised by a sudden and significant price increase of about 30%, the screen does not wrongly report the existence of a cartel in the French market. This result follows from the fact that – unlike in the Italian

and Spanish industries – the price increase in France could reasonably well be explained by changes in the input costs. Robustness checks suggest that the detection of structural instability is reasonably robust to the inclusion of additional independent variables, as long as the main determinants of price dynamics are included. However, the dating algorithm used to determine the start and end of collusion is more susceptible to the specification of the DGP.

The structural break cartel screen is a useful tool for detecting and dating cartels. While a conservative econometric approach can reduce the risk of type I errors, e.g. by requiring different structural break tests to unanimously report a break across different specifications, a change in the DGP should not be seen as definite evidence of a cartel. Instead, structural instability calls for a more in-depth analysis of the industry. Only if a supplementary analysis of the market shows that the change in industry pricing cannot be attributed to other factors not considered in the econometric model can the screen be seen to provide substantial evidence for collusion. Further, it should be seen as complementary to the other behavioural cartel screens. First, it is not always possible to obtain all data necessary for a particular type of screen. Second, in some situations behavioural screens which are specifically tailored to a market might provide more compelling evidence based on a technically simpler analysis.

The results show that the development of cartel screens is a promising field of research that should receive more work. Recent developments in the behavioural cartel screen literature provide the opportunity to further strengthen antitrust enforcement. Given the new technology at hand, competition authorities could increase the proactive search for cartels to increase the risk of detection. This renders collusion less viable and attractive to cartels, and could complement other efforts to increase deterrence that are not based on the problematic unilateral increases of cartel fines.

2.7 Appendix

Figure 2.6: OLS-CUSUM EFPs of Italy and Spain

(a) OLS-CUSUM Italy **(b)** OLS-CUSUM Spain
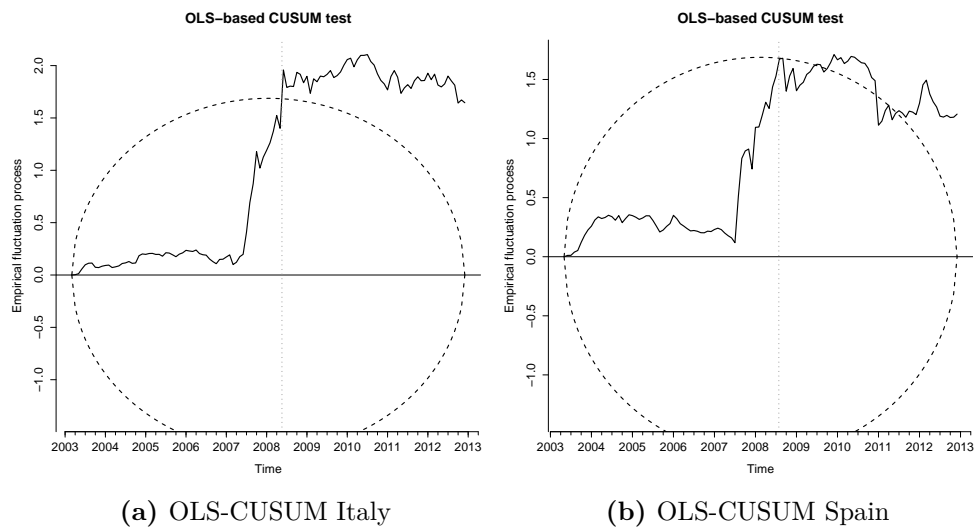
Table 2.7: Data sources and definition of variables

Variable/Country	Source	Description
Pasta price index		
IT	ISTAT	Producer price index for industrial products: manufacture of macaroni, noodles, couscous, base year 2005.
ES	INE	Producer price index for industrial products: manufacture of macaroni, noodles, couscous and similar farinaceous products, base year 2010.
FR	INSEE	Producer price index for industrial products: pasta products, base year 2010.
Domestic durum		
IT, ES, FR	EC	Durum wheat price in the domestic market, in €/tonne. This variable is country-specific.
International durum		
IT, ES, FR	INSEE	International price for imported durum wheat, in US cents / bushel of 60 pounds.
Labour costs		
IT	ISTAT	Labour costs per full time equivalent unit in the manufacturing industry, base year 2010.
ES	INE	Total wage cost per effective hour of work in the manufacturing industry, base year 2012.
FR	INSEE	Labour cost index including wages and payroll taxes in the manufacturing industry, base year 2012.
Energy		
IT	ISTAT	Producer price index for electricity, gas, steam and air conditioning supply, base year 2010.
ES	INE	Industrial price index for electric power generation, transmission and distribution, base year 2010.
FR	INSEE	Industrial market price index for energy, base year 2005.
Borrowing costs		
IT, FR	ECB	Monthly average interest rate of borrowing capital for non-financial cooperations with new businesses coverage. This variable is country-specific.
ES	INE	Average interest rates for loans over 1 million € to non-financial corporations.
Household expenditure		
IT	ISTAT	Quarterly national expenditure of households and non-profit institutions serving households, base year 2010.
ES	INE	Quarterly final household consumption expenditure, base year 2010.
FR	INSEE	Monthly household consumption expenditure on goods, base year 2005.

Notes: IT, ES, and FR denote the variables for the Italian, Spanish, and French markets. EC denotes the European Commission, ECB the European Central Bank, INE the Spanish Statistical Office, INSEE the French Statistical Office, and ISTAT the Italian Statistical Office. Variables are selected according to being the best available proxy variable for the price shifters between 2003 and 2012.

Table 2.8: Descriptive statistics

Variable	Italy			Spain			France		
	Mean (Std. Dev.)	Min	Max	Mean (Std. Dev.)	Min	Max	Mean (Std. Dev.)	Min	Max
Pasta price index	0.505 (2.692)	-7.20	16.70	0.250 (1.978)	-7.25	10.15	0.075 (2.295)	-12.80	8.73
Domestic durum	0.797 (19.574)	-76.10	105.73	0.976 (16.086)	-62.50	57.12	0.784 (14.809)	-65.80	71.50
International durum	4.475 (53.312)	-210.10	201.50	4.767 (53.668)	-210.10	201.50	4.064 (53.269)	-210.10	201.50
Labour costs	0.044 (5.635)	-8.68	9.29	0.050 (2.610)	-5.30	4.13	0.004 (0.143)	-0.57	0.26
Energy costs	0.462 (1.558)	-4.70	4.50	0.443 (1.816)	-7.39	9.26	0.407 (2.043)	-7.60	5.00
Borrowing costs	-0.009 (0.145)	-0.69	0.29	-0.009 (0.207)	-0.81	0.42	-0.013 (0.147)	-0.81	0.23
Household expenditure	-19.595 (468.154)	-1374.17	647.48	-0.018 (1.132)	-1.66	2.28	0.002 (0.390)	-1.36	1.23
Observations	117			115			118		

Notes: All variables are detrended and in case of France seasonality has been removed with the STL approach of Cleveland et al. (1990). The values are based on the first differences.

Table 2.9: Variance test: GARCH regressions

	Italy	Spain	France
Pasta Price Equation			
Domestic durum	0.018***	0.045***	0.021
	(0.005)	(0.012)	(0.025)
	4	3	4
International durum	0.005**	0.006*	0.000
	(0.002)	(0.003)	(0.004)
	3	3	0
	–	–	0.006
			(0.005)
			4
Labour costs	0.004	0.084	1.436
	(0.009)	(0.056)	(0.993)
	0	2	5
Energy costs	0.146***	0.040	1.013
	(0.036)	(0.113)	(0.842)
	3	1	6
Energy costs sq.	–	–	-0.004
			(0.004)
			6
Borrowing costs	-0.268	–	1.731
	(0.606)		(1.377)
	2	4	1
Household expenditure	0.000	0.249**	0.253
	(0.000)	(0.124)	(0.387)
	4	6	2
Phase 2	1.496*	0.787	0.684
	(0.832)	(0.504)	(0.595)
Phase 3	0.0276	-0.2128	-0.134
	(0.190)	(0.246)	(0.372)
Pasta Price Variance Equation			
Phase 2	5.294***	3.531***	2.136***
	(0.432)	(0.476)	(0.692)
Phase 3	2.710***	2.282***	1.251*
	(0.341)	(0.731)	(0.659)
Constant	-2.275***	-1.513***	0.245
	(0.241)	(0.302)	(0.429)
Prob > chi2	0.000	0.017	0.251
Observations	117	115	118

Notes: The dependent variables are the domestic producer prices for pasta products. All variables are included as first differences. The first row for each variable features the coefficient, with ***, **, and * showing significance of the coefficient at a 1%, 5%, and 10% significance level. The second row contains robust standard errors in brackets. The third row indicates which lag of the variable is used with 0 denoting the contemporaneous value.

Chapter 3

Post-cartel tacit collusion: determinants, consequences, and prevention

3.1 Introduction

[1] Post-cartel tacit collusion (PCTC) occurs when firms tacitly collude after an explicit cartel that they were involved in breaks down. Such PCTC intensifies the negative welfare effects of collusion and at the same time undermines the deterrence of cartels. Due to PCTC prices do not immediately return to the level of competition even after the cartel is detected. As a result, firms continue to earn supernormal profits and the harm induced by the cartel on welfare extends to post-cartel periods. Moreover, fines that are imposed on detected cartels can deter collusion only if the fines are based on the conspiring firms' gains from the cartel. These gains are estimated as the cartel overcharge and are used by some antitrust authorities to impose fines. Further, they are also used in private damage litigation to calculate damages awarded to the cartel customers. PCTC results in underestimated cartel overcharges if the supernormal markup created by PCTC is not accounted for. This leads to fines that are insufficient to deter collusion and fully compensate customers. This downward bias in overcharge estimates is in particular a problem in some of the price-based approaches commonly used in court cases, in which post-cartel periods are used as competitive counterfactuals to establish the cartel overcharge (see, e.g. Davis and Garcés, 2009; Harrington, 2004b).[2] Despite these important consequences of PCTC, little is known under which circumstances PCTC might occur, to

---

[1]This chapter is based on joint work with my supervisor Subhasish Modak Chowdhury and builds on Chowdhury and Crede (2015).

[2]In the last 30 years, private damage litigation related to cartels grew significantly in the United States. Currently about 90% of all cartel court cases are based on private litigation representing an important source of cartel deterrence (Lande and Davis, 2008; Wils, 2003). A similar development is in process in Europe triggered by the European Commission's Directive on Antitrust Damage Actions (December 2014).

what extent the overcharge estimates may be biased due to PCTC, and how antitrust law can be designed to obstruct or prevent it. Thus, a better understanding of the determinants of PCTC and of potential tools aimed at its prevention is vital. This study aims to add to this knowledge.

PCTC has been observed or at least suspected in various industries and studied based on different methodologies. Harrington (2004b) provides a theoretical model, Fonseca and Normann (2012) experimental results, and Connor (2001, 1998), de Roos (2006), Ordóñez-de Haro and Torres (2014), and Kovacic *et al.* (2007) empirical observations that point towards the occurrence of tacit collusion after the end of cartels. Connor (1998) notes that prices in the citric acid industry did not decline significantly even 18 months after the breakdown of the cartel. However, it is not certain whether this observation was triggered by increases in input prices or by tacit collusion. Similar suspicions are raised in Connor (2001) and de Roos (2006) for the lysine cartel. de Roos (2006) provides two potential explanations for the lack of post-cartel price reductions in the lysine industry, in which prices actually rose after the detection of the cartel. It could have been possible that the conspirators learnt enough about each others' behaviour through several years of explicit communication and cooperation that enabled them to collude tacitly. Knowing that communication to dissolve disputes was no longer possible after breakdown, the firms were particularly careful to prevent a price war. However, it is also possible that the firms simply continued to set collusive prices to reduce fines to be paid under the U.S. antitrust sentencing guidelines (that refers to post-cartel prices to determine the cartel overcharge). Harrington (2004b) shows that firms have the strategic interest to keep the prices high after cartel detection during litigation, such that overcharge estimates based on post-cartel prices underestimate the true harm caused by the cartel. Erutku (2012) provides empirical evidence in support of this idea. Ordóñez-de Haro and Torres (2014) examine the breakup of several Spanish food cartels that relied on the signals of trade associations. Significant levels of price hysteresis (i.e. prices remained high and were subject to a reduced variance) can be observed in most of the cartels after antitrust intervention. This evidence suggests that the firms may have continued to post prices based on past signals from their trade associations. Fonseca and Normann (2012) provide experimental evidence for the existence of tacit collusion after periods of explicit communication that suggests that the chance of PCTC to arise in industries as well as its magnitude are negatively correlated with the number of firms in the market. Similar findings are reported by Kovacic *et al.* (2007), who empirically study multiple markets that were engaged in the Vitamins cartels.[3]

Although these studies provide potential explanations regarding the existence and potential sources of PCTC, these theories have not been formally tested. This lack of empirical evidence inhibits the creation of knowledge that can be used to prevent inappropriate overcharge esti-

---

[3]Isaac and Walker (1988) are the first to test the effects of communication on coordination after communication is disallowed in public goods experiments. They find that preceding communication has a negative but diminishing effect on free-riding in periods without communication.

mates and the development of policies aimed at preventing or obstructing PCTC. Therefore, the aim of this study is to concentrate on tacit collusion after periods of explicit communication,[4] and to shed light on the following research questions: (1) Is the existence of PCTC robust to differences in competition regimes (in terms of fines, leniency programmes etc.)? (2) What are the determinants of PCTC? (3) What consequences does PCTC have on attempts to estimate cartel overcharges? (4) Can policy measures be implemented to prevent or reduce PCTC?

To our knowledge, this is the first study to systematically investigate the driving factors and related consequences of PCTC as well as possible preventive measures aimed against it. For this, we carry out a laboratory experiment that allows for an analysis of the marginal contribution of different market characteristics to tacit collusion in a controlled environment. Lack of data prevents to carry out a similar exercise with field data. Results show that PCTC is a robust phenomenon across competition regimes. Learning about other players' types through successful cartel formation and collusive price hysteresis are found to be determinants of PCTC. Further, the downward bias in cartel damage estimates induced by PCTC increases with the preceding cartel success. Rematching is proposed and found to be a promising measure to prevent or reduce PCTC.

3.2 Sources of post-cartel tacit collusion

Although an important legal difference exists between explicit and tacit collusion, the standard theory of collusion does not differentiate between the two. Only recently have scholars begun to close this gap with theoretical models (Bos *et al.*, 2015; Harrington, 2012; Martin, 2006). An important function of communication in collusion is that it reduces uncertainty about present and past actions (Mouraviev, 2006). Throughout this essay, we refer to explicit communication as *communication*, and implicit communication as *price signalling*. Price signalling enables subjects to express their intention to collude by setting prices above the market level (see, e.g. Davis *et al.*, 2010; Cason, 1995). Although there are other forms of implicit communication, signalling with price choices is the only means to express intentions to collude apart from (explicit) communication in this experiment.

Despite the importance of communication for collusion, the aforementioned empirical evidence indicates that tacit collusion can be sustained after periods of communication, such that communication can have intertemporal spillover effects on collusion. It might not only reduce uncertainty in the period in which it is used, but also in the periods afterwards. Then, PCTC can be induced through two distinct channels.

First, former cartelists abstain from price reductions in attempts to prevent triggering a price war that ends collusion (de Roos, 2006). We refer to this source of PCTC as *collusive price hysteresis*. Prime examples for this source of tacit collusion are the Spanish food

---

[4]Therefore, we are not interested in pure tacit collusion, i.e. collusion established without any communication (see, e.g. Martin, 2006; Ivaldi *et al.*, 2003).

cartels observed by Ordóñez-de Haro and Torres (2014). Second, past actions in periods with communication allow firms to learn about their competitors' types in terms of discount factors. This knowledge helps to sustain collusion by reducing the uncertainty about the other cartel members. Hence, given successful explicit collusion, the perceived profitability of playing collusive strategies in the post-cartel periods increases. We refer to this effect as *learning in cartels.* This argument is provided by de Roos (2006) as one possible explanation for the observed tacit collusion following the detection of the lysine cartel. More formally, deviation is an important source of risk to colluders that can only be observed *a posteriori.* As such, a firm that considers collusion needs to form subjective beliefs about this risk and incorporate them into the decision problem. The observed history of play is important and shapes the subjective beliefs and therefore a firm's decisions. *Ceteris paribus*, firms with a longer history of successful collusion should assign a higher subjective probability to other firms' actions of continuing to abide to the collusive agreement. Such belief-updating as a reaction to risk has been studied theoretically in the context of tacit collusion by Harrington and Zhao (2012) and in generic multi-agent learning models (see, e.g. Young, 2007; Foster and Young, 2003). Thus, PCTC might be a function of the preceding cartel success, and markets colluding more successfully in the past are more likely to engage in and sustain PCTC.

Fonseca and Normann (2012) provide experimental evidence for the effect of communication on collusion after the end of communication, and point out that the effect's magnitude depends on the number of firms in the market. In their experiment, the gains for firms are characterised by an inverted U-shaped curve and are highest for markets with four firms. Furthermore, they find that these gains diminish over time. Fonseca and Normann (2014) find a higher level of cartel recidivism for markets with four firms than with duopolies, as the four-firm-markets profit more from re-engaging in communication after the breakdown of collusion. These two studies are the only ones to provide experimental evidence on PCTC. However, they focus on the link between PCTC and the number of firms in the market. Thus, they neither investigate the reasons for and consequences of PCTC nor strategies that can be used to prevent it.

3.3 Experiment

3.3.1 Experimental procedure

The experiment was conducted at the Centre for Behavioural and Experimental Social Science (CBESS) at the University of East Anglia, UK. It was programmed with z-Tree (Fischbacher, 2007) and the recruitment of subjects was done using ORSEE (Greiner, 2015). The subjects were allocated into groups of three and interacted with the same two other participants throughout the experiment (except for a treatment in which subjects at some point in time with their knowledge are rematched into new groups). We recruited 228 students with no prior experience in oligopoly experiments. 36 subjects participated in each treatment to obtain 12 independent market observations.[5]

---

[5]42 subjects participated in the Fine and the Rematching treatments. Hence, each of them features 14 independent markets.

Subjects were randomly seated in the laboratory at the start of each session. Each participant received a printed copy of the instructions, which were also displayed on the computer screen and were read aloud by an experimenter at the beginning of the session. Questions about the instructions could be asked in private by subjects raising their hands. The experiment was comprised of two parts. The first part consisted of a risk elicitation task whereas the second part was the market game. In the market game, subjects interacted in markets for 20 (30 in one treatment) regular periods, i.e. periods that are played with certainty before a random stoppage rule applies. To prevent potential end-game effects and to reflect the infinitely repeated game with discounting, a random stopping rule in the spirit of Dal Bó (2005) was implemented. After the end of the regular periods, in each period there was a 20% chance that the experiment ends. Subjects' understanding of the instructions was tested with a questionnaire, in which all values used in the questions were randomised across subjects to prevent example numbers to systematically influence decisions in the experiment.[6] An example of the instructions can be found in the appendix in Section 3.6.2.

Sessions lasted between 25 and 50 minutes and subjects were allowed to participate in one session only. Earnings in part one were denoted in British Pounds, whereas earnings in the second part were labelled as "experimental points". Each experimental point gained in the market game was converted into £0.15 at the end of the experiment. Payments varied from £5.63 to £28.90 with a mean of £11.35.

3.3.2   Experimental design

In this experiment three subjects, each representing a firm in a market, engage in homogeneous goods price competition with perfectly inelastic demand as proposed by Dufwenberg and Gneezy (2000). This oligopoly market design is similar to that of Gillet *et al.* (2011) and is combined with a variation of the "Talk-NoTalk" design of Fonseca and Normann (2012). We implement a three firm homogeneous goods rather than a two firms differentiated goods market (e.g. Bigoni *et al.*, 2012), as this significantly reduces the complexity of the decision making process for subjects as well as the subjects' learning effects on outcomes. Finally, triopolies are used because previous studies find that three firms are sufficient to prevent significant levels of collusion without communication in markets with both price (Wellford, 2002; Dufwenberg and Gneezy, 2000) and quantity (Huck *et al.*, 2004) competition.

Unless stated otherwise, the experiment consists of four stages. In the first stage, subjects need to decide whether they would like to attempt to reach a price agreement with the other subjects in the market. In the instructions, they are informed that in the experiment they "... may decide to agree with the other firms to set the highest price of **102** and share the earnings". On the computer screen, subjects in this stage are asked "*Do you want to agree on prices?*" and need to click on option "Yes" to signal their intention to form a price agreement with the others. An agreement is only reached if all three subjects in the market confirm

---

[6]Risk preferences of the subjects were elicited using a risk elicitation task based on Holt and Laury (2002) before the market game, and an anonymous questionnaire followed at the end of the experiment.

that they want to agree on prices. If it is reached, a message is displayed that all subjects agreed to set the price of 102. However, the agreement is non-binding, i.e. subjects are not required to follow the price agreement. In the second stage, subjects are asked to make a price decision. Each subject can charge a price between 90 and 102 (integer values only) facing costs of 90 if she sells the good. Therefore, a subject's profit equals **Price - 90** if she sells the good and is **0** otherwise (when another subject charges a lower price). If either two or all three subjects charge the same lowest price, the profits are shared equally. Thus, the demand is characterised by a computerised buyer that buys either 1 or 0 units from each subject depending on whether the subject sets the lowest price in that round. Subsequently, we refer to the price entered by subjects as the *asking price*, and to the lowest price in a market as the *market price*. There are several Nash equilibria in this framework. In one equilibrium two subjects charge 90 and the remaining subject charges any of the available prices including 90. Alternatively, all subjects charge 91. However, the latter is both the payoff-dominant equilibrium as well as the unique equilibrium in strategies that are not weakly dominated. In the third stage, the subjects learn about each others' prices. In this stage, they also receive additional information and face further choices that are treatment-specific. In the last stage, subjects learn their profits in that period.

Figure 3.1: Sequence of the experiment

Stage 1: Collusion decision	Stage 2: Price decision	Stage 3: Feedback	Stage 4: Final outcome
• First 10 periods only • Yes/No question whether agreement shall be attempted	• Info. whether cartel formed • Price choice required	• Info. on price choices of all subjects • Info. on the min. price	• Profits are reported • Info. about potential detection and fines

Figure 3.1 depicts the sequence of the experiment and shows the four stages as well as the main feedback provided to subjects in each of them. In all treatments (except for the Baseline treatment introduced below), subjects were told in the instructions that they *may* agree on prices in the market game (i.e. the option of communication might or might not be given). Then they were allowed to communicate in the first 10 periods only – which we call the *Communication phase*. Then, without prior notice, the communication is disallowed for the rest of the game – which we call the *No Communication phase*. As such, while subjects know that at some points they might be able to communicate with others with respect to price agreements, they also know that this option might not always be available.[7] The uncertainty with respect to the possibility to communicate ends at the beginning of period 11 when subjects are informed that from this point onwards communication is not possible any more

---

[7]The word *may* is applied deliberately as it is defined in the Cambridge Dictionary as "used to express possibility". In contrast to other words such as *will* ("used to talk about what someone or something is able or willing to do" and *can* ("to be able to"), the word *may* does not imply that communication is always possible. Taken from the Cambridge dictionary, available online at *http://dictionary.cambridge.org/*.

and that previous agreements cannot be detected for the rest of the experiment. This design prevents strategic behaviour of subjects in the transition from explicit to tacit collusion and assures that no cheating is triggered by the anticipation of the end of communication in period 10. In one treatment (ExtComm), the Communication phase is preceded by 10 additional periods in which no communication is possible. An overview of the possibility to communicate in all treatments can be found in Table 3.1.

Table 3.1: Communication in treatments

Treatments	No Communication phase	Communication phase	No Communication phase
Baseline	-	×	×
Comm	-	✓	×
ExtComm	×	✓	×
Fine	-	✓	×
Leniency	-	✓	×
Rematching	-	✓	×
Periods	-9 to 0	1 to 10	11 to 20

Notes : A ✓ indicates that communication is possible in the time periods, and in periods denoted with × subjects cannot communicate. The dash (-) denotes that in all but the ExtComm treatment directly start with communication in the Communication phase.

Instead of implementing an exogenously given cutoff point for communication after 10 periods, an alternative design could have been to stop communication after the first incidence of cheating or detection in a market. We have decided against such a design for several reasons. First, both re-emergence of collusion after temporal breakdown as well as cartel recidivism are common observations in the field. Our design allows us to observe whether PCTC occurs despite both such forms of interruptions. Second, collusion in the lab has been noted to be very unstable, especially when it is not based on free-form communication. Removing the possibility to communicate after the first incidence of failure of collusion would therefore significantly limit the scope for learning. This would in turn undermine the analysis of learning, one of the main determinants mentioned in the literature. Third, our design provides a common cutoff point for all groups as well as all treatments, which greatly simplifies the analysis and allows for a clean identification of the sources of PCTC. In particular, it allows us to separate the effects of changing the expected length of interaction in the Rematching treatment introduced below from the effects of disrupting PCTC by ending the possibility to communicate.

We introduce several treatments pertaining to our research questions:

Baseline: Subjects cannot communicate at any point and each round starts directly with the price decision in the *Baseline* treatment. It serves as the benchmark for tacit collusion that can be obtained without communication. Any difference in price levels between this and the other treatments in which subjects can communicate represents the effect of communication.

Comm: Subjects can agree on prices as described above for the first 10 periods during the

Communication phase in the *Comm* treatment, but not afterwards in the No Communication phase. This is the equivalent of the relevant treatment in Fonseca and Normann (2012).

ExtComm: In this treatment, the Communication and No Communication phases of the Comm treatment are supplemented by 10 additional, initial periods without communication. Subsequently, we do not analyse these initial 10 periods (-9 to 0) but in line with the other treatments focus on the periods with and without communication that follow. The treatment is introduced to test whether experiencing competition before communication affects PCTC. Subjects can learn about the Nash equilibrium in the initial periods and revert back to it quickly after the end of communication. Furthermore, they might have a better understanding of the benefits of communication because of preceding exposure to low profits during competition.

Fine: The *Fine* treatment replicates the effect of an antitrust authority on illegal communication. Subjects face an exogenous detection probability of 16% if they agree to fix a price in the Communication phase. This probability is in the range of the estimated detection probabilities of cartels of between 13%-17% provided by Bryant and Eckard (1991). Detection is possible either in the period in which the agreement is formed or in subsequent periods provided that it has not been detected before. Detected subjects have to pay a fine of 5 experimental points irrespective of the number of agreements that they reached before. Past agreements can only (jointly) be detected and fined once, so additional fines are not possible unless another agreement is formed later.

Leniency: The *Leniency* treatment is an extension of the Fine treatment. It implements a leniency programme by offering subjects the option to report price agreements. This leads to the immediate detection and to fines of the other cartel members in return for a (partial or full) reduction of the fine for the reporter(s). If a cartel is formed in the same or a previous period and so far has remained undetected, subjects can report it after learning about each others' prices in Stage 3. Such a fine reduction procedure for leniency applications is standard in the experimental literature (Bigoni *et al.*, 2012; Hamaguchi *et al.*, 2009). If only one subject submits a leniency application, she is not fined but the other two subjects pay the full fine of 5 experimental points.[8] If two subjects submit leniency applications, both pay only half of the fine while the third pays the full fine. If all three subjects use the leniency scheme, they all pay 1/3 of the fine. A cartel is always detected if at least one leniency application is submitted, but subjects are not informed whether the detection occurred due to the exogenous detection probability or because of a leniency application.

Rematching: The *Rematching* treatment introduces a mechanism aimed at disrupting PCTC by targeting its source of learning. Similar to the Comm treatment, here each subject starts in a group with two other subjects but they are informed that they will be rematched with

---

[8]The fine is chosen such that the incentive compatibility constraints (ICC) for the infinitely repeated games that characterise the incentives to collude in the Fine and in the Leniency treatments are similar (given collusion on the price of 102, the critical discount factors necessary to support collusion are approximately 0.66 and 0.68, respectively, if only one subject deviates to price 101 and is the only one to submit a leniency application).

two new randomly chosen subjects at some point in the experiment. The point in which they are rematched is not revealed beforehand; it is announced immediately before the rematching is carried out. The rematching takes place at the beginning of period 11, in which communication ends too. This ensures that subjects cannot learn about the types of the new group members. Hence, any change in behaviour observed in this treatment from period 11 onwards compared to the Comm and ExtComm treatments comes from the disruption of the effects of learning. Further, from a supergame perspective, this should yield lower rates of cooperation by reducing the horizon for cooperation itself. The uncertainty due to different expectations of the duration of cooperation in the supergame may also destabilise collusion.

This treatment is new to the literature. The mechanism in Rematching replicates one of the indirect enforcement effects that (criminal) sanctions against managers involved in cartels have on PCTC. Sanctions against cartel managers in the form of imprisonment or debarment, i.e. disqualification from taking up managing positions in the same or similar industries after conviction, remove convicted managers from the market. While we do not attempt to exactly replicate such sanctions, the Rematching has a similar disruptive effect with respect to learning in cartels. Just as sanctions against managers do in the field, rematching creates instability in the lab through shortening the expected length of interaction between sanctions. Moreover, rematching eliminates any knowledge about the strategies and likely actions of the other subjects, as is the likely effect of removing key managers involved in operating a cartel in the field. Hence, we regard this mechanism as a preventive measure against PCTC.

3.4 Results

3.4.1 Sources of post-cartel tacit collusion

As a first step, we test the existence and determinants of PCTC across the treatments that approximate various competition regimes. All observations after the 20th period are excluded from the analysis to prevent potential end-game effects and an unequal sampling of the groups (towards the end of the experiment) to influence the results.[9] We distinguish between the asking and market prices as defined in the previous section. The market price serves the whole market in homogeneous goods price competition and is the relevant market outcome from a welfare perspective. The asking price captures additional information such as price signalling or failed attempts to collude. This is in particular important for periods without communication because subjects can signal their intentions to establish collusion by deviating from the Nash equilibrium and setting a price of 102.

Table 3.2 contains the average absolute margins (Average price – 90) for both the asking and market prices separated by treatment in the Communication and No Communication phases. For Baseline, for ease of exposition we include periods 1-10 and 11-20 into the Communication and the No Communication columns, respectively. As the market prices are

---

[9]Given the random stoppage-rule, actual termination varies between the 20th and the 25th period across sessions.

the market-clearing prices, they are at least as low as the asking prices in all treatments. Based on the magnitude of price margins, the ranking of treatments with respect to asking prices in the Communication phase is as follows: ExtComm features the highest price margins followed by Comm, Rematching, Leniency, Fine, and Baseline.

Table 3.2: Asking and market price margins by communication possibility

| | Communication phase | | | | No Communication phase | | | |
| | Asking prices | | Market prices | | Asking prices | | Market prices | |
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Baseline	3.328	3.324	1.925	2.338	3.436	3.600	2.125	2.410
Comm	7.744	4.816	5.958	5.004	6.925	4.968	5.725	5.042
Fine	4.978	4.019	3.508	3.498	4.206	4.229	3.042	3.487
Leniency	5.276	4.784	3.429	4.125	4.595	4.699	3.021	3.888
ExtComm	8.078	4.769	6.533	5.002	5.817	4.979	4.667	4.731
Rematching	6.874	4.730	4.507	4.365	5.238	4.725	2.557	3.232

This ranking coincides with the number of markets successfully engaged in collusion in the Communication phase. Successful collusion, i.e. a cartel is formed and all subjects abide to the agreement in a period, occurs at least once in 7 markets in ExtComm, 6 in Comm, 5 in Rematching, 4 in Leniency, and 2 in Fine. This strong link between price agreements and asking and market price margins provides a first indication of the importance of collusion for generating positive margins.

Comparing prices between the Communication and the No Communication phases shows a strong correlation of price margins across the two phases. Price margins in the No Communication phase are significantly higher in the treatments with communication compared to Baseline and the order of treatments remains the same apart from Rematching. The price margin in the No Communication phase relative to Baseline is an indicator of PCTC, as it is enabled by subjects' preceding ability to communicate. Therefore, the occurrence and magnitude of PCTC appears to be correlated with successful collusion in the Communication phase. This correlation does not exist in the Rematching treatment. Whereas market prices in Rematching are close to those of Comm and ExtComm in the Communication phase, they are subject to a significant decline in the No Communication phase and then are very close to Baseline. Thus, unlike in the other treatments, PCTC appears to be absent in Rematching. This provides first evidence of the disruptive effect of rematching on collusion by eliminating learning that apparently drives PCTC.

The link between PCTC and successful collusion in the Communication phase becomes clearer if it is distinguished between the markets in which price agreements were successfully implemented and those in which no such successful collusion occurred. In Figure 3.2, market prices are separated by treatments and markets are divided into two groups. The group "Successful cartels" contains the average market prices for markets which successfully established a market price of 102 based on a price agreement *at least once* in the Communication phase. The group

Figure 3.2: Market prices by preceding cartel success
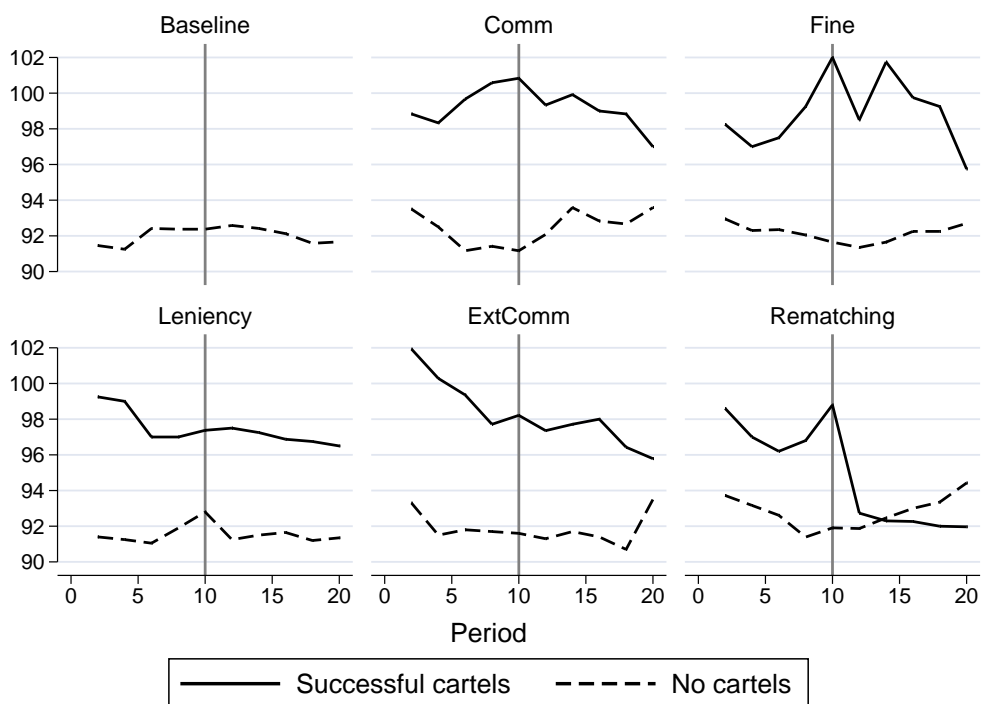
"No cartels" contains all other markets, i.e. those in which the subjects did not manage once to reach a market price of 102 based on a price agreement. The vertical grey line marks the last period of communication, and market prices are averaged over two periods.

Note that the particular shape of the price paths in the treatments should be interpreted with care, as only 2 and 4 cartels are formed in Fine and Leniency, respectively. Yet, while the data does not allow to assess whether PCTC occurs to a larger or smaller extent in Fine and Leniency compared to Comm and ExtComm, Figure 3.2 clearly shows that PCTC does occur in all four treatments. Subjects successfully forming a cartel are able to charge higher prices throughout the experiment in all four treatments.

For the Rematching treatment, market prices in the No Communication phase are separated between subjects previously engaging in successful collusion and those who did not. Note that market prices in Rematching immediately collapse in the No Communication phase after subjects are matched into new groups. This sudden decline in market prices does not occur in the other treatments with communication. This suggests that the positive effect of communication on PCTC is removed in the Rematching treatment.[10]

We turn to regression analysis to formally test the observations regarding the sources of PCTC and its absence in the Rematching treatment. This allows us to distinguish the sources, control for the dynamics, and run analyses that capture the marginal contributions

---

[10]As we argue below, market prices recover during the end of the experiment in the Rematching treatment because of an increased stability of tacit collusion in the Rematching treatment compared to Baseline. We attribute this stability to the subjects' preceding experience that a return to collusion after deviation is hard to achieve after the possibility to communicate ceases to exist.

of different key determinants of prices. Asking and market prices in the No Communication phase are regressed on other market outcomes and the results are reported in Table 3.3. To distinguish PCTC from any tacit collusion that is established by price signalling only, we include variables aimed at capturing the sources of PCTC, i.e. learning in cartels and collusive price hysteresis. The regressions are calculated using the random effects model.[11] For all estimations, cluster-robust standard errors based on pairs cluster bootstrapping with 500 iterations are used to account for clustering at the market level. The small numbers of cartels in Fine and Leniency do not allow for producing reliable treatment-specific estimates. Hence, we pool them with Comm and ExtComm to estimate average effects for these treatments. Results based on all treatments excluding Rematching are presented in Columns I and III using the asking and market prices as the dependent variables. We analyse the Rematching treatment separately and report the results in Columns II and IV due to the potentially very different nature of tacit collusion in this treatment. The regressions are based on Periods 12-20 only, such that despite the use of lagged variables the analysis is limited to data from the No Communication phase.

We include the following independent variables in the regressions. *Lag price* represents a subject's own asking price (Columns I and II) or the market price (Columns III and IV) from the previous period. *Max price others* and *Min price others* contain the higher and lower of the other two subjects' asking prices in the previous period and are included in the asking price regressions only. If both competitors set the same asking price, both variables contain that price. Inclusion of lags of the dependent and independent variables yields the autoregressive distributed-lag model, which is a widespread model in applied econometrics to model dynamics in time series and panel data (Pesaran and Shin, 1998; Banerjee *et al.*, 1990). *Comm*, *Fine*, *Leniency*, and *ExtComm* are treatment indicators, with the Baseline treatment being the baseline category for the regressions in Columns I and III. Therefore, the treatment dummies control for treatment-specific effects on PCTC that are not captured by any of the other included regressors. The variable *Period* measures a time trend.

We include two independent variables to measure the effect of preceding collusion on pricing, which repesent the two sources of PCTC described above. *Lag collusion* is an indicator variable that takes the value 1 if all three subjects charged the collusive price of 102 in the previous period and is 0 otherwise. Given the *ceteris paribus* character of the analysis, this variable measures collusive price hysteresis. *No. of successful cartel periods* contains the market's number of periods of successful cartelisation (i.e. all subjects agreed to fix prices and did not cheat) in the preceding Communication phase.[12] For all treatments except for

---

[11]As a robustness check, we have also used the correlated random effects model (Wooldridge, 2010; Mundlak, 1978). It is less restrictive with respect to unobserved heterogeneity than random effects models and does not require the random effects to be uncorrelated with the included level 2 variables (e.g. variables that vary by subject but not over time). The results are reported in Table 3.5 in the appendix in Section 3.6.1 and are robust to the choice of the estimator.

[12]In the Rematching treatment, the three subjects in a market in the No Communication phase come from markets with different histories of collusion. Therefore, in this treatment we use the average value of the variable across the three markets that the subjects come from. This allows us to control

Rematching, it captures the effect of preceding cartel success on PCTC and corresponds to the effect of learning in cartels on subsequent tacit collusion. For the Rematching treatment the interpretation is different, as rematching has a strong and immediate negative impact on PCTC. The coefficient of *No. of successful cartel periods* in Column II shows whether a subject's intention to establish collusion with price signalling is driven by preceding experience of collusion. Our rematching procedure allows us to observe how subjects with a history of engaging in collusion behave in a new market environment. In Column IV, the coefficient captures the average collusive experience in the new market and shows how price signalling triggered by former collusion contributes to market prices. We also include an interaction of *Period* with the measure of preceding cartel success, *Period × No. of successful cartel periods*. The interaction term measures whether the contribution of learning as proxied by preceding cartel success deteriorates faster in markets with a stronger history of collusion.

Column I provides strong evidence that collusion in the preceding period has a significant positive effect on price choices. This suggests that PCTC is indeed partly caused by collusive price hysteresis, and collusion in the preceding periods increases asking prices in the current period. The significance and high magnitude of the positive effect of preceding cartel success on the asking prices provides evidence for the effect of learning in cartels on PCTC. The lack of significance of the treatment dummies provides support for the intuition that treatment differences with respect to tacit collusion arise from differences in the formation and stability of cartels in the Communication phase. In line with the previous analysis, this suggests that other than through their effect on cartel success, communication between subjects in preceding periods does not affect prices. Moreover, the insignificant coefficient of *ExtComm* shows that our experimental design with respect to the Communication phase followed by the No Communication phase is robust to experimental learning effects.

The estimates based on market prices in Column III show that the results are robust to the choice of the price variable and lead to the same qualitative findings. Given these results, we conclude that learning as proxied by preceding cartel success fosters PCTC through two distinct channels. First, markets with former cartels inherit supercompetitive prices that only slowly erode back towards levels of competition due to collusive price hysteresis. Second, learning about the other players' types contributes to the existence and stability of tacit collusion. This result has important consequences for the estimation of cartel damage overcharges, which are discussed in detail in Section 3.4.2.

> **Result 1**: PCTC is determined by both collusive price hysteresis and learning about the other players' types and strategies.

Turning to the Rematching treatment in Columns II and IV, the large positive coefficient of *Lag collusion* suggests that collusion is more stable after rematching. It may be because

---

for the effect of the average level of preceding experience of successful collusion of subjects on PCTC after rematching.

Table 3.3: Prices in the No Communication phase – Random effects model

Baseline: Baseline treatment	I	II	III	IV
	Asking price		Market price	
	Coeff.	Coeff.	Coeff.	Coeff.
	(Std.E.)	(Std.E.)	(Std.E.)	(Std.E.)
Lag price	$0.516^\dagger$	$0.539^\dagger$	$0.709^\dagger$	0.309***
	(0.040)	(0.044)	(0.069)	(0.096)
Min price others	0.045	0.019	–	–
	(0.048)	(0.085)		
Max price others	$0.207^\dagger$	0.099	–	–
	(0.030)	(0.071)		
Lag collusion	$1.691^\dagger$	$2.647^\dagger$	$2.273^\dagger$	$6.503^\dagger$
	(0.446)	(0.687)	(0.658)	(1.238)
No. of successful cartel periods	0.213**	0.532	0.271**	0.427
	(0.094)	(0.698)	(0.127)	(0.510)
Period	0.024	0.030	−0.008	−0.019
	(0.030)	(0.101)	(0.026)	(0.068)
Period × No. of successful	−0.010**	−0.031	−0.014**	−0.013
cartel periods	(0.005)	(0.046)	(0.007)	(0.038)
Comm	0.275	–	0.112	–
	(0.395)		(0.379)	
Fine	0.209	–	0.102	–
	(0.324)		(0.311)	
Leniency	0.004	–	−0.168	–
	(0.268)		(0.222)	
ExtComm	0.331	–	0.086	–
	(0.394)		(0.332)	
Constant	$21.008^\dagger$	$31.419^\dagger$	$26.910^\dagger$	$62.590^\dagger$
	(5.104)	(6.130)	(6.437)	(8.888)
R^2 overall	0.714	0.416	0.863	0.595
R^2 between	0.961	0.930	0.981	0.939
R^2 within	0.140	0.169	0.408	0.396
Observations	1,674	378	558	126

Notes : ∗ 10% level, ∗∗ 5% level, ∗∗∗ 1% level, † 0.1 % level. – Cluster and autocorrelation-robust standard errors are based on pairs cluster bootstraps with 500 iterations. Columns I and III contain observations for all treatments except for Rematching, and Columns II and IV are based on Rematching treatment observations only. Random intercepts are included at the subject level in Columns I and II, and at the market level in Columns III and IV.

subjects are aware that re-establishing collusion after cheating is harder to achieve without communication. However, as collusion on price 102 only arises in about 6% of the observations in the No Communication phase under Rematching, the magnitude of the coefficient might be overstated due to unrepresentative outliers. In Rematching, the coefficient of *No. of successful cartel periods* is insignificant and implies that the positive effect of learning on PCTC is eliminated by being rematched with other subjects. This is consistent with the idea that the information obtained with past successful collusion about competitors becomes redundant due to a change in group composition. Therefore, the regression results are consistent with the descriptive analysis above that suggests that PCTC is virtually absent in Rematching.

Result 2: PCTC occurs in all treatments except for Rematching.

However, notice that the coefficient of the interaction term *Period × No. of successful cartel periods* is negative. It suggests that the positive impact of learning in cartels deteriorates faster for previously more successful cartels. For example, in period 11, *ceteris paribus*, the regression results in Column III show that successful collusion in 5 periods in the Communication phase lead to an increase in market prices of, on average, 0.585 experimental points. Yet, in period 15 its effect has gone down to 0.305 experimental points.

3.4.2 Implications for cartel overcharge estimations

We use the "before-after approach" to calculate the damages caused by all cartels formed in the experiment and study the relationship between preceding cartel success and overcharge bias. This estimator is one of the most common methods used in the field to estimate cartel overcharges. The cartel overcharge is established by comparing the price during the cartel with a counterfactual price under competition from a benchmark period. Three different variants of this approach are commonly used (see, e.g. Davis *et al.*, 2010; Baker and Rubinfeld, 1999). *Pre-Cartel* denotes the overcharge estimate that compares the price during periods of cartelisation to a price benchmark based on prices before the cartel. *Post-Cartel* denotes the estimate based on post-cartel prices serving as benchmark prices, and *Whole sample* uses prices both before and after a cartel as the counterfactual for competition.

As we observe pre-communication prices only for the Baseline and ExtComm treatments, we use the average market price of the ExtComm treatment observations from periods -9 to 0 as the benchmark of competition for all treatments.[13] To calculate the overcharges, a reasonable assessment has to be made about the periods that should be regarded as cartel periods. In the Comm, ExtComm, and Rematching treatments we include only those periods

---

[13]Market price margins are considerably larger in ExtComm with 3.475 than in Baseline with 1.925. Hence, the Baseline treatment is not a good benchmark for the calculation of the cartel overcharge as it lacks comparability with the other treatments with respect to the attainable profits before communication has taken place. Thus, we use only the ExtComm treatment for such purposes.

in which subjects communicate and reach a price-fixing agreement as cartel periods.[14] Fine and Leniency feature periods in which either a cartel forms or a previous cartel is undetected in the Communication phase. These differences in the composition of cartel periods reflect the underlying differences in incentives for cartel formation and pricing. Given that detection is possible in Fine and Leniency even if no cartel is formed in a certain period but a previous price agreement so far has remained undetected, subject behaviour might be affected by the presence of a chance of detection.

Table 3.4: Overcharge estimates and biases

		Overcharge estimate			Overcharge bias	
	Obs.	Pre-Cartel	Post-Cartel	Whole sample	Post-Cartel	Whole sample
Comm	6	64.45	19.91	41.12	-77.75%	-40.73%
Fine	2	55.41	10.73	32.01	-68.01%	-35.62%
Leniency	4	48.91	8.82	27.91	-20.30%	-10.63%
ExtComm	7	46.89	22.12	33.91	-24.67%	-12.92%
Pooled	19	53.76	17.42	34.72	-45.07%	-23.61%
Rematching	5	40.63	53.86	47.25	129.73%	64.87%

Notes: Pre-Cartel, Post-Cartel, and Whole sample overcharge estimates represent average values of estimated cumulated cartel overcharges by cartel based on competitive price benchmarks including periods before, after, and before and after the cartel. Pre-cartel prices serve as the counterfactuals for the calculation of overcharges biases. Pooled includes the average values of the columns excluding the Rematching treatment.

Table 3.4 reports the average of the estimated cartel overcharges using the different benchmark prices in the first three columns by treatment. The prices in Pre-Cartel that are not affected by any communication represent the correct counterfactual of competition. Unlike post-cartel prices, they are untainted by tacit collusion enabled by preceding communication. The last two columns report the average overcharge bias. The results show that the Post-Cartel and the Whole sample overcharge estimates are biased downwards for all treatments except for Rematching. Hence, PCTC leads to a significant underestimation of cartel damages by econometric techniques that rely on post-cartel data. It is not possible to rank the treatments with respect to the severity of the downward bias due to the limited sample size. Yet, the main implication that the problem of underestimating cartel damages does not exist in the Rematching treatment because of an absence of PCTC remains valid.[15]

As has been shown with Table 3.3, post-cartel prices are correlated with preceding cartel

---

[14]Periods without price agreements that lie between periods with price agreements could have also been included here. Whether exclusion of such periods with potential tacit collusion increases or decreases the overcharge estimate depends on the market outcome in these periods. If the subjects collude tacitly (compete fiercely) between periods with price agreements, then the true damage would be higher (lower).

[15]In fact, the estimations point to a large overestimation of damages in this treatment. However, these results should be treated with caution, as the competitive counterfactual of ExtComm prices in periods -9 to 0 might not be good counterfactuals for Rematching. Given the destabilising effect of informing about rematching in the future on collusion, a proper counterfactual for this treatment would likely contain lower prices.

Figure 3.3: Post-Cartel overcharge bias by cartel success
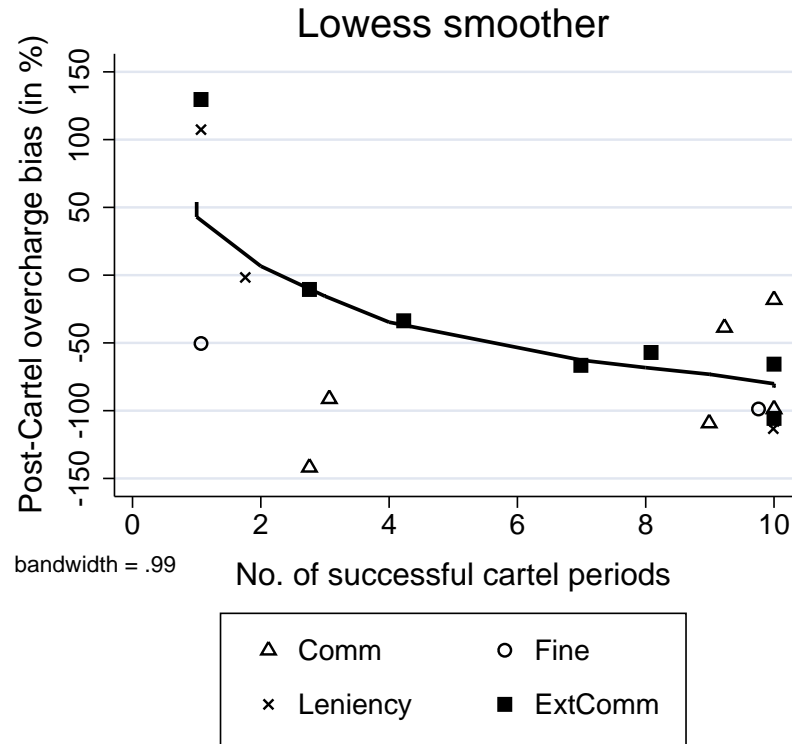
success. Hence, the downward bias of the estimates should be increasing with the number of preceding cartel success. Figure 3.3 plots the relationship between the number of successful cartel periods and the bias of the Post-Cartel estimates with a lowess smoother excluding the Rematching treatment (the overcharge estimates are jittered to improve readability). Indeed, the downward bias is increasing with preceding cartel success.

> **Result 3**: There is a downward bias in overcharge estimates based on the before-after approach when post-cartel prices are considered as benchmark prices. The bias increases with preceding cartel success without rematching.

3.4.3 The impact of rematching on explicit collusion

Our final investigation centres on the effects of rematching on the performance and stability of cartels. The absolute price margin based on market prices in the Communication phase in the Rematching treatment appears to be lower than in the Communication treatment (4.507 vs. 5.958; Table 3.2) for markets with at least one successful cartel period. As the two treatments are identical in the Communication phase aside from the announcement of future rematching of the groups in the Rematching treatment, we can attribute the lower market prices in Rematching to a negative effect of anticipated rematching on collusion.

To determine how rematching affects cartels, we compare the incidence of collusion and cheating in the Communication phase between the treatments. Figures 3.4a and 3.4b show

Figure 3.4: Incidence of cartelisation and cheating in the Communication phase
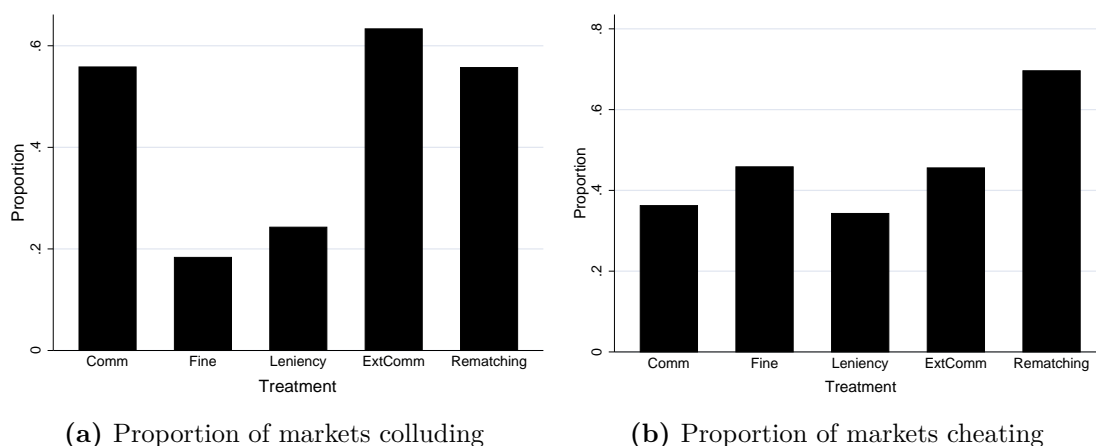
(a) Proportion of markets colluding   **(b)** Proportion of markets cheating

differences in the proportions of markets with price agreements and with cheating on existing agreements, respectively.[16] We define cheating as any subject's decision to charge a price below 102 when either an agreement was reached in the same period or a previous periods' agreement has not yet been undercut by any other subject. Thus, a higher level of cheating shows a lower level of stability of cartels.

In line with the literature, Fine and Leniency feature lower levels of collusion by rendering collusion less attractive, although collusion and cheating are not much different between the two treatments. Rematching does not reduce attempts to collude (as cartel formation is unchanged) in the Rematching treatment compared to the Comm treatment (a two-sample t-test testing for differences in the proportion of the subjects colluding in Comm and Rematching reports a p-value of 0.497). Yet, the incidence of cheating in the Rematching treatment is higher than in the Comm treatment. A two-sample t-test comparing cheating between the Comm and the Rematching treatments shows a weakly significant difference between the treatments (p-value = 0.058).[17] Thus, rematching does not reduce attempts to collude, but it significantly increases the incidence of cheating. This destabilising effect is very pronounced with the proportion of firms cheating rising from 36.2% in Comm to 69.6% in Rematching.

> **Result 4**: Rematching reduces explicit collusion through its negative effect on cartel stability.

---

[16]Attempts to collude that fail are implied the difference between agreements that were reached and the number of observed cases of cheating.

[17]The t-tests use cluster- and autocorrelation-robust standard errors based on pairs cluster bootstrapping with 500 iterations and compare the incidence of collusion and cheating at a market and period level. They are derived from a linear probability model. The t-tests are preferred here to Mann Whitney U tests, as the latter cannot take sample weights into account. As different markets engage to different extents in collusion and cheating, markets more active in collusion and cheating are more informative. Using this information with weighting leads to efficiency gains of the test statistic compared to non-parametric tests.

3.5 Conclusion

Although it is a conventional wisdom that firms may resort to tacit collusion after a cartel breaks down, little is known about the conditions under which this happens and about the determinants that drive the level and persistence of such behaviour. As a result, it is hard to assess implications of such firm behaviour for competition policy and how to counteract it. Given the importance of PCTC for deterrent fines, welfare effects of cartels, and the right design of antitrust legislation, this essay aims at adding to the knowledge on the existence, determinants, consequences, and prevention of PCTC.

We run experiments in which groups of three firms, each controlled by a subject, compete in homogeneous goods price competition and can establish price agreements. These price agreements can be renewed in the following periods or remain active absent new agreements that were neither detected nor cheated upon. After this initial phase of communication, the ability to agree on price-fixing ends and subjects are only able to collude tacitly. Such an approach contributes to our understanding on how cartels react to detection when continued communication is deemed too risky. We test the existence of PCTC in different competition regimes to establish whether it is a common phenomenon unrelated to particular policy tools. Conducting an econometric analysis we study the different sources of PCTC. We then show how under PCTC the standard procedures to estimate cartel damages may be biased. Furthermore, the use of rematching to disrupt the positive effects of learning on PCTC is tested.

The results suggest that firms might be able to profit frequently from PCTC irrespective of different antitrust laws. We identify two sources of PCTC: collusive price hysteresis and learning in cartels. The former describes a firm's strategy to continue charging preceding cartel prices after the end of the cartel in order to avoid triggering a price war resulting in lower prices of competition. The latter describes how communication and a cooperative history facilitate PCTC by reducing uncertainty about the actions of the other firms. Moreover, the magnitude of PCTC is positively linked to preceding cartel success. In line with Bigoni *et al.* (2015), this stresses the importance of beliefs for successful collusion in infinitely repeated games. Rematching in the experiment is found to be an effective mechanism to prevent PCTC as well as to reduce cartel stability. The Rematching treatment emulates one indirect enforcement effect that debarment, i.e. disqualification orders for convicted cartel managers and imprisonment, have on collusion. Note, however, that we do not fully replicate such mechanisms.[18] Our focus is on the indirect enforcement effects of sanctions against managers. As such, stronger (deterrent) effects are likely to arise if the *direct enforcement effects* linked to these punishment mechanisms were modelled as well.

---

[18]See e.g. the December 2016 disqualification of Daniel Ashton by the Competition and Markets Authority in the UK (https://www.gov.uk/government/news/cma-secures-director-disqualification-for-competition-law-breach), last retrieved on 06/08/2017.

Several implications arise from our analyses. Antitrust laws that reduce the formation and stability of cartels lessen the negative welfare effects of PCTC, as the incidence of tacit collusion is linked to the preceding cartelisation of the industry. Cartels that do not break down due to cheating but are detected exogenously might realise supercompetitive profits long after the end of communication. Therefore, competition agencies should rely on leniency programmes to reduce cartel formation as much as possible to reduce the negative welfare effects of PCTC. In addition, provided that debarment programmes and imprisonment have similar disruptive indirect enforcement effects on collusion in the field as indicated by the Rematching treatment in the lab, these policy tools may help to minimise the harm caused by PCTC. In particular, the debarment of managers so far has been limited to few countries such as the USA, UK, Sweden, and Slovenia (Ginsburg and Wright, 2010). Our results suggest that this policy tool might offer the potential to reduce the damage caused by cartels in other ways than the direct effect on individuals that has been discussed in the literature, and should receive greater attention by the antitrust authorities. Finally, our analyses show that post-cartel prices should not be used as competitive counterfactuals to determine cartel overcharges as this creates the risk of a downward bias in these estimates that increases with preceding cartel success. As such, the most harmful cartels might be those deterred the least.

There are several ways to extend our analysis. We focus on learning as a source of PCTC abstracting from focal points in the spirit of Scherer (1967) as a source of collusion. After rematching, subjects could try to establish tacit collusion by setting the price last charged in markets in previous periods with collusion. Furthermore, the effects of the variations of market characteristics including firm numbers and product differentiation on PCTC and its identified sources should be studied. Finally, the complete effects of debarment are not tested experimentally because our implementation only captures an indirect effect.

3.6 Appendix

3.6.1 Auxiliary tables

Table 3.5: Prices in the No Communication phase – Correlated random effects model

Baseline: Baseline treatment	I	II	III	IV
	Asking price		**Market price**	
	Coeff.	Coeff.	Coeff.	Coeff.
	(Std.E.)	(Std.E.)	(Std.E.)	(Std.E.)
Lag price	$0.257^{\dagger}$	$0.336^{\dagger}$	$0.467^{\dagger}$	0.207^{**}
	(0.054)	(0.046)	(0.091)	(0.087)
Min price others	0.009	-0.018	–	–
	(0.064)	(0.083)		
Max price others	$0.134^{\dagger}$	0.078	–	–
	(0.040)	(0.082)		
Lag collusion	$2.088^{\dagger}$	$2.473^{\dagger}$	2.654^{***}	$6.059^{\dagger}$
	(0.510)	(0.749)	(0.833)	(1.115)
No. cases of cheating	0.003	-0.041	-0.016	-0.033
	(0.036)	(0.034)	(0.035)	(0.058)
No. of successful cartel periods	0.171^{*}	0.603	0.272^{**}	0.350
	(0.103)	(0.980)	(0.127)	(0.669)
Period	0.007	0.021	0.008	0.021
	(0.042)	(0.124)	(0.032)	(0.084)
Period $\times$ No. of successful	-0.014^{**}	-0.045	-0.021^{***}	-0.025
cartel periods	(0.007)	(0.063)	(0.007)	(0.042)
Comm	-0.036	–	-0.087	–
	(0.256)		(0.294)	
Fine	0.085	–	-0.021	–
	(0.140)		(0.173)	
Leniency	-0.074	–	-0.127	–
	(0.126)		(0.171)	
ExtComm	0.121	–	-0.045	–
	(0.208)		(0.235)	
$\varnothing$ Lag price	$0.642^{\dagger}$	$0.639^{\dagger}$	$0.521^{\dagger}$	$0.671^{\dagger}$
	(0.052)	(0.063)	(0.053)	(0.130)
$\varnothing$ Lag collusion	-1.923^{***}	-0.154	-2.284^{**}	-1.827^{**}
	(0.686)	(0.577)	(0.965)	(0.792)
$\varnothing$ Min price others	0.054	-0.080	–	–
	(0.093)	(0.104)		
$\varnothing$ Max price others	-0.079	0.017	–	–
	(0.053)	(0.112)		
Constant	-1.651	1.952	1.064	11.083
	(4.431)	(5.259)	(6.101)	(9.554)
R^2 overall	0.758	0.493	0.880	0.635
R^2 between	0.987	0.970	0.990	0.993
R^2 within	0.149	0.172	0.418	0.399
Observations	1,674	378	558	126

Notes: $*$ 10% level, $**$ 5% level, $***$ 1% level, $\dagger$ 0.1 % level. – Cluster and autocorrelation-robust standard errors are based on pairs cluster bootstraps with 500 iterations. Columns I and III contain observations for all treatments except for Rematching, and Columns II and IV are based on Rematching treatment observations only. Random intercepts are included at the subject level in Columns I and II, and at the market level in Columns III and IV. Variables without a $\varnothing$ symbol control for within-group effects. Group mean-centred variations are measured by coefficients with a $\varnothing$ symbol, which capture the effect of the cluster-specific mean. This coefficient captures both the between effects in the sample as well as the captured unobserved heterogeneity (the fixed effect). Their significance indicates evidence of between variation with respect to the variable under consideration. No such decomposition is necessary for variables that do not vary over time or between the groups.

3.6.2 Instructions (Leniency)

Instructions

Welcome and thank you for taking part in this experiment. In this experiment you can earn money. How much money you will earn depends on your decision and on the decision made by other participants in this room. The experiment will proceed in two parts. The currency used in Part 1 of the experiment is Pound Sterling (GBP). The currency used in Part 2 is experimental points. Each experimental point is worth 15 pence. All earnings will be paid to you in cash at the end of the experiment.

Every participant receives exactly the same instructions. All decisions will be anonymous. It is very important that you remain silent. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you.

Instructions for Part 1

In the first part of the experiment you will be asked to make 15 decisions. For each line in the table that you will see on the computer screen there is a paired choice between two options ("Option A" and "Option B"). Only one of these 15 lines will be used in the end to determine your earnings. You will only know which one at the end of the experiment. Each line is equally likely to be chosen, so you should pay equal attention to the choice you make in every line. At the end of the experiment a computerized random number (between 1 and 15) determines which line is going to be paid.

Your earnings for the paid line depend on which option you chose: If you chose Option A in that line, you will receive £1. If you chose Option B in that line, you will receive either £2 or £0. To determine your earnings in the case you chose Option B there will be second computerized random number (between 1 and 20). Both computerized random numbers will be the same for all participants in the room.

Instructions for Part 2

In this part of the experiment you will form a group with two other randomly chosen participants in this room. Throughout the experiment you are matched with the same two participants. All groups of three participants act independently of each other. This part of the experiment will be repeated for at least 20 rounds. From the 20th round onwards, in each round there is a **one in five** (**20%**) chance that the experiment will end.

Your job:

You are in the role of a firm that is in a market with two other firms. In each round, you will have to choose a price for your product. This price must be one of the following prices:

90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102.

You will only sell the product if your price is the lowest of the three prices chosen by you and the other two firms in that round. If you sell the product, your earnings are equal to the difference between the price and the cost, which is 90:

Earnings = Price - 90.

If you do not sell the product, you will not get any earnings but you will not incur costs either. If two or more firms sell at the same lowest price, the earnings will be shared equally between them. Before you choose your price, you may decide to agree with the other firms to set the highest price of **102** and share the earnings. This agreement is only valid if all three firms want to agree on it. After you made your choice, you will be informed whether the price agreement is reached. However, the price agreement is not binding and firms are not required to set the agreed price. After your price choice, you will be told whether you have selected the lowest price as well as the prices of the other firms.

The price agreement may be discovered by the computer. In that case, a fine of **5** points has to be paid. The computer can detect it in 16 out of 100 cases (a chance of 16%). A price agreement remains valid – and can be discovered – as long as it has not been discovered in a previous round. Once this has happened, you will not be fined in the future, unless you make a price agreement again. If you have reached a price agreement in this period, or a past agreement has not been detected by the computer, you must decide whether to report it. You can do this by choosing between the "Report" and "Not report" buttons. If you report it, you are charged additional costs of **1**.

In case one or more group members reports the agreement, it is discovered and a penalty of **5** has to be paid by all group members. However, in case you report your penalty gets reduced as follows:

- If you are the only one to report, you will not pay the penalty but the others will pay the full penalty.

- If you report and exactly one of the other two reports, then your penalty is reduced by half (50%). The other reporting participant has to pay only half of his penalty, while the remaining participant will pay his full penalty.

- If you report and both the other two also report, then the penalty is reduced by one third (33%) for all three of you.

At the end of each round, you will be told the earnings you made in this round. If you agreed on prices, you will also be told whether the agreement was detected by the computer (either because it was detected by chance or by reports).

Final Payment:

At the beginning of the experiment you start with an initial endowment of 40 points = 6 GBP. If the sum of your profits from Part B is below 0, the difference is being covered by the initial endowment. The earnings you earned in each round minus any fine and penalty that you paid will be converted into cash. Each point is worth 15 pence, and we will round up the final payment to the next 10 pence. We guarantee a minimum earning of 2 GBP.

Chapter 4

The effects of endogenous enforcement on strategic uncertainty and cartel deterrence

4.1 Introduction

[1] The fight against cartels fixing prices or allocating customers remains a priority of antitrust authorities around the globe. The main tool of antitrust authorities to deter cartels is to punish detected cartelists with fines, which feature both fixed and variable elements. Cartels are usually prosecuted for their wrongdoings *per se*, i.e. they receive an exogenous, fixed fine for the mere attempt to collude irrespective of the actual cartel harm. Yet, in some jurisdictions fines are partially adjusted to the magnitude of harm induced by the cartel. Furthermore, a growing part of costs of detection for cartelists stems from private damage litigation, in which damages are awarded based on estimated overcharges, and therefore are endogenous and determined by cartel conduct. Subsequently, we refer to enforcement in which the detection probability and/or the fine depend on the cartel overcharge as *endogenous enforcement*. In contrast, *exogenous enforcement* is characterised by fines and detection probabilities that do not change with the cartel overcharge.

The deterrent effect of antitrust enforcement can be decomposed into two elements: frequency and composition deterrence. *Frequency deterrence* refers to the prevention of cartel formation and is often measured as the number of cartels that are not formed in the presence of antitrust laws, but would have formed otherwise. *Composition deterrence*

_____

[1]This chapter is based on joint work with Liang Lu and builds on previously released research in Crede and Lu (2016) and Lu (2016).

refs to the mitigation of harm caused by the cartels that are formed nonetheless. It is achieved through inducing a change in cartel behaviour in response to the enforcement, e.g. a reduced cartel overcharge (Bos *et al.*, 2016; Harrington, 2005, 2004a; Block *et al.*, 1981). It is usually assumed that composition deterrence results from a change in the cartel's optimisation problem in the presence of endogenous enforcement and is treated as profitability-driven. Yet, cartel behaviour may not only be influenced by profitability concerns, but also by strategic responses of cartelists to uncertainty about the actions of the other cartel members. Drawing on Heinemann *et al.* (2009), we refer to this type of uncertainty as *strategic uncertainty*. *Ex ante*, the probabilities of the specific actions of the other cartel members are unknown to a cartelist. Such strategic uncertainty between cartel members might be affected by endogenous enforcement, e.g. if the enforcement contributes to mistrust among cartelists by improving the relative profitability of unilateral deviations from the agreement. Thus, strategic uncertainty potentially generates an indirect effect of endogenous enforcement and contributes to the deterrence of cartels.

In this essay, we seek to determine whether endogenous enforcement can induce composition deterrence through increasing strategic uncertainty in a non-cooperative infinitely repeated game market experiment. In doing so, we abstract from the direct profitability-driven effects.[2] Instead, we focus on the indirect effects induced by antitrust law on strategic uncertainty. We contribute to the experimental literature on collusion in the presence of an antitrust authority by proposing a framework that allows us to study the impact of endogenous enforcement on equilibrium cartel price selection. For this purpose, we allow the expected punishment to increase with the cartel price, so that cartels formed on a high price and a low price are identical in expected profitability but different in the *riskiness of collusion* (Blonski *et al.*, 2011; Blonski and Spagnolo, 2015). As a result, the two collusive equilibria are equally payoff-dominant, but the low cartel price is the equilibrium with a lower riskiness of collusion.

While we expect endogenous enforcement to add to composition deterrence, it may produce adverse effects on frequency deterrence. The enforcement regime featured in our design renders the low cartel price a less risky collusive equilibrium, and thus may encourage more collusive agreements on the low cartel price as well as stabilise such agreements. Therefore, endogenous enforcement may produce frequency and composition deterrence that move in opposite directions.

The results show that – consistent with our theoretical predictions – endogenous enforcement produces composition deterrence through its effect on strategic uncertainty.

---

[2]See, e.g. the aforementioned studies by Bos *et al.* (2016), Harrington (2005, 2004a), and Block *et al.* (1981).

Subjects tend to suggest the low cartel price with the lower riskiness of collusion in the presence of fines and detection probabilities that increase overproportionately with the cartel overcharge. This tendency is not affected by different combinations of fines and detection probabilities that feature the same expected payoffs and an almost identical riskiness of collusion. Yet, we find an adverse effect on welfare induced by a reduction of frequency deterrence due to an increased stability of cartels formed on the low price. The overall effect of endogenous enforcement on market outcomes in the experiment is unclear, as the positive effect of strategic uncertainty, *ceteris paribus*, is neutralised by the adverse effect on frequency deterrence. Furthermore, our results show that subjects' preferences over cartel prices are not driven by their risk attitudes, suggesting that subjects behave strategically in such repeated cooperative games.

4.2 Literature review

Until now, most cartel experiments that explore how enforcement regimes affect frequency deterrence focus on the impact of exogenous fines and detection probabilities on collusion. Examples with exogenous fines include studies on the effects of leniency programmes by Hamaguchi *et al.* (2009), Bigoni *et al.* (2012, 2015), Clemens and Rau (2014), and Hinloopen and Onderstal (2014), but also on other aspects such as the substitutability of fines and detection probabilities (Chowdhury and Wandschneider, 2017), avoidance activities of cartels (Chowdhury and Wandschneider, 2016), and tacit collusion induced by previous cartel activities (Chowdhury and Crede, 2015).

Exceptions are the studies of Apesteguia *et al.* (2007) and Hinloopen and Soetevent (2008) on the effect of leniency programmes and of Fonseca and Normann (2014), who study how firm numbers affect the necessity to engage in repeated communication to preserve collusion. In these three studies, fines are endogenous and depend on either the chosen cartel price and/or the preceding collusive history. Yet, they do not study how the endogeneity of fines affects cartel prices or stability such that the issue of composition deterrence remains unaddressed.

To the authors' knowledge, endogenous detection probabilities (apart from detection triggered by leniency applications) have not yet been experimentally implemented or studied in the context of cartels. However, in the field cartelists decide not only whether to collude, but also which price to collude on. Although it is reasonable to assume that the risk of detection of cartels increases with the magnitude of collusive price changes (Crede, 2015; Harrington and Chen, 2006; Harrington, 2005, 2004a), cartel experiments so far have forgone to analyse composition deterrence due to the exogenous design of detection probabilities (and often of the fines) in the implemented enforcement regimes.

Although the breakdown of all-inclusive cartels is, in general, either triggered internally by coordination failure of cartelists or externally through the detection by antitrust authorities, the experimental literature on antitrust enforcement against cartels tends to focus on the latter channel. In a recent study, Bigoni *et al.* (2015) suggest that deterrence can be achieved not only through imposing severe punishments, but also by worsening the incentive and trust problems faced by cartelists. As a cartel is a type of collective crime, when reaching the decision whether to collude each member cares not only about the expected profitability of a collusive agreement, but also about its stability and sustainability. As severe coordination complexity might itself render collusion infeasible, the formation of a cartel fundamentally depends on the cartelists' ability to reach and sustain cooperative equilibria.

The relevant literature of coordination games can loosely be divided into two classes: finitely or infinitely repeated non-cooperative games and cooperative coordination games that are classified according to the access to information. Both strands of literature point out that the Nash equilibrium concept fails to predict a unique outcome in such games (e.g., Dal Bó and Fréchette, 2011; Fudenberg and Maskin, 1993; Cooper *et al.*, 1990). Coordination failure may arise not only from the absence of a payoff-dominant cooperative equilibrium, but also from Nash equilibria that are not self-enforcing (Van Huyck *et al.*, 1990). The multiplicity of equilibria calls for equilibrium refinement criteria to select the equilibria that are most likely to arise. While Harsanyi and Selten (1988) suggest that payoff dominance performs better at predicting choices than alternative deductive selection criteria, more recent theoretical and experimental studies (e.g. Van Huyck *et al.*, 1990) tend to support the risk dominance criterion. An underlying reason for coordination failure that has been suggested in the literature by Van Huyck *et al.* (1990) and formalised by Heinemann *et al.* (2009) is strategic uncertainty that arises in a socially interactive decision situation.

Knight (1921) distinguishes between risk and uncertainty. The former is usually referred to as Knightian uncertainty or exogenous uncertainty, whereas the latter is referred to as strategic uncertainty or endogenous uncertainty. Heinemann *et al.* define exogenous uncertainty as "*a priori* given and known probabilities for all possible states of the world" (Heinemann *et al.*, 2009, p.182). In the context of cartel experiments, the probability of detection by the competition authority usually is assumed to belong to this source of uncertainty.[3] Strategic uncertainty, on the other hand, describes situations in which the probabilities with which the states occur are not exogenously given or known *a priori*.

---

[3]Fines for collusion in many jurisdictions follow fining guidelines, such that they can be anticipated by cartels. While detection probabilities are generally not known, there is a wide consensus that the mean detection probability most likely lies between 10% to 20% (Bos *et al.*, 2016; Ormosi, 2014; Hyytinen *et al.*, 2011; Bryant and Eckard, 1991). Taken together, cartels are able to form expectations about the expected punishment for explicit collusion.

Each subject is uncertain about the strategies and actions of the others within the same group, and thus has to take decisions according to subjectively assigned probabilities based on her own beliefs. The trustworthiness of the other group members is a crucial element in these assigned probabilities. Based on a behavioural definition of trust and betrayal aversion, Fehr (2009) suggests that trust is captured by preferences and beliefs of individuals that partly are shaped by social interactions and influence them too. Intuitively, as the degree of strategic uncertainty faced by decision makers increases, it becomes more difficult to trust and subjects tend to choose inefficient but secure actions.

In the evolutionary game theory literature, Dal Bó and Fréchette (2011) conclude that the games' fundamentals, which are structural determinants contributing to exogenous uncertainty such as the attractiveness and riskiness of alternative actions, the length of the games, the form of communication, and subjects' experience may affect cooperative behaviour. Although repeated play of the game allows for inductive selection of strategies and learning, long run stochastically stable equilibria that emerge tend to be those that satisfy the concept of risk dominance (Kandori *et al.*, 1993; Young, 1993). Blonski *et al.* (2011) and Blonski and Spagnolo (2015) introduce the concept of riskiness of collusion of a cooperative equilibrium in non-cooperative infinitely repeated games, which is heuristically related to the concept of risk dominance.[4] In doing so, they consider strategic uncertainty by taking cognitive and behavioural determinants into account with a particular emphasis on the sucker's payoff, i.e. the payoff a subject receives who is playing a collusive strategy but is cheated upon by others. They take an axiomatic approach and offer a strategic uncertainty-based measure of the critical discount factor, which allows for a comparison of the riskiness of several cooperative equilibria.

Cooperative games have been applied in the lab to study illegal activities and the optimal design of enforcement. For example, Berninghaus *et al.* (2013) and Tan and Yim (2014) use coordination games to model corruption and tax evasion. Both studies highlight the role of trust and beliefs and find higher degrees of uncertainty to be an effective device to deter illegal activities. However, unlike tax evasion a cartel is a type of collective crime, such that the incentive and trust issues are more pronounced. This follows from the risk of punishment that is attached to attempted collusion. Bigoni *et al.* (2015) suggest that a crucial part of the deterrent effect offered by leniency programmes is driven by the "distrust" that they create. They measure the minimum level of trust required to sustain collusion and find that it increases with strategic uncertainty, such that collusion is less likely to be established.

---

[4]The riskiness of cooperation measured by Blonski and Spagnolo (2015) is related to strategic uncertainty that we discuss above, rather than exogenous risk, although they refer to it as "strategic risk".

4.3 Experiment

4.3.1 Experimental procedure

The experiment was carried out at the Centre for Behavioural and Experimental Social Science (CBESS) at the University of East Anglia, UK, in February and March 2015. Recruitment of subjects was carried out with hRoot (Bock *et al.*, 2014) and the experiment was programmed and run with zTree (Fischbacher, 2007). 144 students from a variety of backgrounds and nationalities and without prior experience in oligopoly experiments participated in the experiment. 36 subjects were allocated in groups of three, providing 12 independent market observations in each of the four treatments. Group composition was fixed throughout the session. At the start of each session, subjects were randomly seated in the laboratory at workstations separated by modular walls. Each subject received a printed copy of the instructions, which were also read aloud by an experimenter and displayed on each computer screen at the beginning of each session. Subjects' understanding of the instructions was tested with a questionnaire before they participated in the main part of the experiment.

The experiment consisted of two parts. In the first part, the risk preferences of subjects were tested with a risk elicitation task similar to that in Eckel and Grossman (2008).[5] Subjects indicated the choices of their preferred lotteries on their computers, and then an experimenter determined the outcome of the lottery with a coin toss monitored by a volunteering subject. In the second part, subjects each represented a firm and played an oligopoly game in markets of three firms as described below in Section 4.3.2 for 20 regular periods. A random stopping rule was implemented to avoid end-game effects as described in Dal Bó (2005): there was a 20% chance in each additional period after the end of the regular periods that the experiment ends.

At the end of each session, subjects filled out an anonymous demographic survey before being called out of the laboratory and paid in private. Earnings are denoted as "experimental points" and each point was converted into £0.12 for cash payment. Based on subjects' performance, payments varied from £4.00 to £13.20 with a mean of £8.19. Sessions lasted between 40 to 60 minutes. Sample instructions can be found in the appendix in Section 4.6.2.

---

[5]The risk elicitation task can be found in the sample instructions in the appendix in Section 4.6.2.

4.3.2 Experimental design

In the experiment, each subject represents a firm and engages in competition in a market with two other subjects. The market is defined as a homogeneous goods price competition triopoly with (discontinuous) inelastic demand as introduced by Dufwenberg and Gneezy (2000) and with market characteristics similar to those in Gillet *et al.* (2011). Previous experimental evidence indicates that three firms are sufficient to ensure that collusion can only be sustained effectively with communication (Fonseca and Normann, 2012; Wellford, 2002; Dufwenberg and Gneezy, 2000).

In each period, all subjects simultaneously make price decisions by choosing any price p from the choice set $p \in \{40, 41, ..., 52\}$. A subject sells one unit of the good and incurs costs of 40 experimental points if her price is not higher than that of any other subject, and does not sell anything nor incur any costs otherwise. The subject who sells the good at a price lower than her two competitors in a period therefore makes a profit of $p - 40$ experimental points whereas the other two subjects make a profit of zero. If two or three subjects choose the same lowest price, the profit is evenly divided among them.

Before subjects engage in the price competition as described above, in each period they first have to simultaneously decide whether they wish to enter a non-binding price agreement, and if so, which price to agree on. There are two price agreements to choose from: one on the high cartel price $p_C^h = 52$ or one on the low cartel price $p_C^l = 46$.[6] We allow subjects to choose between two cartel prices to facilitate the identification of effects in the econometric analysis of price agreements and to ensure that fines and detection probabilities are transparent to the subjects. Subjects wishing to cooperate can choose to suggest their preferred cartel price, or they can suggest both prices if they are indifferent between the two. However, a successful price agreement can be detected by the computer with a positive probability resulting in a fine that is deducted from the profit. If there is detection, only the most recent price agreement is detected. Further, an agreement can only be detected once. The fine does not depend on the length of the agreement.

Subjects need to indicate their intention to cooperate by answering the question "Do you want to agree on prices?". If all three subjects wish to agree on prices and a unique common price exists among the three subjects' suggested cartel prices, an agreement is reached on that price. If all three subjects choose both prices, implying that they wish to agree on either price, then in half of the groups in each treatment a price agreement

---

[6] 52 is the highest possible price and 46 is the mean price in the choice set. These prices represent markups of 30% and 15%, respectively. This roughly corresponds to the estimated median of cartel overcharges, which depending on samples and geographic and historical scope usually vary between around 15% to 25% (Boyer and Kotchoni, 2015; Connor and Lande, 2012; Bolotova, 2009).

on 46 and in the other half a price agreement on 52 is reached automatically. This is done to control for any potential effect of this protocol on cartel price choice.[7] The default price remains the same within a group throughout the experiment. No price agreement is reached if there is no common price among the suggested cartel prices of all subjects.[8] For example, this is the case if two subjects choose a suggested price of 46 and one subject chooses a suggested price of 52. If subjects do not wish to cooperate, they can express so by choosing the option "No". If at least one subject chooses "No", no price agreement is reached in that period. To avoid potential effects arising from the order of items on the screen on subjects' choices, their order is randomised across subjects and periods.[9]

With our design, we expect coordination complexity arising from endogenous enforcement to render reaching price agreements more difficult than in the presence of free-form communication. Although a free-form communication protocol may help to facilitate agreements, we restrict communication to a limited-form protocol for several reasons. First, our focus is to obtain clear measures of subjects' desired price choices. Second, limited-form communication prevents the social dimension to influence results (see, e.g. Cooper and Kühn, 2014) and ensures that communication is limited to pure signalling and cheap talk. Furthermore, our communication protocol allows for a clear identification and econometric analysis of how strategic uncertainty increases the complexity of cartel coordination. Given that in our experiment bargaining is only possible through signalling over the course of several periods of the game, observations on cartel price choices at the subject level may not directly translate into similar patterns at the market level. Yet, we focus primarily on subjects' decisions to study the uncertainty-driven behavioural patterns in the choices of subjects.

The sequence of the market experiment is summarised below:

1. Subjects simultaneously indicate their decisions with respect to cooperation by answering the question "Do you want to agree on prices?". Subjects willing to cooperate can choose either "Yes, with price 46", "Yes, with price 52", or "No" otherwise.

2. Subjects are informed about whether a price agreement has been reached. If this is the case, they learn about the agreed price as well.

---

[7]It is shown in Section 4 that this default agreement price rule does not affect subjects' price choices.

[8]This is in contrast to the field, where usually the mere attempt to establish collusion is illegal and results in a fine if detected. Yet, we refrain from introducing potential punishment for the mere attempt to collude. This greatly simplifies the decision problem faced by subjects as well as the experimental design. Thus, it facilitates the identification of our effects of interest in the experiment.

[9]A computer screenshot of the communication protocol can be found in the instructions in the appendix in Section 4.6.2.

3. Subjects simultaneously make price decisions for their goods by choosing a price from the discrete choice set $p \in \{40, 41, ..., 52\}$. Any price agreements reached are not binding for the subjects' price decisions in this stage.

4. Subjects learn about each others' prices and whether they sell a good in the current period.

5. Subjects are informed whether they are detected by the computer provided that they have reached a price agreement in the current period and/or their price agreement(s) in previous periods have not been detected yet.

6. Finally, subjects learn about their profits in the current period minus any potential fines, as well as about their accumulated profits.

This experiment is based on a 2x2 between-subjects design. *Baseline* is the control treatment and serves to capture coordination failure that occurs because a set of cooperative actions are not supported by an equilibrium. The other treatments assess the explanatory power of the concept of riskiness of collusion when two cooperative equilibria are payoff equivalent, which is useful to determine the uncertainty-driven channel of composition deterrence. We compare the three treatments with endogenous enforcement with each other to examine – given a constant expected profitability of 1.6 experimental points among the two collusive prices and across those treatments – whether there is a limited substitutability between fines and detection probabilities that affects cartel pricing as well as deterrence. The detection probabilities (D) and the fines (F) differ in the treatments as shown in Table 4.1, and the resulting expected profits are denoted as $E(\pi)$.

<div align="center">

Table 4.1: Treatments and key parameters

</div>

	Exogenous detection prob.	Endogenous detection prob.
Exogenous fine	**Baseline**	**EndoD**
	46: $D = 20\%$, $F = 12$, $E(\pi) = -0.4$	46: $D = 3.3\%$, $F = 12$, $E(\pi) = 1.6$
	52: $D = 20\%$, $F = 12$, $E(\pi) = 1.6$	52: $D = 20\%$, $F = 12$, $E(\pi) = 1.6$
Endogenous fine	**EndoF**	**BothEndo**
	46: $D = 20\%$, $F = 2$, $E(\pi) = 1.6$	46: $D = 10\%$, $F = 4$, $E(\pi) = 1.6$
	52: $D = 20\%$, $F = 12$, $E(\pi) = 1.6$	52: $D = 20\%$, $F = 12$, $E(\pi) = 1.6$

In Baseline, the detection probability (20%) and the fine (12 experimental points) are the same for collusion on the high price of 52 and the low price of 46. It represents a homogeneous goods price competition cartel experiment with exogenous fines and detection parameters independent of the cartel price. In *EndoF*, the detection probability is held fixed but the fine is endogenous. The magnitude of the fine is a function of the

agreed cartel overcharge: collusion on price 52 yields a higher per period profit and hence requires a higher potential fine of 12, whereas collusion on price 46 requires a lower potential fine of 2 to achieve the same expected punishment. Similarly, in *EndoD* the fine is held fixed but the detection probability increases with the cartel overcharge: collusion on the low price 46 and on the high price 52 feature detection probabilities of 3.3% and 20%, respectively. In *BothEndo*, the fine and detection probability attached to collusion on price 52 are the same as in Baseline, whereas those attached to collusion on price 46 are lower with $F = 4$ and $D = 10\%$, i.e. both elements are endogenous in this treatment.[10]

Parameters and the experimental design are chosen such that the expected profitability and the incentive compatibility constraints (ICCs) are equal across the treatments with endogenous enforcement and the two cartel prices as is shown in Section 4.3.3. This is crucial for answering the research questions: it rules out the profitability-driven effects of endogenous enforcement on cartel behaviour as analysed in Katsoulacos *et al.* (2015), and allows us to focus on the strategic uncertainty-driven effects.

4.3.3 Theoretical background

Before we describe the theoretical model that underlies our framework and experimental design, we briefly discuss the main insights from the theoretical literature on endogenous punishment on cartel prices. Consider a homogeneous goods market with $n \geq 2$ firms that face identical unit costs of production c and compete in prices. The demand function is $Q(p)$, where p denotes price and Q is the quantity supplied to the market. The industry profits $\pi(p, c)$ can be expressed as

$$\pi(p, c) = (p - c)Q(p). \tag{4.1}$$

If price collusion is not possible, competition gives rise to the price $p^B = c$. If price collusion is possible and all firms agree to collude, then in the absence of antitrust enforcement, each firm sets $p^M = arg\ max_p\ \pi(p, c)$ yielding each firm a profit of π^M/n.

In this context, the existing literature (e.g. Katsoulacos *et al.*, 2015; Block *et al.*, 1981) has shown that antitrust enforcement that is increasing in the cartel overcharge can mitigate the harm caused by the illegal activity by contributing to composition deterrence. To illustrate this result, we include an expected punishment function DF

---

[10]Detection probabilities of 10% and 20% are in range of the mean detection probabilities reported in Bryant and Eckard (1991), Hyytinen *et al.* (2011), Ormosi (2014), and Bos *et al.* (2016). Although the detection probability in EndoD of 3.3% is somewhat low, it is not ruled out by the estimation results of the aforementioned studies, which represent *mean* estimated detection probabilities.

that is a product of D denoting the detection probability and F that is the penalty imposed upon detection. Let the function DF depend on cartel prices with $\partial DF/\partial p > 0$ and $\partial^2 DF/\partial p^2 \geq 0$, such that it is weakly convex with respect to cartel prices.[11] With this general functional form, we allow either the fine, the detection probability, or both elements to increase with the cartel overcharge, which is a reasonable assumption given observations from the field.[12] Figure 4.1 plots the cartel gross profit against the price in the absence of the expected punishment (π^M) as well as the net profit under consideration of the expected punishment presence (π^{DF}). As can be seen, such endogenous antitrust enforcement leads to composition deterrence by altering the expected profitability of a cartel. As a result, cartels may not be deterred altogether, but they are formed at lower prices, i.e. at p^{DF} instead of p^M.

Figure 4.1: Profitability-driven composition deterrence
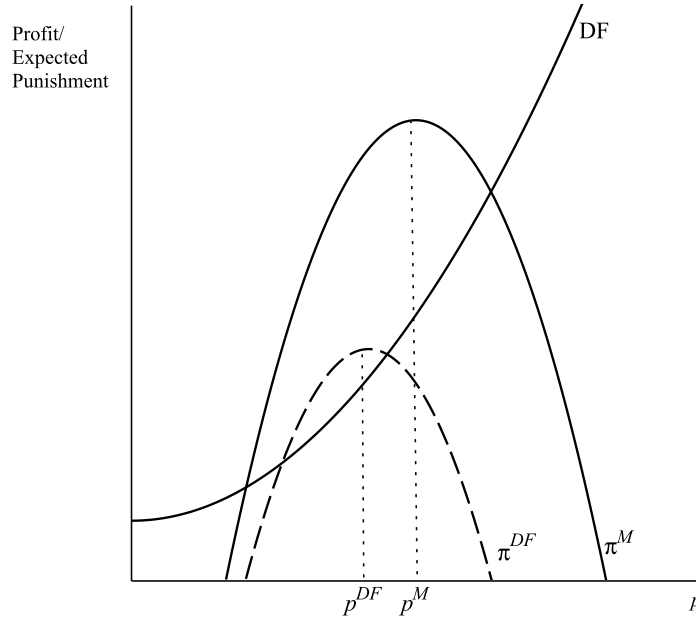
Changing the cartel's optimisation problem, however, is not the only channel through which composition deterrence can be achieved by endogenous enforcement. Price collusion is in nature a cooperative problem. Therefore, in this analysis we wish to assess whether pricing below the monopoly level could also be an optimal reaction to strategic uncertainty in non-cooperative infinitely repeated games. For this purpose, we introduce an antitrust regime with endogenous punishment, in which several cartel prices are equally profitable but carry different levels of riskiness of collusion as is

---

[11]We consider our findings based on this assumption to hold in most cases, in particular with respect to the counteracting effects of frequency and composition deterrence. However, special scenarios may exist in which the assumption does not hold. See Bos *et al.* (2015) for an example in which antitrust punishment benefits cartels.

[12]As outlined before in Sections 4.1 and 4.2, fines increase with the overcharge, e.g. in the presence of private enforcement against cartels, in which compensations are based on overcharge estimates. Further, detection is more likely to occur the more pronounced the effect of the cartel is on market outcomes (Crede, 2015; Harrington and Chen, 2006; Harrington, 2005, 2004a).

explained below. As such, we rule out any incentive that arises from the relative size of the expected profitability of the cartel, and make sure that cartelists' collusive price choices are driven by their decisions when facing strategic uncertainty. As mentioned in Section 3.3.2, this constitutes the crucial property of our experimental design, which is explained further below.

In the framework of the experiment, several price equilibria exist. The non-cooperative equilibrium that tends to dominate choices in case of coordination failure is characterised by each subject choosing $p^B = 41$ and obtaining a competitive profit of $\pi^B = (41 - 40)/3 = 0.33$ in each period. However, subjects reaching price agreements – depending on which price they have agreed on – can each earn a high or a low cartel profit, $\pi_C^h = (52 - 40)/3 = 4$ or $\pi_C^l = (46 - 40)/3 = 2$, both of which are strictly higher than π^B. In the treatment Baseline, the expected punishment is exogenous and fixed at $0.2 \times 12 = 2.4$, therefore the relevant expected payoffs of forming cartels are given by

$$
\left.\begin{aligned}
E(\pi_C^h) &= 4 - 2.4 &\rightarrow& \quad 1.6 > \pi_B, \\
E(\pi_C^l) &= 2 - 2.4 &\rightarrow& \quad -0.4 < \pi_B.
\end{aligned}\right\} \text{ for Baseline} \tag{4.2}
$$

The expected payoff of colluding on p_C^h is strictly higher than the payoff π_B associated with the non-cooperative equilibrium, whereas it is negative for p_C^l. Hence cartels can only be profitably formed on price 52 in expected terms, and rational subjects would – absent other strategic considerations – either collude on price 52 or not collude at all.

In treatments EndoF, EndoD, and BothEndo, the expected punishment of forming a cartel increases with the cartel price, i.e. it is endogenous. The expected punishment DF is now different for the two cartel prices: it is lower with 0.4 for price 46 and higher with 2.4 for price 52. Consequently, the expected payoffs of colluding on the high and the low cartel prices are the same, i.e.

$$
\left.\begin{aligned}
E(\pi_C^h) &= 4 - 2.4 \rightarrow 1.6 > \pi_B, \\
E(\pi_C^l) &= 2 - 0.4 \rightarrow 1.6 > \pi_B.
\end{aligned}\right\} \text{ for EndoD, EndoF, and BothEndo} \tag{4.3}
$$

In these treatments, the two cartel prices thus represent two payoff-equivalent Pareto-dominant collusive equilibria. In the following, we characterise the equilibrium conditions, which may vary under different treatments and with different cartel prices.

Suppose that firms react to cheating with a grim-trigger strategy. Thus, a firm that is slightly undercutting the cartel price obtains a one shot deviation profit of π_{dev}, while the others earn a profit of zero. The corresponding ICC for sustaining collusion

infinitely is given by

$$\frac{\pi_C}{1-\delta} - \frac{DF}{1-\delta(1-D)} \geq \pi_{dev} + \frac{\delta\pi_B}{1-\delta} - \frac{DF}{1-\delta(1-D)}, \tag{4.4}$$

where δ denotes the discount factor. Note that the punishment is linked to the agreed cartel price in the experiment, irrespective of whether deviation occurs afterwards or not.[13] As such, the term measuring expected punishment, $DF/(1-\delta(1-D))$, appears on both sides of the ICC and cancels out. Thus, the ICC is not affected by the variations in fines and detection probabilities and the discount factor derived from the ICC is given by

$$\delta \geq (\pi_{dev} - \pi_C)/(\pi_{dev} - \pi_B), \tag{4.5}$$

which is almost identical for p_C^h and p_C^l (0.66 and 0.64). Therefore, cartel formation should not be driven by the relative size of the ICCs: when firms are able to cooperate, they should be equally likely to collude on either of the two cartel prices.[14]

However, although the ICCs ensure the existence of a cooperative equilibrium and that the firms' incentives are not affected by variations in the exogenous uncertainty DF, the cooperative difficulties and trust issues arising from strategic uncertainty may render collusion infeasible. Thus, firms may not stick to a price agreement even if it represents an equilibrium. Recent theoretical and experimental studies suggest that the riskiness of collusion criterion selects more self-enforcing and sustainable equilibria. We use the prisoner's other dilemma by Blonski and Spagnolo (2015) to compare the riskiness of collusion when firms collude on different prices. This measure of risk can be calculated for each cartel price as the difference between two squared expressions:

$$(\pi_B - \pi_S)^2 - (\pi_C - \pi_{dev})^2. \tag{4.6}$$

This is closely related to the comparison of Nash products in static 2x2 games to determine the risk-dominant equilibrium as proposed by Harsanyi and Selten (1988). The first squared expression captures the difference between the non-cooperative equilibrium profit of π_B and the profit π_S obtained in case of being cheated upon. The second

---

[13]Absence of the sensitivity of fines to cheating is in line with fining practices in the field in most jurisdictions, in which the mere attempt to collude is illegal and is fined. Unlike it is common in the field, unsuccessful attempts to establish a price agreement are not subject to a fine in the experiment. This prevents attempts of subjects to merely indicate indifference between both prices to avoid being fined in case of unsuccessful attempts to collude from affecting the results with respect to their signalled indifference between both prices.

[14]Strictly speaking, the critical discount factors necessary to support collusion are not identical. Yet, the difference is not pronounced enough to expect it to have any (measurable) effect on subject behaviour.

squared expression contains the difference between the collusive profit and the deviation profit. *Ceteris paribus*, the more pronounced the difference between the two Nash products, the higher is the riskiness of a collusive equilibrium, which requires a higher discount factor to sustain collusion. In line with standard prisoner dilemma payoffs, in the experiment the order of profits is $\pi_{dev} > \pi_C > \pi_B > \pi_S$. The corresponding net present value of the riskiness of collusion in the experiment when the game is infinitely repeated is given by

$$\left(\frac{\pi_B}{1 - \delta^*} - \pi_S + \frac{DF}{1 - \delta^*(1 - D)} - \frac{\delta^*\pi_B}{1 - \delta^*} \right)^2 - \left(\frac{\pi_C}{1 - \delta^*} - \pi_{dev} - \frac{\delta^*\pi_{dev}}{1 - \delta^*} \right)^2, \quad (4.7)$$

where δ^* denotes the discount factor incorporating the riskiness of each collusive equilibrium. Equation 4.7 can be used to calculate the riskiness of collusion at different cartel prices and in different treatments for any given discount factor of the firms.

Figure 4.2: The riskiness of collusion
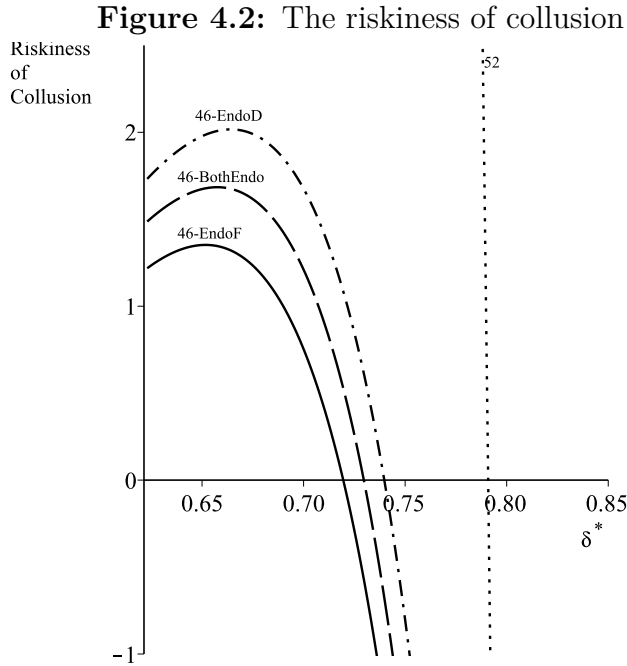
Figure 4.2 illustrates the riskiness of collusion attributed to the two cartel prices as a function of the discount factor. The riskiness of collusion is identical for price 52 in all treatments because of the identical fines and detection probabilities attached to it across treatments, whereas the riskiness of collusion for price 46 varies. Note that we do not consider collusion on price 46 in Baseline, as it is an off-equilibrium option. Given an identical payoff, the risk constraint is the tightest for collusion on price 52, whereas it does not differ much for collusion on price 46 across the treatments. The high riskiness of collusion on price 52 compared to price 46 follows from the fact that collusion on price 52 is associated with higher cheating incentives than collusion on price 46, whereas the sucker's payoff of 0 remains the same in both cases.

4.3.4 Hypotheses

We primarily seek to examine how endogenous enforcement produces composition deterrence through affecting strategic uncertainty in non-cooperative infinitely repeated games. The hypotheses relate to the behaviour of subjects rather than to markets. This follows from the focus of the experimental design on identifying individuals' preferences, which might not translate to the same behaviour at a group level due to coordination difficulties.

First, as collusion on price 46 is not a payoff-dominant collusive equilibrium in Baseline, we want to verify that collusion on price 46 is seldom chosen by subjects. The potential failure of (stable) collusion on price 46 provides evidence that payoff dominance is an important equilibrium selection criterion and fundamentally affects subjects' decisions. As the only feasible collusive equilibrium in Baseline is characterised by collusion on price 52, subjects who are willing to collude should tend to do so on price 52.

> **Hypothesis 1**: In Baseline, the low cartel price is suggested less often than the high cartel price.

In itself this is not a very interesting hypothesis to test, as the existing literature has already shown the importance of payoff dominance in equilibrium selection (see, e.g. Dal Bó and Fréchette, 2011). Yet, providing evidence in support of Hypothesis 1 is necessary for confirming subjects' understanding of the game and for the interpretation of the results. Subjects who adjust their behaviour to payoff dominance provide an indication of the presence of rational behaviour and strategic considerations. Given such behaviour, reactions to changes of the expected punishment on price 46 in the other treatments are likely to reflect strategic choices.

Next, we address the role of strategic uncertainty in the choice of cartel prices. In treatments EndoF, EndoD, and BothEndo, endogenous enforcement equalises the expected payoff of colluding on both prices, whereas the expected punishment is imposed in a way such that collusion on 46 is the equilibrium with the lower riskiness of collusion. As shown in Figure 4.2, collusion on price 46 is subject to a significantly lower level of riskiness than on price 52. If strategic uncertainty plays a role in collusive decisions, subjects should tend to suggest the low cartel price of 46. Further, disruptions of collusion by previous cheating of cartelists and detection by the antitrust authority should discourage collusion on price 52 in favour of price 46. We therefore expect subjects to show a tendency to switch from the high to the low cartel price, but not vice versa.

> **Hypothesis 2**: In treatments EndoF, EndoD and BothEndo, the low cartel price is suggested more often than the high cartel price.

Recall that the expected payoffs from collusion and the riskiness of collusion on price 52 stay the same across treatments. Further, for collusion on price 46 the expected payoffs are identical in the treatments with endogenous enforcement and the riskiness of collusion does not differ much between these treatments (as shown in Figure 4.2). As such, subjects' price choices should not significantly differ across treatments with endogenous enforcement.

> **Hypothesis 3**: Subjects' suggested cartel prices do not differ significantly between EndoF, EndoD, and BothEndo.

As cartels formed on prices 46 and 52 yield identical expected payoffs, but differ in the risk of detection, it renders the choice between the two prices similar to a choice between two lotteries with different exogenous risks. Therefore, *ceteris paribus*, risk averse subjects might be more likely to choose the cartel price 46 or not to join a cartel at all, as opposed to choosing price 52 or being indifferent between both prices. Studies on the relationship between risk aversion and cooperation do not provide conclusive findings. Sabater-Grande and Georgantzis (2002) suggest the relationship to be negative. Reuben and Suetens (2012) show that the majority of subjects behave strategically in infinitely repeated games. Dreber *et al.* (2014) conclude that the primary determinant of subjects' behaviour in these games is payoff maximisation and find no conclusive pattern between cooperation and risk attitude.[15]

Recent studies that distinguish between subjects' risk attitudes towards exogenous uncertainty and their beliefs driven by strategic uncertainty measure both elements separately (Tan and Yim, 2014; Berninghaus *et al.*, 2013; Heinemann *et al.*, 2009). Tan and Yim (2014) suggest that strategic uncertainty might be more important than exogenous uncertainty for coordination and Berninghaus *et al.* (2013) find that beliefs rather than risk attitudes explain subjects' choices. The importance of beliefs for cooperation in infinitely repeated games has also been stressed by Dal Bó and Fréchette (2011). Similarly, we conjecture that strategic uncertainty is the main driver of subjects' choices and that risk attitude does not drive subjects' price choices.

> **Hypothesis 4**: Risk attitude does not determine subjects' suggested cartel prices.

---

[15]There are two major differences between the experiment of Dreber *et al.* (2014) and ours, which might affect the results with respect to risk attitudes. First, the average cooperation period in our experiment (23.3 rounds) is significantly longer than that in theirs (10.7-11.5 rounds). Second, one feature in their experiment is that in all analysed treatments, an "execution error" occurs with 12.5% probability and alters subjects' chosen strategies. This makes it difficult for subjects to form beliefs about each other. Our experiment does not feature such a design. Thus, subjects in our experiment are more likely to be able to form beliefs on their opponents and make informed decisions based on the history of cooperation and cheating.

The four hypotheses above focus on individual subject's desired cartel prices. Yet, the overall welfare effects of endogenous enforcement is also worth examining with respect to deriving potential policy implications. As mentioned in Section 4.1, a potential trade-off may exist between frequency and composition deterrence, which makes it difficult to predict the efficiency of endogenous enforcement *ex ante.* As the overall effect on welfare is important as well, it is examined in Section 4.4.2.

4.4   Results

All reported results are based on the first 20 periods to prevent potential end-game effects from affecting the results. Further, the consideration of all markets based on the same number of observations prevents an unequal sampling of the groups from affecting the results.[16]

First, we present the results relating to our main outcome variables of interest, i.e. subjects' suggested cartel prices and the price agreements that were reached.[17] Figure 4.3a shows the proportions of suggested cartel prices across all subjects separated by treatment, where 0 denotes subjects' preference not to engage in price agreements and *Indiff.* refers to instances in which subjects suggest both prices for an agreement; Figure 4.3b shows the proportions of new price agreements that are successfully reached in a period across all markets, and 0 indicates that no agreement is reached.[18]

At first glance, observed patterns appear to be in line with the predictions of the theoretical model. The first observation is relevant to Hypothesis 1: among all treatments, the collusive price of 46 is suggested and agreed on the least often in Baseline. As such, expected profitability appears to represent an important driver of cartel price choice. The second observation is relevant to Hypothesis 3: the distributions of suggested cartel prices and price agreements reached do not seem to vary substantially across treatments with endogenous enforcement. However, as the (suggested) cartel prices are driven by a substantial number of factors and a static perspective does not allow to identify potentially important dynamics in subjects' choices, we draw further inference based on the regression analysis below.

---

[16]Given the random stoppage-rule, actual termination varies across sessions between the 20th and the 25th period.

[17]Figures on the distribution of asking prices and market prices, as well as of the asking prices conditional on cheating in the previous round can be found in the appendix in Section 4.6.1.

[18]We refer to new agreements reached in a period because agreements from previous periods that are not detected yet in the absence of new price agreements are usually being cheated upon and do not bind subjects' price choices. As such, focusing on non-detected agreements that are in place obscures subjects' behaviour and cartel success in the market.

Figure 4.3: Choice of agreements by treatment

(a) Subjects' suggested cartel prices

(b) New price agreements reached
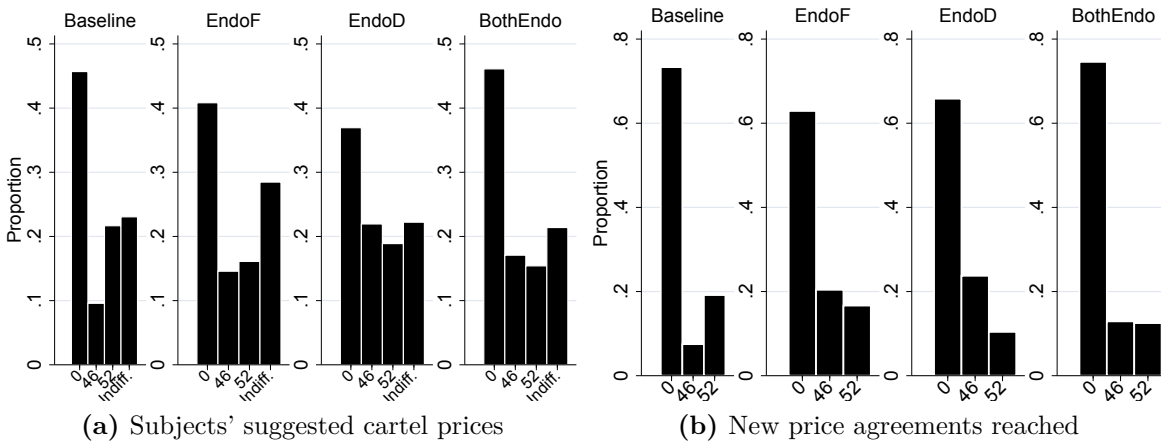
Table 4.2 provides means or proportions for additional important variables that are used in the regression analysis below together with their respective standard deviations. It is complementary to the results presented in Figure 4.3. *Market prices* represents the prices at which goods are sold and consists of the lowest asking prices in each market in each period. *Asking prices* represent the selling prices that subjects individually set in each market in each period. *Prop. new agreement* denotes the proportion of markets with newly reached price agreements, i.e. they are formed or renewed in the period irrespective of the chosen cartel price in the previous period. *Prop. cheating on 46 (52)* reports the proportion of markets with the observed occurrence of cheating on an active agreement of 46 (52), i.e. agreements that so far have not been detected. Finally, *Prop. detection on 46 (52)* is the observed proportion of markets in which agreements on 46 (52) are detected by the computer.

Table 4.2: Descriptive statistics

Variable	Baseline		EndoF		EndoD		BothEndo	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Asking prices	43.801	3.962	44.133	3.744	43.856	3.242	43.543	3.870
Market prices	41.817	2.352	42.713	3.007	42.208	1.863	42.038	2.688
Prop. new agreement	0.117	0.322	0.167	0.373	0.096	0.295	0.108	0.311
Prop. cheating 46	1.000	0.000	0.429	0.507	0.750	0.452	0.818	0.405
Prop. cheating 52	0.818	0.395	0.684	0.478	0.818	0.405	0.533	0.516
Prop. detection 46	0.000	0.000	0.190	0.402	0.000	0.000	0.091	0.302
Prop. detection 52	0.227	0.429	0.316	0.478	0.273	0.467	0.200	0.414

Several observations can be made in Table 4.2. First, asking and market prices do not differ substantially across treatments, and they both appear to be above the non-cooperative equilibrium price of 41.[19] This implies that prices rise with firms' attempts to collude. Second, combining the results in Table 4.2 with Figure 4.3 shows that

---

[19]Pairwise Mann-Whitney U (MWU) tests show that asking and market prices do not differ significantly between the treatments. See Section 4.4.2 for a detailed discussion of market prices.

cheating occurred with 100% probability on the few agreements that were reached on price 46 in Baseline. This suggests that subjects appear to have a clear understanding about the fact that collusion on price 46 is not an equilibrium in Baseline. Furthermore, no cartel formed on price 46 was detected in EndoD because of the low detection probability of 3.3% in this treatment.

A comparison with the results of the *Fine* treatment in the experiment in Chapter 3 of this thesis shows that the more complicated coordination problem in this essay's experiment reduces the prevalence of price agreements. The comparability is given because both experiments feature the same three firm homogeneous goods price competition game with very similar discount factors (0.66 in the Fine treatment in Chapter 3 compared to 0.66 for collusion on price 52 and 0.64 for collusion on price 46 in this experiment). The main difference arises from the difficulty of reaching price agreements. Whereas in Chapter 3 price agreements merely require subjects to unanimously express the support for such an agreement with a pre-determined cartel price, in this essay's experiment they need to agree on the cartel price as well. Comparing the results in the two experiments, the additional layer of complexity results in a much higher incidence of cheating than in the Fine treatment in Chapter 3, in which it is just above 40%.

4.4.1 Subjects' suggested cartel prices

The hypotheses are tested with nonparametric hypothesis tests as well as with regression analysis. The hypotheses tests allow for simple first tests of Hypotheses 1 to 3, but feature low power as they disregard more than 98% of the data. This follows because the Mann-Whitney U (MWU) test requires to condense the 20 observations of each subject that contains information on subject-, market-, and time-specific determinants of suggested cartel prices and interactions into a single (static) observation for each market. Thus, unlike the regression analysis it does not allow for a *ceteris paribus* analysis of the different actions and effects that determine the results and hence cannot properly capture important dynamics that arise from subjects' learning over time. Therefore, the nonparametric hypothesis tests are complementary to the regression analysis, which represents our main tool for the analysis.

We examine individual subjects' choices of suggested cartel prices with a multi-level multinomial logit model with random effects (random coefficients) at the subject and the market levels. The results are presented in Table 4.3. Columns II to V present the estimated average marginal effects (with their cluster-robust standard errors in brackets) of the regressors on subjects' choices of price agreements, with the choices being no agreement (0), an agreement on 46, an agreement on 52, or an agreement on either price (indifference between 46 and 52). The inclusion of lagged dependent

Table 4.3: Suggested cartel agreement – Multi-level multinomial logit results

	0	46	52	46/52
Lag asking price	0.003	0.001	0.006*	−0.010**
	(0.003)	(0.001)	(0.003)	(0.004)
Lag lowest seller	−0.046*	0.010	0.016	0.020
	(0.026)	(0.016)	(0.019)	(0.022)
Lag agreement 46	−0.150	−0.033	0.097	0.085
	(0.151)	(0.059)	(0.102)	(0.124)
Lag agreement 52	−0.075	−0.086***	0.087	0.074
	(0.087)	(0.023)	(0.062)	(0.074)
Lag choice 46	−0.231***	0.143***	−0.010	0.098***
	(0.042)	(0.033)	(0.031)	(0.037)
Lag choice 52	−0.309***	0.058**	0.146***	0.105***
	(0.026)	(0.025)	(0.037)	(0.033)
Lag choice 46/52	−0.427***	−0.053***	−0.069**	0.549***
	(0.040)	(0.018)	(0.029)	(0.050)
Lag detection 46	0.169**	−0.028	−0.024	−0.118***
	(0.068)	(0.035)	(0.057)	(0.033)
Lag detection 52	0.150**	−0.030	−0.056	−0.064
	(0.071)	(0.030)	(0.034)	(0.048)
Lag cheating 46	0.186	0.026	−0.115***	−0.097
	(0.160)	(0.104)	(0.034)	(0.075)
Lag cheating 52	0.015	0.127*	−0.091***	−0.051
	(0.069)	(0.070)	(0.020)	(0.057)
EndoF	−0.068	0.091*	−0.070	0.047
	(0.076)	(0.053)	(0.044)	(0.054)
EndoD	−0.099	0.154**	−0.016	−0.040
	(0.072)	(0.063)	(0.046)	(0.055)
BothEndo	−0.005	0.098*	−0.046	−0.047
	(0.065)	(0.058)	(0.048)	(0.047)
Period	0.014***	−0.001	−0.008***	−0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
Automatic price 52	−0.042	0.029	−0.035	0.049
	(0.048)	(0.025)	(0.038)	(0.039)
Risk attitude – 2	−0.084	0.011	0.080	−0.007
	(0.115)	(0.056)	(0.089)	(0.070)
Risk attitude – 3	0.007	−0.005	0.047	−0.049
	(0.104)	(0.068)	(0.089)	(0.072)
Risk attitude – 4	−0.202	0.073	0.164	−0.034
	(0.127)	(0.075)	(0.126)	(0.069)
Risk attitude – 5	−0.009	−0.015	0.030	−0.006
	(0.117)	(0.072)	(0.097)	(0.067)
Risk attitude – 6	−0.039	−0.003	0.034	0.008
	(0.110)	(0.061)	(0.080)	(0.075)
Pseudo-LL		-2135.055		
Observations		2736		

Notes: Values represent average marginal effects. Cluster-robust standard errors in parentheses. ***, **, and * denote 1%, 5%, and 10% significance levels. Random effects (coefficients) are included at the subject and the market levels.

and independent variables allows to capture dynamics in subjects' behaviour and represents an autoregressive distributed-lag model, which is a workhorse model in applied econometrics to model dynamics in time series and panel data (Pesaran and Shin, 1998; Banerjee *et al.*, 1990). Their coefficients measure how subjects react to choices and outcomes in the previous period when making a choice with respect to their suggested cartel prices in the present period.

Lag asking price denotes the price that a subject set in the previous period. *Lag lowest seller* is an indicator variable showing whether the subject set the lowest price in the previous period (such that she made a positive profit). *Lag agreement 46 (52)* controls for the effects of previous periods' active price agreement on 46 (52) on subjects' suggested cartel prices in this period. *Lag choice 46, 52*, and *46/52* are indicator variables of a subject's suggested agreement price in the previous period, with no agreement (0) serving as the baseline. Given that a price agreement was in place in the previous period, *Lag cheating 46 (52)* and *Lag detection 46 (52)* indicate whether cheating or detection occurred on the respective agreement. *EndoF*, *EndoD*, and *BothEndo* are treatment indicator variables that measure the treatment effects, and Baseline serves as the baseline. *Period* measures the period effects. *Automatic price 52* is an indicator variable for the markets in which a price of 52 is chosen automatically if all three subjects in the group are indifferent between both cartel prices. Finally, *Risk attitude* measures subjects' risk attitudes elicited in the risk elicitation task based on Eckel and Grossman (2008). This variable contains discrete values between 1 and 6 where the value of 1 represents the highest level of risk aversion and serves as the baseline. In order to rule out potential biases resulting from the specification of the functional form of the risk attitude measure, we include 5 indicator variables for the different degrees of risk aversion and use risk attitude with a level of 1 as the baseline.

The multi-level multinomial logit model does not allow for a direct test of the hypotheses related to particular included variables. Instead, we need to rely on *indirect tests* that are derived from different patterns in the data that are likely to arise if the hypotheses hold true. *Ceteris paribus*, we would expect Hypothesis 1 to lead to more agreements to be reached on price 46 in EndoF, EndoD, and BothEndo than in Baseline. This is the case if the marginal effects of the three indicator variables indicating these treatments are positive and significant in Column 46. If Hypothesis 3 holds true, *ceteris paribus*, we would expect no significant differences in the marginal effects of the treatments EndoF, EndoD, and BothEndo with respect to the choice of the suggested cartel price 46. If Hypothesis 4 is correct and risk attitudes do not determine the suggested cartel price, the marginal effects of the risk attitude indicator variables should be insignificant.

Finding evidence in support of Hypothesis 2 necessitates relying on observing more

complex patterns for deriving any inference. Indications in support of Hypothesis 2 can be derived from asymmetries in reactions of subjects to cheating and detection on prices 46 and 52 in the previous period in suggesting cartel prices in the present period. In the presence of strategic uncertainty, subjects should react to cheating and detection on price agreements on price 52 by becoming more likely to suggest the low cartel price 46. Similarly, persistence in subjects' choices of the suggested cartel price over time or reactions to agreements on prices 46 or 52 reached in the previous period can point towards behaviour in line with Hypothesis 2. This is the case if subjects over time tend to adjust their suggested cartel prices from 52 to 46. These asymmetries can be identified by comparing signs and significant of the respective variables on cheating, detection, past suggested cartel prices, and cartel prices reached in Columns 46 and 52.

We first address Hypothesis 1. Taken together, the results in Figure 4.3a and Table 4.2 suggest that relatively fewer price agreements are reached on price 46 in Baseline and that all of them are cheated upon. We use the Wilcoxon matched-pairs signed-rank test to test whether subjects are less likely to suggest the cartel price 46 than the cartel price 52 in Baseline. It tests the H_0 of an equal proportion of both suggested cartel prices across all markets within a treatment and separated by treatments. For this, the proportion of periods a subject suggests 46 and 52 as the cartel price in the 20 regular periods are calculated and then compared with each other. The p-values of these tests are reported in Table 4.4.

Table 4.4: Wilcoxon matched-pairs signed-rank test:
Prop. of subjects suggesting cartel price 46 compared to 52

Baseline	EndoF	EndoD	BothEndo
0.007	0.937	0.556	0.480

As can be seen from Figure 4.3a and Table 4.4, there is evidence that subjects in treatment Baseline suggest 46 as a cartel price significantly less often than cartel price 52. This result is also evident from the regression analysis. The marginal effects of the treatment indicators in Table 4.3 show that, *ceteris paribus*, subjects are between 9.1% to 15.4% more likely to suggest the low price 46 in the treatments with endogenous enforcement than in Baseline. An additional indirect way to test Hypothesis 1 is to compare the proportion of suggested agreements on price 46 across markets between the treatments using the MWU test as done in Table 4.5. This draws on the result in Figure 4.3a that the proportion of suggested cartel prices on price 46 appears to be larger in the three treatments with endogenous enforcement than in Baseline. The p-values of pairwise comparisons across the treatments both for suggested cartel prices 46 and 52 are reported in Table 4.5. For this, the proportions of periods a subject suggests 46 and 52 as the cartel price in the 20 regular periods are calculated as the

average across time and markets within a treatment. These averages are then compared across the treatments. Indeed, in line with the other evidence the indirect test shows that the proportion of suggested cartel prices of 46 are lower in Baseline than in EndoD and BothEndo. Taken together, we find strong evidence in support of Hypothesis 1.

Table 4.5: MWU test p-values: Prop. of markets' suggested price agreements

Variable		EndoF	EndoD	BothEndo
	Baseline	0.111	0.013	0.082
Sugg. agreements on price 46	**EndoF**		0.111	0.707
	EndoD			0.224
	Baseline	0.285	0.623	0.235
Sugg. agreements on price 52	**EndoF**		0.602	0.931
	EndoD			0.469

> **Result 1**: In Baseline, the low cartel price is rarely chosen and results in unstable price agreements.

The intuition behind Result 1 is straightforward: the low cartel price does not produce stable cartels in Baseline because it is not a payoff-dominant collusive equilibrium. It follows that, *ceteris paribus*, positive payoffs are necessary for collusion but payoff dominance does not have to be the most suitable equilibria selection condition to describe subjects' choices. Furthermore, it confirms that subjects show an understanding of the game.

Next, we test Hypothesis 2 that subjects choose the low cartel price more often than the high cartel price in the treatments with endogenous enforcement. No clear preference for the low cartel price is evident from Figure 4.3b for the treatments with endogenous enforcement. This is in line with the results of the Wilcoxon matched-pairs signed-rank test in Table 4.4 that tests whether the proportion of suggested agreements on prices 46 and 52 are different from each other. This is in contrast to the regression results in Table 4.3, in which the dynamics and evolution of subjects' preferences with respect to the suggested cartel prices can be captured. First, the treatment indicator variables show that, *ceteris paribus*, subjects in the treatments with endogenous enforcement are between 9.8% and 15.4% more likely to suggest the low cartel price of 46 than in Baseline. Second, the *Lag cheating* variables suggest that cheating in the previous period strongly discourages subjects from choosing price 52 as the agreement price choice in the present period. However, cheating on an agreement on price 52 in the previous period increases the probability of subjects choosing 46 in the present period. Hence, subjects learn about the different levels of riskiness and tend to prefer the low cartel price with the lower riskiness of collusion to the higher cartel price as a reaction to cheating. Third, the *Lag choice* variables suggest that, *ceteris paribus*, subjects show

a 5.8% probability to switch from price 52 to 46, but no such substitution pattern can be observed for the opposite direction.

The divergence in results might follow from the static analysis conducted in the nonparametric hypothesis tests. If subjects need several periods to learn about the riskiness of collusion attached to both cartel prices, a preference for the low cartel price of 46 might only develop over time. This might especially be the case when attempts of (successful) collusion only occur for the first time after several periods into the game. The regression analysis makes more efficient use of the data and can capture these dynamics in the game. We therefore prefer the regression results to the nonparametric tests in Table 4.4. Thus, we find evidence in support of Hypothesis 2, which indicates composition deterrence in line with the model predictions.

> **Result 2**: Subjects tend to suggest the low cartel price in the treatments
> with endogenous enforcement.

The substitutability between fines and detection probabilities in the treatments with endogenous enforcement is addressed next. In line with Hypothesis 3, comparisons of the proportions of the suggested cartel prices of both 46 and 52 across the treatments in Table 4.5 show that there are no significant differences in subjects' choices in the treatments with endogenous enforcement. This is in line with the results of the regression analysis in Table 4.3, as is indicated by a lack of significance of the marginal effects of the treatment indicator variables for the suggested cartel price of 52 in Column 52. A similar result arises for the low cartel price of 46. To formally compare the relevant treatment effects on the suggestion of cartel price 46, we conduct pairwise Wald tests of the treatment indicator coefficients in Column 46 in Table 4.3. The test results do not reject the Null hypothesis of equality of coefficients with p-values of 0.169 (EndoF/EndoD), 0.765 (EndoF/BothEndo), and 0.103 (EndoD/BothEndo). Therefore, the results are in line with the predictions based on the experimental design establishing equality of the fines and detection probabilities for prices 46 and 52 across all treatments. Further, the regression results show that there are no significant differences between treatments for subjects' choices not to collude or to be indifferent between the two cartel prices. As such, several tests provide evidence in support of Hypothesis 3.

> **Result 3**: Subjects' suggested cartel prices are not sensitive to variations
> in the fine and the detection probability that feature the same expected
> punishment and riskiness of collusion.

As Baseline serves as the baseline in the regressions, concerns may arise due to the different design and unique collusive equilibrium of the price 52 in Baseline such that the coefficients for this treatment might differ from those in the other treatments.

Hence, to ensure the robustness of the regression results, we re-estimate the multi-level multinomial logit model but exclude Baseline and use EndoF as the baseline. This potentially provides a more homogeneous sub-sample, increases data precision, and reduces the standard errors of the estimates. The results of this robustness check can be found in Table 4.8 in the appendix in Section 4.6.1. Note that almost all qualitative findings shown in Table 4.3 are supported by the regression estimates based on the sub-sample in Table 4.8, and that the quantitative results only change marginally. As such, Result 3 and the estimated marginal effects are not driven by pooling all treatments for the estimations.[20]

Finally, we test Hypothesis 4 relating to the impact of the risk attitude on subjects' suggested cartel prices using the regression analysis in Table 4.3. Indicator variables control for the different degrees of risk aversion without specifying a particular functional form for its effect on subjects' choices. Recall that we measure 6 different degrees of risk aversion using an Eckel and Grossman (2008) style risk elicitation task, and that the measure is negatively correlated with risk aversion (a value of 1 indicates the highest level of risk aversion). Strikingly, despite this very conservative approach of including the different levels as indicator variables, not a single marginal effect is significant. Yet, it also has to be tested whether the different coefficients are jointly significant. For this purpose, we conduct joint F-tests of the coefficients for the risk attitude indicator variables for all outcome regressions. However, in each case the joint F-test reports that the variables are jointly insignificant with p-values of 0.252 for outcome regression 46, 0.327 for 52, and 0.790 for 46/52.[21] Therefore, Hypothesis 4 is supported by the data.

Result 4: Risk attitude does not determine subjects' cartel price choices.

A possible explanation for the above result may be the length of interaction of 20 periods, which enables subjects to engage in learning about their partners' strategies through repeated interactions. This learning and the associated effects on beliefs might dominate the effects of the risk attitude on choice. However, when information is limited and learning could not take place yet, subjects' actions in the presence of coordination difficulties may be affected by their risk attitudes. This might be the case especially

---

[20]Note that we obtain the above result in a framework that is relevant to the field and features an overproportionate increase in the punishment compared to an increase in the cartel profit resulting from a higher overcharge. The result may therefore not be generalised to enforcement regimes that have different punishment strategies, i.e. punishment that increases linearly or underproportionately to the increase in cartel profit, which have little relevance for the field.

[21]Similarly, the risk attitude coefficients are jointly insignificant in the specification that excludes Baseline, with p-values of 0.144 for 52 and 0.579 for 46/52. Only for the choice of 46, the risk indicator variables are jointly significant (p-value 0.038). This result is driven by the single significant coefficient of *Risk attitude – 4*. If risk attitude indeed plays a role, then the other coefficients should be individually significant as well. We therefore do not consider this as sufficient and robust evidence for the significance of risk attitude for price choices.

at the beginning of the game. Therefore, we re-run the same model again, but limit it to the first three periods of the game. However, the results stay the same and joint F-tests again indicate that the coefficients are jointly insignificant.[22] This suggests that risk attitude does not drive the results in the first few periods either.

4.4.2 Cartel prices reached and welfare effects

In the final part of this section, the welfare effects of endogenous enforcement are analysed at the market level. First, pairwise comparisons of market prices and the overall proportions of cartel formation between treatments are conducted using MWU tests.[23] Although Table 4.2 reports that average market prices are slightly higher in the treatments with endogenous enforcement (in particular in EndoF), the p-values shown in Table 4.6 suggest that there are no significant differences in market prices across all treatments.

Similarly, the p-values suggest that the proportions of cartel formation do not differ across the treatments. Recall that with exogenous enforcement in Baseline and endogenous enforcement in the other treatments, a lack of overall differences might provide evidence for the absence of significant welfare changes induced by endogenous enforcement. Yet, this lack of differences might be due to the trade-off between frequency and composition deterrence.

Table 4.6: MWU test p-values: Market prices and prop. of markets with agreements

Variable		EndoF	EndoD	BothEndo
	Baseline	0.368	0.312	0.751
Market prices	**EndoF**		0.840	0.214
	EndoD			0.452
	Baseline	1.000	0.702	0.768
Price agreements	**EndoF**		0.836	0.883
	EndoD			0.977

To address the factors underlying this lack of differences, regression analysis on cartel price agreements that were reached is conducted using a multinomial logit model with random effects at the market level. The estimated average marginal effects (and cluster-robust standard errors in brackets) are reported in Table 4.7. Unlike the multi-level multinomial logit model used to analyse subjects' cartel price choices, the regression table of cartels' price choices does not feature a column for the indifference between

---

[22]The results can be found in Table 4.9 in the appendix in Section 4.6.1.

[23]Given the instability of agreements observed in Table 4.2, the proportion of markets with newly reached price agreements is regarded as a more insightful measure of cartel formation for this purpose. Yet, the same results arise with the alternative measure of active cartel agreements.

the two prices, as price agreements can only be reached on a single price. Most of the regressors in Table 4.7 are introduced before in the discussion of Table 4.3. As the regressions in Table 4.7 are based on market level observations, we use *Lag market price* to instead of *Lag asking price* included in Table 4.3. Further, due to the fact that the observations are at the market level, it cannot be controlled for individual subjects' past suggested cartel prices. Instead, we control for the effect of previous periods' price agreements that were reached on prices 46 and 52 on the choice of cartel price agreements in the present period with *Lag agreement on 46* and *Lag agreement on 52*, respectively.[24]

We start with the uncertainty-driven composition deterrence, as it is our primary focus. The marginal effects of the treatment indicator variables in Table 4.7 suggest that, compared to Baseline, fewer cartels are formed on the high cartel price 52 in EndoD and BothEndo. As the overall proportions of cartel formation do not differ across the treatments, this indicates that more cartels are formed on the low cartel price 46 in these two treatments. As such, we observe that endogenous enforcement leads to composition deterrence that, on average, reduces the cartel overcharge. A similar treatment effect is missing in EndoF, which is in line with the observation from Figure 4.3a that there are no large differences in the choice of 52 between Baseline and EndoF. Yet, a significant effect of EndoF might only be missing because the treatment effect could be captured by other coefficients already.

---

[24]Unlike in Table 4.3, we do not include separate effects of cheating on prices 46 and 52 in Table 4.7, as covariance matrices for the marginal effects cannot be calculated if both effects are determined separately.

Table 4.7: Choice of price agreements – Multinomial logit

	0	46	52
Lag agreement on 46	-0.825^{***}	0.743^{***}	0.082
	(0.066)	(0.198)	(0.150)
Lag agreement on 52	-0.828^{***}	0.318	0.511
	(0.057)	(0.318)	(0.321)
Lag cheating	0.138^{***}	-0.067^{***}	-0.071^{***}
	(0.013)	(0.019)	(0.016)
Lag detection 46	0.064^{***}	-0.007	-0.056^{***}
	(0.016)	(0.011)	(0.017)
Lag detection 52	0.093^{***}	-0.051^{***}	-0.042^{**}
	(0.019)	(0.014)	(0.019)
Period	0.007^{***}	-0.003^{**}	-0.004^{***}
	(0.002)	(0.001)	(0.002)
Automatic price 52	-0.005	-0.006	0.011
	(0.020)	(0.013)	(0.017)
Lag market price	0.009	-0.008	-0.001
	(0.006)	(0.006)	(0.003)
EndoF	-0.030	0.041	-0.011
	(0.041)	(0.036)	(0.017)
EndoD	0.002	0.032	-0.034^{**}
	(0.033)	(0.033)	(0.017)
BothEndo	0.001	0.026	-0.027^{*}
	(0.033)	(0.032)	(0.016)
Log pseudolikelihood		-290.758	
Observations		912	

Notes: Values represent average marginal effects. Cluster-robust standard errors in parentheses. ***, **, and * denote 1%, 5%, and 10% significance levels. Random effects (coefficients) are included at the market level.

Given the evidence of composition deterrence in EndoD and BothEndo, *ceteris paribus*, market prices should be lower in these treatments than in Baseline. The fact that market prices do not differ significantly implies that there must be an opposing effect that is driving market prices up. The dynamics of cartel prices that are agreed upon provide some indications that the frequency deterrence is reduced. As shown in Table 4.7, the *Lag agreement* marginal effects indicate that price agreements on price 46 are persistent whereas those on price 52 are not. Furthermore, the *Lag detection* variables suggest that detection of an agreement on 46 in the previous period does not reduce the probability that an agreement is reached on 46 in this period, but significantly reduces the probability of reaching an agreement on 52. Hence, while endogenous enforcement induces composition deterrence towards the low price with the lower riskiness of collusion, it also stabilises collusion on this price.

To conclude, we find strong evidence in support of the trade-off generated by endogenous enforcement on deterrence and welfare. First, it reduces cartel overcharges and increases the complexity of cartel coordination rendering reaching cartel agreements more difficult. Second, it leads to an adverse effect on frequency deterrence as it encourages additional collusion on the low cartel price and stabilises collusion on the low cartel price. As the result of these counteracting effects, market prices and the proportions of cartel formation do not differ significantly between treatments with exogenous and endogenous enforcement, and the effects of endogenous enforcement on overall cartel harm are unclear.

There are additional reasons for this lack of a clear result to arise. First, as mentioned in Section 3.2, the experiment is designed to examine the effects of strategic uncertainty on cartel price choices. Fines and detection probabilities are chosen such that the expected profitability between cartel prices 46 and 52 is equal, which allows us to focus purely on the uncertainty-driven incentives. Yet, one would expect stronger effects of endogenous enforcement if it imposes a penalty regime that leads to a relatively higher expected profitability for collusion on price 46 than on 52. Thus, as a result of the focus of this study on the indirect effects of endogenous enforcement on cartel price choices through strategic uncertainty, only lower bounds for potential overall effects of endogenous punishment on cartels are estimated. Second, the communication protocol is designed to clearly capture individual subjects' preferences for cartel prices and to measure the effects of strategic uncertainty on decision-making. This comes at the cost of rendering coordination more difficult. Alternative communication protocols such as free-form chat may be more effective in establishing and sustaining collusion, as they allow bargaining to precede cartel price choices. This is in contrast to our experimental design, in which bargaining is only possible through signalling over the course of several periods of the game.

4.5 Conclusion

As antitrust enforcement cannot deter cartels altogether because to do so would be prohibitively costly, composition deterrence might offer the prospect to reduce the harm caused by cartels. In addition, it represents an important determinant of the overall welfare effects of antitrust laws. To better understand the properties of endogenous antitrust enforcement, this essay seeks to establish its effects on composition deterrence that stem from its impact on cartel prices through strategic uncertainty. For this purpose, we conduct a laboratory experiment with non-cooperative infinitely repeated games, in which subjects engage in competition in homogeneous goods price competition

triopolies. Subjects can coordinate on prices but face the risk of being detected and fined if they collude. The experimental design and chosen parameters allow us to abstract from the profitability-driven incentives induced by endogenous enforcement and focus on the indirect effects of strategic uncertainty on subjects' decision-making. The implemented endogenous enforcement regime features an overproportionate increase in the expected punishment compared to an increase in the cartel profit resulting from a higher overcharge, and gives rise to payoff-equivalent collusive equilibria that bear different levels of the riskiness of collusion.

The results confirm that payoff dominance is an insufficient criterion to explain cartel prices in the presence of endogenous punishment. In line with the prisoner's other dilemma framework of Blonski *et al.* (2011) and Blonski and Spagnolo (2015), subjects are found to tend to select the collusive equilibrium with the lower riskiness of collusion in the presence of two Pareto-dominant collusive equilibria. This highlights the role of strategic uncertainty in a cartel's coordination problem. Overproportionately increasing punishment steers cartels towards a lower cartel price by aggravating the cooperation risks. In addition, we find that cartel price selection is insensitive to different combinations of fines and detection probabilities and is not driven by subjects' risk attitudes. Further, the results show a trade-off between increased composition deterrence and reduced frequency deterrence in the presence of endogenous enforcement. On the one hand, we find strong evidence of the strategic uncertainty-driven overcharge reducing composition deterrence. On the other hand, the frequency deterrence properties of the enforcement are reduced. This follows from the fact that cartels can self-select themselves into weaker expected enforcement by colluding on low prices only. As a result of these opposing effects and the study's focus on the indirect effects through strategic uncertainty, the overall welfare effects of endogenous enforcement remain unclear.

Future research should study the trade-off between frequency and composition deterrence in the context of cartels further. Although a large body of literature exists for the two forms of deterrence, little is known about how they interact under different enforcement regimes. For example, while the enforcement featured in our study focuses purely on uncertainty-driven effects, the trade-off may differ when profitability-driven effects are present as well. Several studies (e.g. Alm *et al.*, 1995, 1992) have found that subjects do not react to punishment in accordance to its functional form. Therefore, the observed behaviour might differ from the optimal behaviour as predicted by theory.

4.6 Appendix

4.6.1 Auxiliary figures and tables

Figure 4.4: Choice of prices

(a) Asking and market prices   **(b)** Asking prices and previous cheating
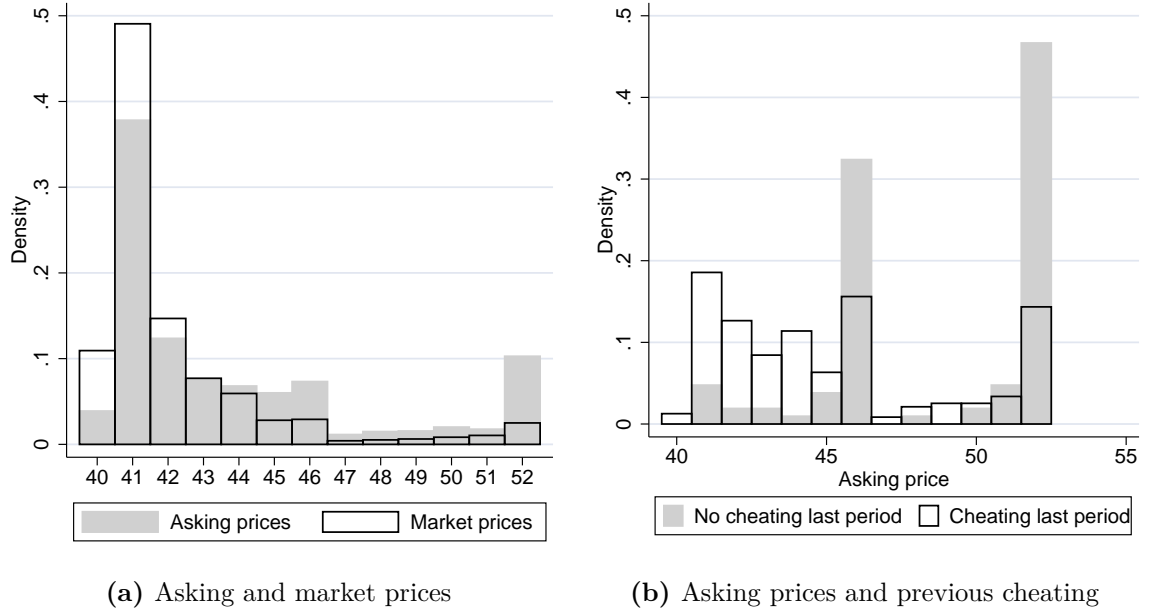
Figure 4.4a plots the distributions of both asking and market prices across treatments in the first 20 periods. As can be seen, homogeneous goods price competition and cheating leads market prices to be below the asking prices. As stated above, the predominant price set by individuals is 41, which is the only symmetric non-cooperative equilibrium that yields positive payoffs for all subjects. Figure 4.4b plots the distribution of asking prices dependent on whether cheating occurred in the previous period on a price agreement reached in that period. Remarkably, although many agreements were reached on 52, cheating primarily leads subjects to charge prices in the range between 41 and 45. This is evident from Figure 4.4a as well. In other words, subjects that deviate from agreements on price 52 tend to undercut prices by charging prices below the low cartel price of 46.

Table 4.8: Suggested cartel agreement excluding Baseline – Multi-level multinomial logit results

	0	46	52	46/52
Lag asking price	0.004	0.002	0.004	−0.010**
	(0.003)	(0.001)	(0.003)	(0.004)
Lag lowest seller	−0.045	0.018	0.000	0.027
	(0.033)	(0.020)	(0.022)	(0.023)
Lag agreement 46	−0.126	−0.039	0.101	0.064
	(0.147)	(0.074)	(0.100)	(0.107)
Lag agreement 52	0.039	−0.109***	0.039	0.031
	(0.076)	(0.025)	(0.036)	(0.074)
Lag choice 46	−0.227***	0.172***	−0.028	0.082**
	(0.046)	(0.032)	(0.034)	(0.036)
Lag choice 52	−0.321***	0.063*	0.120***	0.138***
	(0.030)	(0.027)	(0.037)	(0.056)
Lag choice 46/52	−0.443***	−0.072***	−0.058*	0.573***
	(0.052)	(0.023)	(0.035)	(0.068)
Lag detection 46	0.140**	−0.029	−0.013	−0.098***
	(0.068)	(0.044)	(0.059)	(0.033)
Lag detection 52	0.106	−0.035	−0.041	−0.066
	(0.084)	(0.035)	(0.029)	(0.060)
Lag cheating 46	0.181	0.033	−0.109***	−0.104
	(0.166)	(0.121)	(0.037)	(0.069)
Lag cheating 52	−0.095*	0.197**	−0.066***	−0.035
	(0.055)	(0.083)	(0.018)	(0.056)
EndoD	−0.029	0.055	0.055	−0.081*
	(0.074)	(0.041)	(0.057)	(0.046)
BothEndo	0.070	−0.003	0.020	−0.086**
	(0.070)	(0.041)	(0.056)	(0.040)
Period	0.015***	−0.002	−0.008***	−0.005
	(0.003)	(0.002)	(0.002)	(0.002)
Automatic price 52	−0.010	0.011	−0.041	0.040
	(0.054)	(0.033)	(0.042)	(0.045)
Risk attitude – 2	−0.019	0.022	0.031	−0.034
	(0.112)	(0.052)	(0.079)	(0.068)
Risk attitude – 3	0.037	0.026	0.027	−0.090
	(0.107)	(0.084)	(0.085)	(0.075)
Risk attitude – 4	−0.278**	0.213**	0.121	−0.056
	(0.085)	(0.085)	(0.115)	(0.074)
Risk attitude – 5	0.007	0.058	−0.009	−0.056
	(0.152)	(0.118)	(0.100)	(0.070)
Risk attitude – 6	−0.004	0.048	−0.056	0.012
	(0.112)	(0.078)	(0.052)	(0.087)
Pseudo-LL		-1587.218		
Observations		2052		

*Notes: Values represent average marginal effects. Cluster-robust standard errors in parentheses. ***, **, and * denote 1%, 5%, and 10% significance levels. The Baseline treatment observations are excluded, and EndoF observations serve as the baseline. Random effects (coefficients) are included at the subject and the market levels.*

Table 4.9: Suggested cartel agreement in the first three periods – Multi-level multinomial logit results

	0	46	52	46/52
Lag asking price	0.017*	−0.016*	−0.007	0.006
	(0.010)	(0.009)	(0.013)	(0.009)
Lag lowest seller	0.061	−0.053	−0.095*	0.086
	(0.062)	(0.053)	(0.054)	(0.060)
Lag agreement 46	0.059	−0.188***	0.297**	−0.168**
	(0.120)	(0.018)	(0.102)	(0.065)
Lag agreement 52	−0.065	−0.222***	0.361***	−0.073
	(0.104)	(0.020)	(0.127)	(0.088)
Lag choice 46	−0.227***	0.067	0.011	0.149**
	(0.049)	(0.058)	(0.073)	(0.066)
Lag choice 52	−0.355***	0.030	0.279***	0.046
	(0.043)	(0.058)	(0.066)	(0.068)
Lag choice 46/52	−0.387***	−0.136***	−0.168***	0.691***
	(0.042)	(0.036)	(0.056)	(0.060)
Lag detection 52	0.406***	−0.107**	−0.149*	−0.150**
	(0.125)	(0.045)	(0.082)	(0.066)
Lag cheating 46	−0.194***	0.709***	−0.286***	−0.229***
	(0.028)	(0.020)	(0.031)	(0.075)
Lag cheating 52	−0.215***	0.658***	−0.239***	−0.204***
	(0.035)	(0.022)	(0.036)	(0.032)
EndoF	0.076	−0.020	−0.088	0.032
	(0.068)	(0.066)	(0.065)	(0.056)
EndoD	0.011	0.161**	−0.075	−0.098*
	(0.068)	(0.076)	(0.054)	(0.054)
BothEndo	0.001	0.203***	−0.117*	−0.086**
	(0.075)	(0.077)	(0.067)	(0.044)
Period	0.073	−0.082**	−0.015	0.025
	(0.048)	(0.037)	(0.057)	(0.044)
Automatic price 52	0.004	−0.003	−0.047	0.046
	(0.040)	(0.039)	(0.046)	(0.042)
Risk attitude – 2	0.009	0.059	0.019	−0.087
	(0.070)	(0.083)	(0.081)	(0.060)
Risk attitude – 3	0.068	−0.056	−0.038	0.026
	(0.087)	(0.072)	(0.076)	(0.080)
Risk attitude – 4	−0.196***	0.146	0.101	−0.052
	(0.066)	(0.102)	(0.090)	(0.059)
Risk attitude – 5	−0.020	0.063	0.008	−0.051
	(0.076)	(0.095)	(0.104)	(0.061)
Risk attitude – 6	0.019	0.086	−0.022	−0.083
	(0.084)	(0.094)	(0.077)	(0.058)
Pseudo-LL		-296.673		
Observations		288		

Notes: Values represent average marginal effects. Cluster-robust standard errors in parentheses. ***, **, and * denote 1%, 5%, and 10% significance levels. Includes only the first three periods. Random effects (coefficients) are included at the subject and the market levels. The variable Lag detection 46 could not be included, as detection on this price did not occur in the first three periods.

4.6.2 Instructions (BothEndo – automatic price of 52)

Instructions

Welcome and thank you for taking part in this experiment. In this experiment you can earn money. How much money you will earn depends on your decision and on the decision made by other participants in this room. The experiment will proceed in two parts. The currency used in the experiment is experimental points. Each experimental point is worth 12 pence. All earnings will be paid to you in cash at the end of the experiment.

Every participant receives exactly the same instructions. All decisions will be anonymous, that is, your identity will not be revealed to other participants at any time during or after the experiment. It is very important that you remain silent. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you.

Instructions for Part 1

In the first part of the experiment you will be asked to choose from six different gambles (as shown below). Each circle represents a different gamble and you must **choose the one that you prefer**. Each circle is divided in half. The two numbers in each circle represent the amount of experimental points the gamble will give you.
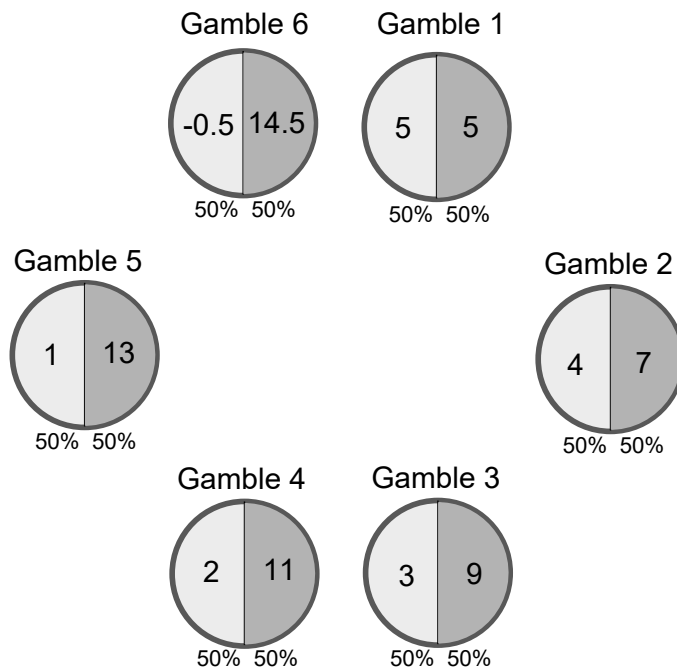
An experimenter will toss a coin to determine which half of the circle is chosen. A

volunteer will come to the front of the room and confirm the result of the coin toss. If the outcome is heads, you will receive the number of points in the light grey area of the circle you have chosen. If the outcome is tails, you will receive the number of points shown in the dark grey area of the circle you have chosen. Note that no matter which gamble you pick, each outcome will occur with a 50% chance.

Please select the gamble of your choice by **entering the number** of your gamble (1, 2, 3, 4, 5, or 6) in the field "I choose Gamble" and press OK.

Once everyone has made their decision, Part 1 will end and we will move on to Part 2 of the experiment.

Instructions for Part 2

In this part of the experiment you will form a group with two other randomly chosen participants in this room. Throughout the experiment you are matched with the same two participants. All groups of three participants act independently of each other. This part of the experiment will be repeated for at least 20 rounds. From the 20th round onwards, in each round there is a **20 out of 100 (20%)** chance that the experiment will end.

Your job:

You are in the role of a firm that is in a market with two other firms. In each round, you will have to set a price for your product. This price must be one of the following prices:

40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52

You will only sell the product if the other firms do not set a lower price than you in that round. If you sell the product, your earnings are equal to the difference between the price and the cost, which is 40:

Earnings = Price - 40.

Therefore, you will not make any profit if you set a price of 40. If you do not sell the product, you will not get any earnings but you will not incur costs either. If two or more firms set the same lowest price, the earnings will be shared equally between them.

Before you set your price, you may decide to agree with the other firms to set the same price and share the earnings. There are two prices you can agree on, **46** and **52**. If you agree with the other two firms to set the price of 46 and all firms set 46, each firm will get a profit of **2** experimental points. If you agree with the other two firms to set the price of 52 and all firms set 52, each firm will get a profit of **4** experimental points.

The picture below shows how this will look on the computer. All firms get asked whether they want to agree on prices.

If you want to agree on prices, you can indicate so by choosing the price you want to agree on. You can choose either 46 or 52, or you can choose both prices. The other two firms will do the same. If all three firms wish to agree on prices, and there exists a common price among the three firms' chosen prices, an agreement is reached on that common price.

If all three firms choose both prices, implying that they are fine agreeing on both prices, then a price agreement on 52 will be reached automatically. If there is no common price among firms' choices, no price agreement is reached. For example, no common price is reached if two firms suggest price 46 and one firm suggest price 52.

If you do not wish to agree on prices with the other two firms, you can indicate so by choosing the option No. If at least one firm chooses No, there will be no price agreement.

The following table summarizes how price agreement can be reached:

Price agreement is reached, if	All three firms wish to agree on prices and they reach one common price	Agreed price is the common price (46 or 52)
	All three firms wish to agree on prices and they reach two common prices (when all firms choose both 46 and 52)	Agreed price is 52
Price agreement is not reached, if	All three firms wish to agree on prices and they reach no common price	
	At least one firm does not wish to agree on prices	

After deciding whether you would like to form a price agreement, you have to set a price by filling the "Choose a price" box shown below. If a price agreement is reached, a message will appear above the "Choose a price" box showing the price that you agreed

on. If no price agreement is reached in that round, no message will appear and you have to set a price without being able to coordinate with the other firms.

Choose a price

However, the agreement is not binding and you are not required to set the agreed price. After your price choice, you will be informed about the prices that you and the other firms set in that round. If you successfully reach a price agreement, the agreement may be discovered by the computer. The computer can discover the agreement on price 46 with a **10 out of 100 (10%)** chance, and can discover the agreement on price 52 with a **20 out of 100 (20%)** chance. If the agreement on price 46 is discovered, a fine of **4** experimental points has to be paid. If the agreement on price 52 is discovered, a fine of **12** experimental points has to be paid. If no price agreement is reached, you cannot be discovered or receive a fine.

The chance of being discovered and the fine depend on the price agreement reached, but not on the prices you set afterwards. The table below summarizes the chance of being discovered by the computer and the associated fines:

A price agreement can be discovered as long as it has not been discovered in a previous round. Once this has happened, you will not be fined in the future, unless you make a price agreement again. If you and the other two firms had several price agreements and none of them has been discovered, the chance of being discovered and fine always depend on the latest price agreement.

The picture below summarizes the structure of Part 2 of the experiment.

| 1. Question: Agree on price? |
| 2. Set a price for your product |
| 3. Learn about the prices set by all firms |
| 4. Learn about potential detection and final profits |
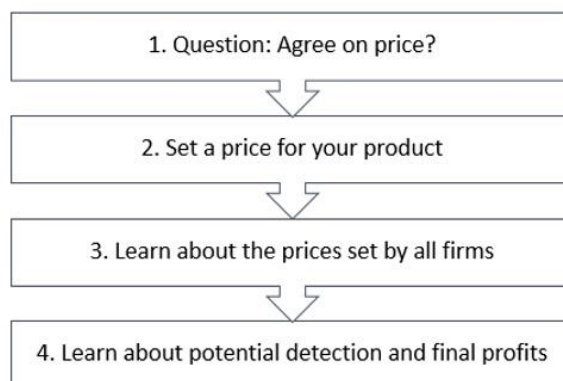
At the end of each round, you will be told the earnings you have made in this round. If you have reached a price agreement, you will also be told whether the agreement has been discovered by the computer.

Final payment:

At the beginning of the experiment you start with an initial endowment of 50 points = 6 GBP. The earnings you make in each round will be converted into cash. Each point is worth 12 pence, and we will round up the final payment to the next 10 pence. We guarantee a minimum earning of 2 GBP.

Chapter 5

An ex-post evaluation of multi-market merger decisions by the European Commission

5.1 Introduction

[1] The operation of a firm – and with it its business strategy – often depends on the interrelation between its activity in different product and geographic markets. Such interdependencies exist for many reasons including the strategic allocation of resources for their most efficient use or firms' attempts to leverage market power across markets. Despite the fact that such a *multi-market* presence represents the norm and not the exception, its effects on firm behaviour are seldom considered. These effects often come in the form of *spillover effects*, in which a firm's presence or behaviour in one market affects its operation in other markets. In competition economics, spillover effects play an important role in the analysis of mergers and acquisitions.[2] For example, the effects of a merger on post-merger price equilibria and their dependence on the multi-market presence of a firm are often of high interest to competition authorities. Prominent examples for the importance of spillover effects in merger control are some high-profile mergers in the airlines industry, in which multi-market contact and networks, two common sources of spillover effects, have played major roles in merger control. Nevertheless, scope and scale of spillover effects in practice remain elusive.

The lack of consideration given to spillover effects poses the question to what extent they potentially undermine the validity of predictions about (anticompetitive) effects of

---

[1]This chapter is based on joint work with my supervisor Stephen Davies and Farasat Bokhari.

[2]For ease of exposition, we will subsequently refer to both mergers and acquisitions as mergers unless stated otherwise.

mergers and should receive more scrutiny. A prime example of an area in which this question arises is the merger control case practice of the European Commission (EC) in the pharmaceutical industry. Virtually all large mergers evaluated by the EC affect more than one member state and product market, i.e. they are multi-market mergers. At the same time, the pharmaceutical industry is characterised by large expenditures on research & development (R&D) and a high capital intensity, which provide much scope for spillover effects. In this article, we seek to determine whether the case practice developed by the EC is effective at handling such merger cases, in which the market investigation usually is focused only on markets, in which the parties have both an overlapping market presence and a joint market share above 35%. This narrow focus might be undermined if spillover effects commonly create anticompetitive effects in other markets.

The evaluation of mergers based on market-level instead of firm-level observations can come in the form of two different types of studies. First, the effects of several mergers that affect the same or different markets can be analysed. Our study falls into this category. In such studies, clusters of firms, in which all observations of a firm are cross-sectionally and intertemporally interdependent, exist both in the treated and non-treated groups. Second, the effects of a single merger on several separate geographic (local) markets can be studied. The clustering structure is similar to that occurring if several mergers are studied, but only one firm can be found in the treatment group. In both types of studies, similar identification and estimation issues arise. Evaluating mergers with respect to their effects at the market-level rather than the commonly used firm-level allows us to study price effects at the market-level. The term market-level here refers to a firm's individual business presence in a specific product (or geographic) market, which is a more disaggregated perspective than the overall firm-level. We refer to all of an individual firm's business presence (i.e. all its sales) in a market as the *business unit* (BU) of this firm in the market.

We develop a novel framework to estimate the effects of multi-market mergers at the market-level that addresses the clustering structure in these environments using the PSM-DiD estimator that uses weights derived from propensity scores for DiD regressions. The new framework is then used to evaluate the EC's case practice. Our novel empirical framework allows us to assess whether spillover effects across a firm's BUs lead to significant price effects of mergers in markets, in which the involved parties had no overlapping market presence pre-merger. We find evidence for such spillover effects that cast doubts on the approach chosen by the EC to evaluate multi-market mergers. Further, with regards to the *ex post* evaluation of mergers, we show how control groups need to be carefully constructed to avoid their contamination in the presence of such spillovers to ensure the identification of the treatment effect.

Not only the standard difference-in-differences (DiD) estimator, which is commonly used in the literature to estimate the effects of mergers, breaks down in the presence of spillover effects, but also matching methods used to address the self-selection of firms into treatment, i.e. the decision to merge. We propose adequate strategies to estimate propensity scores in merger analysis at the market-level, which require not only BU-level variables, but also overall firm-level variables to be included in their estimation. We show how failure to do so leads to self-selection bias not removed by propensity score matching (PSM) and biased treatment effects.

The price effects on BUs of the firms involved in mergers are tested for two types of spillover effects. First, we distinguish between markets in which the parties have different levels of market share. Second, spillover effects are distinguished between markets with and without overlaps by the parties involved in the mergers. The results are complex but provide some evidence for both positive and negative spillovers. In particular, significant price reductions are found in markets in which the parties had market shares exceeding 35% prior to the merger. On the other hand, significant price increases are detected in markets, in which they had market shares between above 15% and 35%. Most importantly for our purposes, the price reductions primarily appear in those markets in which only one of the parties was active prior to the merger. This can only be explained by the existence of spillover effects, and suggests that merger appraisal should not just be confined to markets in which parties overlap.

5.2 Literature review and background

5.2.1 *Ex post* merger evaluation studies

A large body of literature has empirically studied the effects of mergers on different measures of firm performance and market outcomes. Most closely related to this study are *ex post* studies that analyse price effects of mergers on either the involved firms (see, e.g. Krishnan, 2001) or on their competitors (see, e.g. Prager and Hannan, 1998). The literature started with case studies of single mergers and often with price effects in single markets (Schumann *et al.*, 1992; Werden *et al.*, 1991; Borenstein, 1990; Barton and Sherman, 1984), and multi-merger analysis with overall price effects only followed later (Prager and Hannan, 1998; Kim and Singal, 1993). Weinberg (2007) and Kwoka *et al.* (2015) provide surveys of the related literature and the chosen estimation methods.

The literature can further be classified according to their treatment of mergers as exogenous or endogenous events: a merger can be viewed as exogenous, i.e. if the assumption is imposed that it occurs completely at random across the population of

firms. Yet, if it is seen as endogenous and firms decide to merge in anticipation of the expected gains from doing so (i.e. they self-select themselves into the group of involved firms), the involved firms represent a non-random subset of the population of firms in an industry. This self-selection has important consequences for the empirical approach required to estimate the causal effects of mergers.

In particular older studies treat mergers as exogenous events (see, e.g. Ho and Hamilton, 2000; Kim and Singal, 1993), i.e. they assume that self-selection does not occur. Thus, they are based on the assumption that merging firms are a random draw of the population of all firms and that mergers occur by chance and not because of profitability considerations. Such an assumption must be motivated well, as its violation can induce significant biases of the estimated causal effects. More recent studies tend to treat mergers as endogenous events (see, e.g. Danzon *et al.*, 2007; Dranove and Lindrooth, 2003), as firms self-select into the decision to merge based on the anticipation of profit gains to be realised from the transactions. Endogeneity of the decision to engage in mergers has important implications on the estimation strategy. While exogenous mergers can be analysed with DiD estimators based on the treatment evaluation literature (see, e.g. Kwoka *et al.*, 2015), self-selection requires different approaches, as it leads to biased estimates of the treatment effects in standard DiD estimators. Depending on the availability of data and the research questions at hand, different approaches can be used to estimate treatment effects in the presence of endogenous mergers: merger simulation (Verboven and Björnerstedt, 2016), instrumental variable estimators (Haucap and Stiebale, 2016), regression discontinuity design (Saarimaa and Tukiainen, 2010), PSM (Behr and Heid, 2011), or a combination of PSM with DiD (PSM-DiD) estimators (Ornaghi, 2009).[3]

The large majority of *ex post* studies either estimate the effects of mergers at the firm-level, i.e. they are concerned with overall profits, or they only study single markets that are affected. Yet, fuelled by an increased availability of detailed disaggregated data, recent studies have started to analyse the effects of a single merger across different markets. These studies include in particular a growing literature on the effects of mergers on local markets. For example, Villas-Boas *et al.* (2016) analyse the impact of a French supermarket merger on prices of the merging firms and their competitors in different cities across France, each of which constitutes an own local market. Again, merger studies aimed at local markets can be classified into those assuming exogenous

---

[3]Most closely related to this essay is the study of Ornaghi (2009), who studies the effects of mergers in the pharmaceutical industry on R&D activity as well as the financial performance of the consolidated firms. In line with this essay, Ornaghi employs PSM-DiD estimators to correct for self-selection in the estimation of the treatment effect. However, his analysis is based on firm-level rather than market-level data, such that it does not deal with the additional layers of difficulties relating to the selection of the control group and identification of the treatment effect in more disaggregated data.

mergers (Argentesi *et al.*, 2016; Ashenfelter *et al.*, 2015; Hosken *et al.*, 2012) and others assuming endogenous mergers and correcting for self-selection (Villas-Boas *et al.*, 2016; Aguzzoni *et al.*, 2016). Both groups of studies rely on DiD-type estimators to estimate treatment effects. What they all have in common is their varying reliance on the standard Rubin causal model and estimation strategies developed for (approximating) randomised controlled trials. Therefore, they all suffer to varying extents from potential problems in the construction of valid control groups and endogeneity resulting from unsatisfactory attempts to estimate propensity scores.

As we argue below, in such a type of analysis, in which a treatment is occurring at the firm-level but the units of observation are BUs at the market-level, a different set of assumptions and estimation methods is required. This follows from the fact that this involves a cluster structure in which each unit of observation depends on other units of the firm in the sample. This cluster structure does not represent a "standard" randomised controlled trial. As a result, unless a number of restrictive assumptions are met, the chosen approaches lead to wrong inference.[4]

5.2.2   Spillover effects

Multi-market mergers warrant the consideration of all affected markets because of potential spillover effects across separate markets, i.e. the effect of a merger in one market potentially affects other markets as well.[5] These spillover effects can occur for a multitude of reasons, and can be classified into two broad categories according to their potential effect on market prices.

Spillover effects with potentially price-reducing effects include increases in efficiency and cost reductions following from a reallocation of resources within the company towards the most efficient use (Jovanovic and Rousseau, 2008; Harris *et al.*, 2005; Peristiani, 1997), product diversification that reduces portfolio risks and therefore financing costs (Perold, 2005), economies of scope in production and innovation (Cockburn and Henderson, 2001), and economies of scale, e.g. cost savings realised from large joint input purchases (Crook and Combs, 2007; Mabert and Venkataramanan, 1998).

On the other hand, spillover effects with potentially price-increasing effects in turn include the anticipation of innovation competition by competitors leading them to

---

[4]Problems include the violation of the underlying stable unit treatment value assumption leading to biased treatment effects, PSM that does not fully eliminate the self-selection bias, and underestimated standard errors because of wrongly specified covariance matrices. This is discussed in greater detail below.

[5]A classification of spillover effects of mergers in multi-market environments is provided in Section 5.3.3.1.

limit their commitment to the market (Farrell and Shapiro, 2010), coordinated effects resulting from increased multi-market contact (Baker, 2002; Bernheim and Whinston, 1990), endogenous barriers to entry arising from increased marketing expenditure (Sutton, 1998; European Commission, 1998), opportunities to leverage market power across markets resulting from an increased size of the product portfolio (European Commission, 1998), and an increased financial strength that can be used to weaken rivals, e.g. by enabling aggressive pricing or by serving as a pre-emptive device to limit rivals' entry and capacity expansion (Fresard, 2010; Campello, 2006; Bolton and Scharfstein, 1990; Benoit, 1984). An overview of some of the identified effects together with sources discussing them is provided in Table 5.1.

The importance of spillover effects in merger control has been acknowledged in the literature for several decades. Yet, due to the large number of potential procompetitive (increases in efficiency) and anticompetitive (abuse of dominance, foreclosure, predatory pricing) effects, especially in conglomerate mergers spillovers have remained elusive and guidelines for merger control are hard to derive. As the overall effect of such mergers can be positive or negative depending on which of the effects dominate, the issue is frequently debated among scholars, lawyers, and competition authorities. One of the most spectacular cases related to this was the proposed General Electric/Honeywell merger, which was cleared by the DOJ in the USA but blocked by the EC in 2001 based on concerns over the establishment of a dominant position arising from portfolio effects. Discussing the case, Kolasky (2002) points out that spillover effects played almost no role in U.S. merger control in the 80's and 90's, and that in Europe only three cases (including the General Electric/Honeywell merger) were blocked by the EC based on suspected anticompetitive effects arising from spillovers, two of which were later cleared by the courts due to a lack of evidence backing up such concerns.

Neven (2008) assesses the application of different concepts of spillover effects by the EC in recent cases. He argues that economies of scale and scope as a result of multi-market presence, portfolio effects inducing changes of demand triggered by reduced search and transaction costs, and the bundling of brands to strengthen the weaker brand do not represent meaningful sources of anticompetitive behaviour. He finds that only the reduction of potential competition by strategically exploiting substitution patterns between separate markets appears to be problematic. Further, he suggests that potential competition can be softened by tying goods together from separate but competitive markets, and that exclusionary practices can arise in case the tied products are complements.

Nevertheless, at least some of the findings of Neven (2008) remain contested. For example, Chung and Jeon (2014) examine conglomerate mergers in Korean beer and

Table 5.1: Spillover effects and their implications on prices

Factor	Impact on price	Literature
Increase in efficiency & cost reduction following from reallocation of resources for more efficient use	- / ?	Jovanovic & Rousseau (2008), Harris *et al.* (2005), Peristiani (1997)
Product diversification: reduces risk and financing costs	- / ?	Perold (2005)
Economies of scope (in innovation)	- / ?	Cockburn & Henderson (2001)
Economies of scale (e.g. cost savings from joint input purchases)	- / ?	Crook & Combs (1997) Mabert & Venkataramanan (1998)
Innovation competition & anticipation by competitors	+	Farrell & Shapiro (2010)
Coordinated effects & multi-market contact	+	Bernheim & Whinston (1990), Baker (2002)
Barriers to entry – marketing expenditure	+	Sutton (1998), European Commission (1998)
Product portfolio & leveraging market power	+	European Commission (1998)
Financial strength & weaker competitors	+	Fresland (2010), Benoit (1984), Campello (2006), Bolton and Scharfstein (1990)

liquor markets and find that the merged firms could leverage market power across markets due to portfolio effects. Yet, despite dominant market positions, they could not find dominant conglomerate firms engaging in any foreclosure activity. They conclude that market conditions need to be evaluated in merger control to anticipate gains in market power arising from increased leverage related to portfolio effects. Chen and Rey (2015) provide a theoretical model on how portfolio effects can strategically be used to soften competition using bundling and tying, and propose that the potential for such practices needs to be scrutinised in merger control.

The ambiguity surrounding the assessment of spillover effects has lead to controversies among competition authorities and divergent practices in merger control (see, e.g. Fox, 2007). As a result, many competition authorities are reluctant to oppose mergers on the grounds of spillover effects. The need for further studies is self-evident. It is hoped that this essay helps to fill the gap by suggesting a methodology that can be used to shed further light on the issue. Moreover, in the application in Section 5.4 we find evidence of spillover effects.

5.3   Estimation strategy

5.3.1   Novel approach and contribution

In this section, we introduce the empirical methodology to carry out BU-level treatment evaluation in the context of multi-market mergers. First, we introduce the concept of cluster-randomised trials previously applied in the fields of sociology, medicine, and educational research. This provides the potential for a theoretical framework that can be applied in our study to identify causal effects. Second, and again from other disciplines, we introduce concepts for the estimation of propensity scores in cluster-randomised trials, which we will use for matching and the construction of control groups to address self-selection of specific firms into merger activities. Third, we operationalise the methodology so it may be applied in the present context by introducing classifications and exclusion restrictions that allow the construction of valid treatment and control groups.

We shall follow the prevailing approach of most *ex post* merger evaluation studies to use DiD models that draw on the treatment evaluation methodology. Further, like many recent studies in this area, we make use of PSM to address self-selection bias and combine the two methodologies using the PSM-DiD estimator. This choice is motivated by several factors. First, DiD approaches are relatively simple to apply to a large number of markets. Second, PSM allows to correct for self-selection of firms engaging in

mergers, such that unbiased, causal effects of mergers can be estimated. Third, unlike other methods based on propensity scores, PSM yields comparably little bias in the outcome regression results in case the propensity score is misspecified (see, e.g. Zhao, 2008; Rubin, 2004). Fourth, relying on the PSM-DiD model in the outcome regressions ensures that unobserved time-invariant heterogeneity does not bias the estimates. Fifth, the PSM-DiD estimator can be adapted to the special form of randomised controlled trials introduced below that needs to be approximated in the econometric design to capture the causal effects of mergers.

5.3.2   Modifying the Rubin causal model to conduct a cluster-randomised trial

The Rubin causal model

The foundation of most of the treatment evaluation literature is the Rubin causal model (Rubin, 1974). It is based on the concept of *randomised controlled trials* (RCT). The RCT framework provides a set of conditions needed to identify the causal effects of an intervention. These include the assumptions that all observations used in the analysis are statistically independent from each other. In addition, the intervention needs to be assigned randomly to the observations, such that the treatment does not correlate with characteristics of the treated units (e.g. firms). In case this assignment is non-random, the *treatment effect on the treated* (ATT) can be estimated nonetheless (e.g. using PSM) provided that the selection process can be controlled for – this is ensured by the *conditional independence assumption* (CIA). A central element to identify the causal effect of a treatment is the *stable unit treatment value assumption* (SUTVA). It requires that the fact that a unit (such as a firm) receives the treatment does not affect the outcomes (i.e. prices) of any of the other observations (firms) in the sample. A more in-depth look at the standard methodology can be found in the appendix in Section 5.6.1.

Identification of the treatment effect in the context of multi-market mergers is more difficult if the level of observation/aggregation is more disaggregated than the firm-level. A particular cluster structure exists in such data. All business units of a firm form a cluster because their outcome observations such as prices are correlated with each other and across time. As a result, the SUTVA is fundamentally violated because the effect on prices of a merger on the merging firm in market A is most likely not independent from the effect on prices in market B. Further, to satisfy the SUTVA the construction of the control group needs to be adapted to the correlations between BU observations

of treated firms across markets that affect competitors. Due to these spillover effects, prices of competitors of merging firms can be affected by the treatment (the merger) in markets, in which only one of the merging firms is active pre-merger. This follows from the fact that the competitors' prices in different markets are correlated with each other within "their" clusters. Then, the SUTVA is only satisfied if these observations of the competitors are not included into the control group.

In addition, self-selection is not only potentially correlated with market-specific characteristics, but also with overall firm-level characteristics. In other words, not every single BU of a firm independently decides whether it merges with another firm, but the firm reaches this decision based on both its position in the individual markets in which it is active as well as based on its overall firm-level position and characteristics. As such, the CIA does not hold because the probability of a BU to receive the treatment does not only depend on its BU-level variables, but also on the overall firm-level variables.

The cluster-randomised trial

A cluster-structure as described above requires a refinement of the SUTVA in order to take correlations and spillovers into account. Further, in case self-selection occurs in an observational study, the steps that have to be taken to control for non-random treatment assignment are aimed at replicating conditions of the RCT by correcting for self-selection to fulfil the CIA. In the context of multi-market mergers the controlled trial that needs to be approximated is fundamentally different to the standard RCT as assumed by most studies. Specifically, unlike in RCTs, the treatment (the merger) is assigned either to *all* (a firm involved in a merger) or to *none* (firms not engaging in any merger) of the units (BUs) within a cluster. As a result, the SUTVA is violated as the prices in different markets served by the involved firms are not statistically independent, and self-selection bias needs to be corrected differently to account for both market- and firm-level effects. The type of underlying randomised controlled trial that must be approximated is known as the *cluster-controlled trial*, also known as *group-controlled trial* or *randomised community trial* (see, e.g. Giraudeau and Ravaud, 2009).[6]

> **Definition (CRT):** *Cluster-randomised trials* are a form of randomised controlled trials in which the treatment is administered to either all or none of the units within a cluster, and clusters both with and without treatments assigned to them exist.

---

[6]This is not to be confused with multisite randomised trials, in which the treatment is assigned to some units within a cluster only: they require different approaches to the estimation of propensity scores, matching, and outcome analysis compared to CRTs (see, e.g. Leon *et al.*, 2013; Thoemmes and West, 2011).

VanderWeele (2008) extends the concept of RCTs introduced above to CRTs. CRTs (as well as multisite randomised trials) are largely unknown in economics, but have been developed in the sociological research of neighbourhood interventions, education, as well as in medicine. In order to allow causal inference in CRTs, five fundamental assumptions need to be fulfilled.

First, the SUTVA assumption underlying RCTs needs to be altered. In our context, this is because spillover effects lead an involved firm's outcome (e.g. the price) in a particular market to potentially depend on the outcomes in its other product markets. VanderWeele introduces:

> **Definition (CL-SUTVA):** *The cluster-level stable unit treatment value assumption.* This is a modified SUTVA assumption that allows for the identification of causal effects in CRTs.

Although he borrows the term from sociological research, VanderWeele (2008)'s article relates to the field of medicine.[7] Nevertheless, the concept is general and equally applies to economic applications. In the present analysis, it requires that the price of a firm in a given market does not depend on the treatment assignment of other firms. Put differently, the outcomes of a firm's BUs do not depend on the merger activity of other firms.[8]

Importantly, however, the CL-SUTVA explicitly allows for spillover effects between a firm's different BUs. Interdependencies between the different BUs through spillover effects lead to a treatment effect, which depends not only on the causal effect of a merger on the outcome variable in a market, but also on the effect it has on other markets, in which the involved firms are present. Thus, the estimated treatment effect is a *total effect*, which consists of the *direct effect* of the merger on a BU as well as the *indirect effect* on it originating from spillover effects from the firm's other BUs.

Second, a *modified conditional independence assumption* is required. If X denotes BU-level-specific and Z overall firm-level-specific variables with indices i and k denoting markets and firms, respectively, it can be expressed as:

$$Y_{ikd} \perp D_k | X_{ik}, Z_k, \tag{5.1}$$

where Y_{ikd} denotes the outcome variable, D_k the treatment, and index d takes the value 1 for treated units and 0 otherwise. If propensity scores are used to eliminate self-selection bias, failure to control for firm-level variables Z_k leads to a violation of

---

[7]He labels the CL-SUTVA the neighbourhood-level stable unit treatment assumption.

[8]This has important consequences for the construction of control groups and is discussed below.

the modified conditional independence assumption. Thus, propensity scores need to include all relevant firm-level variables Z_k that determine both the decision to merge as well as the outcome variable, and not only BU-specific variables X_{ik}. Therefore, an observational study that attempts to approximate a CRT necessitates stronger data and identification requirements than "standard" RCTs.

Third, the *overlap assumption* – also known as the *common support assumption* – needs to be defined as

$$0 < P(D_k = d | X_{ik} = x, Z_k = z) < 1 \tag{5.2}$$

for all realisations of d, x and z.[9] This means that BU observations with any given combinations of values of firm-market and overall firm-level variables X and Z have a positive probability to be in the treatment and the control groups. In other words, there need to be observations in the control group that are sufficiently similar to those in the treatment group both with respect to the BU-level and the firm-level variables.

Fourth, the *consistency assumption* requires $D_k = d \Rightarrow Y_{ikd} = Y_{ik}$. In other words, the outcome Y (price) that is observed for an observation if (or if not) assigned to the treatment D is equal to the outcome that is in fact observed for the given treatment assignment. This condition is usually fulfilled in observational studies and relates to the absence of measurement errors.

Fifth, VanderWeele (2010) adds the assumption of *intact clusters* as introduced by Hong and Raudenbush (2006) in the context of sociological multisite randomised trials. This requires that treatments do not alter the cluster membership of the BUs. In the present context, the two parties of a merger have to be pooled not only after the merger or acquisition, but also before it takes place. Only then can the observed difference in the outcome over time in a treated unit be related to the treatment effect.

5.3.3   Constructing treatment and control groups in CRTs

5.3.3.1   A classification of treatment effects in multi-market mergers

The treatment effect is identified by comparing the BUs that are affected by mergers with a control group consisting of BUs of firms not affected by any merger. To guide

---

[9]This extension is necessary for both treated and non-treated observations for the *average treatment effect* (ATE), but only for treated observations for the ATT. Therefore, the overlap assumption does not need to hold for non-treated units in order to obtain a valid ATT. Further, it follows that common support must be satisfied not only for BU specific variables X, but also for the overall firm-specific variables Z.

the construction of the control group, the following example contains a list of treatment effects that can arise in multi-market mergers if the analysis is based on market-level observations. Consider the hypothetical case shown in Figure 5.1. There are four firms with BUs active in five markets A,B,C,D, and E. Firms 1 and 2 merge and taken together are active in markets A, B, C, and D both before and after the merger. They have an overlapping market presence only in market B. Firm 3 is active in markets A, B, D, E, and F. Therefore, it is a competitor to the merging parties because of overlaps in market presence in markets A, B, and D.

> **Definition:** A firm that is active in the same markets as the merging firms is a *competitor* to those firms (in these markets). The observations of the competitor in these overlapping markets are referred to as competitor observations.

Another firm, 4, is active in markets F and G, which are not (directly) affected by the merger and it is referred to as an *outsider*.

> **Definition:** An *outsider* is defined as a firm that is neither directly (through treated BUs) nor under certain assumptions indirectly (through competitors' BUs) affected by mergers. Its BUs are referred to as outsider observations.

Under certain assumptions about spillover effects (see below), Firm 3's BUs in markets E and F can also be treated as outsider BUs.

Figure 5.1: Construction of treatment and control groups
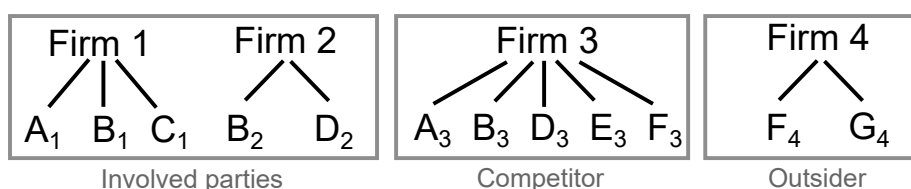
A classification of treatment effects that can be estimated in such a multi-market environment is presented in Table 5.2. Treatment effects can be estimated both for the merging parties and their competitors.

> **Definition:** Denote an *acquirer* as the firm that fully acquires another firm or parts of another firm's assets and the firm that is being bought as a *target*.

In case only parts of a firm are bought by the acquirer, the term target refers to the parts subject to the transaction only. Mergers and acquisitions are pooled in the empirical analysis and it is assumed that both types of transactions generate similar effects, as they both give rise to mostly the same spillover effects.

Depending on which observations are used from the involved parties, different treatment effects for the acquirer and the target (in case of acquisitions), their combined presence, or the BUs in which the involved parties had or did not have an overlapping market presence pre-treatment can be estimated. For the competitors, estimation is possible of the direct spillover effects of mergers for markets in which the mergers occur and of the indirect spillover effects that arise in the other markets in which the competitors are active. Theoretically, in the example above it is perceivable that firm 1 and 2's merger creates *indirect spillover effects* for firm 3's market behaviour in non-affected markets E and F as well through altering its behaviour in the affected markets A, B, and D. In that case, firm 3's observations in market E and F can be used to estimate indirect merger effects on competitors.

Table 5.2: Classification of treatment effects

Estimated treatment effect	Treatment group	Control group
All BU's of merging parties	A_1, B_{1+2}, C_1, D_2	$[(E_3, F_3), F_4], G_4$
Overlapping BUs of merging parties	B_{1+2}	$[(E_3, F_3), F_4], G_4$
Non-overlapping BUs of merging parties	A_1, C_1, D_2	$[(E_3, F_3), F_4], G_4$
Acquirer's BUs	A_1, C_1	$[(E_3, F_3), F_4], G_4$
Target's BUs	D_2	$[(E_3, F_3), F_4], G_4$
Direct effect on BUs of competitors	A_3, B_3, D_3	$[F_4], G_4$
Indirect effect on BUs of competitors	E_3, F_3	G_4

Notes: BU observations in round (squared) brackets have to be dropped in the presence of spillover effects of type 1 (2).

In the application below, we focus on the effects of the mergers on all BUs of the merging parties, and further analyse those with and without overlaps in market presence. Therefore, we can refrain from imposing somewhat arbitrary choices with respect to the classification of firms into acquirers and targets in mergers among equals.

5.3.3.2   Exclusion restrictions

Cross-sectional exclusion restrictions

Construction of the control group depends on the limits of the indirect spillover effects. In the context of this study, the CL-SUTVA as defined by VanderWeele (2008) requires that a firm's BUs should only be affected by its own merger, but not by other mergers. Thus, it allows for spillovers across a firm's BUs created by the treatment, which are captured in the total treatment effect as indirect effects. They add to the total effect together with the direct effects that consist of the treatment effect on a BU in a market that originates from that market. Continuing to use the above example, we classify potential spillover effects into several groups.

Definition: A *spillover effect of type 1* is an effect of a merger on the competitors' BUs in markets in which the merging parties have no presence.

For example, observations E_3 and F_3 would be affected by type 1 spillover effects. This effect might arise if the competitor's exposure to the merger in markets A, B, and D lead it to change its behaviour in markets E and F. If such indirect spillover effects exist, observations inside the **round brackets** in Table 5.2 need to be dropped from the control group pool.

Definition: A *spillover effects of type 2* exists if the competitors' exposure to a merger has an effect on outsiders in markets in which the competitor is active but the merging parties are not.

Such a spillover effect of type 2 can affect BU F_4. Put differently, being affected by a merger in markets A, B, and D, Firm 3 might change its behaviour in markets E and F. This in turn might affect outsider Firm 4 and its BU in market F. If such indirect spillover effects exist, all observations contained in the **squared brackets** in Table 5.2 need to be dropped from the control group pool as well. The existence of type 2 spillover effects implies stronger assumptions than type 1. This follows from the fact that type 1 effects are necessary for type 2 effects to occur.[10]

In assessing how to address such spillover effects we face an inevitable trade-off. Spillover effects of type 1 might be unlikely, e.g. if merger control ensures that the involved parties do not have strongly overlapping market activities that could generate large merger effects. In that case, observations E_3 and F_3 can safely be included into the control group pool. A larger pool from which the control group can be constructed is desirable, as it potentially increases the quality of matches as well as the efficiency of the econometric estimators suggested below. Outsider observations in markets in which the competitor is active but the involved parties are absent need to be dropped (observation F_4) from the control group pool if one assumes spillover effects of type 2 to be present.

Provided that spillover effects of type 1 are unlikely to arise and are limited in magnitude in case of their existence, type 2 spillover effects are very unlikely to arise.[11] In the

---

[10]Strictly speaking, the indirect spillover effect of the merger on the competitors' markets in which the merging parties are not active could in turn produce an indirect spillover effect of observations of outsiders in which neither the involved parties nor the competitors are active (*spillover effect of type 3*, observation G_4). We rule out such an effect as it is very unlikely to arise but would put very strong demands to the availability of data.

[11]An indirect way to test the existence of type 1 spillover effects is to test for indirect merger effects on the competitors' BUs. If no such merger effects can be found, it could be argued that neither type 1 nor type 2 spillover effects exist, such that the potentially affected BU observations can be included into the control group pool. Yet, such a test is imperfect as the lack of significant treatment effects, e.g. for BUs E_3 and F_3, might be the result of misspecification of the econometric models used for the

application below, we will assume that neither spillover effects of type 1 or type 2 are present. This choice is motivated by our limitied control group pool, and allows us to add competitor observations to the control group pool which are more similar to the treated BUs than the other available observations increasing matching quality.

In practical applications, researchers often face a trade-off between potential bias and efficiency when determining exclusion restrictions, especially in a panel data context, in which several mergers might affect a market over time. While stronger exclusion restrictions reduce the chance that the CL-SUTVA is violated, it might result in many observations being dropped from the analysis. This can have undesirable effects on the analysis. First, the number of treated units for which no suitable control group can be constructed is likely to increase. This potentially leads to treatment effects that are only representative for particular sub-samples of the treated units. Second, the reduced sample size used in the estimation of the ATT reduces the statistical power of the treatment effect estimators. This problem does not only follow from treated units that are dropped because no suitable control group matches can be found, but also from reductions of the control group pool.

Therefore, two further restrictions are imposed to better address the above trade-off. We assume that mergers do not have any effects unless the parties have a joint market share above a given threshold. Such an assumption of a *de minimis* threshold of a critical market share is proposed by Hosken *et al.* (2012). More specifically, in this study we ignore the parties' BU observations in markets in which their pre-treatment joint market share is below 3%. This assumption has two positive effects on the sample composition. First, it removes markets from the sample, which in the predominant number of cases should not feature any effects and would conceal significant effects in the other markets by reducing the estimated treatment effects and increasing standard errors if included. Second, it reduces the number of merger markets with significant effects that need to be dropped because of overlapping mergers in the treatment window as described below.

In addition, for similar reasons a restriction based on the relative size of the smaller party is imposed. In this study, mergers are not considered if the smaller party's revenue does not exceed 10% of the total revenue of the merged entity. Such mergers are very unlikely to generate any notable (spillover) effects on prices across markets, as the target is relatively small and unimportant for the overall firm performance of the joint entity. An example for the use of these cross-sectional exclusion restrictions can be found in the appendix in Section 5.6.2.

---

estimation or follow from a lack of power induced by an insufficient sample size.

Intertemporal exclusion restrictions

A further complication arises from the panel data structure and the timing of mergers. The above discussion is based on the assumption that, in any one merger, none of the BU observations is affected by other mergers, which is required by the CL-SUTVA. As a result, in case two mergers affect the same markets at the same point in time, the causal effect of neither of them can be identified. The extent of this problem depends on fundamentally untestable assumptions about the number of time periods that have to pass until all effects of a merger are realised. In this study, we assume that in the two post-treatment years all merger effects are realised.[12]

Unlike competitors' observations, for which the dates of treatment are given, control group observations can potentially be picked from any period for which sufficient data is available. To increase the control group pool as well as the matching quality, the control group pool should be constructed from different time periods available in the data. Therefore, in this study outsider observations are considered from all time periods in which mergers occur.[13] Construction of the control group pool is conducted based on the same intertemporal exclusion restrictions.

BUs of firms that at some point in time are in the treatment group because they are engaging in merger activity can at other points in time be included into the control group pool. This follows from the assumption imposed above that treatment effects are realised after, e.g. two years. Such BUs can be used in other time periods if – around those time periods – the estimation window does not include the time periods of their own post-treatment windows "contaminated" with the treatment effect. An example for the use of intertemporal exclusion restrictions in the study can be found in the appendix in Section 5.6.2. Two further intertemporal exclusion restrictions are imposed to ensure that the common trend and common support assumptions needed for the PSM-DiD estimator are not violated. Merger BU observations are excluded (i) if they have gaps in the treatment window because the firm temporarily left the market, (ii) if there is market entry or exit occurring in the treatment window.[14]

---

[12]Thus, in the application the PSM-DiD regressions are based on a one year pre-treatment and two years post-treatment window. These choices are admittedly somewhat arbitrary but reflect our pragmatic judgement of what is sufficient to capture the non-instantaneous effects of a merger, e.g. efficiency improvements, whilst confining the post-treatment period to a relative short time scale to minimise the chance of confounding factors to impact the results. This choice is in line with many *ex post* merger studies: Mariuzzo and Ormosi (2016) report that in their sample of 55 *ex post* estimates contained in 37 published merger studies, 56% of the estimates are limited to two years and 87% to three years after the merger.

[13]The mergers are roughly spread evenly across the available time periods. Considering outsiders at different points in time improves matching quality as outlined above. At the same time, determining the selection of periods from which the outsider observations are chosen from the time periods at which the mergers occurred increases the likelihood that the common trend assumption necessary for the DiD estimator holds.

[14]Note that overlapping observations of BUs involved in mergers in a market (after the *de minimis*

5.3.4 Model specification

5.3.4.1 Estimation of propensity scores

PSM-DiD estimators can be applied to the sample once the BUs of firms engaged in mergers, their competitors, as well as their outsiders are identified and problematic observations are removed based on the exclusion restrictions. In a first step, the propensity scores, which are used to construct the weights in the PSM-DiD estimator, need to be estimated.[15] The propensity scores usually are predicted based on running a logit model that estimates a BU's probability to be subject to a merger. As the unit of observation features observations of individual firms' market presence in a specific market (i.e. the BU), the data usually contains multiple observations for each firm (one observation for each market in which the firm is active).

The dependent variable in the logit model is a binary variable that takes the value 1 for each BU observation of a firm subject to a merger in that period and 0 for all BUs of firms not engaged in mergers at that point in time.[16] To deal with the time dimensionality in the panel data and to apply the intertemporal exclusion restrictions, the dependent variable is set to missing for all time periods without mergers, such that these time periods are not included in the logit estimation. We pool all time periods that are considered for the logit model (i.e. those in which mergers occur) for the estimation of the logit model.[17] Observations that need to be excluded because of any of the cross-sectional or intertemporal exclusion restrictions introduced in Section 5.3.3.2 are not used in the estimation of the logit model.

Thoemmes and West (2011) provide a discussion of different approaches to estimate

---

rule has been applied) are dropped before they are checked for any form of incomplete data. This ensures that the CL-SUTVA is not violated by merger BU observations that overlap with other observations of BUs involved in mergers which are not used because of incomplete data.

[15]Inverse probability weighting (IPW) represents an alternative way to use propensity scores to address self-selection, which has recently been extended to be used with DiD estimators (Stuart *et al.*, 2014). However, we do not use it here for two reasons. First and foremost, while being more efficient than matching, it is subject to larger levels of bias if the propensity scores are not specified correctly (see, e.g. Zhao, 2008; Rubin, 2004). Second, it is problematic to use in the context of panel data with treatments that occur at different points in time because it is based on the concept of including all available data into the analysis of the outcome variable. However, because of the intertemporal exclusion restrictions that need to be imposed and the common trend assumption in the DiD regression, inclusion of control group observations should be limited to the few best control group observations and to the periods of the pre- and post-treatment window used in the DiD regression.

[16]As such, the dependent variable for a firm that merges in one period takes the value 1 in that period for all its BUs and the value 0 in the other included periods in which it is not merging. Thus, the group of outsiders/the potential control group includes observations of firms that engage in mergers at other points in time, provided that these observations are not contaminated by treatment effects as outlined at the end of Section 5.3.3.2

[17]The time periods in which mergers occur are spread roughly equal across the data set. This approach allows to prevent ambiguities and arbitrariness in the application of the intertemporal exclusion restrictions.

propensity scores for clustered data in the context of sociological research. They show that the specification to estimate propensity scores fundamentally depends on the type of RCT that is approximated in observational studies. Further, the optimal specification depends on whether matching is to be performed within or across clusters. In case of CRTs, matching has to be performed across clusters (Leon *et al.*, 2013; Leyrat *et al.*, 2013). This follows from the fact that all units in a treated cluster are receiving the treatment, such that no control group observation are available within it. For matching across clusters in CRTs, Thoemmes and West (2011) suggest to estimate a logit model to obtain propensity scores – which expressed as a generalised linear model using the logit function as the link function – takes the (linear) form

$$logit\left(\pi_{ikt}(X,Z)\right) = \alpha + \sum_{a=1}^{A} \beta_a X_{aikt} + \sum_{b=1}^{B} \gamma_b Z_{bkt} + \sum_{\tau=t+1}^{C} \theta_\tau, \tag{5.3}$$

where $\pi_{ikt}(X,Z)$ is a BU's probability to be affected by a merger depending on firm-market variables X and overall firm-level characteristics Z, $logit\left(\pi_{ikt}(X,Z)\right)$ is the logit of probability $\pi_{ikt}(X,Z)$, and θ_τ denotes time-dummies controlling for period fixed effects. To obtain $\pi_{ikt}(X,Z)$, the dependent variable in the logit regression is a variable that takes the value 1 for BUs affected by mergers in the periods, in which these mergers occur and is 0 otherwise. Adding θ_τ to the specification controls for general differences in the probability to engage in mergers across different time periods that might arise from, e.g. the business cycle. $\sum_{a=1}^{A} \beta_a X_{aikt}$ denotes the sum of regression coefficients β_a and covariates X_{aikt}, where $a, ..., A$ represents the number of BU-specific covariates included in the specification. Similarly, $\sum_{b=1}^{B} \gamma_b Z_{bkt}$ denotes the sum of regression coefficients γ_b and covariates Z_{bkt}, where $b, ..., B$ depends on the number of overall firm-level-specific covariates included in the specification.

The specification proposed in Equation 5.3 is a direct product of the CL-SUTVA introduced in Equation 5.1 to allow for causal inference in CRTs. It is aimed at approximating the modified conditional independence assumption belonging to the CL-SUTVA, which takes the form of $Y_{ikdt} \perp D_{kt}|X_{ikt}, Z_{kt}$ if the time dimension is included and in which Y denotes the outcome variables used in the main analysis. As such, not only BU-level variables X that determine both the probability that the BU observation is subject to a merger and the outcome variable (in our case prices) over time need to be included, but so also do overall firm-level variables Z that affect both stages.

In determining the specification of the propensity score, it is important to keep in mind that in particular in small datasets a good specification is not a model that includes all variables that determine treatment assignment but one that includes only those

variables that are strongly correlated both with treatment assignment and the outcome variable (the price). In other words, an inconsistent model to predict the propensity score is often preferred to a fully specified and consistent model. This follows from the fact that the propensity score in matching is merely a tool to achieve covariate balance between a specific set of variables.[18] A complete and consistent estimation of the treatment assignment in turn is not the objective (see, e.g. Clarke *et al.*, 2015; Wyss *et al.*, 2013; Brookhart *et al.*, 2006; Augurzky and Schmidt, 2001). As such, the validity of the estimated propensity scores is quite robust to the misspecification of the error distribution in the logit model. An exception to this is endogeneity bias resulting from a correlation between unobserved or excluded variables and included variables that are correlated with the treatment assignment as well. However, Augurzky and Schmidt (2001) show that including higher order terms of or interaction effects between the included explanatory variables can substantially alleviate this problem.[19]

5.3.4.2 Conducting the matching

After Equation 5.3 has been estimated, it is used to predict the probability for each BU observation to be affected by mergers. This probability constitutes the propensity score, which is then used to match the BUs affected by mergers to those unaffected by mergers from the control group pool to obtain the matched sample. The matching algorithm determines a weight for each observation. Depending on the matching algorithm that is used, a treated observation is matched to several observations, and observations that are closer and more comparable to the treated unit receive higher weights than those observations not that comparable to the treated unit. Observations that cannot be matched are not considered in the outcome analysis that follows.

There are two possible strategies to carry out matching given the panel dimension of the data. Very restrictively, one could match a treated observation only to control group observations originating from the same point in time but from an unaffected market, e.g. by determining the best match for a treated observation. If all pharmaceutical mergers are affected by the same time-variant unobservable shocks, such an approach ensures that the common trend assumption required by the DiD model as outlined below is fulfilled. Yet, such a strict matching requirement negatively affects the matching quality, as it significantly reduces the size of the control group pool. This creates several problems. First, this renders it harder to remove the self-selection bias, as the reduced number of potential control group observations increases the difficulty of achieving

---

[18]In our case these need to fulfil the modified conditional independence assumption in Eq. 5.1 and are the BU-level variables X and overall firm-level variables Z that correlate both with the probability to receive the treatment and the outcome measure (the price).

[19]In the application, as a safeguard against potential endogeneity in the logit model we follow this approach and include higher order terms for the main explanatory variables.

common support. Second, some treated observations might have to be dropped because no suitable control group observations are available in the same period. In case the treatment effect varies across units with different propensity scores – which is likely to be the case in most applications – this creates the risk of obtaining very specific treatment effects only valid for particular sub-groups in the sample. Third, a reduced matching quality directly threatens to violate the common trend assumption, as it is not only driven by time-variant unobservables, but also by time-variant observables that are used to carry out the matching.

Therefore, we do not restrict matching to simultaneous observations, although we do apply such a restriction as a robustness check only. As reported in Section 5.4.5 below, this does not affect the qualitative findings in our application.[20] An example for this approach is provided in Figure 5.2. Consider first the left part of the Figure. Assume that we seek to determine the best match for a treated observation T, which originates from 2008. Further, assume that the control group pool consists of the three control group observations C1, C2, and C3 originating from 2004, 2008, and 2011, respectively. Crosses represent the point of time from which the observations are originating, and the horizontal lines show the price paths a year before to two years after the considered matching periods. Suppose control group observation C1 is the best match for treated observation T (as can be seen from the closest match in a common trend pre-treatment), and is matched to T accordingly.

Figure 5.2: Absolute and relative treatment times
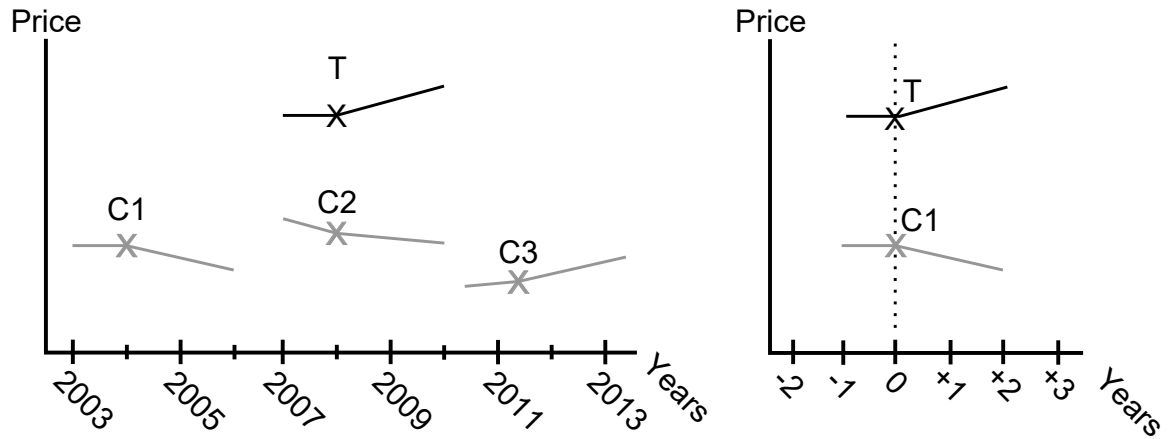
The outcome analysis is based on relative time rather than absolute time, as depicted in the right part of the Figure. Although originating from different years, treated and control units are considered to be from time period 0 in the DiD outcome regression, which is the time of the treatment, after which the treated and control group prices differ in post-treatment years 1 and 2 because of the effect of the merger. Matching is

[20]However, limiting the matching to control group observations significantly increases the mean bias in the estimates (see Tables 5.6 and 5.29), i.e. it reduces the matching quality.

115

successful if it satisfies the common support assumption as well as generates matches that fulfil the common trend assumption in the (relative) year preceding the treatment. The common trend assumption might be violated if, e.g. macroeconomic shocks affect the observations originating from different time periods differently. To ensure that the intertemporal matching strategy does not lead the common trend assumption to fail, formal tests need to be conducted to validate that this is indeed not the case.[21]

5.3.4.3   Outcome analysis

After matching has been conducted, a regression-based outcome analysis is used on the matched sample. For this purpose, the sample weights created by the matching algorithm are used in the estimation of the weighted DiD regression. The PSM-DiD estimator has been proposed to combine the strengths of both the DiD and matching methods (Heckman *et al.*, 1998, 1997). Smith and Todd (2005) show that the PSM-DiD estimator allows to remove the bias induced by time-invariant observables that matching estimators in isolation cannot eliminate. In other words, the PSM-DiD estimator corrects self-selection bias both originating from time-variant (and time-invariant) observables as well as from time-invariant unobservables, such that only time-variant unobservables cause bias. As such, it critically relaxes the assumptions of both the DiD estimator as well as the matching methods. This is a particularly useful feature in the presence of the stronger requirements towards observable variables imposed by the modified conditional independence assumption in Equation 5.1.

The DiD model compares the difference in the outcome variable between the treatment and control groups after the mergers have taken place with the same difference before the events. The difference in these differences constitutes the effect of the mergers on the observed outcome. Potential endogeneity due to self-selection in this specification is overcome by using the weights derived from PSM. The outcome variable P_{ikt} denotes the price index of all products offered by a firm's BU i in market k at time t. Abstracting from the weighting for simplicity, the DiD model can be estimated based on a fixed effects regression of the form

$$P_{ikt} = \alpha + \delta_1 G_{kt} + \delta_2 D_{kt} + \sum_{\tau=t+1}^{T} \varphi_\tau U_\tau + \epsilon_{ikt}, \qquad (5.4)$$

where G_{kt} denote firms in the treatment group, $\sum_{\tau=t+1}^{T} U_\tau$ denotes time dummies for periods $t+1$ to T with corresponding coefficients φ_τ, the treatment effect is measured by $D_{kt} = G_{kt} \times \top_{kt}$, and $\top_{kt}$ denotes an indicator variable that takes the value 1 for the post-treatment periods.[22] This is an adaptation of the multi-period DiD model allowing

---

[21]In the application, we employ the common trend assumption test suggested by Autor (2003).

[22]Note that treatment group indicator G_{kt} and treatment time indicator $\top_{kt}$ feature both time and

for different treatment times (see, e.g. Wooldridge, 2013) to the multi-level dimension of the data. In this framework, the treatment effect is measured by the coefficient δ_2. Of course, alternative specifications that assume time-variant or heterogeneous treatment effects can be applied in the framework as well.

Matching creates treatment and control groups that are very similar with respect to their BU-level variables X and overall firm-level variables Z. As a result, variables used in the matching procedure have little to no explanatory power in the propensity score-weighted DiD regression. Therefore, in order to reduce standard errors in the DiD regression, the variables used to estimate the propensity scores are not included in the DiD regression. While this does not exclude the inclusion of other time-variant variables in the DiD regression, these variables are not included here for simplicity, because no such inclusions are made in the application below.

5.4 Application

5.4.1 The merger control framework of the European Commission in the pharmaceutical industry

Using the above empirical framework, we analyse the EC merger control case practice in the pharmaceutical industry. For this purpose, an *ex post* evaluation of the price effects of 11 mergers in the United Kingdom that were cleared by the EC is carried out. As the pharmaceutical industry is characterised by a high capital intensity and a large expenditure on R&D, it holds much scope for spillovers to arise. While the focus on the British market does not allow us to estimate the price effects in other countries, the framework allows us to capture price effects in the British market that result from spillovers originating from other countries.[23]

In order to focus resources in the assessment of multi-market mergers in the pharmaceutical industry and cope with the large number of affected markets, the EC has developed a case practice that limits the number of markets that are considered in-depth in the merger control procedure. Pharmaceutical markets are usually national in scope, as prices are regulated at a national level. Product market definition in the industry commonly is based on the Anatomical Therapeutic Chemical (ATC) classification system. This employs a code based on 5 levels: level 1 denotes the anatomical main group

---

firm indices k and t, as treated units can at other times serve as control units, and the post-treatment time for a (control group) observation depends on the time of the merger (that the control group observation is matched to).

[23]This follows from the CL-SUTVA that yields the total treatment effect. We subsequently assume that these inter-country spillovers do not lead to systematic positive or negative price effects in the British market such that we can generalise the treatment effects.

(such as N for nervous system), level 2 the therapeutic usage (e.g. N01 for anaesthetics treating the nervous system), level 3 the pharmacological sub-group (e.g. N01 B local anaesthetics), level 4 the broad chemical sub-group within the pharmacological sub-group, and level 5 the chemical substance of the main active ingredient. In many cases, the EC defines a market to consist of all pharmaceuticals offered in an ATC3 class, but sometimes a definition based on the ATC4 classes or a combination of ATC3 or ATC4 classes is more appropriate. The EC's case practice is characterised by requiring the involved parties to report their market shares in all relevant markets with an overlap of business activity together with an own assessment with respect to the geographical and (ATC-based) product market definition. The market definition is then established by comparing the involved parties' definition with those of competitors and customers.

The Commission sorts the markets into three groups, in which potential anticompetitive practices might arise. This grouping is based on the joint market shares of the involved parties (see, e.g European Commission, 2011a) and is labelled as

- **Group 1** if they are above 35% and increase through the merger by more than 1%,

- **Group 2** if they are above 35% and increases through the merger by less than 1%,

- **Group 3** if they are between 15% and 35%.

As a further simplification of the procedure, the EC does not always consider all markets that fall in any of the three groups in the in-depth assessment. For example, if the involved parties' "*activities overlap and their joint market shares do not exceed 35% under any plausible market definition and/or where the increment is below 1%, competition concerns may be excluded.*" (European Commission, 2011a, p.4). While the groups that are considered vary between the cases, in line with the previous example the exclusive focus in many of them is on Group 1 markets (see, e.g European Commission, 2011b).

Such a merger control practice relies on several implicit assumptions. First and foremost, it rules out potentially significant spillover effects of mergers between markets with and without an overlapping market presence of the parties prior to the transaction. Second, it assumes that anticompetitive effects are unlikely to arise unless the joint market shares of the parties involved in the mergers exceed a threshold (in particular when only Group 1 is considered in an investigation). This focused market investigation limits the analysis to the markets potentially being affected the most by a merger, but risks missing out anticompetitive effects in non-overlapping markets or those, in which the parties have a low market share. This study seeks to determine whether this merger

control practice is effective.

5.4.2 Analysed markets and the data

5.4.2.1 Analysed markets

The empirical analysis focuses on pharmaceutical markets and prices in the United Kingdom. Pharmaceutical prices in the United Kingdom are regulated in the Pharmaceutical Price Regulation Scheme (PPRS), which is negotiated approximately every five years between the National Health Service (NHS) and the industry body ABPI. Prices are regulated based on a rate of return cap that is set in the PPRS, and pharmaceutical companies can freely set prices for newly introduced branded drugs. Price increases for existing drugs are possible, but need to be justified and approved by the Department of Health, whereas price reductions do not require approval (Vogler *et al.*, 2009; Maynard and Bloor, 2003).

The necessary data for the analysis is derived from the British Pharmaceutical Index (BPI) provided by the Intercontinental Marketing Service (IMS).[24] The data contains monthly information on sales and prices at the package level by manufacturer and markets as defined by the ATC classification for all drugs sold in the United Kingdom between April 2003 and March 2013. Apart from sales and prices, the data further includes information on product names, main active molecules, strength and form, as well as information whether the product is branded or a generic. For the analysis, the data is aggregated to quarterly observations at the BU-level, i.e. all products and packages sold by a manufacturer in a quarter are summed up to a single observation. Aggregation to quarterly data is conducted for two reasons. First, this takes into account that sales are stable for most products, such that little additional information is gained at the monthly level. Second, for a small minority of products, unsteady sales including arbitrary fluctuations and temporary exit might lead to misleading inference in the empirical analysis. Thus, the aggregation removes noise in the data that might otherwise undermine the identification of effects in the data.

In the application, all medicines that belong to an ATC3 class constitute a market. This market definition results in some markets to be wider than defined by the EC, which establishes a market to consist in some cases not of all products in an ATC3

---

[24]The focus on mergers assessed at the European level potentially misses out mergers checked by national competition authorities. However, this is unproblematic for this analysis. First, there were no mergers in the United Kingdom in the analysed time periods that could significantly affect market prices. Second, smaller mergers cleared by other European competition authorities are unlikely to have a substantive impact on prices in the United Kingdom. IMS allocates sales of a small number of manufacturer-level joint ventures equally among the international corporations that own them. This hardly affects the analysis, as data aggregation is conducted at the international corporation level.

class, but of some of the ATC4 levels within an ATC3 class.[25] This approximate market definition is necessitated by the fact that market definitions by the EC only exist for a fraction of the markets in the data, all of which are subject to mergers. Thus, relying on these market definitions results in a sample, in which no proper control group can be constructed, as most control group observations would have to be excluded because of the exclusion restrictions introduced above. The approximate market definition used here in turn allows us to significantly extend the control group pool in the analysis. As our market definition has a tendency to be wider than that of the EC, the estimated treatment effects represent a lower bound for the effects that can be expected for the affected markets.

Sales of non-branded products by generic producers are not separable by manufacturer in the data, such that generic products are excluded from the analysis. Pooling the sales of all non-branded products would imply joint profit maximisation by the generic producers within and across markets. To prevent potential bias and distortions in market power by these firms, generic producers are excluded from the analysis but are considered in the calculation of market shares of other firms and of the overall sales in a market.

Excessive fluctuations in the dependent variable (prices) in a small number of outliers threaten the common trend assumption required by the DiD estimator. They further risk to induce inflated standard errors from potential measurement errors as well as spurious results. To tackle this issue, a firm's observations in a market are excluded if the associated price index features a coefficient of variation that is in the top 10% range of the data, i.e. it features a coefficient of variation greater than or equal to 0.513.[26] After aggregation and data cleaning, the data contains 35,607 quarterly observations of 213 firms active in 242 product markets at defined by ATC3 classification codes ranging from the second quarter of 2003 to the first quarter of 2013.

All mergers considered in this study were assessed by the EC between 2003 and 2013. Case selection is based on mergers, in which both involved firms are pharmaceutical companies and the mergers are classified according to NACE classification code C21 "Manufacture of basic pharmaceutical products and pharmaceutical preparations" or its related sub-groups. Acquisitions of pharmaceutical firms by non-pharmaceutical

---

[25]Some products with multiple applications appear in several ATC3 classes, such that in few cases a market definition based on the ATC3 class might be to narrow as well. However, this is less common than markets that consist of some of the ATC4 levels within an ATC3 class. As such, potential biases resulting from these cases in our market definition should be small in scale and scope.

[26]In the data set, some BU observations limited to a few small firms feature very large price fluctuations over time with the highest case having a coefficient of variation of 3.843. Excluding these observations with excessive fluctuations – which in most cases are likely to be caused by artefacts in the data – makes it more likely that the common trend assumption holds and increases the power of the PSM-DiD estimator.

firms are not considered: these cases are neither changing the market structure and competition nor the R&D activities other than through potential changes in financial strength and investment behaviour of a company, which in these cases are not likely to arise. Further, acquisitions are not included if they only affect few product markets, e.g. when the licence of a single branded product is sold from one pharmaceutical company to another.[27] Furthermore, both firms had to be active in markets in the United Kingdom at the time of the merger, such that changes in market power and competition due to the event can be expected in the market.[28] A list of the 11 merger cases that are considered in the study can be found in Table 5.12 in the appendix.

5.4.2.2 Descriptive statistics

Table 5.3 contains descriptive statistics for key variables used in the analysis. *Avg. price* is the unweighted average price of all products sold by a firm's BU in an ATC3 market. *HHI* denotes the ATC3 market's Herfindahl-Hirschman Index. *Share in market* measures the market share as defined by the relative sales revenue of a firm's BU compared to all sales in the market, and *Share generics* represents the joint market share of all generic producers in the market. *No. of active markets* counts the number of ATC3 product markets, in which a firm is active in the United Kingdom. *BU's no. of* unique packages summed up across all products and forms (liquid, tablet, gels, etc.) by a firm's BU in a market and the firm in total across all markets, respectively. *No. of uniq. product form packs in market* contains the same information but aggregated across all firms' sales in an ATC3 market. *No. of uniq. molecules in market* counts the number of unique molecules (active ingredients) offered in an ATC3 market.

The first two numeric columns with the title *Whole sample* in Table 5.3 show the mean and standard deviation for the whole sample of BUs that can be used after data cleaning. Columns 4 and 5 named *Treatment obs.* contain the same statistics for the treated observations that are used in the analysis, and Columns 6 and 7 labelled *Other obs.* show the same statistics for the BU observations that are not subject to any measurable merger in the data.[29] The last two columns named *Diff.* contain the results of two-sample t-tests with unequal variance, with which the observations of the treatment observations and all other observations are compared and tested for equality.

---

[27]Note that such mergers would be excluded because of the exclusion restrictions anyway, as the target's sales are likely to be below 10% of the total sales of the joint entity.

[28]As a result, the study does not analyse spillover effects arising from overlaps in other, geographically separate markets on firms in the British market if only one of the involved firms is active in the United Kingdom.

[29]Table 5.3 is based on the treatment and control group observations that according to kernel matching are on support and can be considered in the analysis. As 74 treated BUs can be matched to control group observations and are included with a three year window of quarterly observations, this yields (up to) 888 observations in the PSM-DiD regressions.

Significant differences between the two samples provide an indication of a non-random selection into the treatment group, i.e. of self-selection.

Table 5.3: Descriptive statistics

	Whole sample		Treatment obs.		Other obs.		Diff.	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean diff.	T-Stat.
Avg. price	7.311	33.708	8.867	42.887	7.271	33.440	-1.60	(-1.100)
HHI	0.443	0.239	0.486	0.247	0.442	0.238	-0.04***	(-5.315)
Share in market	0.166	0.281	0.318	0.326	0.162	0.279	-0.16***	(-14.137)
Share generics	0.241	0.266	0.174	0.233	0.242	0.266	0.07***	(8.556)
Total sales in market (Mill. £)	18.973	30.179	12.249	20.467	19.145	30.367	6.90***	(9.768)
Firm's sales in market (Mill. £)	1.337	5.054	2.426	5.322	1.309	5.044	-1.12***	(-6.180)
No. of active markets	25.903	25.029	58.644	17.275	25.066	24.632	-33.58***	(-56.471)
BU's Num. of uniq. product-form-packs	4.338	5.411	6.325	6.176	4.287	5.381	-2.04***	(-9.741)
Av. no. of product form packs	4.353	2.231	5.821	1.327	4.316	2.237	-1.51***	(-31.697)
Num. of uniq. product-form-packs in market	62.373	67.469	41.027	49.381	62.919	67.781	21.89***	(12.904)
Num of unique molecules by drug-class	9.559	6.826	7.401	5.619	9.614	6.845	2.21***	(11.519)
Observations	35607		888		34719		35607	

Notes: Sample comparisons are based on the two-sample t-test with unequal variance. * p<0.1, ** p<0.05, *** p<0.01.

The results of the t-tests provide strong evidence that self-selection is present with respect to all explanatory variables. BUs of firms engaged in mergers, on average, are active in more concentrated markets. These markets are smaller in size as measured by sales than markets without mergers, and the involved parties' BUs, on average, have higher sales than their non-treated counterparts. The most important difference between the two samples relates to the market presence of the firms that are involved in mergers: whereas they are active in, on average, about 58 markets, firms not affected by mergers are active in 25 markets only. This suggests that firms engaged in mergers tend to be larger multi-market firms than their non-treated counterparts. As these firms are more diversified with respect to geographic and product coverage, they are likely to have stronger financial capabilities and R&D activities and are therefore more likely to be affected by spillovers. Further, Table 5.3 indicates that mergers tend to occur in markets with more limited product varieties (and thus fewer substitutes available) and fewer molecules (active ingredients).

As outlined in Section 5.3.3.2, not all BU observations affected by mergers can be studied due to the exclusion restrictions that need to be applied to guarantee that the causal effect of the treatment can be identified. The effect of these exclusion restrictions on the sub-sample that can be studied is shown in Table 5.13 in the appendix in Section 5.6.4. Of the 423 treated BUs that are identified in the data, only 81 can potentially be used in the analysis (if matching is possible – with kernel matching we are only able to match 74 treated BUs), and 342 need to be excluded. Two-sample t-tests between included and excluded cases indicate that the included cases are not a random draw from the BU observations affected by mergers. Therefore, while the strict exclusion restrictions strengthen the internal validity of the study, they limit the external validity

of the results, such that conclusions need to be drawn accordingly. That said, these findings highlight that a careful construction of the treatment and control groups to satisfy the CL-SUTVA is an important and non-trivial exercise in such clustered data.

5.4.3 Propensity score matching

Two different matching procedures are used to carry out PSM: kernel matching and nearest-neighbour matching. Kernel matching matches each treated unit to all available control group BU observations in a bandwidth of 0.05 or less, i.e. to control group BUs which are sufficiently close to a treated unit with respect to the difference in propensity scores.[30] Nearest-neighbour matching matches each treated unit to (up to) the 5 closest matches from the control group pool with replacement, provided that they lie within a caliper of 0.1, i.e. the difference in the estimated propensity scores between a treated and the control group observations is below 0.1. The latter procedure often provides a better reduction in self-selection bias, as it only matches each unit with a limited number of the closest counterfactual observations available with respect to the propensity score.[31] This might come at the cost of lower efficiency compared to kernel matching, which uses more information, but tends to yield a lower reduction in the bias in many applications (Caliendo and Kopeinig, 2008). Given the advantages of each method, they are complementary and should provide similar results if the matching and empirical results are robust. In both cases, matching is conducted with replacement to reduce the number of treated observations that cannot be matched as they are off support, which is primarily a problem for treated observations with high propensity scores in the analysis. This allows to include those treated observations into the analysis, which are likely to be affected the most by mergers. The results based on kernel matching are presented below, and those based on nearest-neighbour matching can be found in the appendix in Section 5.6.4. The results based on kernel matching are our preferred results, as matching tests show that it achieves the lowest levels of bias (as is evident from a comparison of Tables 5.6 and 5.29) and represents the more efficient method.

Before matching can be carried out, the propensity scores are estimated using a logit model as laid out in Equation 5.3. The dependent variable is an indicator variable that takes the value 1 if a BU observation belongs to a firm that is involved in a merger, and is not excluded from the sample because of the exclusion restrictions laid out in

---

[30]Using these algorithms, 1251 of the 1251 and 74 of the 81 BU observations in the control and treated groups are on support for kernel matching. For nearest-neighbour matching, the respective numbers are 1251 of 1251 and 80 of 81 BU observations, respectively.

[31]As stated further below, in our results using nearest-neighbour matching instead results in a higher bias, potentially because of a large caliper compared to the bandwidth specified for kernel matching. We chose the parameters to obtain roughly the same number of treated BUs that can be matched in the analysis.

Section 5.3.3.2. Similarly, control groups are constructed according to the procedure described in Section 5.3.3.2: outsider firm BU observations from markets unaffected by mergers are included in the logit estimation with the dependent variable taking the value of 0. Independent variables shall include all variables that both determine a BU's propensity to be affected by a merger and market prices, which serve as the dependent variable in the outcome analysis based on the PSM-DiD estimator. Variables that only affect the BU's propensity to be subject to a merger or only the prices are not included, as they potentially reduce matching quality, render fulfilment of the common support assumption harder, and inflate standard errors in the PSM-DiD estimator. As is common in the literature, the propensity to receive treatment is calculated using independent variable values from the period prior to the treatment, i.e. the first lags of independent variables are included in the logit model (see, e.g. Caliendo and Kopeinig, 2008).

The chosen specification for the logit model and the results can be found in Table 5.4. A parsimonious specification is chosen to prevent biased propensity scores to arise from overspecification (Zhao, 2008). Column I contains variable names and indicates the logit specification, Column II displays the estimated coefficients, and Column III contains the corresponding average marginal effects. In line with suggestions in the literature (see, e.g. Caliendo and Kopeinig, 2008; Austin *et al.*, 2007), inclusion of independent variables is limited to those that should affect both a BU's propensity to be affected by a merger as well as the market prices. Further, we include squared terms of the main explanatory variables to reduce the impact of potential endogeneity bias on the estimated propensity scores (Augurzky and Schmidt, 2001).[32]

With respect to market-level-specific variables, *HHI* is included as a measure of market power, as a higher concentration should yield higher market prices. It further increases the profitability of acquiring a BU in the market through its positive correlation with markups. A BU's market share, measured by *Share in market* and *Share in market sq.*, is included both in a simple and squared form to allow for a non-linear inverted U-shaped effect: this captures the tendency of larger firms to complement their market presence with innovative products of smaller specialist producers. In addition, higher shares correlate with higher market power and therefore determine market prices. Yet, the competition authorities' practice to block mergers based on joint market shares that are deemed too large should provide a soft threshold for this effect. The presence of generics, which is controlled for with *Generic prod. market share*, limits the price-cost

---

[32]Augurzky and Schmidt (2001) show that the inclusion of higher order terms of the explanatory variables approximates a Taylor expansion of the specified function that can capture most of the deterministic effects of omitted variables that correlate with the included variables. As a result, the effect of omitted variables on the errors term is mostly limited to stochastic noise.

margin that producers can set for branded products. Therefore, the market share of generics might not only influence prices, but also the profitability of acquiring a target or asset operating in the market.

In addition to the market-level variables outlined above, two overall firm-level variables are included as well. As many of the mergers are undertaken by large multi-national, multi-product firms and spillover effects on prices partly depend on their multi-market presence, information on multi-market activity is essential to obtain valid propensity scores satisfying the CL-SUTVA. It is measured with the variable *No. of active markets* and *No. of active markets sq.*, which count the number of ATC3 markets, in which a firm was active prior to the treatment period. While a positive effect of a multi-market presence on a BU's propensity to be affected by mergers thus can be expected, the effect likely is not linear, as is the case for market shares: a substantial market presence reduces the scope for a merger, as its unconditional clearance, on average, becomes less likely. Multi-market firms are also more likely to charge higher prices, as they can profit from, e.g. portfolio effects in distribution and higher marketing expenditure. *Av. no. of product form packs* is another firm-level variable that is included and controls for portfolio effects as well as product diversification strategies that might increase the value of a BU rendering it more likely to be acquired in a transaction or be affected by a merger. Yet, its effect is unclear *ex ante*, as a negative effect of this variable is equally plausible if one assumes that mergers and acquisitions are predominantly used to complement product portfolios.

The results of the logit model are mostly in line with expectations. The HHI is found to have a positive effect on a BU's probability to be affected by a merger. Similarly, the BU's market share features a strongly positive but diminishing effect with a turning point at a market share of approx. 49%. As expected, the coefficient of the market shares of generics is negative. Multi-market presence has a strong and positive but diminishing effect (with the turning point occurring at approx. 62 markets) on the probability to receive the treatment. We find weak evidence of a negative effect of *Av. no. of product form packs*. The significance of the two overall firm-level variables shows the importance of including these variables into the propensity score estimation in order to satisfy the CL-SUTVA.

The logit model estimated in Table 5.4 is used to predict the propensity scores. In line with suggestions of Heckman and Todd (2009), the logarithm of the odds ratio of the predicted propensity score is subsequently used to carry out matching.[33] Results of the matching, which is carried out using Stata's user-written program *psmatch2* (Leuven

---

[33]Matching on the log odds ratio of the propensity score rather than the score itself ensures that despite unknown sample weights in choice-based sampling – as is the case for this study – the probability of being in the treated group can consistently be estimated (see, e.g. Heckman and Todd, 2009).

Table 5.4: Estimation of propensity scores – Logit model

	Coeff./Std. Dev.	ME/Std. Err.
HHI	1.022*	0.045*
	(0.600)	(0.027)
Share in market	7.106***	0.113***
	(1.289)	(0.041)
Share in market sq.	-7.309***	
	(0.909)	
Generic prod. market share	-1.468*	-0.065**
	(0.754)	(0.030)
No. of active markets	0.373**	0.002**
	(0.153)	(0.001)
No. of active markets sq.	-0.003**	
	(0.001)	
Av. no. of product form packs	-0.600	-0.026*
	(0.409)	(0.016)
Constant	-9.797***	
	(2.924)	
Observations	1332	1332
Pseudo R^2	0.349	

*Notes: Independent variables are included with their value of the last period before the treatment to ensure exogeneity. Column II reports coefficients, and Column III reports average marginal effects. Cluster–robust standard deviations are reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.*

and Sianesi, 2015), are presented for the kernel matching algorithm in Tables 5.5 and 5.6, and for the nearest-neighbour matching algorithm in Tables 5.18 and 5.19 in the appendix in Section 5.6.4.

A comparison of the distribution of the logarithm of the odds ratio pre- and post-matching can be found in Figures 5.6a and 5.6b in the appendix: it shows that treated and control groups have substantially different propensity score distributions indicating the presence of self-selection. Further, it can be seen that matching establishes distributions of propensity scores that are more similar to each other. Table 5.5 shows differences in means for the variables used to calculate the propensity scores both before and after kernel matching. Successful matching removes significant differences in means between the treated and control groups, which can be tested either using a two-sample t-test, or by calculating the variance ratio of the treated over the non-treated as suggested by Austin (2009).[34] Another measure for sample bias is provided by *% bias*,

---

[34]In case of perfect balance, the variance ratio should be one. Values exceeding the 2.5th and 97.5th percentiles of the F-distribution with *number of matched treated observations - 1* and *number of matched treated observations - 1* degrees of freedom offer a rough guideline whether the achievement of balance has failed and are reported at the bottom of the Table 5.5.

which is defined as the % difference of the sample means in the treated and non-treated groups divided by the square root of the average of the sample variances in the groups of treated and non-treated observations. A comparison in the bias pre- and post-matching is provided by *% reduct. |bias|*, which contains the % value of change of the absolute value of the *% bias* achieved by matching. Similar to Table 5.3, a comparison of samples indicates whether self-selection is present. Yet, unlike in Table 5.3, which includes all treated observations selected by the matching algorithm after matching and all non-discarded observations including all time periods in the treatment window, Table 5.5 only contains the means of the variables in the period prior to the merger, which offers a more precise look at self-selection.

Table 5.5: Matching results (kernel matching)

Variable	Unmatched (U) Matched (M)	Mean Treated	Control	% bias	% reduct. |bias|	t-test t	|p>t|	V(T)/ V(C)
HHI	U	0.481	0.481	0.0		0.0	0.998	0.89
	M	0.480	0.475	2.2	-7098.5	0.1	0.890	1.07
Share in market	U	0.333	0.206	40.4		3.5	0.000	0.99
	M	0.327	0.350	-7.3	81.9	-0.4	0.663	1.00
Share in market sq.	U	0.209	0.142	21.1		1.9	0.055	1.19
	M	0.208	0.224	-5.0	76.6	-0.3	0.769	1.15
Share generics	U	0.159	0.231	-28.2		-2.3	0.023	0.67
	M	0.170	0.185	-6.0	78.8	-0.4	0.704	0.89
No. of active markets	U	58.000	28.230	135.2		10.2	0.000	0.45*
	M	58.946	59.605	-3.0	97.8	-0.2	0.824	0.97
No. of active markets sq.	U	3660.7	1466.3	103.8		8.4	0.000	0.70
	M	3789.0	3875.9	-4.1	96	-0.3	0.789	0.98
Av. no. of product form packs	U	5.670	4.522	62.0		4.6	0.000	0.37*
	M	5.781	5.800	-1.0	98.3	-0.09	0.929	1.20

*Notes: * if variance ratio outside [0.64; 1.55] percentiles of the F-distribution for U and [0.64; 1.57] for M.*

Several observations can be made in Table 5.5. First, there is strong evidence for significant differences in the explanatory variables between the treated and control groups and thus for the presence of self-selection. Second, differences in means are much less pronounced after matching has been carried out pointing towards a fulfilment of the covariate balance requirement. The biggest sample differences occur for the firm-level variables *Av. no. of product form packs* and *No. of active markets*, which highlights the importance of controlling for firm-level factors in BU-level merger studies. Third, apart from *HHI*, matching leads to significant reductions of the bias. For *HHI*, the bias increases, but as the post-matching differences are non-significant, this is unproblematic. Fourth, both t-test and variance ratio criteria report that differences in means between treated and control groups are successfully eliminated by matching as well.

The same conclusion with respect to the achievement of covariate balance can be drawn based on five tests of overall covariate balance reported in Table 5.6. Columns *Ps R2*, *LR chi2*, and *p>chi2* report results of a probit estimation of the propensity score on all variables used to estimate the propensity score with the columns featuring the Pseudo R2, the chi2 value of the log likelihood ratio test of joint insignificance of all covariates, as well as the corresponding p-values. As evident from Table 5.6, the variables have no explanatory power post-matching, such that their effect on a BU's propensity to be affected by a merger has successfully been removed using PSM. Columns *Mean Bias* and *Median Bias* contain the mean and median values of the distribution of the absolute values of the bias reported in Table 5.5. Again, matching significantly reduces bias induced by self-selection. *Rubins' B* and *Rubin's R* denote the absolute standardised difference of the means of the linear index of the propensity score of the treated and non-treated groups as well as the ratio of the treated to the non-treated variances of the propensity score index, respectively. Both measures are suggested by Rubin (2001): values of B below 25 and of R between 0.5 and 2 indicate that the two samples are balanced. Both measures report that significant imbalances between the samples are removed successfully using matching. The last column *%Var* contains the percentage of variables with variance ratios exceeding the 2.5th and 97.5th percentiles of the F-distribution reported in Table 5.5. In line with the other tests of overall covariate balance, the test shows that covariate balance has been achieved.

Table 5.6: Matching tests (kernel matching)

Sample	Ps R2	LR chi2	p>chi2	Mean Bias	Median Bias	Rubin's B	Rubin's R	%Var
Unmatched	0.344	209.95	0.000	55.8	40.4	182.3*	0.17*	29
Matched	0.005	1.09	0.993	4.1	4.1	17.1	0.92	0

Notes: * *if B>25% and R outside of [0.5; 2].*

Similar tests of individual and overall covariate balance based on nearest-neighbour matching are reported in Tables 5.18 and 5.19 in the appendix. Most of the tests indicate that covariate balance is achieved by nearest-neighbour matching as well.

5.4.4   PSM-DiD results

Drawing on the matched data obtained from PSM, PSM-DiD estimators can be applied to estimate different ATTs relating to the effects of mergers. For this purpose, the weights obtained using matching are used in the estimation of DiD regressions as presented in Equation 5.4 to obtain the PSM-DiD estimator. All estimates are based on cluster- and autocorrelation-robust standard errors and – in contrast to related studies studying merger effects based on BU-level observations – clustering is done at

the firm-level instead of the BU-level: this explicitly allows for correlations in a firm's BU prices across markets as can be expected due to spillover effects and firm-wide pricing policies.[35] In order to study the research questions non-parametrically, PSM-DiD estimators are not merely estimated based on all of the available sample, but also on different sub-samples. In case sub-samples are analysed, the matching algorithms are re-applied to the identified sub-samples to ensure that correct weights are provided for the control group observations that are selected by the algorithms. These observations can have different weights according to the treated observations sample they are matched with, irrespective of whether replacement as a matching option is allowed or not.

As matching on a set of independent variables creates control groups with similar characteristics in the outcome regression, the explanatory power of those variables in the outcome regression is very limited. Thus, the variables used in the estimation of the propensity score are not included in the PSM-DiD outcome regression (see, e.g. Ornaghi, 2009; Girma and Görg, 2007). For the estimation of the DiD regressions, all variables other than those indicating the treatment effects are partialled out, such that the R^2 and Adjusted R^2 values refer only to the treatment effect variables. Sub-sample analysis to answer different research questions is carried out by re-running the PSM-DiD estimator on the different sub-samples. For this purpose, separate sets of weights obtained from sub-sample-specific matching procedures are used.[36] Note that the observation numbers between the columns and tables vary because of differences in the chosen outcomes, the used matching algorithms and approaches, and the (rules set to create) different sub-samples.

---

[35]Valid standard errors in outcome regressions using PSM is a field that currently receives a lot of research. Many researchers use bootstrap procedures in outcome regressions to take into account that the propensity score used for matching is estimated to prevent downward-biased standard errors. Abadie and Imbens (2008) show that such an approach leads to inconsistent estimates of standard errors and Abadie and Imbens (2006) suggest a consistent estimator for nearest-neighbour matching with a fixed number of matches for non-clustered data as a solution. The extension of this nearest-neighbour estimator with a fixed number of matches to clustered data structures by Hanson and Sunderam (2011) so far has seen little use in the literature. Recent articles by Otsu and Rai (2016) and de Luna *et al.* (2010) propose refined bootstrap procedures for nearest-neighbour matches with a fixed number of matches. The limitation of the approaches to nearest–neighbour matching with a fixed number of matches renders their use problematic, as they cannot be combined with caliper matching or kernel matching to rule out bias from bad matches in case no close control group observations are available. Further, this limitation does not allow for an efficient use of the control group by using more than the nearest-neighbours if feasible. Yet, specifying calipers (or bandwidths) is very important in this study due to the lack of control group observations with large propensity scores (as can be seen in Figure 5.6a), as is also evident from the reduction of the mean bias achieved by kernel matching evident in Table 5.6 compared to nearest-neighbour matching in Table 5.29. Thus, we refrain from using these estimators based on a fixed number of matches as they would either risk increasing bias or require excluding some of the observations which are most likely to feature significant treatment effects.

[36]The calculation of sub-sample-specific weights is necessary, as control group observations might otherwise feature weights that are partly driven by matches to treated observations not present in the sub-sample under consideration.

Common trend assumption

Before turning to the specification and results of the PSM-DiD estimator, it is necessary to assess whether the common trend assumption of the estimator is fulfilled. Only then can it provide reliable estimates of the causal effect of the treatment. Figure 5.3 plots the average price of the treatment and control group BU observations over time for the whole sample of cases that can be analysed after the exclusion restrictions for causal identification are applied. Time period 0 represents the quarter in which the merger occurred. Price trends before the treatment at times appear to be somewhat different, although the differences in absolute terms tend to be small.

Figure 5.3: Difference-in-Differences time trends (kernel matching)
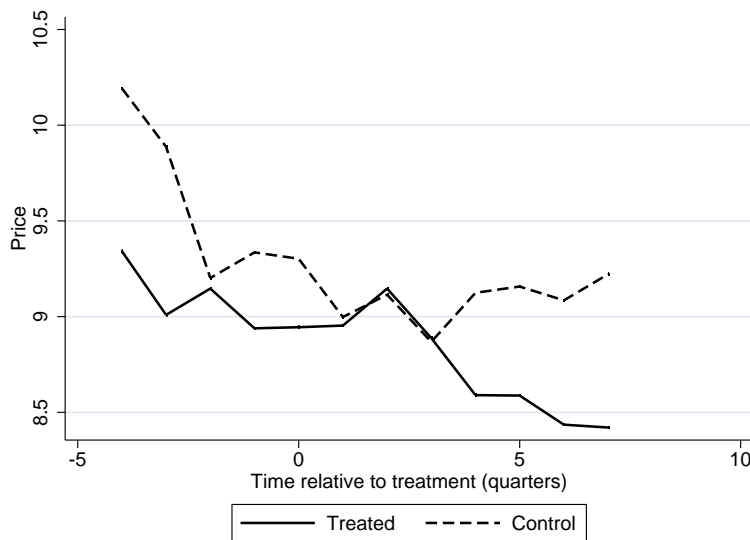
A more formal way to assess the common trend assumption suggested by Autor (2003) is to regress the pre-treatment periods on indicator variables for the periods prior to the treatment that take the value 1 for BUs in the treatment group. These indicator variables are known as leads. They test whether the treatment group is subject to a different trend than the control group prior to the treatment. A lack of significance of the leads indicates that the time trends do not differ between the treated and control groups prior to the mergers and is seen as an indicator that the common trend assumption is satisfied.[37] The sub-sample-based analysis does not only require the common trend to be present in the full sample, but also in the different sub-samples that are used throughout the analysis. The tests show that the common trend assumption appears to be satisfied for the overall sample and for most of the sub-samples.[38]

---

[37]Strictly speaking, the common trend assumption is stronger. It does not only require the trends in the outcome variable of treated and control groups to be the same before the treatment, but also after the treatment, i.e. the treated group would have shown the same trend in the periods after receiving the treatment had it not received it. Yet, this assumption is fundamentally untestable, such that only the proposed indirect way to test the common trend assumption is possible.

[38]The results of the common trend tests for the market shares- and overlap-based sub-samples as

Main results

In the main analysis, the PSM-DID estimator as specified in Equation 5.4 is estimated based on a pooled sample of all mergers that are considered in this essay. As this pooled sample constitutes of a large set of different markets, firms, and business environments, we can study the impact of different factors on the treatment effect. For this purpose, sub-sample analysis is conducted based on two factors. First, it is distinguished between different levels of the (joint) market share of the parties conducting a merger. Second, an overlap-based classification is used that takes into account whether the parties had an overlap in market presence prior to the event. Separating the sample based on different joint market shares allows to analyse whether there are spillovers between markets, in which the parties have a high market share and power, and those, in which their joint market presence is less pronounced. Sub-samples consisting of markets with and without an overlap of the BUs of the parties prior to the mergers in turn allows to study spillovers that exist between markets, which are directly or only indirectly affected by the mergers. Combining the two dimensions to create sub-samples enables us to estimate to what extent these two types of spillovers are interrelated with each other.

The first set of estimates presented in Table 5.7 studies the presence of spillovers across markets with different levels of market shares by the merging parties. *Treatment effect year 1* and *Treatment effect year 2* measure the ATTs in year one and two after the treatment, respectively. The coefficients indicate the relative differences in prices one and two years after the merger compared to the prices in the year before the mergers. Five different sub-samples are analysed: Column II contains the ATT based on all observations that can be used, and Columns III to VI show ATTs for sub-samples according to the joint market share of the involved parties: up to 5%, above 5% to 15%, above 15% to 35%, and above 35%. This sub-sample selection is partly based on the grouping used by the EC used to classify markets with potential anticompetitive effects presented in Section 5.4.1.[39] The validity and representability of the sub-sample regression results depend on two factors. First, the number of treated BUs that make up the treatment group is important. The lower the number of treated BUs, the less representative is the estimated treatment effect. This number is shown in all estimates tables in row "Treated BUs". Second, the validity of the treatment effect depends on whether the common trend assumption is fulfilled for the sub-sample.

well as for combinations of the two can be found in Tables 5.14, 5.15, 5.16, and 5.17 in the appendix in Section 5.6.4, respectively.

[39]The sum of the number of observations in all sub-sample columns does not add up to the number of observations based on the whole sample, as a control group observation can be featured in each of the sub-samples. As such, it is counted several times when the sum of all sub-sample observations is calculated.

Table 5.7: Merger effects by market shares (kernel matching)

	All	<5%	5%-15%	15%-35%	>35%
Treatment effect year 1	-0.579	-0.733	-0.509	0.417	-1.178**
	(0.442)	(0.605)	(0.565)	(0.316)	(0.521)
Treatment effect year 2	-1.049	-0.986	-0.475	1.073**	-1.912**
	(0.681)	(1.656)	(0.664)	(0.515)	(0.906)
Observations	4668	780	2124	1848	2028
Treated BUs	74	7	26	18	23
R^2	0.001	0.010	0.000	0.002	0.030
Adjusted R^2	-0.082	-0.144	-0.101	-0.107	-0.075

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

The results indicate that the mergers significantly reduce market prices of the merging parties in each of the two years after the merger in markets with a joint market share above 35%. On the other hand, their prices increase in the second year after the mergers in markets, in which their market shares lie between above 15% and 35%.

These findings suggest that spillover effects might exist between BUs of the parties in markets with different levels of market shares. Recall that the estimated treatment effect in each case represents the total effect and thus consists of the direct effect that is linked to the market itself and the indirect effect that captures the spillovers. If one conjectures that the treatment effect is linked to market power and consists only of a direct effect that increases with the market share, the observed signs of the coefficients cannot be explained. Alternatively, one could assume that the mergers do not induce any direct effects on prices. In both cases, the results can only be determined by indirect spillover effects. A potential explanation for these results is the use of realised efficiency gains to increase the market power in those markets, in which the parties involved in the transactions have a more limited market presence.

The second source of spillover effects between markets with overlaps in market presence and those without is shown in Table 5.8. Columns II and III contain the results for markets with and without overlaps in market presence of the involved parties. The results suggest an absence of any *per se* effects depending on overlaps in market presence prior to the transactions if all observations are included irrespective of the market shares of the parties involved in the transactions prior to the mergers.

However, in a third set of estimates, the interdependencies of the two sources of spillover effects are analysed. This allows for the identification of the sources of the significant treatment effects in Table 5.7 for the sub-samples of the BUs in which the parties engaged in mergers feature joint market shares of either above 35% or between above

Table 5.8: Merger effects by the presence of overlaps (kernel matching)

	All	Overlap	No overlap
Treatment effect year 1	-0.579	-0.526	-0.661
	(0.442)	(0.656)	(0.492)
Treatment effect year 2	-1.049	-0.161	-1.186
	(0.681)	(1.071)	(0.747)
Observations	4668	900	4452
Treated BUs	74	7	64
R^2	0.001	0.000	0.001
Adjusted R^2	-0.082	-0.140	-0.084

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

15% and 35%. The results for the sub-sample of observations with a joint market share above 35% are reported in Table 5.9, and those for the sub-sample of observations between above 15% and 35% in Table 5.10. As in Table 5.8, Columns II and III contain the results for markets with and without overlaps in market presence of the involved parties, respectively.

Table 5.9: Merger effects by the presence of overlaps for market shares above 35% (kernel matching)

	All	Overlap	No overlap
Treatment effect year 1	-1.178**	-1.471***	-1.044*
	(0.521)	(0.0438)	(0.540)
Treatment effect year 2	-1.912**	1.332***	-1.897**
	(0.906)	(0.0438)	(0.922)
Observations	2028	132	2004
Treated BUs	23	2	21
R^2	0.030	0.032	0.032
Adjusted R^2	-0.075	-0.540	-0.074

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

The results in Table 5.9 provide evidence for significant price reductions in markets, in which only one of the involved parties was active pre-merger with market shares above 35%. However, the effect on markets with an overlapping market presence of above 35% is unclear because this sub-sample only consists of 2 treated BUs.

An analysis of the source of the price increase detected for markets, in which the parties have a joint market share between above 15% and 35%, can be found in Table 5.10. Unlike for non-overlapping markets, in which the party active prior to the merger has a market share above 35%, no significant price reduction is found for markets without an

Table 5.10: Merger effects by the presence of overlaps for market shares between 15% and 35% (kernel matching)

	All	Overlap	No overlap
Treatment effect year 1	0.417	0.547	0.226
	(0.316)	(0.502)	(0.274)
Treatment effect year 2	1.073**	3.496*	0.888
	(0.515)	(2.038)	(0.596)
Observations	1848	420	1452
Treated BUs	18	3	15
R^2	0.002	0.003	0.002
Adjusted R^2	-0.107	-0.203	-0.117

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

overlapping activity if the market share lies between above 15% and 35%. Again, the effect on markets with overlapping market activities is unclear because of an insufficient number of observations in the sub-sample.

Two main results arise from Tables 5.9 and 5.10. First, the presence of significant price effects for the parties' BUs in markets in which the parties have no overlapping market presence points towards the existence of spillover effects between markets with and without overlapping market activities by the parties. As the mergers do not lead to a direct change of the market power of the BUs in these markets, the treatment effect can only be driven by the indirect effect that captures spillovers.

Second, they provide further indications with respect to the presence of spillover effects between BUs of the involved parties in markets with different market shares. As the results of significant price effects in markets with market shares above 35% due to sample composition are primarily driven by non-overlapping markets that are determined by indirect effects consisting of spillover effects, the coefficient patterns observed in Table 5.7 are likely to be driven by these same indirect effects as well.

Taken together, the results provide evidence for the existence of spillovers between the parties' BUs across markets with different market shares and between markets with and without overlapping market activity. Further, both types of spillovers might be interrelated with each other and do not occur in isolation.

5.4.5   Robustness checks

Most of the results are robust to the choice of the matching algorithm. The above analysis is re-estimated based on nearest-neighbour matching with up to 5 matches

within a caliper of 0.1. This matching algorithm mostly passes both the tests of the common support and the common trend assumptions necessary for the DiD model to hold. Exceptions are a failure of the common trend assumption for the sub-sample of treated BUs with a joint market share of up to 5% before the treatment and an indication of a potential violation of the common support assumption in a minority of the common support tests. The results are reported in the appendix in Section 5.6.4 and are mostly in line with those based on kernel matching. Overall, consistent with the lower efficiency of nearest-neighbour matching, they feature fewer significant treatment effects than the results based on kernel matching. The results indicate significant price reductions of the BUs involved in mergers in markets, in which the joint market share exceeds 35%, but suggest that no price increases occur in markets in which the (joint) market shares lie between above 15% and 35%. No price effects are found in the other market share sub-samples or in the overlap-based analysis. Some differences to the results based on kernel matching are present in the analysis of the overlap-based analysis limited to markets, in which the (joint) market shares exceed 35% or lie between above 15% and 35%. In line with those building on kernel matching, the results based on nearest-neighbour matching indicate that price reductions occur in markets without overlaps in market presence, in which the active party has a market share above 35%. For the markets, in which the parties have either overlapping or non-overlapping market shares between above 15% and 35%, no significant price increases are detected.

A second robustness test is conducted with respect to the intertemporal matching of propensity scores laid out in Section 5.3.4.1 and the results are presented in the appendix in Section 5.6.5. As pointed out before, the approach improves the matching quality, but potentially risks undermining the common trend assumption. Therefore, the analysis is conducted again based on the kernel matching estimator limiting matching to control group observations from the same period. However, the tests of the common support and the common trend assumptions report that they might in parts be violated. Further, the mean bias after matching has been conducted is the highest among the three matching approaches conducted in this study (see Table 5.29). Limiting matching to observations from the same period results in a significant reduction in the number of matches. First, more of the treated observations cannot be matched and need to be excluded from the analysis. Second, the number of control group observations that are matched to a treated observation as well as the quality of the matches is reduced. Consequently, the matched control group observations are of lower quality (i.e. the difference in propensity scores between treated and matched control group observations is increased), lead to a reduced efficiency, and produce less representative treatment effects. Nonetheless, the qualitative results are mostly in line with those based on intertemporal matching of the propensity scores.

5.4.6 On the importance of controlling for firm-level variables

Next, we highlight the importance of controlling for firm-level variables for matching as pointed out in the theory discussion in Section 5.3.2. For this purpose, we re-estimate the last column of the results in Table 5.7, i.e. the analysis of treatment effects based on (joint) market shares above 35% of the BUs engaged in mergers, using three different specifications. In the first specification shown in Column *Exog. DiD*, we assume that selection into treatment is exogenous and estimate a standard DiD regression based on the whole sample (after the exclusion restrictions are applied). Second, we assume that self-selection occurs and employ the PSM-DiD estimator but exclude all firm-level variables (i.e. variables Z in Equation 5.1) from the logit specification presented in Table 5.4. We emulate the approach taken in previous studies and assume that each BU on its own decides – independent from all other BUs of the same firm – whether to self-select into treatment by excluding the variables *No. of active markets*, *No. of active markets sq.* and *Av. no. of product form packs*. The results based on this specification are reported in Column *Endo. DiD wo. Z*. Results in Column *Endo. DiD w. Z* are based on the approach proposed in this essay, i.e. they include firm-level variables in the logit specification in order to estimate propensity scores. Thus, the logit specification in this approach equals that in Table 5.4 and the results are identical to the last column in Table 5.7.

Table 5.11: Comparison of different DiD approaches – Merger effect with market share above 35% (kernel matching)

	Exog. DiD	Endo. DiD wo. Z	Endo. DiD w. Z
Treatment effect year 1	0.232	-0.102	-1.178**
	(0.231)	(0.592)	(0.521)
Treatment effect year 2	-1.074	-1.570	-1.912**
	(1.245)	(1.003)	(0.906)
Observations	15628	5052	2028
Treated BUs	26	26	23
R^2	0.000	0.004	0.030
Adjusted R^2	-0.072	-0.075	-0.075

*Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with cluster-ing considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.*

The results in Table 5.11 show that the three specifications produce largely different results. Provided that our preferred set of estimates is unbiased indeed, the treatment effects can be interpreted to be insignificant and biased upwards in case self-selection is incorrectly ignored. Further, taking self-selection into account using a PSM-DiD estimator but wrongly excluding firm-level variables from the logit specification used to estimate propensity scores reduces this bias to a small degree. Yet, as the comparison to

the last column shows, the bias is not eliminated. Further, the treatment effect is only detected when the firm-level variables are taken into account, highlighting the danger of obtaining results that are not only quantitatively, but also qualitatively wrong if the modified conditional independence assumption underlying the analysis is violated.

5.5   Conclusion

In this study, a new empirical approach is developed to carry out *ex post* evaluations of mergers that affect multiple geographic and/or product markets and that focuses on market-level observations as the unit of analysis. Such disaggregated data is characterised by clustering and a correlation between different observations belonging to the same firm. As a result, some of the core underlying assumptions of the treatment evaluation approaches commonly used in the literature are violated. We draw on recent methodological contributions from outside the field of economics to introduce appropriate methodologies to conduct an analysis of the effects of mergers when the outcome is measured at the market-level, but the treatment occurs at the firm-level. Specifically, we introduce appropriate assumptions that modify the treatment evaluation framework and construct a framework to identify treatment effects in clustered data. Further, we show how the approach to correct for self-selection into mergers based on propensity scores needs to be adapted to clustered data and how the subsequent estimation of treatment effects can be conducted.

We use the new framework to assess the case practice used by the EC to prevent anti-competitive effects in multi-market mergers. The application focuses on pharmaceutical markets in the United Kingdom to estimate the price effects of 11 mergers that occurred between 2003 and 2013. As pharmaceutical firms are offering products in a multitude of independent product markets, the level of the analysis is the BU, i.e. a firm's operation within a distinct product market. We find that the EC's case practice to only scrutinise markets in which the involved parties overlap and have a (joint) market share exceeding 35% might be too narrow. While significant price reductions are found for markets, in which their (joint) market shares exceeds 35%, this effect is at least partly offset by price increases in markets, in which their joint market shares lie between above 15% and 35%. Further, significant price reductions are found in markets without overlaps of market activity by the involved parties if the market share of the active party exceeds 35%, but not if this market share lies between above 15% and 35%. Taken together, the results provide evidence of interdependencies induced by spillovers between markets with different levels of market presence of the parties and between markets with and without overlaps in market presence of the involved parties. Both types of spillover

effects are found to be interrelated with each other with respect to their effects on the parties' BU prices. Unfortunately, while we observe some indications that these spillovers lead to price increases in markets with overlaps in market presence, a lack of data prevents us from establishing whether this is indeed the case.

The results show that spillover effects might be a common phenomenon in multi-market mergers that should receive more scrutiny both in research and in merger control. Further, it remains unclear which spillover effects are more likely to be present or which are most important in practice. A current technical limitation of the framework developed in this study is the lack of appropriate econometric estimators to obtain unbiased standard errors. While some researchers have begun to suggest approaches to estimate valid standard errors in the presence of clustered data and matching estimators, current solutions are limited to a small set of matching algorithms often not suitable for the application in specific data sets. Limitations of this study's application are the small sample size as well as the non-random inclusion of affected businesses in this analysis. Both limitations are a product of the strict exclusion restrictions that need to be applied to the data to ensure a causal identification of treatment effects. Thus, the results only apply to the mergers that are included in this study, and might not generalise to other samples or industries. Future research should therefore attempt to control for spillover effects using larger data sets to ensure not only the internal validity, but also the external validity of the results.

5.6 Appendix

5.6.1 The fundamentals of treatment effect identification

The foundation of most of the treatment evaluation literature is the Rubin causal model (Rubin, 1974). It describes the fundamental identification issue of unobservable counterfactuals that any study attempting to evaluate the impact of a treatment faces, as well as conditions under which the effects can be estimated nonetheless. As such, it provides the fundamental core of all studies evaluating mergers using the treatment effects methodology and outlines the conditions under which causal inference on the effects of a merger on a market outcome can be gained. The Rubin causal model is based on the concept of *randomised controlled trials* (RCT) to identify causal effects. The RCT provides a set of (ideal) conditions needed to identify causal effects of an intervention. These are characterised by all observations being statistically independent from each other and the intervention to be assigned randomly to the observations, such that the treatment does not correlate with characteristics of the treated units.

The following description draws on Wooldridge (2010). For simplicity, assume that we attempt to quantify the effects of several mergers on the prices offered by different firms operating in many markets. Assume here that the unit of outcome is a price index of a firm's prices in all markets in which it is active. The outcome Y_{kd} here denotes, e.g. the price index of firm k and subscript $d = \{1, 0\}$ denotes whether the observation is affected by the merger or not, respectively. Here, we ignore the cluster structure of the data, as clustering within firms is not part of the standard RCT framework, and rely on aggregating a firm's prices in different markets into a single price index instead. To ease the notational burden, we ignore the time dimension. We would like to compare the differences in prices after the treatment (i.e. the merger): the difference between the price indices in the presence of the treatment (y_1) and the price indices that would have existed absent the treatment (y_0) constitutes the treatment effect. Calculating the average value of this difference for all firms engaging in mergers yields the *average treatment effect of the treated* (ATT)

$$\tau_{ATT} \equiv E(y_{k1} - y_{k0}|D = 1), \tag{5.5}$$

where subscripts 1 and 0 next to firm subscript k denote the (hypothetical) case of a treated and non-treated outcome of firm k, respectively.[40] Further, the observations are

---

[40]In this study, the estimated treatment effect is the ATT, which requires less strict assumptions to hold than the *average treatment effect* (ATE). First, the ATE using the CIA implies that both potential outcomes y_1 and y_0 are conditionally independent from unobservables in the treatment

limited to the sub-samples of actually treated and non-treated units with $|D = 1$ and $|D = 0$, respectively. Of course the counterfactual outcome, i.e. the effect of a merger on the price had it not taken place, cannot be observed. To overcome this problem, the outcomes of a suitable control group which consists of firms that did not merge can be used for the construction of the counterfactual outcome $Y_{k0}|D = 1$ of a firm engaged in a merger. This solution to construct a counterfactual requires the *stable unit treatment value assumption* (SUTVA) to hold (Rubin, 1977). It implies that the fact that unit k receiving the treatment does not affect the outcomes Y of any of the other units. In the context of multi-market mergers this implies that, e.g. the price of a merging firm in a given market is not affected by the fact that the firm merges in any other market. Further, it rules out any effect of the merger on the control group.

If the decision to merge D is statistically independent from the analysed outcomes (i.e. $D \perp y_{k0}, y_{k1}$), the ATT can simply be estimated by calculating the difference-in-means estimator $\tau_{ATT} = E(y_{k1}|D = 1) - E(y_{k0}|D = 0)$. In the context of panel data analysis, the DiD estimator, which is related to the difference-in-means estimator, can be used to estimate the ATT if the treatment is exogenous (as then the assumption $D \perp y_{k0}, y_{k1}$ is satisfied).

However, if the decision to merge is non-random and correlates with the observed outcomes y_{k0} and y_{k1}, identification using the difference-in-means estimator breaks down as can be seen by re-writing it to (Wooldridge, 2010, p.907):

$$E(y_{kd}|D = 1) - E(y_{kd}|D = 0) = E(y_{k0}|D = 1) - E(y_{k0}|D = 0) + E(y_{k1} - y_{k0}|D = 1)$$
$$= E(y_{k0}|D = 1) - E(y_{k0}|D = 0) + \tau_{ATT}. \qquad (5.6)$$

As shown above, identification is ensured only when y_0 is mean independent of the treatment, i.e. $E(y_{k0}|D) = E(y_{k0})$, which ensures that $E(y_{k0}|D = 1) - E(y_{k0}|D = 0) = 0$. Fortunately, provided sufficient data is available, τ_{ATT} can be estimated nonetheless using the *conditional independence assumption* (*CIA*, also known as *unconfoundedness assumption*) proposed by Rosenbaum and Rubin (1983). Rather than requiring $D \perp y_{k0}, y_{k1}$, the CIA only assumes

$$Y_{kd} \perp D|X_k, \qquad (5.7)$$

---

selection process, which is modelled to obtain the propensity score. Estimation of the ATT relaxes this assumption and only requires the treated outcome y_1 to be conditionally independent from the unobservables in the treatment selection model (Wooldridge, 2010). Second, estimation of the ATT only requires the overlap assumption to apply to covariates of the treated sub-population. As a result, no positive probability of treatment is required for covariate values corresponding to observations that never or seldom receive the treatment.

i.e. conditional on observable variables X, treatment assignment is random. This requires that all variables that determine the decision to engage in a merger and that correlate with the outcome can be controlled for. Several methods have been proposed to control for selection on observables as described above. In this study, propensity scores as proposed by Rosenbaum and Rubin (1983) are used. Another assumption that is necessary, which is usually explicitly tested in studies using propensity scores, is the *common support* assumption that requires all sub-samples with characteristics $X = x$ in the treatment group to have counterfactual observations in the control group with the same characteristics $X = x$, i.e. $0 < P(D_k = d | X_k = x) < 1$ (Lechner, 2010).[41]

5.6.2 On the use of cross-sectional and inter-temporal exclusion restrictions

Figure 5.4 exemplifies the cross-sectional and intertemporal exclusion restrictions imposed in the study. The horizontal line represents time, and the vertical bars depict the beginning and end of the available time periods in the data, respectively. Assume that there are 7 mergers in a given ATC3 market named M1 to M7, whose occurrence in time is marked by crosses and the corresponding treatment windows are indicated by squared brackets. Each merger has the hypothetical joint market share of the merging parties displayed next to the merger name. Mergers in the markets that are written in black can be used in the analysis, whereas those in grey have to be excluded.

Figure 5.4: Exclusion restrictions
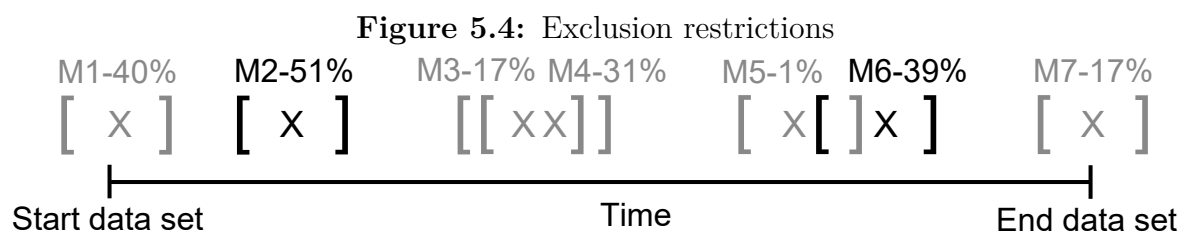
Mergers M1 and M7 have to be excluded, because some data in the considered treatment windows is missing, as it lies outside the time periods available in the data. M2 can be considered, as it does not overlap with any other merger and features a joint market share exceeding the *de minimis* threshold. M3 and M4 both have to be excluded, as they both exceed the *de minimis* threshold and overlap, such that the merger effects cannot be separated and the CL-SUTVA is violated. This is in contrast to mergers M5 and M6, which overlap as well. M5 is ignored because the joint market share is below the *de minimis* threshold. As it is assumed that therefore a merger effect is absent, M6 can be included in the analysis.

---

[41]Limitation of the overlap assumption to the treated group follows from the focus on the ATT.

5.6.3 On the inclusion of treated BUs into the control group pool

With respect to the inclusion of treated units into the control group pool at other points in time, assume that two firms merge in January 2005, and that the effect of the merger is fully realised after two years, i.e. by December 2007. The joint firm does neither engage in nor is affected by another merger for the rest of the time. Then, the observations of the joint firm can be used as control group observations, e.g. around January 2010. At that point in time, they are neither affected by mergers in the one year pre-treatment nor in the two years post-treatment that are included in the estimation window. The same does not hold true when the joint firm's observations are to be used around June 2007: in that case, the estimation window includes the firm's observations from June 2006 to May 2008. Yet, this estimation window overlaps with the post-treatment periods that include the effect of the merger occurring in January 2005, and therefore cannot be used. Figure 5.5 exemplifies the use of firms that engage in merger activity as control group observations.

Figure 5.5: Using firms subject to mergers as control group observations
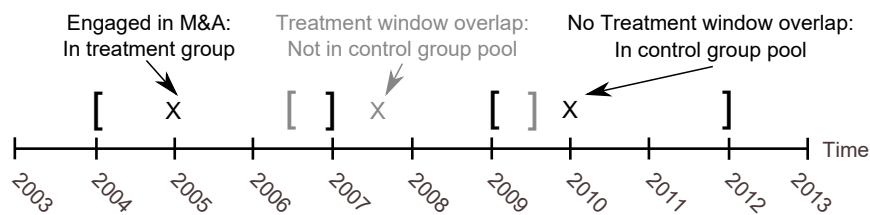
Using observations of firms that engage in mergers as control group observations at other points in time can be crucial to ensure the successful application of matching, especially when the number of independent firm observations is limited and self-selection leads to systematic differences between treated and control group firms. This is likely to be a problem in cluster-randomised trials, as obtaining a large number of independent clusters can be more expensive/data demanding than requiring independent individual units. Therefore, we make efficient use of the data and include treated firms in the control group pool when their observations are not contaminated according to the above logic. This is very beneficial to fulfil the common support assumption of the matching estimator.

5.6.4 Auxiliary tables and figures

Table 5.12: List of mergers

Name	EC Case No.	Date
Sanofi-Synthelabo / Aventis	M.3354	26.04.2004
Teva / IVAX	M.3928	24.11.2005
Novartis / Chiron	M.4049	06.02.2006
UCB / Schwarz Pharma	M.4402	21.11.2006
Abbott / AMO	M.5448	23.02.2009
Sanofi-Aventis / Zentiva	M.5253	20.03.2009
Glaxo Smith Kline / Stiefel Laboratories	M.5530	17.07.2009
Merck / Schering-Plough	M.5502	22.09.2009
Novartis / Alcon	M.5778	09.08.2010
Takeda / Nycomed	M.6278	29.07.2011
Teva / Cephalon	M.6258	13.10.2011

Table 5.13: Descriptive statistics and observations used

	All cases		Considered cases		Excluded cases		Diff.	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean diff.	T-Stat.
Avg. price	7.631	32.430	10.498	43.261	6.953	29.326	-3.55	(-0.700)
HHI	0.447	0.222	0.486	0.242	0.438	0.216	-0.05	(-1.616)
Share in market	0.189	0.275	0.331	0.320	0.156	0.252	-0.18***	(-4.609)
Share generics	0.263	0.285	0.163	0.229	0.287	0.292	0.12***	(4.145)
Total sales in market (Mill. £)	18.333	29.813	13.925	26.118	19.376	30.564	5.45	(1.632)
Firm's sales in market (Mill. £)	1.511	3.680	2.644	5.186	1.243	3.173	-1.40*	(-2.331)
No. of active markets	64.426	26.077	59.370	17.379	65.623	27.626	6.25*	(2.561)
BU's Num. of uniq. product form packs	6.085	6.770	6.272	6.160	6.041	6.914	-0.23	(-0.296)
Av. no. of product form packs	5.707	1.275	5.670	1.355	5.715	1.257	0.05	(0.272)
Num. of uniq. product form packs in market	56.664	63.958	40.000	48.227	60.611	66.596	20.61**	(3.192)
Num of unique molecules by drug-class	8.806	6.319	7.457	5.659	9.126	6.432	1.67*	(2.323)
Observations	423		81		342		423	

*Notes: Sample comparisons are based on the one-sample t-test with unequal variance. * p<0.1, ** p<0.05, *** p<0.01.*

Figure 5.6: Kernel density plots of propensity scores distributions pre- and post-matching (kernel matching)

(a) Pre-matching **(b)** Post-matching
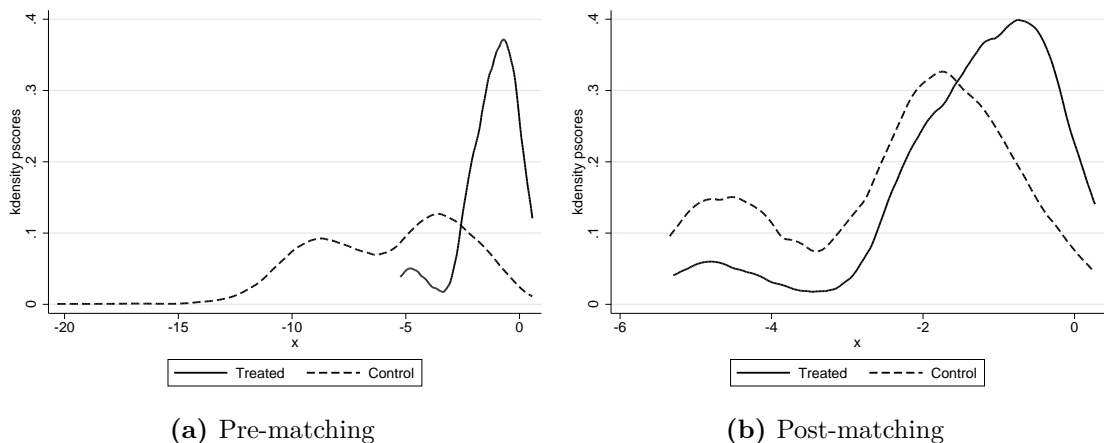
Table 5.14: Common trend assumption test – Merger effects by market shares (kernel matching)

	All	<5%	5%-15%	15%-35%	>35%
Lead quarter 3	-0.594	-1.261*	-1.034	0.117	-0.416
	(0.365)	(0.721)	(0.659)	(0.572)	(0.598)
Lead quarter 2	-0.275	-1.682*	-1.101	0.495	0.245
	(0.419)	(0.875)	(0.709)	(0.521)	(0.517)
Lead quarter 1	-0.851	-0.339**	-1.085*	1.110	-1.818
	(0.673)	(0.142)	(0.639)	(0.927)	(1.468)
Treatment effect year 1	-1.002	-1.574	-1.306	0.825	-1.674*
	(0.699)	(0.973)	(1.042)	(0.666)	(0.888)
Treatment effect year 2	-1.468*	-1.821	-1.277	1.519*	-2.400**
	(0.875)	(2.007)	(1.083)	(0.831)	(1.072)
Observations	4668	780	2124	1848	2028
Treated BUs	74	7	26	18	23
R^2	0.001	0.020	0.000	0.002	0.041
Adjusted R^2	-0.083	-0.137	-0.103	-0.109	-0.064

*Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.*

Table 5.15: Common trend assumption test – Merger effects by the presence of overlaps (kernel matching)

	All	Overlap	No overlap
Lead quarter 3	-0.594	-0.131	-0.712*
	(0.365)	(0.0937)	(0.412)
Lead quarter 2	-0.275	-0.166	-0.379
	(0.419)	(0.193)	(0.478)
Lead quarter 1	-0.851	-0.386	-0.982
	(0.673)	(0.280)	(0.719)
Treatment effect year 1	-1.002	-0.697	-1.170
	(0.699)	(0.553)	(0.788)
Treatment effect year 2	-1.468*	-0.332	-1.691*
	(0.875)	(0.985)	(0.988)
Observations	4668	900	4452
Treated BUs	74	7	64
R^2	0.001	0.000	0.002
Adjusted R^2	-0.083	-0.145	-0.085

*Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.*

Table 5.16: Common trend assumption test – Merger effects by the presence of overlaps for market shares above 35% (kernel matching)

	All	Overlap	No overlap
Lead quarter 3	-0.416	-0.0958	-0.423
	(0.598)	(.)	(0.597)
Lead quarter 2	0.245	-0.190	0.290
	(0.517)	(.)	(0.553)
Lead quarter 1	-1.818	-0.00899	-1.872
	(1.468)	(.)	(1.477)
Treatment effect year 1	-1.674*	-1.545	-1.543*
	(0.888)	(.)	(0.892)
Treatment effect year 2	-2.400**	1.258	-2.388**
	(1.072)	(.)	(1.077)
Observations	2028	132	2004
Treated BUs	23	2	21
R^2	0.041	0.032	0.046
Adjusted R^2	-0.064	-0.598	-0.061

*Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.*

Table 5.17: Common trend assumption test – Merger effects by the presence of overlaps for market shares between 15% and 35% (kernel matching)

	All	Overlap	No overlap
Lead quarter 3	0.117	0.0408*	-0.216
	(0.572)	(0.0244)	(0.507)
Lead quarter 2	0.495	0.162*	0.224
	(0.521)	(0.0879)	(0.332)
Lead quarter 1	1.110	-0.676	0.993
	(0.927)	(0.427)	(0.925)
Treatment effect year 1	0.825	0.429	0.459
	(0.666)	(0.437)	(0.457)
Treatment effect year 2	1.519*	3.378*	1.155
	(0.831)	(1.960)	(0.742)
Observations	1848	420	1452
Treated BUs	18	3	15
R^2	0.002	0.003	0.002
Adjusted R^2	-0.109	-0.213	-0.119

*Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.*

Nearest-neighbour matching results

Table 5.18: Matching results (nearest-neighbour matching)

Variable	Unmatched (U) Matched (M)	Mean Treated	Control	% bias	% reduct. |bias|	t-test t	|p>t|	V(T)/ V(C)
HHI	U	0.481	0.481	0.0		0	0.998	0.89
	M	0.480	0.470	4.2	-13362.9	0.28	0.782	1.10
Share in market	U	0.333	0.206	40.4		3.52	0.000	0.99
	M	0.334	0.381	-14.9	63.2	-0.95	0.346	1.04
Share in market sq.	U	0.209	0.142	21.1		1.92	0.055	1.19
	M	0.210	0.240	-9.4	55.4	-0.6	0.551	1.22
Share generics	U	0.159	0.231	-28.2		-2.27	0.023	0.67
	M	0.161	0.189	-10.9	61.4	-0.72	0.475	0.79
No. of active markets	U	58.000	28.230	135.2		10.2	0.000	0.45*
	M	58.125	59.976	-8.4	93.8	-0.67	0.503	0.99
No. of active markets sq.	U	3660.7	1466.3	103.8		8.41	0.000	0.70
	M	3677.7	3898.9	-10.5	89.9	-0.73	0.469	0.99
Av. no. of product form packs	U	5.670	4.522	62.0		4.55	0.000	0.37*
	M	5.685	5.855	-9.2	85.2	-0.84	0.404	1.28

*Notes: * if the variance ratio is outside the [0.64; 1.55] percentiles of the F-distribution for U and [0.64; 1.56] for M.*

Table 5.19: Matching tests (nearest-neighbour matching)

Sample	Ps R2	LR chi2	p>chi2	Mean Bias	Median Bias	Rubin's B	Rubin's R	%Var
Unmatched	0.344	209.95	0.000	55.8	40.4	182.3*	0.17*	29
Matched	0.023	5.18	0.638	9.6	9.4	36.1*	0.98	0

Notes: * *if B>25% and R outside of [0.5; 2].*

Figure 5.7: Difference-in-Differences time trends (nearest-neighbour matching)

146
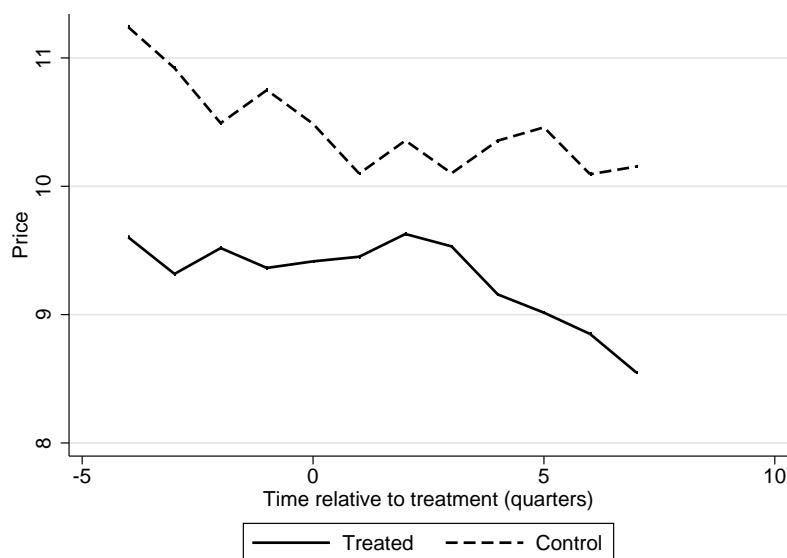
Table 5.20: Merger effects by market shares (nearest-neighbour matching)

	All	<5%	5%-15%	15%-35%	>35%
Treatment effect year 1	-0.321	-0.554	-0.957	0.640	-0.577
	(0.385)	(0.698)	(0.970)	(0.676)	(0.558)
Treatment effect year 2	-0.906	-0.605	-0.674	1.197	-1.950**
	(0.641)	(2.023)	(0.964)	(1.051)	(0.851)
Observations	3336	444	1488	1140	1320
Treated BUs	80	7	27	20	26
R^2	0.001	0.004	0.000	0.002	0.017
Adjusted R^2	-0.088	-0.199	-0.114	-0.124	-0.096

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 5.21: Merger effects by the presence of overlaps (nearest-neighbour matching)

	All	Overlap	No overlap
Treatment effect year 1	-0.321	0.417	-0.490
	(0.385)	(0.697)	(0.447)
Treatment effect year 2	-0.906	1.117	-1.153
	(0.641)	(1.018)	(0.702)
Observations	3336	528	3024
Treated BUs	80	8	72
R^2	0.001	0.001	0.001
Adjusted R^2	-0.088	-0.180	-0.091

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 5.22: Common trend assumption test – Merger effects by market shares (nearest-neighbour matching)

	All	<5%	5%-15%	15%-35%	>35%
Lead quarter 3	-0.619	-1.267*	-1.735	-0.0517	-0.339
	(0.391)	(0.738)	(1.187)	(1.040)	(0.577)
Lead quarter 2	-0.175	-1.633*	-1.931	0.496	0.621
	(0.487)	(0.897)	(1.322)	(1.011)	(0.671)
Lead quarter 1	-0.628	-0.265**	-1.998*	1.549	-1.263
	(0.715)	(0.133)	(1.198)	(1.601)	(1.434)
Treatment effect year 1	-0.669	-1.367	-2.353	1.108	-0.814
	(0.675)	(1.075)	(1.854)	(1.387)	(1.000)
Treatment effect year 2	-1.249	-1.413	-2.079	1.713	-2.178**
	(0.846)	(2.383)	(1.738)	(1.780)	(0.998)
Observations	3336	444	1488	1140	1320
Treated BUs	80	7	27	20	26
R^2	0.001	0.012	0.001	0.003	0.021
Adjusted R^2	-0.089	-0.198	-0.116	-0.126	-0.095

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

Table 5.23: Common trend assumption test – Merger effects by the presence of overlaps (nearest-neighbour matching)

	All	Overlap	No overlap
Lead quarter 3	-0.619	0.950	-0.841*
	(0.391)	(1.604)	(0.432)
Lead quarter 2	-0.175	0.463	-0.387
	(0.487)	(1.519)	(0.518)
Lead quarter 1	-0.628	0.388	-0.885
	(0.715)	(1.516)	(0.750)
Treatment effect year 1	-0.669	0.863	-1.006
	(0.675)	(1.825)	(0.756)
Treatment effect year 2	-1.249	1.563	-1.664*
	(0.846)	(2.121)	(0.951)
Observations	3336	528	3024
Treated BUs	80	8	72
R^2	0.001	0.001	0.002
Adjusted R^2	-0.089	-0.188	-0.092

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

Table 5.24: Merger effects by the presence of overlaps for market shares above 35% (nearest-neighbour matching)

	All	Overlap	No overlap
Treatment effect year 1	-0.577	-0.674***	-0.539
	(0.558)	(0.0246)	(0.578)
Treatment effect year 2	-1.950**	0.534***	-1.987**
	(0.851)	(0.0246)	(0.881)
Observations	1320	144	1260
Treated BUs	26	2	24
R^2	0.017	0.011	0.018
Adjusted R^2	-0.096	-0.515	-0.099

*Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.*

Table 5.25: Merger effects by the presence of overlaps for market shares between 15% and 35% (nearest-neighbour matching)

	All	Overlap	No overlap
Treatment effect year 1	0.640	0.175	0.174
	(0.676)	(0.308)	(0.406)
Treatment effect year 2	1.197	0.486	0.613
	(1.051)	(0.768)	(0.796)
Observations	1140	240	912
Treated BUs	20	4	16
R^2	0.002	0.000	0.001
Adjusted R^2	-0.124	-0.319	-0.139

*Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.*

Table 5.26: Common trend assumption test – Merger effects by the presence of overlaps for market shares above 35% (nearest-neighbour matching)

	All	Overlap	No overlap
Lead quarter 3	-0.339	-0.0958	-0.317
	(0.577)	(.)	(0.584)
Lead quarter 2	0.621	-0.190***	0.679
	(0.671)	(1.58e-15)	(0.709)
Lead quarter 1	-1.263	-0.00899	-1.295
	(1.434)	(.)	(1.441)
Treatment effect year 1	-0.814	-0.747***	-0.764
	(1.000)	(0.0246)	(1.011)
Treatment effect year 2	-2.178**	0.460***	-2.203**
	(0.998)	(0.0246)	(1.022)
Observations	1320	144	1260
Treated BUs	26	2	24
R^2	0.021	0.011	0.022
Adjusted R^2	-0.095	-0.565	-0.097

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

Table 5.27: Common trend assumption test – Merger effects by the presence of overlaps for market shares between 15% and 35% (nearest-neighbour matching)

	All	Overlap	No overlap
Lead quarter 3	-0.0517	0.723	-1.054
	(1.040)	(0.926)	(0.842)
Lead quarter 2	0.496	-0.0308	-0.342
	(1.011)	(0.913)	(0.369)
Lead quarter 1	1.549	-0.121	0.928
	(1.601)	(0.913)	(1.379)
Treatment effect year 1	1.108	0.316	0.0456
	(1.387)	(0.962)	(0.612)
Treatment effect year 2	1.713	0.627	0.524
	(1.780)	(1.106)	(1.027)
Observations	1140	240	912
Treated BUs	20	4	16
R^2	0.003	0.000	0.004
Adjusted R^2	-0.126	-0.340	-0.140

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

5.6.5 Robustness check: matching to same period control group observations only (non-intertemporal kernel matching)

Table 5.28: Matching results (non-intertemporal kernel matching)

Variable	Unmatched (U) Matched (M)	Mean Treated	Mean Control	% bias	% reduct. |bias|	t-test t	t-test |p>t|	V(T)/ V(C)
HHI	U	0.481	0.481	0		0	0.998	0.89
	M	0.456	0.493	-48898.5		-0.69	0.493	0.87
Share in market	U	0.333	0.206	40.4		3.52	0.000	0.99
	M	0.293	0.338	-14.1	65.1	-0.62	0.534	0.8
Share in market sq.	U	0.209	0.142	21.1		1.92	0.055	1.19
	M	0.179	0.229	-16	24.2	-0.69	0.493	0.84
Share generics	U	0.159	0.231	-28.2		-2.27	0.023	0.67
	M	0.210	0.109	39.8	-41.2	2.04	0.044	1.61
No. of active markets	U	58.000	28.230	135.2		10.2	0.000	0.45*
	M	61.381	61.546	-0.7	99.4	-0.04	0.969	0.86
No. of active markets sq.	U	3660.7	1466.3	103.8		8.41	0.000	0.7
	M	4103.2	4180.2	-3.6	96.5	-0.16	0.872	0.9
Av. no. of product form packs	U	5.670	4.522	62		4.55	0.000	0.37*
	M	5.915	5.854	3.3	94.7	0.2	0.841	1.11

*Notes: Results based on matching being limited to control group observations from the same period. * if variance ratio outside [0.64; 1.55] percentiles of the F-distribution for U and [0.54; 1.86] for M.*

Table 5.29: Matching tests (non-intertemporal kernel matching)

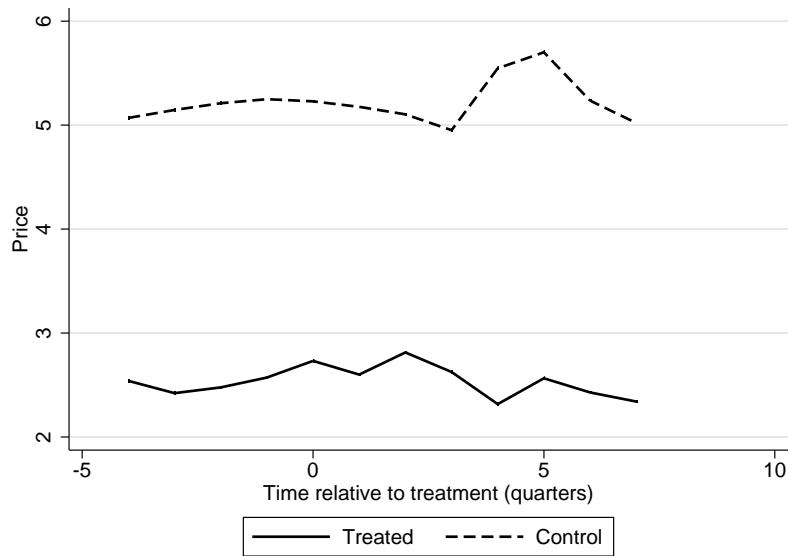| Sample	Ps R2	LR chi2	p>chi2	Mean Bias	Median Bias	Rubin's B	Rubin's R	%Var
Unmatched	0.344	209.95	0.000	55.8	40.4	182.3*	0.17*	29
Matched	0.040	4.63	0.705	13.2	14.1	47.2*	1.42	0

*Notes: Results based on matching being limited to control group observations from the same period. * if B>25% and R outside of [0.5; 2].*

Figure 5.8: Difference-in-Differences time trends (non-intertemporal kernel matching)

Table 5.30: Merger effects by market shares (non-intertemporal kernel matching)

	All	<5%	5%-15%	15%-35%	>35%
Treatment effect year 1	0.173	0.466***	0.0850	0.792	-0.0189
	(0.171)	(0.155)	(0.184)	(0.727)	(0.0876)
Treatment effect year 2	0.435	5.062	-1.012***	1.830*	-0.226*
	(0.710)	(3.192)	(0.390)	(1.061)	(0.126)
Observations	1368	192	576	432	384
Treated BUs	42	4	15	13	10
R^2	0.002	0.067	0.020	0.021	0.008
Adjusted R^2	-0.117	-0.178	-0.134	-0.172	-0.210

*Notes: Results based on matching being limited to control group observations from the same period. Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.*

Table 5.31: Merger effects by the presence of overlaps (non-intertemporal kernel matching)

	All	Overlap	No overlap
Treatment effect year 1	0.173	0.418**	0.179
	(0.171)	(0.180)	(0.341)
Treatment effect year 2	0.435	3.117	0.120
	(0.710)	(3.746)	(0.671)
Observations	1368	228	1428
Treated BUs	42	5	37
R^2	0.002	0.023	0.000
Adjusted R^2	-0.117	-0.143	-0.120

*Notes: Results based on matching being limited to control group observations from the same period. Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.*

Table 5.32: Common trend assumption test – Merger effects by market shares (non-intertemporal kernel matching)

	All	<5%	5%-15%	15%-35%	>35%
Lead quarter 3	-0.174	-0.0408	-0.282*	0.0291	-0.283
	(0.114)	(0.0869)	(0.145)	(0.121)	(0.244)
Lead quarter 2	-0.225	-0.00773	-0.128	-0.0229	-0.216
	(0.262)	(0.194)	(0.125)	(0.244)	(0.161)
Lead quarter 1	-0.333	-1.095**	0.102*	-0.106	-0.413*
	(0.241)	(0.540)	(0.0593)	(0.111)	(0.215)
Treatment effect year 1	-0.00971	0.180***	0.0110	0.768	-0.244***
	(0.192)	(0.0469)	(0.212)	(0.685)	(0.0945)
Treatment effect year 2	0.255	4.776	-1.086***	1.803*	-0.447*
	(0.674)	(3.115)	(0.417)	(1.048)	(0.248)
Observations	1368	192	576	432	384
Treated BUs	42	4	15	13	10
R^2	0.002	0.068	0.020	0.021	0.013
Adjusted R^2	-0.119	-0.201	-0.140	-0.182	-0.214

Notes: Results based on matching being limited to control group observations from the same period. Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

Table 5.33: Common trend assumption test – Merger effects by the presence of overlaps (non-intertemporal kernel matching)

	All	Overlap	No overlap
Lead quarter 3	-0.174	-0.0635	-0.190
	(0.114)	(0.0718)	(0.237)
Lead quarter 2	-0.225	-0.0270	0.396
	(0.262)	(0.158)	(0.560)
Lead quarter 1	-0.333	-0.853	-0.343
	(0.241)	(0.618)	(0.896)
Treatment effect year 1	-0.00971	0.182***	0.147
	(0.192)	(0.0411)	(0.650)
Treatment effect year 2	0.255	2.881	0.0896
	(0.674)	(3.624)	(0.953)
Observations	1368	228	1428
Treated BUs	42	5	37
R^2	0.002	0.023	0.001
Adjusted R^2	-0.119	-0.160	-0.122

Notes: Results based on matching being limited to control group observations from the same period. Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * p<0.1, ** p<0.05, *** p<0.01.

Table 5.34: Merger effects by the presence of overlaps for market shares above 35% (non-intertemporal kernel matching)

	All	Overlap	No overlap
Treatment effect year 1	-0.0189	0.198***	-0.693*
	(0.0876)	(4.04e-17)	(0.396)
Treatment effect year 2	-0.226*	0.0321***	-1.225*
	(0.126)	(4.38e-17)	(0.686)
Observations	384	24	456
Treated BUs	10	1	9
R^2	0.008	0.438	0.023
Adjusted R^2	-0.210	-0.499	-0.169

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 5.35: Merger effects by the presence of overlaps for market shares between 15% and 35% (non-intertemporal kernel matching)

	All	Overlap	No overlap
Treatment effect year 1	0.792	0.557***	0.450
	(0.727)	(0.207)	(0.488)
Treatment effect year 2	1.830*	9.559***	1.384
	(1.061)	(3.673)	(1.280)
Observations	432	60	420
Treated BUs	13	2	11
R^2	0.021	0.137	0.005
Adjusted R^2	-0.172	-0.232	-0.201

Notes: Cluster– and autocorrelation–consistent standard errors in parentheses with clustering considered at the firm–level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Chapter 6

Summary

This thesis consists of four essays that show how the design of competition law and its enforcement might be improved. Chapter 2 develops a new empirical cartel screen to increase the detection of cartels, and Chapter 3 investigates policy measures to reduce the harm that tacit collusion after the end of cartels can cause on welfare. Acknowledging that not all cartels can be detected, Chapter 4 studies how cartel fining rules can be designed to reduce the welfare damage of cartels by lowering the optimal cartel price. Finally, Chapter 5 develops an empirical framework to study the occurrence of spillover effects in multi-market M&As. The framework is used to test for ineffectiveness in the merger control practice of the EC caused by interdependencies of firms' market presences and their effects on market prices.

The new empirical cartel screen developed in Chapter 2 to uncover cartels from analysing market data is applied to three European markets for pasta products. It successfully reports the cartels that were present in the Italian and Spanish markets, but does not wrongly detect signs of collusion in the French market, which was not cartelised. The new screen is less dependent than existing cartel screens on specific hypotheses with respect to how cartel behaviour creates collusive markers, and detects the cartels in Italy and Spain despite the fact that the established price variance-based screens fail to report any signs of collusion. Based on these results, the new methodology offers the promise to be a valuable contribution to the growing toolkit of behavioural cartel screens and to help strengthen deterrence by increasing the risk of detection.

The experimental investigation in Chapter 3 provides insights on the determinants, consequences, and prevention of PCTC. Tacit collusion after cartels is found to be linked to preceding cartel success and collusive price hysteresis. As a result, firms in previously successful cartels appear to be least deterred by exogenous fines based on the cartel periods only in the presence of PCTC. Further, in its presence some of the most common methods to estimate cartel overcharges used in private damage litigation are

subject to a downward bias. As this bias increases with preceding cartel success, private damage litigation cannot compensate for this deficiency in public enforcement. Erasing information about the cartel pricing strategies from the industry through removing key cartel personnel from their positions after detection shows the prospect to disrupt PCTC and reduce its negative effects on welfare.

The use of endogenous enforcement against cartels that depends on their behaviour to reduce the harm induced by undetected cartels is experimentally analysed in Chapter 4. It is found that endogenous expected punishment that increases with the cartel price can contribute to the strategic uncertainty between cartelists and lead them to choose lower cartel prices. This is shown in the experiment by offering cartels two equally profitable cartel prices, of which the lower cartel price features a lower level of riskiness of collusion. As predicted by the equilibrium concept of Blonski *et al.* (2011) and Blonski and Spagnolo (2015), the results show that subjects tend to agree on the low cartel price, as it offers a less negative payoff in case of being cheated on the agreement. The essay shows that enhancing composition deterrence – despite currently largely being overlooked in the context of cartels in favour of frequency deterrence – offers the promise to reduce the harm caused by cartels. As one cannot hope to be able to detect all cartels, the strategic design of antitrust enforcement to reduce the cartel price of undetected cartels can contribute to effective competition policy.

A new empirical framework to conduct *ex post* evaluations of mergers in clustered data, in which the treatment applies to either all or none of the market-level observations by a firm, is developed in Chapter 5. In these environments, spillover effects can induce a dependence of the effect of mergers on a firm's market presence in one market on its presence in other markets. The framework is used to study whether such spillover effects undermine the effectiveness of the EC's case practice to evaluate multi-market mergers in the pharmaceutical industry. The results show that spillovers are present both between markets in which the parties have different levels of market shares, and between markets with and without overlaps in market presence prior to the merger. Significant price reductions are found in markets, in which the market shares of the parties exceed 35%, but they are partly offset by price increases in markets in which they feature market shares between above 15% and 35%. Yet, due to a lack of data, no definite answer can be provided whether significant price reductions found in markets, in which the parties had no overlapping market presence prior to the mergers, lead to increases in market prices in markets with overlaps. The findings suggest that spillover effects need to receive more attention in merger control to ensure that potential anticompetitive effects of mergers are not overlooked. Further, more research is needed on the determinants of spillover effects in practice, as they are *ex ante* hard to predict for competition authorities with respect to their existence and magnitude.

Bibliography

ABADIE, A. and IMBENS, G. W. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, **74** (1), 235–267.

— and — (2008). On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, **76** (6), 1537–1557.

ABRANTES-METZ, R. M. (2014). Recent Successes of Screens for Conspiracies and Manipulations: Why Are There Still Skeptics? *Antitrust Chronicle*, **10** (2).

— and BAJARI, P. (2010). A Symposium on Cartel Sanctions: Screens for Conspiracies and Their Multiple Applications. *Competition Policy International*, **6** (2), 129–253.

—, FROEB, L. M., GEWEKE, J. and TAYLOR, C. T. (2006). A variance screen for collusion. *International Journal of Industrial Organization*, **24** (3), 467–486.

AGUZZONI, L., ARGENTESI, E., CIARI, L., DUSO, T. and TOGNONI, M. (2016). Ex post Merger Evaluation in the U.K. Retail Market for Books. *The Journal of Industrial Economics*, **64** (1), 170–200.

ALM, J., McCLELLAND, G. H. and SCHULZE, W. D. (1992). Why do people pay taxes? *Journal of Public Economics*, **48** (1), 21–38.

—, SANCHEZ, I. and JUAN, A. D. (1995). Economic and Noneconomic Factors in Tax Compliance. *Kyklos*, **48** (1), 1–18.

ANDREOU, E. and GHYSELS, E. (2009). Structural Breaks in Financial Time Series. In T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. V. Mikosch (eds.), *Handbook of financial time series*, Springer Science & Business Media, pp. 839–870.

ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61** (4), 821–856.

— and PLOBERGER, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, **62** (6), 1383–1414.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton: Princeton University Press.

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

Argentesi, E., Buccirossi, P., Cervone, R., Duso, T. and Marrazzo, A. (2016). The effect of retail mergers on variety: An ex-post evaluation. *Mimeo.*

Ashenfelter, O. C., Hosken, D. S. and Weinberg, M. C. (2015). Efficiencies brewed: pricing and consolidation in the US beer industry. *The RAND Journal of Economics*, **46** (2), 328–361.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, **34** (1), 1–16.

Augurzky, B. and Schmidt, C. M. (2001). The propensity score: A means to an end. *IZA Discussion Paper No. 271.*

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, **28** (25), 3083–3107.

—, Grootendorst, P. and Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine*, **26** (4), 734–753.

Autor, D. H. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*, **21** (1), 1–42.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66** (1), 47–78.

— and — (2003). Critical values for multiple structural change tests. *The Econometrics Journal*, **6** (1), 72–78.

Baker, J. B. (2002). Mavericks, mergers, and exclusion: Proving coordinated competitive effects under the antitrust laws. *New York University Law Review*, **77** (1), 135–203.

— and Rubinfeld, D. L. (1999). Empirical methods in antitrust litigation: review and critique. *American Law and Economics Review*, **1** (1), 386–435.

Banerjee, A., Galbraith, J. W. and Dolado, J. (1990). Practitioner's corner: Dynamic Specification and Linear Transformations of the Autoregressive–Distributed Lag Model. *Oxford Bulletin of Economics and Statistics*, **52** (1), 95–104.

Barton, D. M. and Sherman, R. (1984). The price and profit effects of horizontal merger: a case study. *The Journal of Industrial Economics*, **33** (2), 165–177.

Behr, A. and Heid, F. (2011). The success of bank mergers revisited. An assessment based on a matching strategy. *Journal of Empirical Finance*, **18** (1), 117–135.

Benoit, J.-P. (1984). Financially Constrained Entry in a Game with Incomplete Information. *The RAND Journal of Economics*, **15** (4), 490–499.

Bernheim, D. and Whinston, M. D. (1990). Multimarket Contact and Collusive Behavior. *The RAND Journal of Economics*, **21** (1), 1–26.

Berninghaus, S. K., Haller, S., Krüger, T., Neumann, T., Schosser, S. and Vogt, B. (2013). Risk attitude, beliefs, and information in a Corruption Game – An experimental analysis. *Journal of Economic Psychology*, **34**, 46–60.

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

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

Blair, R. D. and Sokal, D. D. (eds.) (2013). *Oxford Handbook on International Antitrust Economics*. Oxford: Oxford University Press.

Blanckenburg, K., Geist, A. and Kholodilin, K. A. (2012). The Influence of Collusion on Price Changes: New Evidence from Major Cartel Cases. *German Economic Review*, **13** (3), 245–256.

Block, M. K., Nold, F. C. and Sidak, J. G. (1981). The deterrent effect of antitrust enforcement. *The Journal of Political Economy*, **89** (3), 429–445.

Blonski, M., Ockenfels, P. and Spagnolo, G. (2011). Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence. *American Economic Journal: Microeconomics*, **3** (3), 164–192.

— and Spagnolo, G. (2015). Prisoners' other Dilemma. *International Journal of Game Theory*, **44** (1), 61–81.

Bock, O., Baetge, I. and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, **71**, 117–120.

BOLOTOVA, Y., CONNOR, J. M. and MILLER, D. J. (2008). The impact of collusion on price behavior: Empirical results from two recent cases. *International Journal of Industrial Organization*, **26** (6), 1290–1307.

BOLOTOVA, Y. V. (2009). Cartel overcharges: An empirical analysis. *Journal of Economic Behavior & Organization*, **70** (1-2), 321–341.

BOLTON, P. and SCHARFSTEIN, D. S. (1990). A theory of predation based on agency problems in financial contracting. *The American Economic Review*, **80** (1), 93–106.

BORENSTEIN, S. (1990). Airline mergers, airport dominance, and market power. *The American Economic Review*, **80** (2), 400–404.

BOS, I., DAVIES, S. W., HARRINGTON, J. E. and ORMOSI, P. L. (2016). Does enforcement deter cartels? A tale of two tails. *CCP Working Paper 14-6 v2*.

—, LETTERIE, W. and VERMEULEN, D. (2015). Antitrust as Facilitating Factor for Collusion. *The B.E. Journal of Economic Analysis & Policy*, **15** (2), 797–814.

BOSWIJK, P., SCHINKEL, M. P. and BUN, M. (2016). Cartel Dating. *UVA Econometrics Discussion Paper: No. 2016/04*.

BOYER, M. and KOTCHONI, R. (2015). How Much Do Cartels Overcharge? *Review of Industrial Organization*, **47** (2), 119–153.

BRANDER, J. A. and ROSS, T. W. (2006). Estimating damages from price-fixing. *Canadian Class Action Review*, **3** (1), 335–369.

BROOKHART, M. A., SCHNEEWEISS, S., ROTHMAN, K. J., GLYNN, R. J., AVORN, J. and STÜRMER, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, **163** (12), 1149–1156.

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, **37** (2), 149–192.

BRYANT, P. G. and ECKARD, E. W. (1991). Price fixing: the probability of getting caught. *The Review of Economics and Statistics*, **73** (3), 531–536.

CALIENDO, M. and KOPEINIG, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, **22** (1), 31–72.

CAMPELLO, M. (2006). Debt financing: Does it boost or hurt firm performance in product markets? *Journal of Financial Economics*, **82** (1), 135–172.

CARLTON, D. W. (2004). Using economics to improve antitrust policy. *Columbia Business Law Review*, **283**, 283–333.

CASON, T. N. (1995). Cheap talk price signaling in laboratory markets. *Information Economics and Policy*, **7** (2), 183–204.

CHEN, Z. and REY, P. (2015). A Theory of Conglomerate Mergers. *Mimeo.*

CHOWDHURY, S. M. and CREDE, C. J. (2015). Post-Cartel Tacit Collusion: Determinants, Consequences, and Prevention. *Centre for Competition Policy Working Paper No. 15-1.*

— and WANDSCHNEIDER, F. (2016). An Experimental Analysis of Anti-trust Enforcement under Avoidance. *Mimeo.*

— and — (2017). Anti-trust and the 'Beckerian Proposition': the Effects of Investigation and Fines on Cartels. *Forthcoming in: Handbook of Behavioral Industrial Organization.*

CHU, C.-S. J., HORNIK, K. and KUAN, C.-M. (1995a). MOSUM tests for parameter constancy. *Biometrika*, **82** (3), 603–617.

—, — and — (1995b). The moving-estimates test for parameter stability. *Econometric Theory*, **11** (4), 699–720.

—, STINCHCOMBE, M. and WHITE, H. (1996). Monitoring Structural Change. *Econometrica*, **64** (5), 1045–1065.

CHUNG, J. and JEON, S. (2014). Portfolio effects in conglomerate mergers: the empirical evidence of leverage effects in Korean liquor market. *Applied Economics*, **46** (35), 4345–4359.

CLARKE, K. A., KENKEL, B. and RUEDA, M. R. (2015). Misspecification and the Propensity Score: The Possibility of Overadjustment. *Mimeo.*

CLEMENS, G. and RAU, H. A. (2014). Do Leniency Policies Facilitate Collusion? Experimental Evidence. *Mimeo.*

CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E. and TERPENNING, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, **6** (1), 3–73.

COCKBURN, I. M. and HENDERSON, R. M. (2001). Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research. *Journal of Health Economics*, **20** (6), 1033–1057.

CONNOR, J. M. (1998). The global citric acid conspiracy: Legal–economic lessons. *Agribusiness*, **14** (6), 435–452.

— (2001). "Our Customers Are Our Enemies": The Lysine Cartel of 1992–1995. *Review of Industrial Organization*, **18** (1), 5–21.

— (2005). Collusion and price dispersion. *Applied Economics Letters*, **12** (6), 335–338.

— and LANDE, R. H. (2012). Cartels as Rational Business Strategy: Crime Pays. *Cardozo Law Review*, **34** (2), 427–490.

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

COOPER, R. W., DEJONG, D. V., FORSYTHE, R. and ROSS, T. W. (1990). Selection Criteria in Coordination Games: Some Experimental Results. *The American Economic Review*, **80** (1), 218–233.

CREDE, C. J. (2015). A structural break cartel screen for dating and detecting collusion. *CCP Working Paper No. 15-11*.

— (2016). Getting a fix on price-fixing cartels. *Significance*, **13** (1), 38–41.

— and LU, L. (2016). The effects of endogenous enforcement on strategic uncertainty and cartel deterrence. *CBESS Discussion Paper No. 16-08*.

CROOK, T. R. and COMBS, J. G. (2007). Sources and consequences of bargaining power in supply chains. *Journal of Operations Management*, **25** (2), 546–555.

DAL BÓ, P. (2005). Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *American Economic Review*, **95** (5), 1591–1604.

DAL BÓ, P. and FRÉCHETTE, G. R. (2011). The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence. *American Economic Review*, **101** (1), 411–429.

DANZON, P. M., EPSTEIN, A. and NICHOLSON, S. (2007). Mergers and acquisitions in the pharmaceutical and biotech industries. *Managerial and Decision Economics*, **28** (4-5), 307–328.

DAVIS, D., KORENOK, O. and REILLY, R. (2010). Cooperation without coordination: signaling, types and tacit collusion in laboratory oligopolies. *Experimental Economics*, **13** (1), 45–65.

DAVIS, P. and GARCÉS, E. (2009). *Quantitative techniques for competition and antitrust analysis*. Princeton: Princeton University Press.

DE LUNA, X., JOHANSSON, P. and SJÖSTEDT-DE LUNA, S. (2010). Bootstrap inference for K-nearest neighbour matching estimators. *Mimeo*.

DE ROOS, N. (2006). Examining models of collusion: The market for lysine. *International Journal of Industrial Organization*, **24** (6), 1083–1107.

DRANOVE, D. and LINDROOTH, R. (2003). Hospital consolidation and costs: another look at the evidence. *Journal of Health Economics*, **22** (6), 983–997.

DREBER, A., FUDENBERG, D. and RAND, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization*, **98**, 41–55.

DUFWENBERG, M. and GNEEZY, U. (2000). Price competition and market concentration: an experimental study. *International Journal of Industrial Organization*, **18** (1), 7–22.

ECKEL, C. C. and GROSSMAN, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, **68** (1), 1–17.

ERUTKU, C. (2012). Testing post-cartel pricing during litigation. *Economics Letters*, **116** (3), 339–342.

ESPOSITO, F. M. and FERRERO, M. (2006). Variance screens for detecting collusion: an application to two cartel cases in Italy. *Mimeo*.

EUROPEAN COMMISSION (1998). Case No IV/M.950 - Hoffmann La Roche/Boehringer Mannheim. *Official Journal of the European Communities*, pp. L 234/14–L 234/38.

EUROPEAN COMMISSION (2011a). Case No COMP/M.6258 - TEVA / CEPHALON. *Public version of the merger procedure C(2011) 7435, final release on 13/10/2011*.

EUROPEAN COMMISSION (2011b). Case No COMP/M.6278 - TAKEDA / NYCOMED. *Public version of the merger procedure document C(2011) 5649, final release on 29/07/2011*.

FARRELL, J. and SHAPIRO, C. (2010). Antitrust evaluation of horizontal mergers: An economic alternative to market definition. *The BE Journal of Theoretical Economics*, **10** (1), Art. 9.

FEHR, E. (2009). On The Economics and Biology of Trust. *Journal of the European Economic Association*, **7** (2-3), 235–266.

FISCHBACHER, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **10** (2), 171–178.

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

— and — (2014). Endogenous cartel formation: Experimental evidence. *Economics Letters*, **125** (2), 223–225.

FOSTER, D. P. and YOUNG, H. (2003). Learning, hypothesis testing, and Nash equilibrium. *Games and Economic Behavior*, **45** (1), 73–96.

FOX, E. M. (2007). GE/Honeywell: The U.S. Merger that Europe Stopped - A Story of the Politics of Convergence. In E. M. Fox and D. Crane (eds.), *Antitrust stories*, New York, N.Y.: Foundation Press.

FRESARD, L. (2010). Financial strength and product market behavior: The real effects of corporate cash holdings. *The Journal of Finance*, **65** (3), 1097–1122.

FUDENBERG, D. and MASKIN, E. (1993). Evolution and repeated games. *Mimeo*.

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

GINSBURG, D. H. and WRIGHT, J. D. (2010). Antitrust Sanctions. *Competition Policy International*, **6** (2).

GIRAUDEAU, B. and RAVAUD, P. (2009). Preventing bias in cluster randomised trials. *PLoS Medicine*, **6** (5), e1000065.

GIRMA, S. and GÖRG, H. (2007). Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *Journal of International Economics*, **72** (1), 97–112.

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

HAMAGUCHI, Y., KAWAGOE, T. and SHIBATA, A. (2009). Group size effects on cartel formation and the enforcement power of leniency programs. *International Journal of Industrial Organization*, **27** (2), 145–165.

HANSON, S. G. and SUNDERAM, A. (2011). The Variance of Non-Parametric Treatment Effect Estimators in the Presence of Clustering. *Review of Economics and Statistics*, **94** (4), 1197–1201.

HARRINGTON, J. E. (2004a). Cartel Pricing Dynamics in the Presence of an Antitrust Authority. *The RAND Journal of Economics*, **35** (4), 651–673.

— (2004b). Post-cartel Pricing during Litigation. *Journal of Industrial Economics*, **52** (4), 517–533.

— (2005). Optimal Cartel Pricing in the Presence of an Antitrust Authority. *International Economic Review*, **46** (1), 145–169.

— (2007). Behavioural Screening and the Detection of Cartels. In C.-D. Ehlermann and I. Atanasiu (eds.), *European Competition Law Review 2006: Enforcement of Prohibition of Cartels*, Oxford/Portland, Oregon: Hart Publishing.

— (2008). Detecting cartels. In P. Buccirossi (ed.), *Handbook of Antitrust Economics*, Cambridge, Mass.: MIT press, pp. 213–258.

— (2012). A Theory of Tacit Collusion. *Johns Hopkins University Department of Economics Working Paper Archive No. 588.*

— and CHEN, J. (2006). Cartel pricing dynamics with cost variability and endogenous buyer detection. *International Journal of Industrial Organization*, **24** (6), 1185–1212.

— and ZHAO, W. (2012). Signaling and tacit collusion in an infinitely repeated Prisoners' Dilemma. *Mathematical Social Sciences*, **64** (3), 277–289.

HARRIS, R., SIEGEL, D. S. and WRIGHT, M. (2005). Assessing the impact of management buyouts on economic efficiency: Plant-level evidence from the United Kingdom. *Review of Economics and Statistics*, **87** (1), 148–153.

HARSANYI, J. C. and SELTEN, R. (1988). *A general theory of equilibrium selection in games.* Cambridge, Mass.: MIT press.

HAUCAP, J. and STIEBALE, J. (2016). How Mergers Affect Innovation: Theory and Evidence from the Pharmaceutical Industry. *DICE discussion paper no. 218.*

HECKMAN, J. J., ICHIMURA, H. and TODD, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, **65** (2), 261–294.

—, — and TODD, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, **64** (4), 605–654.

— and Todd, P. E. (2009). A Note on Adapting Propensity Score Matching and Selection Models to Choice Based Samples. *The Econometrics Journal*, **12** (1), 230–234.

Heijnen, P., Haan, M. A. and Soetevent, A. R. (2015). Screening for collusion: a spatial statistics approach. *Journal of Economic Geography*, **15** (2), 417–448.

Heinemann, F., Nagel, R. and Ockenfels, P. (2009). Measuring Strategic Uncertainty in Coordination Games. *Review of Economic Studies*, **76** (1), 181–221.

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

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

Ho, V. and Hamilton, B. H. (2000). Hospital mergers and acquisitions: does market consolidation harm patients? *Journal of Health Economics*, **19** (5), 767–791.

Holt, C. A. and Laury, S. K. (2002). Risk Aversion and Incentive Effects. *The American Economic Review*, **92** (5), 1644–1655.

Hong, G. and Raudenbush, S. W. (2006). Evaluating Kindergarten Retention Policy. *Journal of the American Statistical Association*, **101** (475), 901–910.

Hosken, D. S., Olson, L. and Smith, L. (2012). Do Retail Mergers Affect Competition? Evidence from Grocery Retailing. *Mimeo*.

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

Hüschelrath, K. and Veith, T. (2013). Cartel Detection in Procurement Markets. *Managerial and Decision Economics*, **35** (6), 404–422.

Hyytinen, A., Steen, F. and Toivanen, O. (2011). Cartels uncovered. *SNF Working Paper No. 08/2011*.

International Pasta Organisation (2012). Annual Report 2012. *Available online at http://www.internationalpasta.org/resources/report/IPOreport2012.pdf, retrieved 19/07/2015.*

Isaac, R. M. and Walker, J. M. (1988). Communication and free–riding behavior: The voluntary contribution mechanism. *Economic Inquiry*, **26** (4), 585–608.

Italian Competition Authority (2009). Desicion of the Autorità Garante della Concorrenza e del Mercato regarding UNIPI – Unione Industriali Pastai Italiani e Union Alimentari – Unione Nazionale della Piccola e Media Industria Alimentare. *Available online at http://www.governo.it/backoffice/allegati/42113-5213.pdf, retrieved 17/05/2014.*

Ivaldi, M., Jullien, B., Rey, P., Seabright, P. and Tirole, J. (2003). The economics of tacit collusion. *IDEI Working Paper No. 186.*

Jovanovic, B. and Rousseau, P. L. (2008). Mergers as reallocation. *The Review of Economics and Statistics*, **90** (4), 765–776.

Kandori, M., Mailath, G. J. and Rob, R. (1993). Learning, Mutation, and Long Run Equilibria in Games. *Econometrica*, **61** (1), 29–56.

Katsoulacos, Y., Motchenkova, E. and Ulph, D. (2015). Penalizing cartels: The case for basing penalties on price overcharge. *International Journal of Industrial Organization*, **42**, 70–80.

Kim, E. H. and Singal, V. (1993). Mergers and market power: Evidence from the airline industry. *The American Economic Review*, **83** (3), 549–569.

Kim, J.-H. (2011). Comparison of Structural Change Tests in Linear Regression Models. *Korean Journal of Applied Statistics*, **24** (6), 1197–1211.

Knight, F. J. (1921). *Risk, Uncertainty and Profit.* Boston: Houghton Mifflin.

Kolasky, W. J. (2002). Conglomerate Mergers and Range Effects: It's a Long Way from Chicago to Brussels. *George Mason Law Review*, **10** (3), 533–550.

Kovacic, W. E., Marshall, R. C., Marx, L. M. and Raiff, M. E. (2007). Lessons for Competition Policy from the Vitamins Cartel. *Contributions to Economic Analysis*, **282**, 149–176.

Krishnan, R. (2001). Market restructuring and pricing in the hospital industry. *Journal of Health Economics*, **20** (2), 213–237.

Kuan, C.-M. and Hornik, K. (1995). The generalized fluctuation test: A unifying view. *Econometric Reviews*, **14** (2), 135–161.

Kwoka, J., Greenfield, D. and Chengyan Gu (2015). *Mergers, Merger Control, and Remedies: A Retrospective Analysis of U.S. Policy.* Cambridge, Mass.: MIT press.

Lande, R. H. and Davis, J. P. (2008). Benefits from private antitrust enforcement: An analysis of forty cases. *University of San Francisco Law Review*, **42**, 879–918.

Lechner, M. (2010). The Estimation of Causal Effects by Difference-in-Difference MethodsEstimation of Spatial Panels. *Foundations and Trends® in Econometrics*, **4** (3), 165–224.

Leisch, F., Hornik, K. and Kuan, C.-M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, **16** (6), 835–854.

Leon, A. C., Demirtas, H., Li, C. and Hedeker, D. (2013). Subject-level matching for imbalance in cluster randomized trials with a small number of clusters. *Pharmaceutical statistics*, **12** (5), 268–274.

Leuven, E. and Sianesi, B. (2015). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components*.

Leyrat, C., Caille, A., Donner, A. and Giraudeau, B. (2013). Propensity scores used for analysis of cluster randomized trials with selection bias: a simulation study. *Statistics in Medicine*, **32** (19), 3357–3372.

Lu, L. (2016). *Essays on Strategic Firms, Vertical Contracts and Horizontal Agreements*. Doctoral thesis, University of East Anglia.

Mabert, V. A. and Venkataramanan, M. A. (1998). Special Research Focus on Supply Chain Linkages: Challenges for Design and Management in the 21st Century. *Decision Sciences*, **29** (3), 537–552.

Mariuzzo, F. and Ormosi, P. L. (2016). Post-merger price variation matters, so why do merger retrospectives ignore it? *Centre for Competition Policy Working Paper No. 16-5*.

Martin, S. (2006). Competition policy, collusion, and tacit collusion. *International Journal of Industrial Organization*, **24** (6), 1299–1332.

Maynard, A. and Bloor, K. (2003). Dilemmas in regulation of the market for pharmaceuticals. *Health Affairs*, **22** (3), 31–41.

Mouraviev, I. (2006). Private Observation, Tacit Collusion and Collusion with Communication. *IFN Working Paper No. 672*.

Mueller, W. F. and Parker, R. C. (1992). The bakers of Washington cartel: Twenty-five years later. *Review of Industrial Organization*, **7** (1), 75–82.

MUNDLAK, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, **46** (1), 69–85.

NEVEN, D. (2008). Analysis of conglomerate effects in EU merger control. In P. Buccirossi (ed.), *Handbook of Antitrust Economics*, Cambridge, Mass.: MIT press, pp. 183–211.

NIEBERDING, J. F. (2006). Estimating overcharges in antitrust cases using a reduced-form approach: Methods and issues. *Journal of Applied Economics*, **9** (2), 361–380.

NOTARO, G. (2014). Methods for quantifying cartel damages: The pasta cartel in Italy. *Journal of Competition Law and Economics*, **10** (1), 87–106.

NYBLOM, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, **84** (405), 223–230.

ORDÓÑEZ-DE HARO, J. M. and TORRES, J. L. (2014). Price hysteresis after antitrust enforcement: Evidence from spanish food markets. *Journal of Competition Law and Economics*, **10** (1), 217–256.

ORMOSI, P. L. (2014). A tip of the iceberg? The probability of catching cartels. *Journal of Applied Econometrics*, **29** (4), 549–566.

ORNAGHI, C. (2009). Mergers and innovation in big pharma. *International Journal of Industrial Organization*, **27** (1), 70–79.

OTSU, T. and RAI, Y. (2016). Bootstrap Inference of Matching Estimators for Average Treatment Effects. *Journal of the American Statistical Association*, **18**, 1–13.

PERISTIANI, S. (1997). Do mergers improve the X-efficiency and scale efficiency of US banks? Evidence from the 1980s. *Journal of Money, Credit, and Banking*, **29** (3), 326–337.

PEROLD, A. F. (2005). Capital allocation in financial firms. *Journal of Applied Corporate Finance*, **17** (3), 110–118.

PERRON, P. (2006). Dealing with structural breaks. In T. C. Mills and K. D. Patterson (eds.), *Palgrave handbook of econometrics*, vol. 1, New York, N.Y.: Palgrave Macmillan, pp. 278–352.

PESARAN, M. H. and SHIN, Y. (1998). An autoregressive distributed-lag modelling approach to cointegration analysis. In S. Strøm (ed.), *Econometrics and Economic Theory in the 20th Century*, Cambridge: Cambridge University Press, pp. 371–413.

PLOBERGER, W. and KRÄMER, W. (1992). The CUSUM test with OLS residuals. *Econometrica*, **60** (2), 271–285.

—, KRÄMER, W. and KONTRUS, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, **40** (2), 307–318.

PORTER, R. H. (1983). A Study of Cartel Stability: The Joint Executive Committee, 1880-1886. *The Bell Journal of Economics*, **14** (2), 301–314.

— (2005). Detecting Collusion. *Review of Industrial Organization*, **26** (2), 147–167.

— and ZONA, J. D. (1993). Detection of Bid Rigging in Procurement Auctions. *Journal of Political Economy*, **101** (3), 518–538.

PRAGER, R. A. and HANNAN, T. H. (1998). Do substantial horizontal mergers generate significant price effects? Evidence from the banking industry. *The Journal of Industrial Economics*, **46** (4), 433–452.

RAMSEY, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, **31** (2), 350–371.

REUBEN, E. and SUETENS, S. (2012). Revisiting strategic versus non-strategic cooperation. *Experimental Economics*, **15** (1), 24–43.

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70** (1), 41–55.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66** (5), 688.

— (1977). Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational and Behavioral statistics*, **2** (1), 1–26.

— (2001). Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services and Outcomes Research Methodology*, **2** (3), 169–188.

— (2004). On principles for modeling propensity scores in medical research. *Pharmacoepidemiology and drug safety*, **13** (12), 855–857.

SAARIMAA, T. and TUKIAINEN, J. (2010). Coalition Formation and Political Decision Making: Evidence from Finnish Municipal Mergers. *Mimeo.*

Sabater-Grande, G. and Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: an experimental test. *Journal of Economic Behavior & Organization*, **48** (1), 37–50.

Scherer, F. M. (1967). Focal point pricing and conscious parallelism. *Antitrust Bulletin*, **12** (2), 495–504.

Schumann, L., Rogers, R. P. and Reitzes, J. D. (1992). Case studies of the price effects of horizontal mergers. *Federal Trade Commission, Bureau of Economics Staff Report.*

Sen, P. K. (1980). Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **52** (2), 203–218.

Smith, J. A. and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, **125** (1), 305–353.

Spanish Competition Authority (2009). Resolución Expte., S/0053/08, FIAB Y ASOCIADOS Y CEOPAN.

Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. and Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health services & outcomes research methodology*, **14** (4), 166–182.

Sutton, J. (1998). *Technology and Market Structure.* Cambridge, Mass.: MIT Press.

Tan, F. and Yim, A. (2014). Can strategic uncertainty help deter tax evasion? An experiment on auditing rules. *Journal of Economic Psychology*, **40**, 161–174.

Thoemmes, F. J. and West, S. G. (2011). The Use of Propensity Scores for Nonrandomized Designs With Clustered Data. *Multivariate behavioral research*, **46** (3), 514–543.

Van Huyck, J. B., Battalio, R. C. and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, **80** (1), 234–248.

VanderWeele, T. J. (2008). Ignorability and stability assumptions in neighborhood effects research. *Statistics in Medicine*, **27** (11), 1934–1943.

— (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological methods & research*, **38** (4), 515–544.

VERBOVEN, F. and BJÖRNERSTEDT, J. (2016). Does Merger Simulation Work? Evidence from the Swedish Analgesics Market. *American Economic Journal: Applied Economics*, **8** (3), 125–164.

VILLAS-BOAS, S. B., TUROLLA, S., CHAMBOLLE, C. and ALLAIN, M.-L. (2016). The Impact of Retail Mergers on Food Prices: Evidence from France. *Mimeo.*

VOGLER, S., ESPIN, J. and HABL, C. (2009). Pharmaceutical pricing and reimbursement information (PPRI) – new PPRI analysis including Spain. *Pharmaceuticals Policy and Law*, **11** (3), 213–234.

WEINBERG, M. (2007). The price effects of horizontal mergers. *Journal of Competition Law and Economics*, **4** (2), 433–447.

WELLFORD, C. P. (2002). Antitrust: Results from the Laboratory. In C. A. Holt and R. Isaac (eds.), *Experiments investigating market power*, *Research in Experimental Economics*, vol. 9, Amsterdam: JAI Elsevier, pp. 1–60.

WERDEN, G. J., JOSKOW, A. S. and JOHNSON, R. L. (1991). The effects of mergers on price and output: Two case studies from the airline industry. *Managerial and Decision Economics*, **12** (5), 341–352.

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

WOOLDRIDGE, J. (2012). *Introductory econometrics: A modern approach.* Mason, Ohio: South-Western, Cengage Learning.

WOOLDRIDGE, J. M. (2010). *Econometric analysis of cross section and panel data.* Cambridge, Mass.: MIT press, 2nd edn.

— (2013). *Introductory econometrics: A modern approach.* Mason, OH: South Western, Cengage Learning, 5th edn.

WYSS, R., GIRMAN, C. J., LOCASALE, R. J., BROOKHART, A. M. and STÜRMER, T. (2013). Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study. *Pharmacoepidemiology and drug safety*, **22** (1), 77–85.

YOUNG, H. P. (1993). The Evolution of Conventions. *Econometrica*, **61** (1), 57–84.

— (2007). The possible and the impossible in multi-agent learning. *Artificial Intelligence*, **171** (7), 429–433.

Zeileis, A. (2004). Alternative boundaries for CUSUM tests. *Statistical Papers*, **45** (1), 123–131.

— (2005). A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals. *Econometric Reviews*, **24** (4), 445–466.

—, Leisch, F., Hornik, K. and Kleiber, C. (2002). strucchange. An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, **7** (2), 1–38.

—, —, Kleiber, C. and Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, **20** (1), 99–121.

Zhao, Z. (2008). Sensitivity of propensity score methods to the specifications. *Economics Letters*, **98** (3), 309–319.

Zivot, E. and Andrews, D. W. K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*, **10** (3), 251–270.