Empirical studies in tourism, public transportation and the construction industry

Dissertation

Zur Erlangung der Würde eines
Doktors der Wirtschafts- und Sozialwissenschaften
Vorgelegt der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität Freiburg (Schweiz)

von

Hannes Lukas Wallimann
aus Alpnach OW

Genehmigt von der Wirtschafts- und Sozialwissenschaftlichen Fakultät
am 20.12.2021 auf Antrag von

Herrn Prof. Martin Huber, Ph.D. (erster Referent) und
Frau Prof. Dr. Costanza Naguib (zweite Referentin)

Freiburg, 2021

Mit der Annahme einer Dissertation beabsichtigt die Wirtschafts- und Sozialwissenschaftliche Fakultät der Universität Freiburg nicht, zu den darin enthaltenen Meinungen des Verfassers Stellung zu nehmen.

Preface

First and foremost, I would like to thank my supervisor, Martin Huber, for his support and guidance. His advice contributed substantially to the present thesis. In particular, I thank him for his positive encouragement and his enthusiasm for empirical research. As co-author and supervisor, working with Martin inspired me professionally and personally and motivated me to continue research in the future. Special thanks also go to Costanza Naguib, who agreed to act as the second supervisor for this thesis. Moreover, I wish to thank my faculty colleagues at the University of Fribourg for their valuable comments and suggestions, for instance, during various PhD seminars of the Department of Economics.

My employer, the Institute for Tourism and Mobility at the Lucerne University of Applied Sciences, enabled me to work as a research associate while pursuing research for my thesis. The topics of this thesis emerged from my involvement with the Institute of Tourism and Mobility. On multiple occasions, my colleagues at the Competence Center Mobility provided me with valuable advice and stimulating ideas. I especially want to thank Widar von Arx, my direct superior, as well as Philipp Wegelin, Timo Ohnmacht and Vu Thi Thao.

I owe special gratitude to my other co-authors, David Imhof and Jonas Meier. The former had already inspired me during an internship at the Swiss Competition Commission. His critical thoughts during extensive discussions improved our joint work considerably. The other co-author, Jonas Meier, I would like to thank for his enriching teamwork and for sharing his opinions with me. I also gratefully acknowledge the helpful assistance of various kinds provided by Kevin Blättler, Christoph Brunner, Pierre Chevalier, Lisa Dang, Bastian Gschwendtner, Sebastian Imhof, Yavuz Karagök, Reto Lüscher, Blaise Melly, Angela Steffen, Jürg Stettler, Philippe Sulger and Niklaus Wallimann.

Finally, I must express my very profound gratitude to my family. I am grateful to my wife Maria Blum for her unconditional support and understanding, despite her not being really interested in my thesis topics. This achievement would have been impossible without her. I would also like to

thank our two lovely children, Lia Leona and Mats Florian, naturally too young to understand what their father has been working on, for enlightening even my most demanding days. Their endless support has been priceless. Moreover, I would like to thank my parents, Heidi Durrer and Leo Wallimann, my sister Lena Wallimann, my aunt Verena Durrer and my mother-in-law Christine Blum. The functioning familial environment they create has been an invaluable foundation for the enriching journey of writing this thesis.

August 2021 Hannes Wallimann

Contents

List of Figures

List of Tables

Introduction

Empirical methods use structured observations from the real world to draw conclusions. It is impossible to imagine economic research without such methods. Empirical research methods can be categorized as either qualitative or quantitative. The former gather information from non-numerical measurements, while the latter use quantitative measurements (mostly numbers). Quantitative economic research can fulfill two different purposes. First, they can indicate causal effects, which usually depend on understanding a counterfactual: What happens with and without a 'treatment'. For instance, in winter sports destinations, knowledge of differences in demand with (treatment) and without (counterfactual) a discount on ski lift passes is important in helping companies decide their pricing strategies. Second, researchers discuss 'prediction policy problems' (Kleinberg, Ludwig, Mullainathan, and Obermeyer, 2015) in order to predict the probability of an event. For instance, the prediction of economically harmful cartels is important in helping competition authorities decide which markets to examine more closely.

This PhD thesis fulfills both purposes of quantitative economic research mentioned above by studying prices in different contexts and research fields, namely tourism, public transportation and the construction industry. The thesis is organized as a collection of four independent chapters. Chapter 1 provides a causal effect, i.e., the effect of a radical discount on ski-lift passes on other companies in a winter sports destination. In recent years, a rapidly growing economic literature has begun using machine learning, a subfield of artificial intelligence. The remaining three chapters add to this literature by analyzing data using predictive and causal machine learning algorithms. Chapters 2 and 3 discuss prediction policy problems by predicting different forms of bid-rigging cartels in the construction industry. Finally, Chapter 4 combines both purposes by assessing the demand effects of railway tickets. The remainder of this introduction summarizes the broader context of each chapter, provides a general description of the methods applied and describes the key results.

Chapter 1 investigates the effect of radical discounts on ski lift passes on accommodation businesses in the same winter sports destination.[1] The interest here is because tourists' choices of a winter sports destination not only depend on the individual supply of a particular company, such as a hotel, but also on the availability of attractive complementary products offered by other companies in the value network of a winter sports destination. Therefore, one company's campaign or attractive offer, such as a price discount, can positively affect partners in the same value network. Since around the turn of the millennium, winter sports destinations in Switzerland have suffered from a decline in visitor numbers. To reverse this decline in demand, in 2016 the ski lift company in one destination, Saas-Fee, offered a yearly ski-lift pass for 222 Swiss francs (the WinterCARD), this being a radical price discount of 80% compared to the previous season. However, the decision to implement it was made independently, and the campaign itself was characterized by weak cooperation, as other businesses in the destination did not contribute to the costs.

To assess the effect of the radical price discount on other partners in the Saas-Fee value network, I use the synthetic control method (see Abadie, Diamond, and Hainmueller, 2010). This method allows comparisons between a single treated unit (Saas-Fee) and a control group (other winter sports destinations in Switzerland). The methodology uses a data-driven procedure to construct a synthetic control unit that adequately mimics the outcome the treated unit would have experienced in the absence of the treatment (WinterCARD). The synthetic control unit is created out of existing units and is not chosen by researchers a priori. Athey and Imbens (2017) state that the synthetic control approach is *"the most important innovation in the policy evaluation literature in the last 15 years"* (p. 9). To the best of my knowledge, this is the first time in tourism research that the method has been used to investigate the impact of a new pricing strategy. The results indicate that a company's practice of radically discounting the prices of its seasonal lift passes had a positive impact on accommodation businesses in the same winter sports destination. More precisely, the impact amounted to about 35% additional overnight stays by domestic tourists per winter season, a seasonal increase of about 32,000 overnight stays. On the other hand, the ski lift company earned its main turnover from transportation. Therefore, by the end of 2018 it concluded that, despite rising frequencies, the campaign had been unsuccessful. Thus, in this chapter, I also show the difficulties associated with such new pricing strategies and emphasize the importance of cooperation between independent companies active in the same destination.

[1] This chapter is based on a paper published in the *Tourism Economics*. It is published as Wallimann (2020).

However, cooperation between firms does not serve the general public's interest in every case. For example, in markets, the joint interests of rival firms are best served by cartelizing the market and keeping prices high. Here, the problem is that the success of the cooperating firms harms the general public's interest: their cooperation leads to reductions in consumer and total welfare and creates inefficiencies. Therefore, governments get into the game and enact antitrust laws making collusion between firms illegal. In Switzerland, for instance, in 2004 the parliament revised the federal Cartel Act and introduced a sanction regime to prevent economically harmful cooperation of rival firms. In order to discover collusive agreements, competition agencies can use insightful information from, for example, compliance programs. However, to reduce their dependence on such sources, researchers have proposed using statistical methods to screen markets. This screening is the first phase of a multi-stage process that may condemn firms cooperating illegally. Chapters 2 and 3, which analyze data from the construction industry, add to the most recent literature on dismantling cartels with predictive machine learning.

The use of machine learning algorithms entails a set of predictors, also features or covariates, to predict an outcome. Implementing such powerful algorithms requires data to be randomly split into independent training and test data. Then predictive models are developed in the training data, where both covariates and outcomes are observed. Finally, the models predict the outcomes in the test data for each observation on the basis of their covariates, so-called 'out of sample predictions'. Machine learning algorithms aim to achieve goodness of fit in the test set by minimizing deviations between the predicted and actual outcomes. In recent years, machine learning algorithms have drawn increasing attention from economists. Varian (2014), for example, states *"I believe that these methods have a lot to offer and should be more widely known and used by economists"* (p. 3).

Chapter 2, a joint work with David Imhof and Martin Huber, proposes a machine learning approach to flag bid rigging, which is particularly useful for detecting incomplete bid-rigging cartels.[2] This approach is essential, as the reality is frequently characterized by a situation in which competitive bidders participate in markets in which a cartel is active. Thus, in public procurement tenders, one observes incomplete cartels due to the presence of competitive bidders. In such a case, the statistical pattern produced by bid rigging is contaminated, rendering its detection more challenging. The approach classifies tenders as collusive using screens as predictors, i.e., statistics derived from the distribution of bids in a tender. The methodological innovation of

[2]Chapter 2 is based on a working paper. It is published as Wallimann, Imhof, and Huber (2020).

this chapter consists in calculating the predictors for a tender in two steps: first, screens are calculated for all possible subgroups of three and four bids in a tender; and second, the summary statistics thereof, i.e., median, mean, maximum and minimum, are used. To evaluate the method's performance, we analyze Swiss data on the Ticino, See-Gaster and Graubünden cartels, in which the incidence of collusive and competitive tenders is known. Our original detection method increases the correct classification rate of collusion vs. competition of previously suggested methods by 3 to 10 percentage points for incomplete bid-rigging cartels.

Chapter 3, a joint work with David Imhof, presents a method for detecting collusive groups of firms.[3] A wide range of methods focuses on discovering tenders affected by bid rigging. Using such a method, agencies must also build concrete suspicions against specific firms in order to flag candidates for further investigations, e.g., house searches. The method we propose considerably simplifies this step and allows cartels or collusive coalitions to be identified directly. Coalitions are formed with three firms, as we aim to discover the smallest possible cartels. Our approach isolates all tenders in which the three firms of a coalition submitted a bid. Then screens (again, statistics derived from the distribution of bids in a tender) exclusively based on the bids of those three firms are calculated for each tender, constituting the so-called 'tender-based screens'. The method proposes a further step by calculating the summary statistics of all the tender-based screens of a coalition. These statistics, so-called 'coalition-based screens', synthesize the distributional features of a specific coalition's bids. We use these coalition-based screens as predictors to flag bid rigging. Using Swiss, Japanese and Italian procurement data, our method (out of sample) correctly classifies 90% of collusive and competitive coalitions for each data set. Moreover, since the auction settings differ in these three countries – i.e., first-price sealed bid procurement mechanisms in Japan and Switzerland, and mean-price sealed bid auctions in Italy – the method applies more generally. Finally, comparing different machine learning algorithms, we find that the super learner (see van der Laan, Polley, and Hubbard, 2008) outperforms the other algorithms in two out of three data sets, making it advisable in our case.

Predictive machine learning algorithms, which we deploy in Chapters 2 and 3, can deal with many predictors and provide precise out of sample predictions. However, predictive machine learning does not frame itself as solving estimation problems, even though the econometric and economic

[3] Chapter 3 is based on a manuscript accepted for publication at the *International Review of Law & Economics* and on a working paper. The working paper is published as Imhof and Wallimann (2021).

literature often focuses on questions that go beyond such problems. More precisely, and as already noted, academics are interested in causal effects. Causal machine learning, being an evolving and promising research field in recent years, combines the theory of causal inference with the advantages of machine learning (see, for instance, Athey, 2019). The last chapter of this thesis contains both predictive and causal machine learning.

Chapter 4, a joint work with Martin Huber and Jonas Meier, assesses the demand effects of discounts on train tickets, so-called 'supersaver tickets', issued by the Swiss Federal Railways.[4] Using supersaver tickets, customers in Switzerland can travel on long-distance public transport routes with a discount of up to 70%. From a business analytics perspective of the railway industry, understanding the demand effects of discounts is primarily relevant for improving the allocation of public transport users. This is due to two facts: capacity constraints in the rush hour, and the low average capacity utilization of the trains. Furthermore, measuring the effectiveness of discounts is important for policy-makers, as taxpayers subsidize public transport. In this study, we provide two cases of business analytics in the railway industry. For this purpose, we combine a unique survey-based sample of buyers of supersaver tickets with data relevant to the supply and calculation of discounts.

In the first use case, we investigate which characteristics are important to predict buying behavior when offering a supersaver ticket. Buying behavior denotes rescheduling a trip (demand shift), booking a trip otherwise not realized by train (additional trip) and buying a first- rather than second-class ticket (upselling). The machine learner random forest (see Breiman, 2001) suggests that demand-related information for a specific connection (like departure time and utilization), the discount level and customer's age are important features to predict buying decisions. The algorithm obtains correct (out of sample) classification rates of 58%, 65% and 82% for demand shift, additional trip and upselling respectively.

In the second use case, we assess the impact of the discount rate on rescheduling a trip among 'always buyers', those who would have booked a journey even at the regular fare. The focus on always buyers is essential, as consumers buying a supersaver ticket with a higher discount generally differ from those who would already buy it at a lower discount in terms of, for instance, their reservation price. On the other hand, the 'always buyers' are homogeneous in terms of their buying decisions. To estimate the causal effects, we assume that (i) the discount rate is quasi-random, conditional on our covariates; and (ii) buying decisions cannot decrease in the discount rate.

[4]Chapter 4 is based on a working paper. It is published as Huber, Meier, and Wallimann (2021). We are grateful to the SBB Research Fund for financial support.

The second assumption we can scrutinize in the data, there being no evidence for its violation.

Our findings are based on the causal machine learning algorithms causal forest (see Wager and Athey, 2018, Athey, Tibshirani, and Wager, 2019) and double machine learning (see Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018). The causal forest finds that, on average, increasing the discount rate by one percentage point increases the share of rescheduled trips by 0.16. The double machine learning approach suggests that discount rates of 30% and more (relative to discounts of less than 30%) on average increase the share of rescheduled trips by 3.8 percentage points. Finally, investigating the effect heterogeneity across pre-selected observables using the causal forest suggests that the effects are more than 0.2 percentage points higher for leisure travelers and during peak hours when controlling several other characteristics.

Chapter 1

A complementary product of a nearby ski-lift company

*

Abstract

The availability of attractive complementary products offered by a nearby company positively affects other companies in the value network of a winter sports destination. We aim to empirically illustrate the positive effects of the campaign of a local ski-lift company in Switzerland on other companies in the same value network. For the first time in tourism research, the synthetic control method is used to investigate the impact of a new pricing strategy. In this case, the company's practice of radically discounting the prices of its seasonal lift passes had a positive impact on accommodation businesses in the same winter sports destination. The impact amounts to about 35% additional overnight stays by domestic tourists per winter season, a seasonal increase of 32,000 overnight stays. However, the ski-lift company concluded that the campaign had been unsuccessful. We, therefore, discuss the difficulties of such a new pricing strategy and emphasize the importance of cooperation between independent companies in the wider destination area.

*This chapter is based on a paper published in the *Tourism Economics*. It is published as Wallimann (2020).

1.1 Introduction

Winter tourism is of prime economic importance to mountain areas (Gonseth, 2013, Koenig and Abegg, 1997, Lohmann and Crasselt, 2012, Moreno-Gené, Sánchez-Pulido, Cristobal-Fransi, and Daries, 2018, Steiger, Scott, Abegg, Pons, and Aall, 2019). Winter sports destinations are characterized by having a number of different independent companies, directly and indirectly, related to tourism. These companies sell complementary products and, because of sales interdependencies between them, form a complementary value network (Baggio, 2011, Lohmann and Crasselt, 2012, Pechlaner, Presenza, and Cipollina, 2010, Wyss, Luthe, and Abegg, 2015). Indeed, tourists' choices of winter sports destinations depend not only on the individual supply of a particular company but also on the availability of complementary products offered by nearby companies (Rigall-I-Torrent and Fluvià, 2010). Therefore, one company's campaign or attractive product can positively affect other companies in the same complementary value network.

As Falk and Steiger (2020) point out, ski lift companies are the backbone of winter tourism, as they transport visitors to the mountain tops. Therefore, the provision of attractive products by a nearby ski lift company impacts the tourist demand in a winter sports resort. In this article, we empirically illustrate the effects of a local ski lift company's campaign on other companies in the value network of the winter sports destination of Saas-Fee, a high-altitude mountain resort in Switzerland. The ski lift company in Saas-Fee tried to reverse a declining demand by offering a radical price discount of 80% on its seasonal lift passes. The decision to do this was made independently, without cooperation with other partners in the Saas-Fee value network. Called WinterCARD, in 2016, the ski pass was offered for 222 Swiss francs.[1] The price of the WinterCARD changed only slightly in the next two seasons, and the campaign lasted three seasons altogether. However, after declines in turnover and earnings, at the end of 2018, the ski lift company concluded that the campaign had been unsuccessful.

In empirically illustrating the effects of the campaign of a local ski lift company on other companies in a value network, we estimate how the WinterCARD affected accommodation businesses in Saas-Fee. As far as we are aware, this is the first time its impact has been estimated for all three winter seasons (2016/2017, 2017/2018, and 2018/2019) in which the WinterCARD was available. Therefore, we use the synthetic control method (Abadie, Diamond, and Hainmueller, 2010, Abadie and Gardeazabal, 2003, Cavallo, Galiani, Noy, and Pantano, 2013), hereafter referred to as SCM, to estimate the causal

[1]This was equivalent to US$ 216 in November 2016.

effect of the WinterCARD on overnight stays in hotels in this resort. The methodology uses a data-driven procedure to construct a synthetic control municipality that adequately reproduces the overnight stays that Saas-Fee would have experienced in the absence of the WinterCARD. Although the SCM has been used in several tourism studies (Addessi, Biagi, and Brandano, 2019, Biagi, Brandano, and Pulina, 2017, Castillo, Figal Garone, Maffioli, and Salazar, 2017, Doerr, Dorn, Gaebler, and Potrafke, 2020, Tkalec, Zilic, and Recher, 2017), the present article is the first in tourism research to use the SCM to investigate the effect of a new pricing strategy of a ski lift company on its partners in a destination's value network.

We find a remarkable impact on accommodation businesses in Saas-Fee of about 35% additional overnight stays by domestic tourists per winter season, a seasonal increase of 32,000 overnight stays. The results for the first two seasons remain robust when we perform several robustness investigations. However, not all robustness tests confirm the third season. By extrapolating the estimates for the first two seasons, we find that the WinterCARD has a considerable economic effect on other companies in the Saas-Fee value network, estimating a value added to be about CHF 25 million. The value added by the WinterCARD includes spending for products and services in the destination minus the value of the intermediate consumption needed to produce these products and services.

The remainder of this article is organized as follows. In section 1.2, we briefly discuss the existing literature on price strategies, tourism value networks, and cooperation, as well as the provision of public goods within such networks. Section 1.3 describes the context of this case study. In section 1.4, we introduce the methodology employed and its implementation. Section 1.5 documents the municipality-level panel data for the whole of Switzerland used for the article. Section 1.6 presents the main results of the empirical analysis and assesses the robustness of our results. Finally, in section 1.7, we discuss the economic effects of the WinterCARD on the value network in Saas-Fee. The article ends with a conclusion.

1.2 Literature review

Pricing strategies play a crucial role in tourism research. For ski lift companies, a distinction must be made between pricing strategies for seasonal passes and strategies focussing on single-day or multi-day tickets. As a famous example of the second strand, dynamic pricing approaches to maximize revenues are frequent strategies in the hotel and airline industries (Abrate, Fraquelli, and Viglia, 2012). Some ski lift companies have started to experiment with a more dynamic approach to pricing as a means of increasing their operating profits (Malasevska and Haugom, 2018). Another strategy in the tourism industry to increase sales and attract guests are price discounts (Becerra, Santaló, and Silva, 2013, Nusair, Yoon, Naipaul, and Parsa, 2010, Yang, Zhang, and Mattila, 2016). Attracting additional visitors is also pursued by ski lift companies introducing radical price discounts on seasonal lift passes. Examples of such radical price discounts on seasonal passes include those introduced in Colorado (Perdue, 2002) and the present case from Switzerland (see also Falk and Scaglione, 2018).

Usually, the ski lift companies compete for guests by applying different pricing strategies. However, horizontal cooperation among ski lift companies in horizontally differentiated markets can be used to avoid this competition between different providers. In doing so, the market power of the ski lift companies involved can also be increased. Cooperation among ski lift companies that have formed an alliance can be used for both selling common single-day or multi-day tickets (Firgo and Kügler, 2018) and seasonal passes (Falk and Scaglione, 2021). A ticket or season pass sold by one ski lift company in an alliance is accepted on a mutual basis by other alliance members.

The price strategy and related demand are not only of interest to the ski lift company itself but also to other service providers in a winter sports destination, with which the ski lift companies form a complementary value network because of sales interdependencies (Flagestad and Hope, 2001). Several studies show that intercompany cooperation in a tourism value network creates a competitive advantage (Beritelli, 2011, Machiavelli, 2001, Thao, von Arx, and Frölicher, 2020, Wilke, Costa, Freire, and Ferreira, 2019, Zehrer and Hallmann, 2015). This is because independent companies in destinations sell complementary products. The importance of cooperation within the value networks of winter sports destinations is supported by several studies (Pikkemaat and Weiermair, 2007, Svensson, Nordin, and Flagestad, 2005). Lohmann and Crasselt (2012) argue that if the snow sports company increases the resort's attractiveness by investing in new infrastructure, this will have a positive impact on accommodation businesses. If the accommodation business does not compensate for the cost, this situation typically creates a problem of underinvestment.

Another stream of literature on which this article draws, linked to the underinvestment problem, examines the provision of public goods (Samuelson, 1954) in tourist destinations (Eppen, Hanson, and Martin, 1991, Garrod and Fyall, 2017, Rigall-I-Torrent and Fluvià, 2007, 2010, Saló, Garriga, Rigall-I-Torrent, Vila, and Fluvià, 2014). Tourists' choices of the destination in which they spend their holidays depend on a company's supply but also on the availability of complementary products offered by nearby companies.

Complementary products can be understood as public goods because they are characterized by a certain degree of non-rivalry (the cost of positively affecting another company is zero) and non-excludability (it is not possible to exclude other companies from the positive effects of the product). The existence of a public good in a value network has a positive external effect on its members. However, this effect is hard to measure and therefore difficult to internalize outside the context of reciprocal shareholdings. In the case of independent companies, the members of a value network may not have enough incentives to contribute to demand-enhancing investments, such as the provision of a price discount for seasonal passes. In the following section, we discuss the radical price discount of the ski lift company in the Saas-Fee destination as an example of a public good in a tourism value network.

1.3 Background

According to Stettler, Zemp, and Steffen (2015), in the 1990s, Saas-Fee was a leading alpine destination not only in Switzerland but also in Europe. The ski slopes at this destination start at 3,600 m above sea level. This, plus the presence of a glacier, ensures that Saas-Fee has snow throughout the winter season. In recent years, however, winter tourism in Switzerland has suffered a decline in visitor numbers (Seilbahnen Schweiz, 2017, Steiger, Scott, Abegg, Pons, and Aall, 2019), Saas-Fee being particularly affected by this trend (Falk and Scaglione, 2018, Stettler, Zemp, and Steffen, 2015). Frequencies on the ski slopes in Saas-Fee fell by one-third between 2006 and 2016 (Saastal Bergbahnen AG, 2017), a decrease of 170,000 visitors. The local ski lift company generated a turnover from tickets sales of about 15 million Swiss francs in the 2015/2016 winter season (Saastal Bergbahnen AG, 2016).

As already noted, in 2016, the ski lift company introduced the WinterCARD, a yearly ski lift offered for 222 Swiss francs. That amounted to a radical price discount of 80% on ski lift passes, which the ski lift company hoped would reverse the decline in visitor numbers. In addition to the yearly season pass, a three-season lift pass for CHF 622 and a 15-year pass for CHF 2,999 were also offered. Implementation of the WinterCARD and its marketing costs amounted to about 4 million Swiss francs (Saastal Bergbahnen AG, 2017). The price of the 1-year passes remained the same in the 2017/2018 season. In 2018/2019, it was increased marginally to 233 Swiss francs.

For other members in the Saas-Fee value network, the WinterCARD came as an attractive complementary product since their customers could also benefit from this offer. However, the ski lift company earned its main turnover from transport, and therefore the company only profited from one-time contributions of 222–233 Swiss francs per customer. Furthermore, the campaign was characterized by weak cooperation, as companies in Saas-Fee other than the ski lift company did not contribute to the costs of the WinterCARD (Saastal Tourismus AG, 2019). Despite rising frequencies, moreover, the ski lift company concluded that the introduction of the WinterCARD had been unsuccessful due to declines in turnover and earnings by the end of 2018 (Saastal Bergbahnen AG, 2018). The operating loss amounted to CHF 4.31 million for the 2017/2018 winter season (Saastal Bergbahnen AG, 2018). Therefore, the WinterCARD campaign ran only for three seasons, 2016/2017 to 2018/2019.

Several other ski lift companies in Switzerland reacted to the campaign launched by the Saas-Fee company. As a first reaction, in the 2017/2018 winter season, 25 companies formed an alliance, whose members, mainly located close to Saas-Fee, introduced a joint ski lift pass called the Magic Pass for 359 Swiss francs (Falk and Scaglione, 2021) to replace their local passes. A Magic Pass owner can use all the ski lifts of the members of the alliance. During the 2018/2019 season, the alliance already contained more than 30 ski lift companies. Second, other ski lift companies in Switzerland switched to dynamic pricing for single-day and multi-day tickets. Famous examples are the ski lift companies of Arosa Lenzerheide and Andermatt Sedrun, which introduced dynamic pricing in the 2017/2018 season.

1.4 Methodology and Implementation

Identifying the effect of the WinterCARD on nearby accommodation businesses in Saas-Fee is equivalent to the overall topic of measuring a causal treatment effect. ‘Treatment’ means an event, a policy or a company decision (here, WinterCARD). In our article, we use the SCM (Abadie, Diamond, and Hainmueller, 2010, 2015, Abadie and Gardeazabal, 2003, Cavallo, Galiani, Noy, and Pantano, 2013) to measure the potential increase in overnight stays at Saas-Fee triggered by the WinterCARD.

The SCM allows comparisons to be made between a single treated unit (here, the municipality of Saas-Fee) and a control group (here, Swiss mountain municipalities other than Saas-Fee). The treated unit under investigation is the unit affected by the treatment. The methodology uses a data-driven procedure to construct a synthetic control unit (here, synthetic Saas-Fee) that adequately mimics the outcome the treated unit would have experienced in the absence of the policy, or the company’s decision. The synthetic control unit is created out of already-existing units and is not chosen by the researchers a priori (Biagi, Brandano, and Pulina, 2017). The SCM is explained in more detail in Appendix 1.A.

Applying the SCM relies on three key assumptions (Bouttell, Craig, Lewsey, Robinson, and Popham, 2018). The *first* assumption states that the outcomes of the treated and the synthetic control unit are similar in the absence of the policy or a decision by the company. The assumption is considered to be given if there exists a similarity between the treated and the synthetic control unit (Abadie and Gardeazabal, 2003, Abadie, Diamond, and Hainmueller, 2010, 2015). The SCM conditions on pre-treatment outcomes of the treatment unit and the synthetic control. However, in contrast to, for example, the difference-in-difference (DiD) approach, the SCM allows the impact of unobserved confounding characteristics with time. The *second* assumption is that the intervention has no spill-over effects, either positive or negative, affecting municipalities other than Saas-Fee. *Third*, we assume that there are no external shocks, other policies, or company decisions that might affect the outcome in the municipalities in the control group. We discuss the assumptions in more detail in the coming sections.

Falk and Scaglione (2018) estimate the effect of the WinterCARD in the 2016/2017 season by using the DiD approach. However, as we have only one treated unit, the choice of the control unit might be problematic (Tkalec, Zilic, and Recher, 2017). Therefore, applying a data-driven construction of a synthetic control unit using the SCM seems more suitable. Furthermore, as far as we are aware, our article is the first in which the effect is estimated for all three seasons during which the WinterCARD was available, making our conclusions regarding the campaign even more reliable.

Most studies applying the SCM use annual panel data sets (Abadie and Gardeazabal, 2003, Abadie, Diamond, and Hainmueller, 2010, Addessi, Biagi, and Brandano, 2019, Biagi, Brandano, and Pulina, 2017). In contrast to these studies, in estimating the effect of the WinterCARD, we use a monthly dependent variable. This approach is in line with the article by Castillo, Figal Garone, Maffioli, and Salazar (2017), investigating the effect of a state tourism development policy on employment in Argentina. Unlike the paper by Castillo, Figal Garone, Maffioli, and Salazar (2017), we fade out the seasonal effects – in our case, for example, the greater demand for winter sports destinations in February compared to December – by considering the difference from the pre-WinterCARD mean instead of the absolute values of overnight stays, the dependent variable of our study we present in the following section. A difference of a municipality, henceforth called the 'difference of overnight stays', is the difference between the monthly overnight stays and the mean of overnight stays in the municipality in the pre-WinterCARD period.

1.5 Data

As we aim to estimate how the WinterCARD affected accommodation businesses in Saas-Fee, using the SCM described in the methodological section above, overnight stays are our basic data source. The Swiss Federal Statistical Office (FSO) provides a survey of tourist accommodations. We use the FSO's monthly municipality-level panel data for overnight stays for the period January 2006 to April 2019. In our paper, we take the winter months as being December to April. Overall, we have data for 69 winter months, therefore covering 54 pre-WinterCARD and 15 WinterCARD periods.

We restrict ourselves to municipalities in Switzerland that are classified as mountain municipalities. Municipalities with incomplete data sets have been removed. Analyzing the effect of the WinterCARD, we assume that no external shocks were affecting the variable of interest in the municipalities in the control group. However, we retain in our control group to ski lift companies that are members of the Magic Pass alliance or that sell single-day and multi-day tickets with dynamic pricing. This, because the price for the Magic Pass is still 50% higher than for the WinterCARD, and dynamic pricing is not applied to season passes. To test for this assumption, in the robustness analysis, we re-estimate the effect without ski lift companies being members of the Magic Pass alliance. Using the SCM, we compare overnight stays in Saas-Fee with overnight stays in all other mountain municipalities in Switzerland other than Saas-Fee, the latter constituting a control group of 88 municipalities. All the municipalities in the control group are listed in Appendix 1.B.

In this research, we take into account overnight stays by domestic but not foreign tourists. The reason for this is the marketing campaign: the WinterCARD was mainly advertised within Switzerland. Table 1.1 gives descriptive statistics for monthly overnight stays in hotels in the pre-WitnterCARD and WinterCARD periods, respectively.[2] The mean of overnight stays by domestic tourists in Saas-Fee amounts to 17,655 before the introduction of the WinterCARD. This contrasts with a lower mean for all municipalities in the control group of 6,656. Conversely, the maximum number of overnight stays in the control group in the pre-WinterCARD period, amounting to 78,855, is higher than the maximum in Saas-Fee. Notably, the mean in Saas-Fee increases to 25,610 monthly overnight stays by domestic tourists after the introduction of the WinterCARD. Simultaneously, the mean for the control group increases to 7,057.

[2] The descriptive statistics for the differences in overnight stays by domestic tourists are presented in Appendix 1.D.

As independent variables, we take the numbers of ski lifts, ski-runs in kilometers, and average elevation per skiing area closest to the municipality from commercial sources (ADAC Ski Atlas; website www.bergfex.com). For federally licensed ski lifts, we have data from the Swiss Federal Office of Transport. We use the distance from Swiss cities with more than 50,000 residents in travel minutes by car and public transport from Google Maps. Regarding the snow dependence of winter tourism (Falk, 2010), we assign our observations to the snow height measured halfway through a month at the nearest station of the Global Climate Observing System (GCOS).[3] In addition, we have FSO data for the number of hotels per municipality. In Appendix 1.C, we present a table with descriptive statistics of the independent variables for all municipalities.

Table 1.1: Descriptive statistics for monthly overnight stays by domestic tourists in Saas-Fee and the control group

Overnight stays (per month)	Mean	SD	Min.	Max.	N
Pre-WinterCARD period					
Saas-Fee	17,656	4,658	7,392	25,791	54
Control group	6,656	9,951	13	78,855	4,752
WinterCARD period					
Saas-Fee	25,610	7,697	13,679	36,327	15
Control group	7,057	11,408	38	88,816	1,320

1.6 Results

In this section, we first present the results obtained from the SCM application to the case of Saas-Fee. In our application, we use the synth package of Abadie, Diamond, and Hainmueller (2011) and the synth_runner package of Galiani and Quistorff (2017) for the statistical software STATA. Secondly, we investigate the robustness of our results.

1.6.1 The effect of the WinterCARD on accommodation businesses

Using the SCM, we construct a synthetic Saas-Fee that best represents Saas-Fee. Therefore, we calculate the root mean squared prediction error, hereafter

[3] A few observations are not available. In these cases, we have assigned the values to the second next station.

referred to as the RMSPE, between Saas-Fee and the synthetic Saas-Fee before the WinterCARD was introduced. To minimize our RMSPE, we use both the independent variables presented in the section above and the data on pre-WinterCARD overnight stays. We discuss this process in detail in Appendix 1.E.

Table 1.2: Municipality weights for the synthetic Saas-Fee

Municipality	Weights
Arosa	0.005
Davos	0.029
Engelberg	0.281
Hasliberg	0.313
Interlaken	0.072
Leukerbad	0.145
Meiringen	0.001
Zermatt	0.153

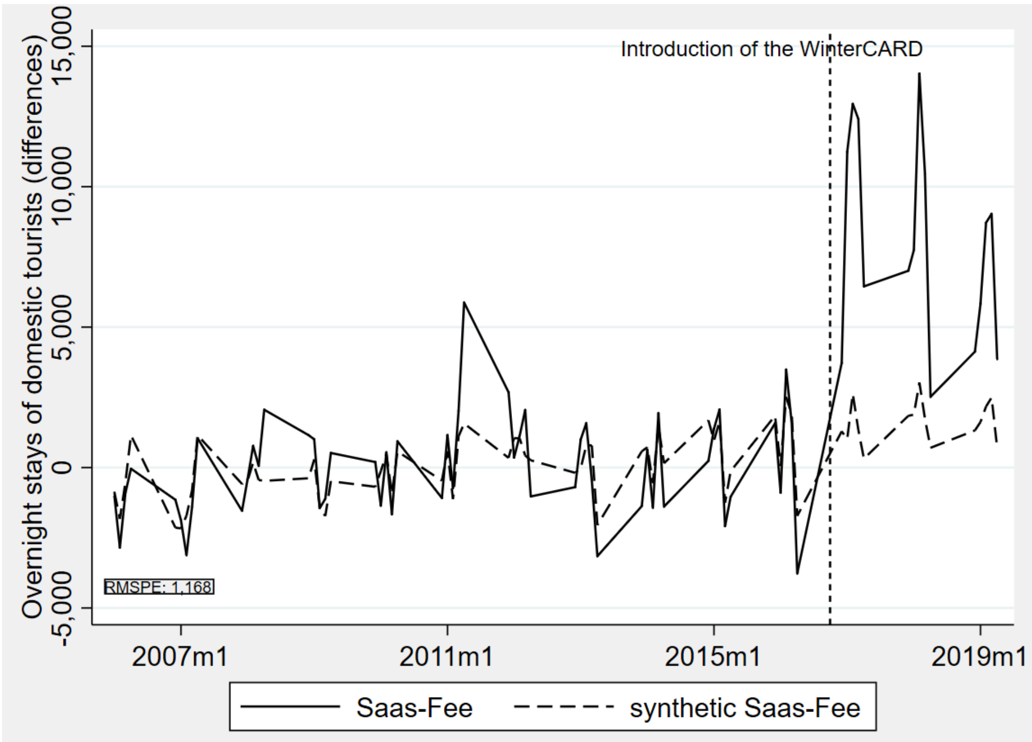
Table 1.2 gives the weights of each control municipality with a weight higher than zero. All the other 80 municipalities receive a weight of zero. The weights indicate that the trends in the differences of overnight stays by domestic tourists in Saas-Fee in the pre-WinterCARD period are best represented by a combination of the Arosa, Davos, Engelberg, Hasliberg, Interlaken, Leukerbad, Meiringen, and Zermatt municipalities. Hasliberg receives the highest weight, amounting to 0.313.

Figure 1.1 plots the differences in overnight stays by domestic tourists for Saas-Fee and the synthetic Saas-Fee during January 2006 and April 2019. As can easily be observed, the two trajectories track each other close in the entire pre-WinterCARD period (pre-WinterCARD RMSPE: 1,168).[4] This suggests that the synthetic Saas-Fee provides a sensible approximation of the number of differences of overnight stays that would have been booked in Saas-Fee in the absence of the price discount. Thus, we conclude that the first assumption of the SCM, namely that the treated and synthetic control unit are similar in the absence of the WinterCARD, is not violated.

To outline the effect of the price discount on overnight stays by domestic tourists in Saas-Fee, we compare the differences in overnight stays by domestic tourists in Saas-Fee and the synthetic Saas-Fee for the period the WinterCARD was available. In Figure 1.1, the two lines diverge after the

[4] See also Table 1.7 in Appendix 1.F giving the figures for the outcome paths for the pre-WinterCARD period for Saas-Fee and synthetic Saas-Fee.

Figure 1.1: Trends in differences in the overnight stays by domestic tourists in Saas-Fee and synthetic Saas-Fee



introduction of the radical price discount. While the two lines have a common trend before the intervention, Saas-Fee has an apparent upward trend in the WinterCARD period. The discrepancy in the differences in overnight stays by domestic tourists suggests a large positive effect of the massive price discount. Furthermore, it can be assumed that the effect is lowest in the third season. This could be due to the introduction of the Magic Pass, which in this case could have affected the municipalities in the control group. However, we assume that this is not the main reason, as we do not assign weights to a member of the Magic Pass alliance.[5] Another explanation could be that the good 2018/2019 winter season positively affected all ski lift companies, not only that in Saas-Fee.

Table 1.3: Effect of the WinterCARD in Saas-Fee (rounded)

Per month	6,500	December	3,500
Per season	32,00	January	6,500
2016/17 season	40,000	February	9,000
2017/18 season	32,500	March	9,000
2018/19 season	23,500	April	3,500

Table 1.3 summarizes the key effects of the WinterCARD on overnight stays by domestic tourists in Saas-Fee. The effect amounts to an average of about 6,500 additional overnight stays per month in the Saas-Fee municipality in the 2016/2017, 2017/2018, and 2018/2019 winter seasons. The average seasonal effect of 32,000[6] overnight stays amounts to an increase of 35% compared to overnight stays by domestic tourists in the 2015/2016 winter season in the Saas-Fee municipality, the season before the price discount was introduced. In addition to comparing the different seasons, thanks to our monthly approach, we can differentiate between the effects on a monthly basis. This can be valuable information, as there may exist differences in the monthly price elasticity of demand. The effect is greatest in February and March, with an average of 9,000 additional overnight stays by domestic tourists.[7] Furthermore, as also shown graphically in Figure 1.1, the effect, with 23,500 additional overnight stays, is the lowest in the 2018/2019 season.

[5] Leukerbad joined the alliance in the 2019/2020 season and is therefore not treated as a member in this article.

[6] The small difference comes from the rounded values.

[7] Since Easter fell in mid-April in 2017 and 2019, the April figures could be positively affected and therefore must be treated with some caution. As already noted, this is because the Saas-Fee ski slopes have snow throughout the whole winter season and lie higher than the municipalities with weights higher than zero in the control group except Zermatt.

To investigate the significance of our results, we run placebo tests (see e.g. Abadie, Diamond, and Hainmueller, 2010, 2015): that is, we use municipalities known to be unaffected by the WinterCARD to evaluate the presence of hidden biases. Assuming a unit of the control group was treated, we estimate the same model as for the treated unit to arrive at the placebo effects. It is likely that the estimated effect of the WinterCARD on accommodation businesses in Saas-Fee, our treated unit, was observed by chance. That would be the case if the distribution of placebo effects yielded many effects as large as the effect in Saas-Fee.[8]

By comparing the distribution of the placebo effects and the effect in Saas-Fee, we construct two meaningful p-values. We calculate the fraction of placebo effects with an absolute value greater than or equal to the effect estimated for Saas-Fee. To determine the joint effect across all WinterCARD periods, we use the RMSPE. In our case, the RMSPE measures the quality of fit between the path of the overnight stays for any particular municipality and its synthetic control municipality. The first p-value can be interpreted as the proportion of control units with an estimated effect at least as large as that of the treated unit. Second, to control for the pre-WinterCARD RMSPE, we divide all post-treatment RMSPEs by the corresponding RMSPE for the pre-WinterCARD period.[9]

The first p-value amounts to 0.01. That is, one unit of the 88 units in the control group has a higher WinterCARD RMSPE than Saas-Fee (1/88 = 0.011). The second p-value amounts to 0.023 (2/88 = 0.023). We have two municipalities with a higher ratio of WinterCARD RMSPE over pre-WinterCARD RMSPE than Saas-Fee. Concluding our low p-values, the positive effect on the differences of overnight stays by domestic tourists in Saas-Fee is highly significant.

1.6.2 Robustness analysis

To confirm the previous results, we perform several robustness checks in this section. In doing so, we test the assumptions of the SCM, presented in section 1.4. Further, we debate the municipality in the control group that is assigned the highest weight and investigate the study-specific applications of the method. The robustness checks are summarised in Table 1.4.

[8] We also conduct placebo tests by reassigning the time when the treatment took place (see, e.g. Castillo, Figal Garone, Maffioli, and Salazar, 2017). We do not find any evidence that there would be any effect prior to the massive price discount.

[9] In Appendix 1.G, we present the graphical results of the placebo tests.

Table 1.4: Robustness checks

	Gap: Original vs. synthetic (rounded)	p-values		Pre-treatment RMSPE
		Post-RMSPE	Post/pre-RMSPE	
(i) *Dropping municipalities of*				
Canton Valais	6,500	0.029	0.015	1,260
Canton Valais and Berne	6,000	0.040	0.020	1,329
(ii) *Omitting municipalities of*				
Magic Pass alliance	6,500	0.013	0.013	1,230
(iii) *Replacing Saas-Fee by*				
Hasliberg	0	0.425	0.724	538
(iv) *Replacing monthly panel data with*				
Seasonal panel data	30,500	0.023	0.125	2,038
(v) *Replacing domestic with*				
Foreign overnight stays	-3,500	0.079	0.114	1,646

Notes: The numbers (1) to (5) refer to the different robustness investigations. The gap in the fourth robustness check differs from the others, mainly because it is a consideration on a seasonal level.

The second assumption of the SCM states that the treatment (WinterCARD) does not affect municipalities other than Saas-Fee. Ski destinations near Saas-Fee could be adversely affected by the massive price discount (also discussed by Falk and Scaglione (2018)). If this were the case, giving weights to municipalities near Saas-Fee would mean overestimating the discount effect. As shown above, we give weights to the Leukerbad and Zermatt, municipalities located in the Canton Valais like Saas-Fee. Therefore, we (i) investigate the robustness of our results with respect to this assumption. In doing so, we discard municipalities in Valais from the donor pool. We obtain similar results of additional overnight stays per month, amounting to about 6,500 and 6,000, respectively. Further, we omit municipalities in Canton Berne because its skiing areas are relatively close to Saas-Fee. For both robustness investigations, the results do not change significantly. Furthermore, other control municipalities in the donor pool become more important and 'step in' to construct a similarly good synthetic Saas-Fee with pre-WinterCARD RMSPEs of 1,260 and 1,329, respectively.

Although members of the Magic Pass alliance are included in our estimates of the effect of the WinterCARD, the introduction of the Magic Pass could constitute an external shock to our analysis, as overnight stays in the control group could be affected. If this were the case, the third assumption of the SCM would be violated. Therefore, in robustness investigation (ii), we exclude all municipalities near a ski lift company being part of the Magic Pass alliance.[10] Again, the result of about 6,500 additional overnight stays differs only slightly from the original application. This is because none of the excluded municipalities are assigned a weight greater than zero in the above results. We conclude that the three assumptions of the SCM in our application are fulfilled. This bases on the results of our first two robustness investigations complementing the assumption that our synthetic Saas-Fee is similar to the original in the pre-WinterCARD period.

In robustness test (iii), we evaluate the goodness of our main control unit in the control group. We replace Saas-Fee with Hasliberg, which received a weight of 0.313, to see whether there is any unexpected development in overnight stays. The estimates indicate a monthly average increase of zero overnight stays by domestic tourists in Hasliberg. In addition, the effect is not significant regarding our two p-values. Therefore, having a non-significant result amounting to zero confirms the goodness of the primary selected unit in the control group.

In the article, we use monthly panel data corrected for seasonal effects.

[10] These municipalities are marked in Appendix 1.B. We also remove municipalities that joined the alliance in the 2018/2019 season.

As already noted, in using this approach, our article differs from many others using the SCM. Therefore, robustness investigation (iv), we apply the SCM on a seasonal level. The effect of additional 30,500 overnight stays by domestic tourists triggered by the WinterCARD is more or less identical. Regarding the first p-value, the result is significant. In contrast, the second p-value amounts to 0.125 (11/88 = 0.125), indicating that 11 municipalities have a higher post-WinterCARD RMSPE divided by the pre-WinterCARD RMSPE. This is mainly because Saas-Fee has a much worse pre-treatment RMSPE than other control municipalities, amounting to 2,030.[11] Two (Ayent and Leytron) of these 11 municipalities are members of the Magic Pass alliance; therefore, the nonsignificance of the result is not only driven by the introduction of the Magic Pass. As with the original calculation, the effect declines in the third season.

As already noted, in this article, we take into account only overnight stays by domestic tourists, as the WinterCARD was mainly advertised within Switzerland. In robustness investigation (v), we therefore also estimate the effect on overnight stays by foreigners.[12] We do not find similar significant effects as for overnight stays by domestic tourists. Interestingly, the results even show a negative effect of the WinterCARD on overnight stays by foreign tourists. One reason for this negative result might be that the campaign was mainly marketed domestically and that the Saas-Fee destination continued to suffer from the declining trend in foreign tourists. However, the effect is not significant at the 5% level given both p-values. Furthermore, we see that the pre-WinterCARD RMSPE, amounting to 2,030, is worse than estimating the effect on domestic guests.

Summarizing our robustness tests, they broadly confirm the previous results. However, to take into account the uncertainties related to the 2018/2019 season and the (not significant) negative effect on foreign guests, in the next section, we calculate the economic effects for the 2016/2017 and 2017/2018 seasons alone.

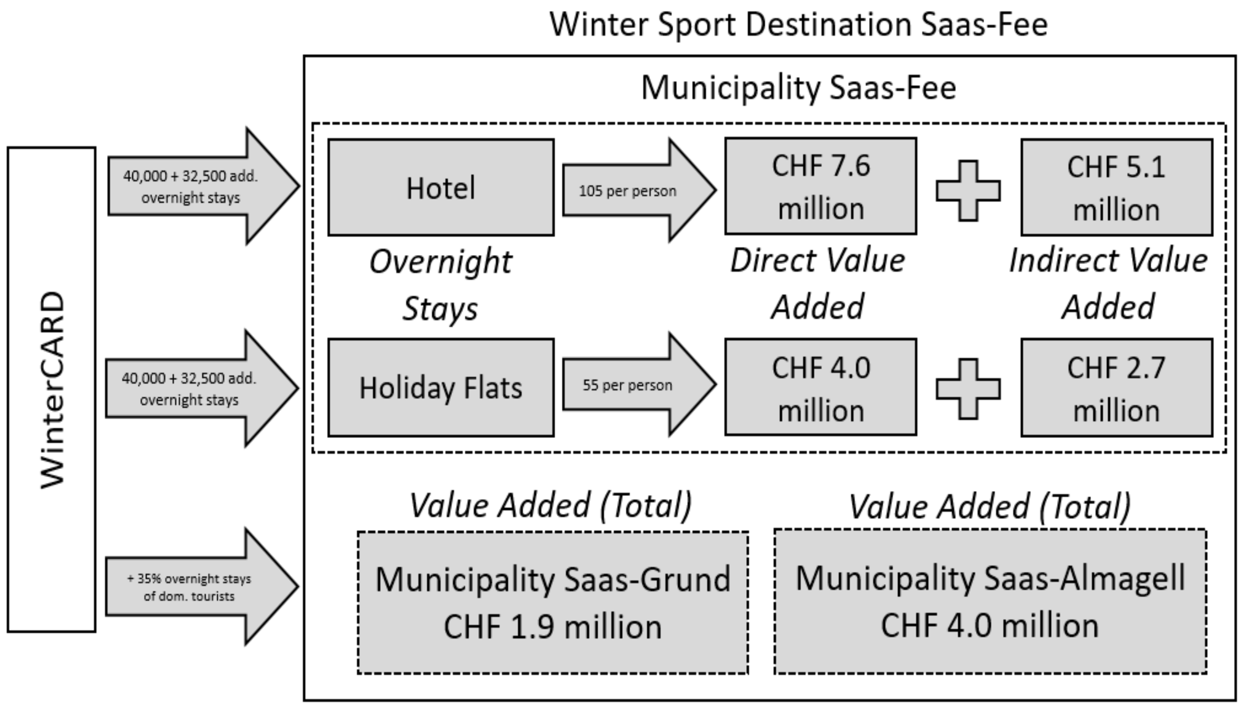
[11] 10 of these 11 municipalities having a higher post-treatment root mean squared prediction error (RMSPE) divided by the pre-treatment RMSPE have a pre-treatment RMSPE below 200, and 5 of these 11 have one below 10.

[12] Similar to the overnight stays by domestic tourists, we consider the difference from the pre-WinterCARD mean instead of the absolute values of overnight stays to fade out seasonal effects.

1.7 The economic effect of the nearby complementary product

Companies at winter sports destinations form a complementary value network. In the previous sections, we illustrated empirically the positive effect on accommodation businesses triggered by the product of the nearby ski lift company. These additional guests spend money in a destination. However, it is not only hotel guests who are attracted by such a product. In this section, we extrapolate the economic effects of the WinterCARD on all companies other than the ski lift company in the value network based on strong assumptions that the figures in studies and reports for where Saas-Fee is located (see e.g. HES-SO Wallis, 2016, Rütter, Rütter-Fischbacher, Berwert, and Landolt, 2001, general classification of economic activities according to the FSO) are valid for our case. We calculate the value added by the campaign in the winter sports destination of Saas-Fee in the 2016/2017 and 2017/2018 winter seasons. Value added is the spending of additional guests on products and services in Saas-Fee less the value of the intermediate consumption required to produce these products and services. Figure 1.2 visualizes our calculations of the value added by the WinterCARD in the Saas-Fee destination.

Figure 1.2: Value added of the WinterCARD for the other companies in the value network of the Saas-Fee winter sports destination in the 2016/2017 and 2017/2018 winter seasons (based on strong assumptions)



To calculate the value added, we start with the reliable figures of 40,000 and 32,500 additional seasonal hotel guests in the Saas-Fee tourism value network for the 2016/2017 and 2017/2018 seasons, respectively. We first calculate the expenses of these 72,500 (40,000 plus 32,500) additional guests in the tourism value network. On average, a hotel guest generates direct value added of CHF 105 in the value network of a winter sports destination. Thus, we extrapolate a direct value added of CHF 7.6 million in the Saas-Fee municipality. Secondly, we estimate the effect on holiday flats, assuming a similar effect of additional overnight stays on this sort of accommodation.[13] If we assume an additional CHF 55 in direct value added from spending by these guests, we arrive at a direct value added of CHF 4 million for the Saas-Fee municipality's value network. Going even further, we can calculate value added triggered by tourism demand for the upstream product or service company, called 'indirect value added'. In doing so, we estimate an indirect value added of about CHF 7.8 million for the Saas-Fee municipality, given the relative amount of indirect effects of tourism reported in Switzerland. As the WinterCARD was a complementary product for the whole of the Saas-Fee destination, in a fourth step, we extrapolate the 35% of additional local hotel overnight stays (compared to the winter season of 2015/2016) to the Saas-Grund and Saas-Almagell municipalities to arrive at additional values added of CHF 1.9 million and CHF 4 million, respectively.[14]

In total, we arrive at a figure of CHF 25.3 million for the value added to the value network of the Saas-Fee destination that the WinterCARD triggered in the 2016/2017 and 2017/2018 winter seasons together. That is one and a half times the total seasonal income from ticket sales to the ski lift company in the 2015/2016 winter season. Therefore, given the high value added from an overall short-term perspective, even based on strong assumptions, we conclude that the WinterCARD was a success for companies in the Saas-Fee value network except for the ski lift company, at least in the 2016/2017 and 2017/2018 seasons.

1.8 Discussion and Conclusion

Tourists' choices not only depend on the individual supply of a particular company in the value network of a destination. The availability of the WinterCARD, in this case, a complementary product offered by the local ski

[13] The Saas-Fee destination has about the same number of guests in holiday flats as in hotels (Saastal Tourismus AG, 2020).

[14] We omit the municipality of Saas-Balen, as we do not know the overnight stays in this municipality.

lift company, positively affected accommodation businesses and companies other than the ski lift company in the Saas-Fee value network. Being the first article to investigate the impact of a ski lift company's new pricing strategy on its partners in a destination's value network using the SCM, an effect of 35% additional overnight stays by domestic tourists generated by a radical price discount introduced by the ski lift company was estimated. This result was supported by several robustness tests, especially for the first two seasons after introducing the discount. By extrapolating, we obtained a figure for the economic effect of the ski lift company's campaign on the local winter sports destination of about CHF 25 million in the 2016/2017 and 2017/2018 seasons taken together. Thus, the introduction of a complementary product by a nearby company positively affected other companies in the value network, as stated in other tourism papers (Rigall-I-Torrent and Fluvià, 2007, 2010).

However, the ski lift company decided to conduct its campaign regardless of the positive effects for the other companies. Without receiving compensation from them for its investments, at the end of 2018, the company concluded that the introduction of the WinterCARD had been unsuccessful. There are two reasons for this. First, the ski lift company earns its main turnover by providing transportation. Second, and more importantly, the company made its decision independently, without the cooperation of local companies. This meant that the other companies, such as the accommodation businesses that were positively affected by the campaign, did not contribute to the costs of the WinterCARD. Therefore, this case represents a downside of a complementary value network with independent companies. Cooperation between the partners in the value network could have led the destination on a path to sustainable development (Beritelli, 2017, Bramwell and Lane, 2000, Machiavelli, 2001). Nonetheless, our empirical illustration of the positive effects will help companies in the value networks of winter sports destinations cooperate to launch similar campaigns in the future. This is because we empirically illustrated how to measure the effect of introducing a new price strategy using the SCM. After ending the WinterCARD campaign, the ski lift company in Saas-Fee decided to go for horizontal cooperation. Since the 2019/2020 winter season, it has been a member of the Magic Pass alliance (Saastal Bergbahnen AG, 2018).

One limitation of this article is its short-term perspective. Due to the short lifespan of the campaign, we only calculated its effects over three seasons. Any generalizations from our findings should therefore consider this factor. Also, we do not use econometric methods to calculate the economic impact of the campaign. Instead, based on strong assumptions, we have extrapolated our figures to arrive at the value added. At the same time, this limitation may encourage further studies to investigate in more detail the value added of campaigns in winter sports destinations using econometric methods.

In the near future, attracting domestic guests should be given priority, given the travel restrictions caused by the COVID-19 outbreak. By aiming at local guests, perhaps with campaigns like the WinterCARD, but with cooperation, the risk of long-term or subsequent travel restrictions can be avoided. However, shifts in demand from guests in terms of peak hours, days, and months will become more important to avoid overcrowding. Therefore, campaigns other than those offering radical price discounts will be needed. Using the SCM to measure the company's decisions in avoiding overcrowding is on the agenda for future research.

Appendices

1.A Synthetic control method

Suppose we observe $J+1$ units, the treated unit and J remaining units in the control group. Let T_0 be the number of pre-treatment periods, with $1 \leq T_0 < T$. Let Y^I_{jt} be the overnight stays that would be observed for unit j at period t if unit j is exposed to the treatment in periods T_{0+1} to T. The treatment has no effect on overnight stays during the pre-treatment periods 1 to T_0. Let D_{jt} be an indicator for treatment for municipality j at time t. In our paper, the first unit (Saas-Fee) is exposed to the intervention, and only after period T_0, we have that

$$D_{jt} = \begin{cases} 1 & \text{if } j = 1 \text{ and } t > T_0, \\ 0 & \text{otherwise.} \end{cases} \tag{1.A.1}$$

For unit j at time t, the observed outcome is

$$Y_{jt} = Y^N_{jt} + \alpha_{jt} D_{jt}. \tag{1.A.2}$$

The effect of the treatment can be written as

$$\alpha_{jt} = Y^I_{jt} - Y^N_{jt}. \tag{1.A.3}$$

Since only the first unit is uninterruptedly exposed to the treatment, we can apply the standard method of Abadie, Diamond, and Hainmueller (2010). Thus, we aim to estimate $(\alpha_{jT_0+1}, \dots, \alpha_{jT})$ for $t > T_0$ and an observed Y^I_{1t},

$$\alpha_{1t} = Y^I_{jt} - Y^N_{jt}. \tag{1.A.4}$$

As we observe Y^I_{1t}, we merely need to estimate Y^N_{1t}, the outcome of the synthetic control unit. We suppose that Y^N_{jt} is given by the factor model.

$$Y^N_{jt} = \delta_t + \theta_t Z_j + \lambda_t \mu_j + \epsilon_{jt}, \tag{1.A.5}$$

where δ_t is an unknown common factor with constant factor loadings across municipalities. Z_j is a $(r \times 1)$ vector of observed covariates and pre-WinterCARD overnight stays unaffected by the WinterCARD, and θ_t is a $(1 \times r)$ vector of unknown parameters. λ_t is a $(1 \times F)$ vector of unobserved common factors, and μ_j is a $(F \times 1)$ vector of unknown factor loadings. We can think, for instance, of λ_t as the appreciation of the Swiss franc (a common shock across municipalities) and μ_j as the heterogeneous impact of the appreciation of the Swiss franc on municipality j according to its touristic potential. We assume the error term ϵ_{jt} to be independent across municipalities and time with zero mean.

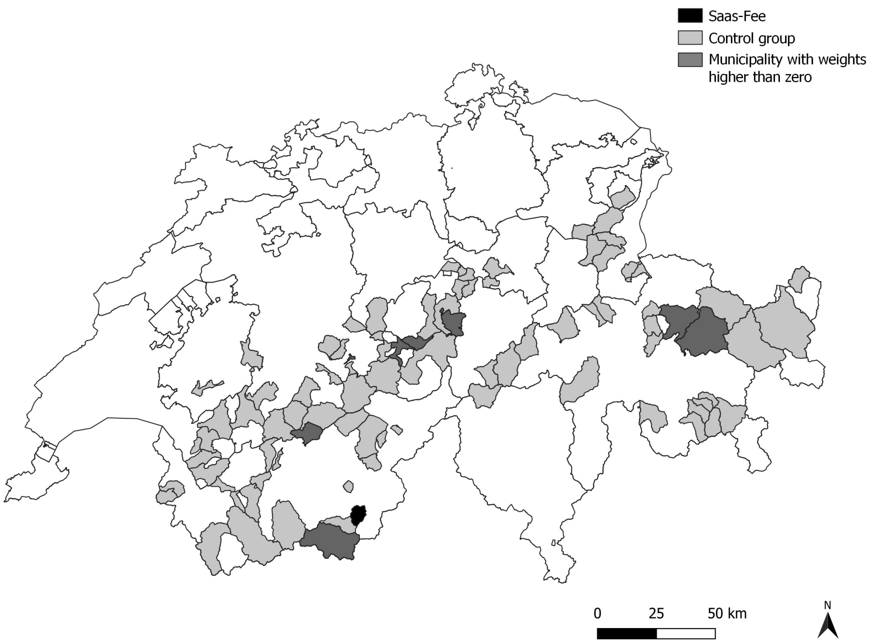
Each unit in the control group is weighted by $W = (w_2, w_3, \dots, w_J, w_{J+1})$, which is a $(J \times 1)$ matrix of non-negative weights that sum to one. Units with a large predictive power on the outcome for the treated unit are assigned large weights. In other words, we choose the weights to minimize the distance between the treated and the synthetic control unit in the pre-treatment period. Therefore, we calculate the root mean squared prediction error (RMSPE) of the dependent variable between the treated and the synthetic control unit. The RMSPE will be small if the Saas-Fee outcome is close to those of the synthetic Saas-Fee and is defined as follows:

$$RMSPE = (\frac{1}{T_0}\sum_{t=1}^{T_0}(Y_{1t} - \sum_{j=2}^{J+1} w_j Y_{jt})^2)^{\frac{1}{2}}. \tag{1.A.6}$$

1.B The municipalities in the control group

Adelboden (Berne), Aeschi bei Spiez (Berne), Airolo, Amden, Andermatt, Arosa, Avers, Ayent (Valais; Magic Pass), Bad Ragaz, Bagnes (Valais), Beatenberg (Berne), Beckenried, Bex (Magic Pass), Blatten (Valais), Breil/Brigels, Brienz (Berne), Brig-Glis (Valais), Celerina/Schlarigna, Champéry (Valais), Chur, Churwalden, Château-d'Oex, Davos, Disentis/Mustér, Emmetten, Engelberg, Evolène (Valais; Magic Pass), Fiesch (Valais), Flims, Flums, Flühli, Frutigen (Berne), Gersau, Grindelwald (Berne), Gruyères (Magic Pass), Grächen (Valais), Hasliberg (Berne), Innertkirchen (Berne), Interlaken (Berne), Kandersteg (Berne), Kerns, Klosters-Serneus, Laax, Lauterbrunnen (Berne), Lenk, Lens (Valais; Magic Pass), Leukerbad (Valais), Leysin (Magic Pass), Leytron (Valais; Magic Pass), Meiringen (Berne), Morschach, Naters (Valais), Nendaz (Valais), Ollon (Magic Pass), Ormont-Dessous (Magic Pass), Orsières (Valais), Plaffeien (Magic Pass), Pontresina, Quarten, Riddes, Riederalp (Valais), Saanen (Berne), Samedan, Samnaun, Schangnau (Berne), Schwende, Schwyz, Scuol, Sigriswil (Berne), Sils im Engadin/Segl, Silvaplana, Sion (Valais), St-Moritz, Tujetsch, Täsch, Val-d'Illiez (Valais), Vals, Vaz/Obervaz, Vilters-Wangs, Vitznau, Walenstadt, Weggis, Wilderswil (Berne), Wildhaus-Alt St-Johann, Wolfenschiessen, Zermatt (Valais), Zernez and Zweisimmen (Berne)

Figure 1.B.1: Map of Saas-Fee and the control group



1.C Descriptive statistics of the independent variables

Table 1.5 provides descriptive statistics of the independent variables for all municipalities.

Table 1.5: Descriptive statistics of the independent variables for all municipalities

Variable	Mean	SD	Minima	Maxima	N
Ski-lifts	25.35	34.09	2	216	89
Ski-runs in km	110.01	133.66	3	650	89
Average elevation of lift station (meters)	1,889.78	374.35	1,122	2,760	89
Fast Licensed Lifts	8.18	6.75	0	25	89
Distance Swiss city car (minutes)	83.21	36.99	20	166	89
Distance Swiss city public transport (minutes)	109.37	45.30	42	253	89
Hotels per municipality (2018)	14.04	13.38	3.75	96.92	89
Snow depth (cm)	80.88	52.27	0	307	5,767

1.D Descriptive statistics of monthly differences in overnight stays by domestic tourists in Saas-Fee and the control group

Table 1.6: Descriptive statistics for monthly overnight stays by domestic tourists in Saas-Fee and the control group

Overnight stays (per month)	Mean	SD	Min.	Max.	N
Pre-WinterCARD period					
Saas-Fee	0	1,791	-3,769	5,875	54
Control group	0	1,437	-13,224	18,592	4,752
WinterCARD period					
Saas-Fee	8,005	3,654	2,518	14,025	15
Control group	424	2,743	-8,491	29,791	1,320

1.E Constructing the synthetic Saas-Fee

We construct a synthetic control that best represents Saas-Fee in the pre-WinterCARD (pre-treatment) period. To minimize our RMSPE, we use both the independent variables presented in Section 5 and pre-treatment overnight stays. It is essential to assess the pre-WinterCARD goodness of fit of the synthetic control compared to our treated unit, Saas-Fee. Constructing our synthetic Saas-Fee only using pre-WinterCARD overnight stays as predictors (RMSPE: 1,168) better represents overnight hotel stays in Saas-Fee than using the independent variables (RMSPE: 1,603). Pre-WinterCARD overnight stays together with the independent variables do not minimize the RMSPE so that it is lower than only using all pre-WinterCARD outcomes as predictors. Using all pre-WinterCARD outcome variables as separate predictors together with covariates, we obtain almost the same RMSPE as when only all pre-WinterCARD outcomes are predictors. Kaul, Klößner, Pfeifer, and Schieler (2017) suggest that using all pre-treatment (here, pre-WinterCARD) outcomes as predictors renders all independent variables irrelevant. Ignoring truly influential independent variables for future outcome values could cause a potential bias in the estimated treatment effect. Ignored observed independent variables are no different from unobserved confounders. According to Abadie, Diamond, and Hainmueller (2010), the SCM is asymptotically unbiased even in the presence of unobserved independent variables. Thus, according to Kaul, Klößner, Pfeifer, and Schieler (2017), optimizing only the fit with respect to many lags of the pre-WinterCARD outcome may, to some extent, be beneficial. Considering our RMSPEs, we assume that our present independent variables are not truly influential. Based on the discussion in Kaul, Klößner, Pfeifer, and Schieler (2017), we do not use both independent variables and all outcomes for the pre-WinterCARD period. Instead, we use only overnight stays for the pre-WinterCARD as predictors in the construction of the synthetic control and also assume that we take unobserved factors into account (see also Cavallo, Galiani, Noy, and Pantano, 2013, Powell, 2017).

1.F Pre-WinterCARD characteristics

Table 1.7: Pre-WinterCARD characteristics of Saas-Fee and synthetic Saas-Fee (differences in overnight stays)

Variables	Saas-Fee		Variables	Saas-Fee	
	Treated	Synthetic		Treated	Synthetic
2006m1	-985	-859	2011m3	2,017	1,112
2006m2	-2,852	-1,826	2011m4	5,875	1,570
2006m3	-904	-528	2011m12	2,680	344
2006m4	-41	1,148	2012m1	356	1,049
2006m12	-1,148	-2,139	2012m2	1,127	1,040
2007m1	-1,871	-2,167	2012m3	2,049	428
2007m2	-3,119	-1,713	2012m4	-1,029	259
2007m3	-1,408	-810	2012m12	-697	-188
2007m4	1,043	1,150	2013m1	997	-35
2007m12	-1,540	-586	2013m2	1,578	794
2008m1	-420	-467	2013m3	-461	787
2008m2	771	174	2013m4	-3,156	-2,049
2008m3	58	-444	2013m12	-1,373	579
2008m4	2,058	-473	2014m1	627	722
2008m12	1,152	-381	2014m2	-1,428	-556
2009m1	1,005	290	2014m3	1,939	1,470
2009m2	-1,445	-1,430	2014m4	-1,391	153
2009m3	-1,111	-1,726	2014m12	239	1,680
2009m4	519	-486	2015m1	1,381	959
2009m12	196	-684	2015m2	2,064	1,641
2010m1	-1,357	-126	2015m3	-2,086	-1,255
2010m2	543	487	2015m4	-1,040	-127
2010m3	-1,666	-943	2015m12	1,575	1,843
2010m4	933	574	2016m1	-890	54
2010m12	-1,084	-467	2016m2	3,489	2,511
2011m1	1,154	578	2016m3	1,576	1,907
2011m2	-726	-1,122	2016m4	-3,769	-1,719

1.G Placebo tests

Figure 1.G.1 displays the results of the placebo tests graphically. The grey lines represent the placebos, meaning they show the gap between the differences in overnight stays by domestic tourists for each municipality in the donor pool and its synthetic control municipality. The black line shows the gap estimated for Saas-Fee. As Figure 1.G.1 shows, the estimated gap for Saas-Fee in the WinterCARD period is unusually large relative to the distribution of the gaps for the municipalities in the control group.

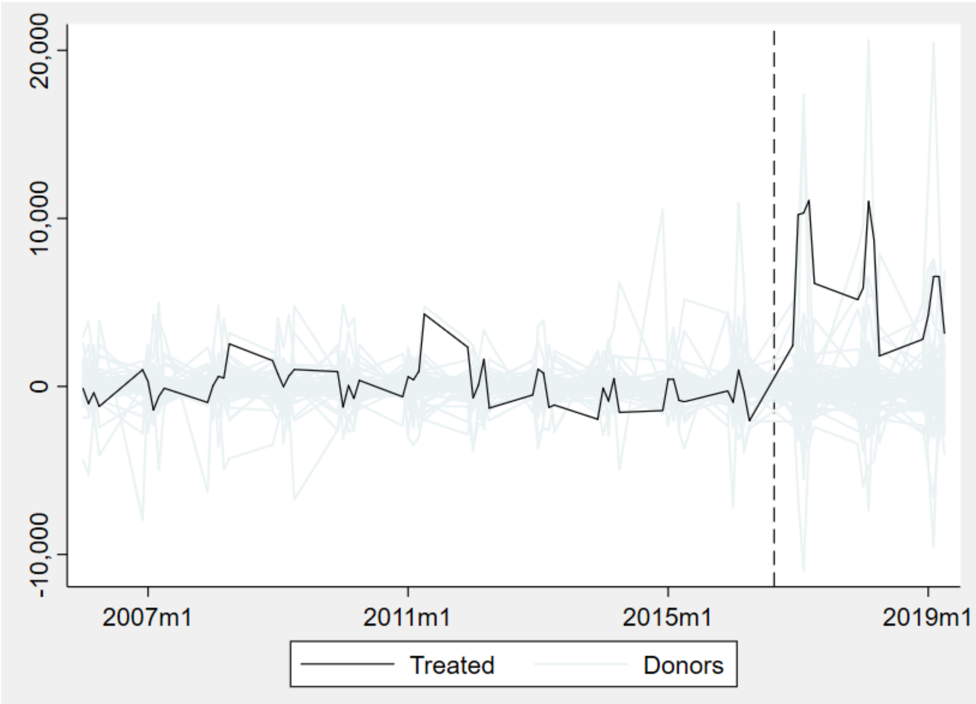


Figure 1.G.1: Distribution of placebo tests of 88 control municipalities plus Saas-Fee

Chapter 2

Flagging incomplete bid-rigging cartels

A machine learning approach

joint with **David Imhof** and **Martin Huber**[*]

Abstract

We propose a new method for flagging bid rigging, which is particularly useful for detecting incomplete bid-rigging cartels. Our approach combines screens, i.e. statistics derived from the distribution of bids in a tender, with machine learning to predict the probability of collusion. As a methodological innovation, we calculate such screens for all possible subgroups of three or four bids within a tender and use summary statistics like the mean, median, maximum, and minimum of each screen as predictors in the machine learning algorithm. This approach tackles the issue that competitive bids in incomplete cartels distort the statistical signals produced by bid rigging. We demonstrate that our algorithm outperforms previously suggested methods in applications to incomplete cartels based on empirical data from Switzerland.

2.1 Introduction

When firms deviate from competitive behavior and form a cartel, they secretly conspire to raise prices or lower the quality of goods or services. As such,

[*] Chapter 2 is based on a working paper. It is published as Wallimann, Imhof, and Huber (2020).

conspiracies directly harm taxpayers, buyers, or sellers. Cartel formation remains a pervasive problem and has been considered in a range of studies. See for instance the Swedish asphalt cartel described in Bergman, Lundberg, Lundberg, and Stake (2020), collusion among seafood processors in the US (Abrantes-Metz, Froeb, Geweke, and Taylor, 2006), bid rigging in public procurement auctions for construction works in Japan (Ishii, 2014), in Poland (Foremny, Kulejewski, Anysz, and Nicał, 2018), in Canada (Clark, Coviello, Gauthier, and Shneyerov, 2018) and in the US (Porter and Zona, 1993, Feinstein, Block, and Nold, 1985) and bid rigging for school milk contracts in Ohio (Porter and Zona, 1999), Florida and Texas (Pesendorfer, 2000). To enhance the fight against cartels, the OECD recommends competition agencies to promote pro-active methods for uncovering conspiracies, as such methods may help to discover cartels where leniency is unlikely to be sought (OECD, 2013). Answering the need for statistical tools in this context, Porter and Zona (1993), Bajari and Ye (2003), Harrington (2008), Jiménez and Perdiguero (2012), Imhof, Karagök, and Rutz (2018), Crede (2019) and Bergman, Lundberg, Lundberg, and Stake (2020), among others, have proposed different methods for uncovering cartels. However, the detection of cartels might be more challenging in the presence of competitive bidders participating in markets in which a cartel is active (McAfee and McMillan, 1992, Asker, 2010, Bos and Harrington, 2010, Conley and Decarolis, 2016, Decarolis, Goldmanis, and Penta, 2020). When a cartel is incomplete due to competitive bidders, it weakens the statistical pattern produced by bid rigging in the distribution of the bids, increasing the difficulty of detecting a cartel.

In this paper, we offer an original method based on screens and machine learning, which can detect incomplete and complete bid-rigging cartels. Screens are statistics derived from the distribution of bids in a tender aiming to capture the distributional changes produced by bid rigging (see Abrantes-Metz, Froeb, Geweke, and Taylor, 2006, Hueschelrath and Veith, 2014, Abrantes-Metz, Kraten, Metz, and Seow, 2012, Jiménez and Perdiguero, 2012, Imhof, Karagök, and Rutz, 2018, Imhof, 2019). Our novel approach consists of calculating screens for all possible subgroups of three or four bids in a tender, and not only for all bids in a tender. We then use the screens calculated for all the subgroups in a particular tender to calculate descriptive statistics of each screen, which synthesize the properties of the distribution of bids in a tender. Those descriptive statistics of screens, henceforth called 'summary screens', circumvent the distortion that competitive bidders generate in the statistical signals produced by bid rigging in a tender, rendering our suggested detection method robust to the presence of competitive bidders.

In our study, we combine the summary screens with machine learning for a prediction policy problem (see Kleinberg, Ludwig, Mullainathan, and Obermeyer, 2015), aiming to predict the probability of a cartel. Machine learning has been applied in a rapidly increasing number of studies (García Rodríguez, Rodríguez Montequín, Ortega Fernández, and Villanueva Balsera, 2020, Rabuzin and Modrusan, 2019, Silveira, Vasconcelos, Resende, and Cajueiro, 2021) and aims at finding the optimal combination of covariates that best predicts the presence or absence of bid rigging in a tender. As we focus on the predictive performance of our models, we do not have to construct explicit structural models for collusion. To train and evaluate models, we focus on the random forest (see Breiman, 2001) as machine learner because it provides a flexible prediction method that does not impose any parametric (e.g., linearity) assumptions when considering our large set of screens. Furthermore, random forests do, in contrast to many other machine learners, not require tuning specific penalty terms, see the discussion in Athey and Imbens (2019), and are therefore arguably relatively user-friendly. That appears desirable if a competition agency aims at reproducing our detection method for screening procurement markets.

Calculating screens for subgroups as in our approach is also considered in Conley and Decarolis (2016) and Chassang, Kawai, Nakabayashi, and Ortner (2020). First, Conley and Decarolis (2016) investigate subgroups to detect cartels in collusive auctions in Italy, but in contrast to our method (which considers all possible subgroups in a tender), exploit firm-specific covariates (such as, e.g., common owner, municipality, or country) to form subgroups. Relying on firm-specific covariates could impede a broad screening activity if firm-specific data is unavailable or the time needed to collect them in secrecy without raising the attention of potential cartel participants is lacking. Chassang, Kawai, Nakabayashi, and Ortner (2020) show that winning bids tend to be isolated in terms of value when bidders collude. They calculate the difference between a bidder's bid and the lowest bid submitted in a tender, focusing on such subgroups of two bids to calculate the distribution of differences. We, however, do not focus only on subgroups formed with the lowest bid in a tender and one of the other bids, but on all possible subgroups formed with three and four bids.

Two important arguments are in favor of our approach based on machine learning and summary screens. First, it exclusively relies on information about bids rather than firm-specific characteristics or cost-related variables required for econometric tests (see for instance Bajari and Ye, 2003, Aryal and Gabrielli, 2013). Our suggested method does not even need the identity of the bidders but only the bids, generally accessible from the bid summaries, which are either public or readily accessible for competition agencies and thus not as costly to acquire as firm- or cost-specific information. The necessity to gather firm-level information can attract, in some cases, the attention of the cartel, decreasing the chance of success to act against it. Second, machine learning relies on the hypothesis that bid rigging affects the distribution of bids in a tender (also common to other methods for flagging bid-rigging cartels as the econometric tests suggested by Bajari and Ye (2003)) but remains agnostic about how the distribution is affected. It is sufficient that bid rigging produces statistical signals in the distribution of bids that the screens can capture.

Our study investigates the correct classification rates of different methods in the context of incomplete cartels. We first apply a benchmark method. Here we consider the approach suggested by Imhof, Karagök, and Rutz (2018), which implements two screens with benchmarks, i.e., a rule of thumb, for classifying a tender as collusive or competitive. The second method applies machine learning using a set of screens, calculated based on all bids in a tender, so-called 'tender-based screens', to predict collusion. Finally, the third method is the novel approach suggested in this paper, which includes summary statistics of the screens (median, mean, maximum, and minimum) calculated for all possible subgroups of bids in a tender as predictors in the random forest.

We use data from Switzerland in which the incidence of collusive and competitive tenders is known. First, we apply our approach to the Ticino bid-rigging cartel (hereafter: the Ticino cartel), which was a complete cartel involving all firms active in road construction in the canton of Ticino (see Imhof, 2019). Since we also have data from the post-cartel period, we use them to simulate competitive bids that we add to the collusive tenders. To ensure that the simulated competitive bids have been adequately generated to reproduce the competitive bids of the post-cartel period, we calculate the screens based on the simulated competitive bids. These bids we use to check whether they are statistically significantly different from the screens calculated with the competitive bids of the post-cartel period. The checks confirm that simulated competitive bids are not different from the competitive bids of the post-cartel period.

However, the simulation exercise based on the Ticino case is limited to examine the correct classification rates of the three different methods. It does not account for the reaction of competitive and collusive bidders when they are aware of their reciprocal existence. Competitive bidders might try to benefit from the umbrella effect of the cartel by bidding higher than they would have done in a competitive situation (Bos and Harrington, 2010). In contrast to the potential umbrella effect, too many competitive bidders can destabilize the formation of cartels. We empirically address these two potential issues by considering data from two investigations of the Swiss competition commission (hereafter: COMCO): See-Gaster and Strassenbau Graubünden. Both cases were characterized by well-organized bid-rigging cartels, which faced competition from outsiders from time to time.

We find that the benchmarking method performs poorly for the Ticino cartel when the number of competitive bids increases. The application with the tender-based screens entails a correct classification of between 72% and 84% (depending on the sample). The classification rate is again decreasing when the number of competitive bids increases. The approach suggested in this paper with summary screens exhibits the highest correct classification rate of 77% to 86%. The difference between the tender-based screens and the summary screens amounts to 10 percentage points with five simulated bids. If we consider the error rate, defined as one minus the correct classification rate, it decreases by 43% using our approach. Cutting the error rate almost by half is substantial concerning the heavy legal consequences of flagging firms as bid-rigging cartels. This result suggests that the summary screens proposed in this paper can detect complete and incomplete bid-rigging cartels with a decent correct classification rate. Therefore, they are reliable for competition agencies and have the potential for a broader application.

We find that in the See-Gaster and Graubünden data, the benchmarking approach again exhibits a low correct classification rate for incomplete cartels. Considering models with machine learning, we note that when competitive bidders are present, the correct classification rate is higher when using summary screens, amounting to 67% to 84%, than when using the tender-based screens, amounting to 61% to 77%. Further, we note that the performance of the machine learning approaches decreases with the proportion of competitive bids. This result confirms the findings from investigations that cartel participants partially endogenize the presence of competitive bidders by adopting, at least in some cases, a more competitive behavior.

The remainder of this study is organized as follows. Section 2.2 presents the bid-rigging cartels uncovered in Switzerland from which our data are drawn. Section 2.3 outlines the detection methods for flagging both complete and incomplete bid-rigging cartels. Section 2.4 applies our original application to a simulation of incomplete cartels based on data from the Ticino bid-rigging cartel and to empirical data from the cases of See-Gaster and Strassenbau Graubünden. Section 2.5 concludes.

2.2 Bid-rigging cartels and data

The Swiss Parliament revised the federal Cartel Act and introduced a sanction regime in April 2004, with an adaptation period of one year, alongside a compliance program. This legislative change helped in initiating a change in the praxis towards economically harmful bid-rigging cartels. At the end of 2004, COMCO began to investigate the Ticino cartel, releasing its decision in 2007. The Ticino cartel dissolved without sanctions since it had ended its illegal conduct precisely before April 2005, consuming the whole adaptation period. However, it stressed the damage and mischief of a bid-rigging cartel with a price increase of over 30% (see Imhof, 2019). In 2008, COMCO decided to prioritize fighting bid rigging.

Following its decision in the Ticino case, the authority prosecuted many bid-rigging cases. Initially, COMCO was rendering an important decision against bid rigging every other year. From 2015 onwards, however, COMCO rendered more decisions, emphasizing its determination to prosecute bid-rigging conspiracies. Table 2.1 lists COMCO's most important decisions in bid-rigging cases and the sanctions it imposed in each case.

Table 2.1: Decisions of COMCO in bid-rigging cases

Decisions of COMCO (excerpt)	Year	Sanctions (million CHF)	Participants
Road asphalting in Ticino	2007	–	17
Electric Installations Bern	2009	1.2	7
Road Construction and Civil Engineering Aargau	2011	7	18
Road Construction and Civil Engineering Zurich	2013	0.5	12
Tunnel Cleaning	2015	0.16	3
Road Construction and Civil Engineering See-Gaster	2016	5	8
Construction in Val Mustair	2017	–	5
Six short Decisions in Engadine	2017	1	12
Construction in Lower Engadine	2018	7.5	7
One short Decision in Engadine	2019	0.5	3
Road construction Graubünden	2019	11	12

Overall, COMCO opens an investigation if there are reasonable grounds to assume the existence of a bid-rigging cartel. Compliance programs, whistleblowers, and procurement agencies can provide insightful information leading to the opening of an investigation. However, COMCO decided to reduce its dependence on such sources and started to develop statistical methods for detecting bid rigging based on screens (see also Imhof, Karagök, and Rutz, 2018). Based on the latter method, COMCO opened an investigation of bid rigging in the region of See-Gaster in 2013.

Considering the evolution of the cases investigated by the COMCO in recent years, incomplete bid-rigging cartels occur more often than well-organized and complete cartels. Therefore, if COMCO desires to reduce the dependence of external sources to open investigations, it must continue to improve its detection methods. Our approach for flagging both incomplete and complete bid-rigging cartels proposed in this paper responds to that need. It is likely to be of interest to competition agencies around the world.[1]

In the empirical analyses, we use data from Switzerland's three most important cases: the Ticino cartel, the See-Gaster cartel, and the Graubünden asphalt cartel. After discussing procurement in Switzerland, we synthesize the main aspects of Swiss procurement data to each case.

2.2.1 Procurement Data

Procurement agencies of cantons and cities in Switzerland follow the Agreement on Public Markets between cantons and their cantonal laws for public procurement purposes. A procurement agency can choose between four procedures: the open, the invitation, the selection, and the discretionary procedure.[2] In the construction sector, a procurement agency generally uses either the open procedure or the procedure on invitation. The open procedure does not restrict the participation of submitting firms, in contrast to the procedure on invitation, as the procurement agency invites only a small number of firms, in general, three to five, to submit a bid. That changes the nature of the

[1] Another recent example is the new Procurement Collusion Strike Force (PCSF) in the US, which focuses on deterring, detecting, investigating, and prosecuting antitrust crimes, such as bid-rigging conspiracies and related fraudulent schemes. See https://www.justice.gov/opa/pr/justice-department-announces-procurement-collusion-strike-force-coordinated-national-response (accessed on June 6, 2021).

[2] The selection procedure allows the procurement agency to select and qualify a set of bidders for participation in a tender. This procedure is useful when bidders are too numerous, as, for example, in architectural design, where hundreds of architects are interested in submitting to the project. However, such a high number of bidders is rarely a problem in the construction sector.

competition, as the participating firms are aware of the restricted number of potential competitors.

A procurement agency announces future contracts and the deadline for submitting bids (varying according to the procedure) in an official journal. If a firm is interested in submitting, the procurement agency provides the firm with all the relevant documents or information for the contract. Between the time of the announcement and the deadline, firms prepare their bids for submission. Collusive agreements, if any, between firms are typically concluded during this period.

At a pre-announced date, the procurement agency gathers the incoming bids for the contract and opens them. It officially records all the bids received on time in a bid summary or so-called official record of the bid opening and registers the firms' names, addresses, and bids. Having registered the official record of the bid opening, the procurement agency proceeds with a detailed examination of the bids. In awarding the contract, the agency not only considers the price of the bids but also other criteria such as quality, references and the environmental or social aspects. However, as contracts are relatively homogeneous in the construction sector, especially in road construction and associated civil engineering, the price in practice remains the most important criterion for awarding the contract. Furthermore, the differences in firms' criteria other than price are typically small. We, therefore, consider the procurement process as an almost first-price sealed-bid auction.

2.2.2 The Ticino Cartel

The Ticino cartel started in January 1999 and dissolved itself at the end of March 2005, precisely when the adaptation year to the new cartel Act, entered in force in April 2004, terminated. The cartel was well-organized (see Imhof, 2019). All firms active in the road construction sector participated in the cartel and rigged all public tenders and all private contracts above 20,000 Swiss francs.[3] The convention allocated contracts among cartel participants according to different criteria. The first criterion was revenue, putting cartel participants with many contracts recently awarded at the bottom of a priority list for allocating new contracts updated each week and those with few contracts at the top. The geographical distance between the firm and the location of the contract was the second most important criterion, one that played an important role in decisions to allocate small contracts. Ties with private clients were another important criterion in the awarding of private

[3] Approximatively 23,500 USA dollars at the exchange rate of 0.85 (indirect quotation) in March 2005.

contracts. In particular, cartel participants that had already produced a quote for a private client were privileged. After allocating contracts, cartel participants decided the price of the bid that the designated winner by the cartel should submit. COMCO stated in its decision that the Ticino cartel raised prices by 30% for contracts in the road construction and asphalting market.[4]

We consider data from the cartel and the post-cartel periods. Table 2.2 summarizes key information about contracts with four or more bids in our sample. We observe 149 tenders in the collusive period, whose value amounts up to 160 million Swiss francs. In total, we record 974 bids for the collusive period, hereafter referred to as collusive bids. For the post-cartel period, we observe only 33 tenders, accounting for a value of 23 million Swiss francs, in which firms submitted 222 competitive bids. Appendix 2.D presents additional descriptive statistics of the Ticino cartel.

Table 2.2: Overview sample Ticino cartel

Tenders in the cartel period	149
Volume of the collusive tenders in million CHF	160.7
Collusive bids	974
Tenders in the post-cartel period	33
Volume of the competitive tenders in million CHF	22.79
Competitive bids	222

[4] For COMCO's decision see *Strassenbeläge Tessin* (LPC 2008-1, pp. 85-112).

2.2.3 The Cartel in See-Gaster

COMCO opened its investigation in the region of See-Gaster mainly because of a statistical analysis based on procurement data from 2004 to 2010 provided by the canton of St. Gallen (see Imhof, Karagök, and Rutz, 2018).[5] In total, eight firms participated in bid-rigging conspiracies in the region of See-Gaster, including the district of See-Gaster in the canton of St. Gallen and the districts of March and Höfe in the canton of Schwyz.[6] Cartel participants regularly met once or twice a month. In their meetings, they discussed future contracts being put out to tender and exchanged their interest in them. The contracts included road construction, asphalting and civil engineering. Before each meeting, one cartel participant sent an actualized table to all the others, listing all future contracts in the region of See-Gaster. Each cartel participant had a column to put a star to a contract it was interested in obtaining, or two stars if it wished to register a very high interest.[7] When the tender procedure for a contract started, the cartel typically designated the cartel participant that should win it. The allocation mechanism was based on the interests that had been announced and fairness in making allocations to participants to maintain cartel stability.[8] In addition, if two cartel participants had both put two stars for a specific contract, they might have formed a consortium to share the contract, while other participants covered the consortium.[9]

The cartel took decisions of contract allocation during the meetings in which they discussed the list, but they organized separate meetings to discuss the price of the bids.[10] One reason for separate meetings is that not all cartel participants were interested in fixing the price since not all necessarily participated in the tender. Second, discussions about price might have taken up too much time, such that the cartel preferred the designated winner to invite the other bidders at a separate meeting to discuss the price. COMCO found some evidence that from time to time, the cartel used the mechanism of the mean in determining the bid to be made by the designated winner,[11] which implies that the latter had to submit either its own bid or the mean

[5] Report release: see the decision *Bauleistung See-Gaster: Verfügung vom 8. Juli 2016*, available on the following internet page: https://www.weko.admin.ch/weko/fr/home/actualites/dernieres-decisions.html.

[6] See the decision *Bauleistung See-Gaster: Verfügung vom 8. Juli 2016*, available on the following internet page: https://www.weko.admin.ch/weko/fr/home/actualites/dernieres-decisions.html.

[7] See the decision *Bauleistung See-Gaster*, R. 809 ff.

[8] See the decision *Bauleistung See-Gaster*, R 587, R 608 and R 623.

[9] See the decision *Bauleistung See-Gaster*, R 620 ff. and R. 645.

[10] See the decision *Bauleistung See-Gaster*, R. 649 ff.

[11] See the decision *Bauleistung See-Gaster*, R. 714 ff.

of all the exchanged bids in the separate meetings. Using this mechanism, the designated winner had some incentive to provide a relatively high bid to influence the calculated mean in the separate meeting. All the other cartel participants whose announced bids were below the mean or below the winner's bid increased their bids to cover the designated winner. As a result, they generally ensured a minimal price difference of 2% to 3% between the bid of the designated winner and their own bids.[12]

Finally, the cartel also made decisions about contracts that were left free for competitive bidding.[13] This decision was also determined by the presence of external bidders. The more external bidders, the lower was the incentive to collude because of decreasing chances of success. More external bidders were the case for some high-value contracts, for which more non-cartel firms were interested in bidding. Sometimes, the cartel also tried to bring external firms into the agreement.

In June 2009, the cartel ended its illegal conduct after COMCO launched house searches in the canton of Aargau, which to a certain extent explained the breakdown of the cartel. In its decision, COMCO attested that the cartel had discussed more than 400 contracts in the region of See-Gaster from 2004 to 2009 with a value of 198 million Swiss francs. COMCO also proved that the cartel had attempted to rig at least 200 contracts with a value of 67.5 million Swiss francs.[14] In making its decision, COMCO sanctioned the firms involved for bid-rigging conspiracies with more than five million Swiss francs. Two firms applied to the leniency program, and two other firms settled an agreement to close the case. Four firms appealed against the decision.

2.2.4 The Strassenbau Cartel in Graubünden

Members of the local trade association organized the cartel in the canton of Graubünden for road construction. In its decision, COMCO proved that the cartel participants met regularly in the period being investigated, from 2004 to the end of May 2010. The meetings, called "allocation meetings" or "calculation meetings", were mainly held at the beginning of the year since the canton and the local municipalities put most of their contracts out to tender in the spring of each year.[15] The cartel discussed contracts for road construction and asphalting tendered by the canton of Graubünden and the local municipalities. Since mountains and valleys profoundly mark the geography of Graubünden, the cartel was divided into firms operating in the

[12] See the decision *Bauleistung See-Gaster*, R. 714 ff. and R 718.
[13] See the decision *Bauleistung See-Gaster*, R 681 ff. and R. 815 ff.
[14] See the decision *Bauleistung See-Gaster*, R. 797 ff. and table 15.
[15] See the decision *Strassenbau Graubünden*, R. 139.

north and south, respectively. In the north of Graubünden, the cartel mostly organized its meetings in the office of the most important mixing plant in the canton, and to a lesser extent in the offices of the cartel participants. The meetings included either all of the twelve to thirteen cartel participants[16] or two different subgroups.[17] In the south, the total of six cartel participants[18] also organized such meetings, though changing their locations.

COMCO stated in its press release that the cartel decided upon the allocation of contracts based on a contingent determination for all the cartel participants in the canton of Graubünden.[19] The cartel allocated contracts according to the interests of each firm and fixed the price of the designated winner following a specific calculation method.[20] As a result, the price of the designated winner was usually above the minimal bid announced in the respective meeting. The calculation method, therefore, contributed to raising the price.

During the period investigated the cartel distributed 70% to 80% of the total value of the cantonal and communal road construction contracts in the north and south Graubünden among its participants. The cartel rigged approximately 650 road construction contracts concerning with a total value of 190 million Swiss francs of market volume.[21] The cartel ceased its illegal conduct in summer 2010 in both the north and the south, since some firms decided to stop, mainly because of increasing concerns regarding the Cartel Act.[22]

2.2.5 Data from the Cases See-Gaster and Graubünden

We requested data on all bid summaries from the investigations of See-Gaster and Graubünden based on the Federal Act on Freedom of Information in the Administration (Freedom of information Act, FoIA).[23] COMCO approved the request and sent us the data, referred to hereafter as Swiss data. They contain the bids, a running number for each contract, a dummy variable for each of the anonymized cartel participants and a dummy variable indicating whether the tender took place in the cartel period (taking the value of 1 for a

[16] See the decision *Strassenbau Graubünden*, R 247 ff.
[17] See the decision *Strassenbau Graubünden*, R 195 ff.
[18] See the decision *Strassenbau Graubünden*, R 248.
[19] See press release: https://www.newsd.admin.ch/newsd/message/attachments/58229.pdf.
[20] The online published decision *Strassenbau Graubünden* and the press release currently give no details of the calculation method.
[21] See press release: https://www.newsd.admin.ch/newsd/message/attachments/58229.pdf.
[22] See the decision *Strassenbau Graubünden*, R 197.
[23] https://www.admin.ch/opc/en/classified-compilation/20022540/index.html.

cartel and 0 otherwise), a categorical variable for the contract type (taking the value of 1 for contracts in road construction and asphalting, 2 for mixed contracts, including road construction and civil engineering and 3 for civil engineering contracts), as well as an anonymized date and year. The first year in our sample begins with the value of 1 and the last year ends with the value of 14. The first anonymized date equals 42, and the last 4,886. To ensure anonymization of the bids, COMCO multiplied them with a factor between 1 and 1.2. The transformation does not affect the calculation of the screens.

Table 2.3: Overview for the Swiss data

Tenders with complete cartels	310
Volume of tenders with complete cartels in million CHF	111.74
Collusive bids in tenders with complete cartels	2,031
Tenders with incomplete cartels	287
Volume of tenders with incomplete cartels in million CHF	114.73
Competitive bids in tenders with incomplete cartels	650
Collusive bids in tenders with incomplete cartels	1,414
Competitive tenders	2,398
Volume of competitive tenders	1,735.91
Competitive bids in competitive tenders	13,925
Tenders for road construction and asphalting	1,389
Tenders for civil engineering	1,286
Tenders for mixed contracts	273

Table 2.3 provides key information on the Swiss data. As for the data from Ticino, we only consider tenders with four bids or more. In total, there are 310 tenders with complete cartels with a total value of more than 110 million Swiss francs and 2,031 bids submitted by the cartel participants. Furthermore, there are 287 tenders with incomplete cartels with a total value of more than 114 million Swiss francs. In these tenders, cartel participants submitted 1,414 bids and external firms 650 bids. Finally, we observe 2,398 competitive tenders with a value of roughly 1,700 million Swiss francs and 13,925 submitted bids. In Appendix 2.E, we present additional descriptive statistics of the Swiss data.

2.3 Detection methods

This section outlines our novel approach to detect bid rigging. We first describe the concept of a random forest, the machine learning algorithm used for training and testing our predictive models for collusion (see Ho, 1995, Breiman, 2001). Second, we present in detail the screens that enter the algorithm as potential predictors. Finally, we discuss five different predictive models applied to our data that differ in the included screens.

2.3.1 Random forest

We use the random forest as a machine learning algorithm for predicting collusive and competitive tenders. In our data, the outcome is given a value of 1 for collusive tenders, including both incomplete and complete bid-rigging cartels, and 0 for competitive tenders. Note that we intentionally do not distinguish between incomplete and complete cartels, as we aim to construct a reliable method for detecting any form of bid rigging. Tenders are therefore either collusive or competitive.

Machine learning requires the data to be randomly split into the so-called training data, used to develop the predictive model, and the test data, used to evaluate the model's performance. We randomly split the data such that the training and test data consist of 75% and 25% of the observations, respectively. The random forest is a so-called ensemble method that averages over multiple decision trees to predict the outcome. Tree-based methods split the predictor space (according to the values the screens might take) of the training data recursively into a number of non-overlapping regions. Each split aims to maximize the homogeneity of the dependent variable within the newly created regions according to a goodness of fit criterion like the Gini coefficient. The latter measures the average gain in purity (or homogeneity) of outcome values when splitting and is popular for binary variables like our collusion dummy. Splitting is continued until the decision tree reaches a specific stopping rule, e.g. a minimum number of observations in a region or a maximum number of splits. Tree-based predictions of bid rigging (1) or competition (0) are based on whether collusive or competitive tenders dominate in the region that contains the values of the screens for which the outcome is to be predicted.

Importantly, there exists a bias-variance trade-off in out of (training) sample prediction when using such tree-based (and other machine learning) methods when it comes to model generality. More splits reduce the bias and increase the flexibility of the model specification, though at the cost of a greater variance in the unseen data, as the test sample is not used for training, due to the regions being smaller. The issue of a too large variance can be mitigated by repeatedly drawing many subsamples from the initial training data and estimating the predictive model, i.e. the tree (or splitting) structure, in each of the newly generated samples. A random forest consists of predicting the outcome by the majority rule across the individual trees, based on whether the majority of the trees estimated in the various subsamples predict collusion or competition for specific values of the screens. A further feature of the random forest is that at each splitting step in a specific subsample, only a random subsample of possible predictors (i.e. screens) is considered, reducing the correlation of tree structures across the subsamples and thus further reducing the prediction variance. In our application, we use the randomForest package by Liaw and Wiener (2018) for the statistical software R with growing 1,000 trees to estimate the predictive models in the training data and assess their performance in the test data based on the correct classification rate.

Note that we repeat the random sample splitting into 75% training and 25% test data and assess the predictive performance in the latter 100 times. Our reported correct classification rate corresponds to the average of the correct classification rates across the 100 repetitions. This procedure is likely to entail a smaller variance in estimating the correct classification rate than relying on a single random data split.

2.3.2 Predictors

Screens are statistics that permit data analysis intending to flag anomalous outcomes indicating potential anticompetitive issues. The literature usually differentiates structural from behavioral screens in cartel detection (see Harrington, 2008, OECD, 2013, Froeb, Sibley, Doane, and Pinto, 2014). Structural screens focus on the factors facilitating the emergence of collusive agreements and help to identify markets in which collusion is more likely. Among these factors, distinctions are made between market structure, demand-related factors, and supply-related factors (OECD, 2013). In contrast, behavioral screens empirically measure the behavior of market participants and assess whether the observed behavior significantly departs from competitive behavior to flag it as a potential issue worth scrutinizing further. Following Huber and Imhof (2019) we propose using various descriptive statistics as screens and combining them with machine learning, however, to uncover not only complete but also incomplete bid-rigging cartels.[24] We consider three classes

[24] In contrast to the context of causal inference, causality goes from the dependent variable (collusion or competition) to the predictors (screens) rather than the other way around. The incidence of collusion as an explanatory variable affects the distribution of bids and thus the screens in causal terms. As in Huber and Imhof (2019) our prediction problem consists of analyzing a reverse causality. By investigating the screens, we infer the existence

of screens: variance, asymmetry, and uniformity.

As variance screens, we implement the coefficient of variation (CV) and the kurtosis statistic (KURTO), as suggested by Huber and Imhof (2019) and Imhof (2019). In addition, we also implement the spread (SPD) of the distribution of the bids as screen.

The coefficient of variation is widely discussed in the literature (see Abrantes-Metz, Froeb, Geweke, and Taylor, 2006, Esposito and Ferrero, 2006, Jiménez and Perdiguero, 2012, Abrantes-Metz, Kraten, Metz, and Seow, 2012, Imhof, 2019) and is defined as the standard deviation divided by the arithmetic mean of all bids submitted in a tender:

$$CV_t = \frac{s_t}{\bar{b}_t}, \tag{2.3.1}$$

where s_t is the standard deviation and $\bar{b}_t$ is the mean of the bids in some tender t. As the coordination and manipulation of bids by cartel participants might affect the convergence in the distribution of the bids, we also consider the following kurtosis statistic as screen:

$$KURTO_t = \frac{n_t(n_t+1)}{(n_t-1)(n_t-2)(n_t-3)} \sum_{i=1}^{n_t} (\frac{b_{it}-\bar{b}_t}{s_t})^4 - \frac{3(n_t-1)^3}{(n_t-2)(n_t-3)}, \tag{2.3.2}$$

where b_{it} denotes the bid i in tender t, n_t the number of bids in tender t, s_t the standard deviation of bids, and $\bar{b}_t$ the mean of bids in that tender. Note that we calculate the kurtosis statistic only for tenders with four bids or more. Furthermore, we estimate the spread using the following formula:

$$SPD_t = \frac{b_{max,t} - b_{min,t}}{b_{min,t}}, \tag{2.3.3}$$

where $b_{max,t}$ denotes the maximum bid and $b_{min,t}$ the minimum bid in some tender t.

As bid rigging may produce asymmetries in the distribution of the bids, we implement the following cover-bidding screens as in Huber and Imhof (2019): the percentage difference (DIFFP), the skewness (SKEW), the relative distance (RD) and the normalized distance (RDNOR). We also add an alternative measure for calculating the relative difference, namely the alternative relative distance (RDALT).

It seems plausible that cartel participants manipulate the difference between the lowest and second lowest bids to secure awards of contract by the

of their cause: collusion.

cartel's designated winner. To analyze the difference between the two lowest bids, we use the following formula to calculate the percentage difference:

$$DIFFP_t = \frac{b_{2t} - b_{1t}}{b_{1t}}, \tag{2.3.4}$$

where b_{1t} is the lowest bid and b_{2t} the second lowest bid in some tender t. We also consider the absolute difference between the first and second lowest bids $D_t = b_{2t} - b_{1t}$ in the empirical analysis.

The manipulation of the bids by cartel participants can simultaneously affect both the difference between the first and second lowest bids and the differences across the losing bids. Following Imhof, Karagök, and Rutz (2018), we calculate a relative distance (relative to a measure of dispersion) in a tender by dividing the difference between the first and second lowest bids by the standard deviation of the losing bids:

$$RD_t = \frac{b_{2t} - b_{1t}}{s_{losingbids,t}}, \tag{2.3.5}$$

where b_{1t} denotes the lowest bid, b_{2t} the second lowest bid, and $s_{t,losingbids}$ the standard deviation calculated among the losing bids in some tender t. In Huber and Imhof (2019) in terms of its predictive power the RD was outperformed by the difference between the second and first lowest bids divided (or normalized) by the average of the differences between all adjacent bids. We also consider this normalized distance in our study:

$$RDNOR_t = \frac{b_{2t} - b_{1t}}{\frac{(\sum_{i=1,j=i+1}^{n_t-1} b_{jt} - b_{it})}{n_t - 1}}, \tag{2.3.6}$$

where b_{1t} is the lowest bid, b_{2t} the second lowest bid, n_t is the number of bids and b_{it}, b_{jt} are adjacent bids (in terms of price) in tender t, with bids being arranged in increasing order.

We consider a further alternative measure for the relative distance, initially suggested by Imhof, Karagök, and Rutz (2018):

$$RDALT_t = \frac{b_{2t} - b_{1t}}{\frac{(\sum_{i=2,j=i+1}^{n_t-1} b_{jt} - b_{it})}{n_t - 2}}, \tag{2.3.7}$$

where b_{1t} is the lowest bid, b_{2t} the second lowest bid, n_t is the number of bids and b_{it}, b_{jt} are adjacent losing bids in a tender t, with bids being arranged in increasing order. In contrast to the normalized distance, the mean of the differences in the denominator is calculated using only losing bids.

Furthermore, bid manipulation might affect the symmetry of the distribution of bids. We therefore include the skewness as screen:

$$SKEW_t = \frac{n_t}{(n_t-1)(n_t-2)} \sum_{i=1}^{n_t} (\frac{b_{it} - \bar{b}_t}{s_t})^3, \tag{2.3.8}$$

where n_t denotes the number of the bids, b_{it} the i^{th} bid, s_t the standard deviation of the bids, and $\bar{b}_t$ the mean of the bids in tender t.

Finally, we consider the nonparametric Kolmogorov-Smirnov statistic (KS) for verifying whether bid rigging (or competition) transforms the distribution of the bids in a less uniform distribution:

$$D_t^+ = max_i(x_{it} - \frac{i_t}{n_t+1}), D_t^- = max_i(\frac{i_t}{n_t+1} - x_{it}), KS_t = max(D_t^+, D_t^-), \tag{2.3.9}$$

where n_t is the number of bids in a tender, i_t the rank of a bid and x_{it} the standardized bid for the i^{th} rank in tender t. The standardized bids x_{it} are the bids b_{it} divided by the standard deviation of bids in tender t to facilitate the comparison of tenders with different contract values. We suspect that the KS-statistic should generally differ across cartels and competitive periods.

In incomplete cartels, competitive bidders distort the statistical signals produced by bid rigging in the distribution of bids in a tender. Therefore, the tender-based screens can fail to recognize bid rigging if they are calculated for all bids. We circumvent that distortion by calculating the screens not (only) for all the bids in a tender but for all possible subgroups of three and four bids. Table 2.4 gives the number of possible subgroups of three or four bids, respectively, when the total number of bids in a tender varies between four to ten bids.

For instance, in a tender with a total number of six bids, we calculate the same screen but for 15 different subgroups containing four bids and for 20 different subgroups containing three bids. In each tender, we then include summary statistics for each screen: the mean, the median, the minimum, and the maximum of the respective screen across the various subgroups of three or four bids. We use these summary statistics, so-called 'summary screens', as predictors for flagging collusive and competitive tenders and permit comparing tenders with different numbers of bids. We subsequently exemplify the computation of such summary screens by means of the coefficient of variation for subgroups formed on four bids.

The mean of all coefficients of variation calculated for subgroups of four bids in each tender is:

Table 2.4: Example of possible subgroups for three and four bids in a tender

Bids in a tender	Subgroups formed with three bids	Subgroups formed with four bids
4	4	1
5	10	5
6	20	15
7	35	35
8	56	70
9	84	126
10	120	210

$$MEAN4CV_t = \sum_{s=1}^{N_t} (\frac{s_{st}/\bar{b}_{st}}{N_t}), \tag{2.3.10}$$

where s and t denote the indices for some sub-group s and some tender t respectively, N_t is the number of all the possible subgroups of four bids in tender t and s_{st} and $\bar{b}_t$ are the standard deviation and the mean of the bids respectively. Likewise, the minimum and maximum of the coefficients of variation across the subgroups in a tender correspond respectively to:

$$MIN4CV_t = min_s \frac{s_{st}}{\bar{b}_{st}}, \tag{2.3.11}$$

$$MAX4CV_t = max_s \frac{s_{st}}{\bar{b}_{st}}, \tag{2.3.12}$$

In order to calculate the median for subgroups of four bids in each tender, define the coefficient of variation in subgroup s and tender t as $CV_{st} = \frac{s_{st}}{\bar{b}_{st}}$ and order the coefficients in so that

$$CV_{1t} \leq CV_{2t} \leq ... \leq CV_{st} \leq ... \leq CV_{N_t t}.$$

If the number of subgroups N_t in a tender is uneven, the median of the coefficient of variation in tender t is calculated as follows:

$$MEDIAN4CV_t = CV_{(N_t+1)/2,t}, \tag{2.3.13}$$

If the number of subgroups is even, the median corresponds to:

$$MEDIAN4CV_t = \frac{CV_{N_t/2,t} + CV_{N_t/2+1,t}}{2}. \tag{2.3.14}$$

We apply these approaches to all the screens discussed above across the different tenders. Note also that we do not calculate summary screens for subgroups of two bidders because of the impossibility of calculating some screens as RD, RDALT, RDNOR, KURTO, or SKEW. Moreover, cartel participants usually numbered more than two in tenders characterized by incomplete cartels. We also renounce calculating screens for sub-groups of five bidders or more. Using summary screens calculated for subgroups of five bidders only makes sense for tenders with six bids and more. Using tenders with six bids or more would have restricted our sample too much and limited the application of our suggested methods in other cases. Finally, our original application of summary screens does not require the identity of bidders. Instead, we only need the bids in each tender to apply them in many different contexts.

Appendix 2.F presents the descriptive statistics for the samples used in the empirical analyses for both the Ticino simulation and the Swiss data.

2.3.3 Model specification

In the empirical analyses, we consider five different predictive models that vary in terms of screens considered and a benchmarking method. For the latter, we use the benchmarks suggested by Imhof, Karagök, and Rutz (2018), developed for and applied to the Swiss construction market.[25] Model 1 only includes screens calculated for all bids in a tender (rather than summary screens). This approach relates to the one discussed by Huber and Imhof (2019). Still, it extends the set of predictors compared to the study by including the relative measure for the alternative distance (RDALT), the spread (SPD), and the Kolmogorov-Smirnov statistic (KS). In total, we use nine predictors and exclude any screens based on the absolute bid value to consider only scale-invariant screens in model 1.

In contrast, model 2 exclusively includes the summary screens, calculated for all possible subgroups of three bids in a tender. In total, we consider the application of 32 of these summary screens, using all screens of model 1 but the kurtosis (KURTO), which requires at least four bids. Model 3 uses summary screens of all screens presented above for all possible subgroups of four bids in a tender, making a total of 36 predictors (now including the kurtosis). Model 4 considers all predictors included in models 1, 2, and 3, resulting in 77

[25] More precisely, tenders with a CV below 6 and a RD above 1 are classified as conspicuous.

screens in total, mixing the summary screens with the tender-based screens. Finally, model 5 also includes three screens based on absolute bid values (and thus not scale-invariant) and the number of bids in a tender (NBRBIDS),[26] producing 81 predictors in total. The three value-based screens are the mean bid in a tender included as a proxy for the contract value (MEANBIDS), the standard deviation of the bids in a tender (STDBIDS), and the absolute difference between the first and the second lowest bids (D).

2.4 Flagging incomplete bid-rigging cartels

2.4.1 The Ticino simulation

We use the data from the Ticino cartel to investigate how the predictive models presented above perform in detecting bid-rigging cartels when competitive bids distort the statistical pattern produced by bid rigging. Since the Ticino cartel was complete, we use the data from the competitive periods to simulate competitive bids and progressively add them to the collusive tenders, thus creating five additional datasets for the cartel period. The first dataset includes only one simulated competitive bid in each collusive tender, the second two, and the fifth five. This stepwise approach permits investigation of how different degrees of partial collusion affect the performance of each model.

We generate simulated bids from competitive bids using the following formula:

$$b_{t,simulated} = \bar{b}_t\left(1 + \frac{b_{i,drawcomp} - \bar{b}_{drawcomp}}{\bar{b}_{drawcomp}}\right), \tag{2.4.1}$$

i and t are indices for bids and tenders, respectively, $\bar{b}_t$ is the mean bid of tender t (without the simulated bid), while $b_{i,drawcomp}$ and $\bar{b}_{drawcomp}$ are the bid and the mean bid randomly drawn from competitive tenders respectively. We simulate competitive bids to be added to collusive tenders in three steps. First, we calculate the normalized bids for all bids in competitive tenders by (i) subtracting the tender's mean from each bid and (ii) dividing this number by the tender's mean. Normalizing the bids like we do and not dividing the bids with the standard deviation avoids losing information about the dispersion of the bids. In a second step, we pool all the normalized bids together. Then, we randomly draw (with replacement) and assign them to a collusive tender

[26]The motivation for including the number of bids is that it might be easier to settle an agreement in a tender with few bidders than with many.

t. Finally, in a third step, we (i) add the mean bid of the collusive tender t to the normalized bid drawn and (ii) multiply the number with the mean bid.

We end up with seven different datasets for the Ticino cartel: the dataset of the post-cartel period, including only competitive bids; the dataset of the collusive period, including only collusive bids; and five different datasets, including the collusive tenders with one to five competitive bids in each tender.

We verify whether the simulated competitive bids are similar to the competitive bids of the post-cartel period. Since we generate five simulated competitive bids for each tender in the collusive period, we only calculate the screens for those five simulated competitive bids. We test whether the screens based on the simulated competitive bids are statistically significantly different from the screens based on the competitive bids of the post-cartel period. The results presented in Appendix 2.C show that most statistical tests do not reject the null hypothesis of no distributional differences, such that our simulation process adequately generates competitive bids. In other words, the distribution of the simulated competitive bids empirically matches the distribution of the competitive bids in the post-cartel period.

We first apply the benchmarking approach to classify conspicuous tenders and find a correct classification rate of 84.8% in the test data in the absence of competitive bids (see Table 2.5). However, when adding one simulated competitive bid to the collusive tenders, the correct classification rate falls to 66.7%. Furthermore, it continuously decreases in the number of competitive bids added. With five competitive bids in the collusive tenders, the method suggested by the benchmarking method does not perform any better than tossing a coin. As expected, the collusive tenders exclusively drive the decrease in the overall correct classification rate. With the addition of only one competitive bid, the correct classification rate among collusive tenders decreases to 48.5%. With five competitive bids, it falls to 15.2%. This decrease suggests that while the benchmarking approach might be well suited for detecting complete bid-rigging cartels, it seems inappropriate for flagging incomplete bid-rigging cartels due to too many false negative predictions. Note that, concerning the competitive tenders, the correct classification rate of 84.8% remains unaffected since the sample of the post-cartel period does not vary across the various simulations.

We turn now to the results for model 1, which only includes tender-based screens, and we find that its correct classification rate shrinks between 5.3 and 11.3 percentage points depending on the number of competitive bids added to the collusive tenders. Contrasting with model 1, model 4 appears to be the most powerful model for predicting bid rigging, whose correct classification rates varying between -6.3 and +3.1 percentage points with the occurrence of competitive bids. When competitive bids are absent, model 1 performs slightly better than model 4, while model 4 outperforms model 1 by 10.1 and 10.3 percentage points when including four and five competitive bids, respectively. This result illustrates the advantage of our approach, considering summary screens. Considering the error rate, it amounts to 24.2% for model 1 such that almost one tender out of four is incorrectly classified. In model 4, the error rate is only 13.9%. This decrease of 42.6% in the error rate compared to model 1 is substantial, indicating that some tenders rigged by incomplete cartels would not have been detected by a conventional detection method based on tender-based screens. This could be potentially problematic if a competition agency judging the presence of reasonable grounds to open an investigation leaves out some problematic cases by lack of statistical power.

Models 2 and 3 also include summary screens calculated in subgroups of three and four bids, respectively. On average, model 2 performs slightly better than model 3, but the correct classification rates are very similar. The maximum difference in (overall) correct classification rates across model 4 and models 2 or 3 amounts to 2.2 percentage points in the sample with three competitive bids. Overall, the correct classification rates of models 2 and 3 hardly differ from model 4. This result indicates that adding tender-based screens to summary screens, as in model 4, does not significantly increase predictive power. Thus, summary screens mainly explain the increase in the correct classification rates in models 2 to 4.

We examine the variable importance in the random forest for predicting collusive and competitive tenders in each dataset according to the mean decrease in the Gini index (hereafter: MDG)[27] when omitting the respective predictor, which ranks variables according to their predictive power. However, it does not allow direct comparison between models since the MDG depends on the number of predictors. As we use fewer variables in model 1 than in model 4, the MDG of the former model is higher. For each dataset and models 1 to 4, we depict the five most important variables in Table 2.6.

We observe for model 1 almost the same important predictors in all six data sets (zero to five competitive bids), namely the Kolmogorov-Smirnov statistic (KS), the coefficient of variation (CV), the kurtosis statistic (KURTO), and the normalized distance (RDNOR). The order of importance changes with the number of competitive bids. With zero or one competitive bid, the Kolmogorov-Smirnov statistic (KS), the coefficient of variation (CV), and the kurtosis statistic (KURTO) are the most relevant variables with MDGs larger than three. In the presence of two or more competitive bids, the normalized distance (RDNOR) and the alternative distance (RDALT) exhibit a greater

[27]This is a measure of the contribution of each variable to the purity of nodes and leaves in the random forest.

Table 2.5: Correct classification rate for the Ticino simulation

Comp.B	Tenders	Bench.	M1	M2	M3	M4
	All	0.848	0.835	0.832	0.834	0.830
0	Comp.	0.848	0.816	0.811	0.806	0.811
	Coll.	0.848	0.863	0.861	0.87	0.859
	All	0.667	0.722	0.766	0.756	0.767
1	Comp.	0.848	0.701	0.781	0.739	0.777
	Coll.	0.485	0.758	0.765	0.788	0.769
	All	0.652	0.752	0.819	0.786	0.802
2	Comp.	0.848	0.751	0.856	0.786	0.832
	Coll.	0.455	0.758	0.79	0.803	0.784
	Overall	0.576	0.782	0.826	0.795	0.817
3	Comp	0.848	0.796	0.862	0.807	0.838
	Coll.	0.303	0.775	0.804	0.793	0.806
	All	0.561	0.727	0.808	0.832	0.828
4	Comp.	0.848	0.713	0.843	0.83	0.837
	Coll.	0.273	0.747	0.777	0.832	0.819
	All	0.500	0.758	0.871	0.871	0.861
5	Comp.	0.848	0.745	0.885	0.879	0.887
	Coll.	0.152	0.778	0.837	0.862	0.837

Notes: "Comp.B", "Tenders", "Bench.", "M1", "M2", "M3" and "M4" denote the number of competitive bids in the collusive tenders, the type of tenders, the results produced by the screening methods of the benchmarking approach, model 1, model 2, model 3 and model 4 respectively. For the type of tenders, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively.

predictive power along with the Kolmogorov-Smirnov (KS) and the kurtosis statistic (KURTO).

In models 2 to 4, the Kolmogorov-statistic (KS), the spread (SPD), the difference in percentage (DIFFP), and the coefficient of variation (CV) are the best predictors. Contrary to model 1, summary screens based on relative distance (RDNOR, RD, and RDALT) or skewness (SKEW) are less important. At the same time, the difference between the first and second lowest bids in percentage (DIFFP) is more important. As in model 1, the order of importance changes with the number of competitive bids. In models 2 to 4 for zero and one competitive bid, the Kolmogorov-statistic (KS) is the most important variable. Conversely, in the data sets with four and five competitive bids, the difference in percentage (DIFFP) comes first.

For models 2 to 4, we find that the median and mean of certain summary screens are the most predictive in the presence of zero or one competitive bids, while the minima and maxima of summary screens predominate under a larger number of competitive bids. Intuitively, the minimum or maximum of a particular summary screen is likely to exclude competitive bids if the latter distort the distribution of collusive bids and should thus be relatively more predictive as the number of competitive bids increases. We also note that, for three or more competitive bids, the random forest mainly selects summary screens calculated in subgroups of four bids in a tender rather than three.

To sum up the results of the Ticino simulation, we find that our approach based on summary screens can flag bid-rigging cartels even when we add competitive bids. When the number of competitive bids increases, the random forest puts more weight on the minima or the maxima of summary screens across the subgroups. The minima and maxima of summary screens appear to be adequate to eliminate the distortion of competitive bids in the statistical pattern produced by bid rigging. Therefore, our approach is able to detect incomplete and complete bid-rigging cartels. For models 1 to 4, the algorithm selects a mix of screens from the three groups, namely variance, asymmetry, and uniformity. The random forest selects screens based on the uniformity (KS) or the variance (CV, KURTO, and SPD) in the sample with zero or one competitive bid added. When the number of competitive bids rises, screens for asymmetry in the distribution of bids become more important (RDNOR and RDALT for model 1 and DIFFP for models 2 to 4). Thus, screens for asymmetry in the distribution of bids might have a higher weight in our data for detecting bid-rigging cartels when a cartel is incomplete.

Table 2.6: Important predictors for the Ticino simulation

	Model 1		Model 2		Model 3		Model 4	
	IV	MDG	IV	MDG	IV	MDG	IV	MDG
	KS	4.63	MEAN3KS	2.33	MEAN4KS	2.88	MEAN4KS	1.56
	CV	4.6	MEDIAN3SPD	1.88	MEDIAN4SPD	1.89	MEAN3KS	1.37
0	KURTO	3.37	MEAN3SPD	1.85	MEAN4SPD	1.86	MEDIAN3SPD	0.99
	SPD	2.42	MEAN3DIFFP	1.78	MEAN4CV	1.83	MEAN3SPD	0.97
	RDNOR	2.36	MEAN3CV	1.74	MEDIAN4KS	1.82	MEAN3DIFFP	0.97
	KS	3.88	MEAN3KS	1.92	MEAN4KS	2.08	MEAN3KS	1.16
	KURTO	3.65	MEDIAN3SPD	1.18	MEDIAN4SPD	1.15	MEAN4KS	1.08
1	CV	3.61	MEAN3SPD	1.17	MEAN4SPD	1.13	MEDIAN3SPD	0.65
	SPD	2.96	MEDIAN3KS	1.13	MEDIAN4CV	1.1	MEAN3SPD	0.65
	RDNOR	2.72	MEAN3CV	1.08	MIN4SPD	1.08	MEDIAN3KS	0.63
	RDNOR	4.2	MEAN3KS	1.45	MIN4DIFFP	1.45	MIN3DIFFP	0.82
	RDALT	3.37	MIN3DIFFP	1.4	MEAN4KS	1.4	MEAN3KS	0.81
2	KS	3.1	MIN3SPD	1.14	MIN4CV	1.14	MIN4DIFFP	0.76
	KURTO	2.99	MAX3KS	1.1	MIN4SPD	1.1	MEAN4KS	0.74
	CV	2.92	MIN3CV	1.1	MAX4KS	1.1	RDNOR	0.72
	RDNOR	3.67	MEAN3KS	1.64	MIN4CV	1.76	MAX4KS	0.99
	KURTO	3.47	MAX3KS	1.46	MAX4KS	1.75	MIN4CV	0.99
3	KS	3.45	MIN3CV	1.44	MIN4SPD	1.74	MIN4SPD	0.95
	RDALT	3.12	MIN3SPD	1.44	MIN4DIFFP	1.71	MEAN3KS	0.9
	CV	3.08	MIN3DIFFP	1.33	MEAN4KS	1.33	MIN4DIFFP	0.89
	KURTO	4.49	MIN3DIFFP	2.04	MIN4DIFFP	2.34	MIN4DIFFP	1.34
	RDNOR	3.72	MIN3SPD	1.69	MAX4KS	2.1	MIN4CV	1.27
4	KS	3.16	MIN3CV	1.59	MIN4CV	2.09	MAX4KS	1.27
	RDALT	2.68	MAX3KS	1.57	MIN4SPD	2.02	MIN4SPD	1.23
	CV	2.52	MAX3RDNOR	1.47	MAX4RDALT	1.28	MIN3DIFFP	1.07
	KURTO	5.08	MIN3DIFFP	2.15	MIN4DIFFP	2.44	MIN4DIFFP	1.53
	RDNOR	3.87	MIN3SPD	1.62	MIN4SPD	2.31	MIN4SPD	1.51
5	RDALT	2.77	MAX3KS	1.52	MAX4KS	2.05	MAX4KS	1.35
	KS	2.71	MIN3CV	1.5	MIN4CV	2.03	MIN4CV	1.34
	CV	2.15	MEAN3KS	1.5	MAX4RDALT	1.54	MIN3DIFFP	1.06

Notes: "IV" and "MDG" denote the important variables selected by the random forest and the mean decrease in Gini index. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP" and "KURTO" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference and the kurtosis statistic respectively.

2.4.2 Application to the Swiss data

We now apply our detection method to data on the cases of See-Gaster and Strassenbau Graubünden, characterized by well-organized bid-rigging cartels, which, however, faced competition from outsiders from time to time. In these 'real' cases, competitive and collusive bidders were aware of their reciprocal existence. Evidence from COMCO's investigations has pointed out that cartel participants adopted a more competitive behavior in the presence of competitive bidders by deciding not to collude in some tenders. The poor chance of success of an agreement due to the number of (potential) competitive bidders motivated such decisions to bid free for diverse contracts. In other cases, the cartel faced only one competitive bidder and tried to enroll him in the agreement. Moreover, competitive bidders aware of the existence of the cartel might have tried – if not enrolled in the agreement – to benefit from the umbrella effect of a cartel by bidding higher in a fully competitive situation. As a consequence of the umbrella effect, bids of the competitive bidder fall nearer to bids of collusive bidders, such as competitive bids distort less the statistical pattern produced by bid rigging.

Similar to the Ticino simulation, we construct different samples of collusive tenders. Sample 1 includes all tenders with incomplete bid-rigging cartels and at least two cartel participants. As stated in Table 2.7, the average percentage of cartel participants in sample 1 amounts to 71%. Sample 2 includes tenders with incomplete bid-rigging cartels formed with at least three cartel participants. The average rate of cartel participants of 75% in sample 2 is superior to sample 1 since sample 2 excludes tenders with only two cartel participants. The logic is the same for samples 3 to 5. Consequently, sample 5 has the highest average percentage of cartel participants and contains fewer competitive bidders but at least one per tender. In addition, we construct a sample including all tenders with complete cartels.

First, we investigate the performance of our various predictive models starting with complete cartels. As shown in Table 2.7, the correct classification rates do not differ notably across machine learning-based models 1 to 5, the range being 81.3% to 83.3%. However, for the benchmarking method, the correct classification rate of 61.7% is clearly below that of models 1 to 5. In addition, it differs strongly between competitive and collusive tenders, amounting to only 33.4% in the latter case. Possible explanations for this poor performance are the reliance on only two screens, which are not necessarily the optimal predictors, and on benchmark values for these two screens from two previous investigations, which are not necessarily optimal in the dataset being considered. In contrast, machine learning approaches use a more extensive set of screens and weight their importance in a data-driven way.

However, if we adjust the benchmarks of our benchmarking approach applied, we can achieve better prediction rates for complete cartels. In Appendix 2.A, we depict a decision tree on Figure 2.A.1 corresponding to the minimal cross-validation error. Our pruned tree, using as predictors only the RD and the CV as in Imhof, Karagök, and Rutz (2018), shows a correct classification rate of 81.6% for complete cartels. Since Imhof, Karagök, and Rutz (2018) have drawn their benchmarks from two previous investigations, one of them the Ticino cartel, it is therefore coherent that the benchmarks produce better results for detecting complete bid-rigging cartels in Ticino than in the Swiss data. This discrepancy illustrates the fundamental difference between a benchmark method and machine learning: benchmarks are exogenous, whereas machine learning outperforms benchmarks since it chooses the best predictors in each case. While a benchmark can still be adapted to different cases, machine learning algorithms are far more precise. Nonetheless, a benchmark method has the advantage of requiring less information to be implemented and therefore remains a simple (first) step in flagging cartels.

Considering models 1 to 5, the correct classification rates vary between 61.2% and 84.1%, depending on the sample and the model. When the proportion of competitive bidders increases, the correct predictions generally decrease, as depicted in Table 2.7. This result suggests that cartel participants anticipated competitive bids and decided not to collude in some peculiar tenders, as attested, for example, in the case of See-Gaster. The models with summary screens calculated for subgroups outperform model 1. Among them, models 3 and 4, unlike in the Ticino simulation, slightly outperform model 2, indicating that in our case, summary screens calculated for subgroups of four bids exhibit a higher predictive power than those calculated for subgroups of three bids. The fact that we have four cartel participants per tender in most cases likely explains this result. In contrast, summary screens calculated for subgroups of three bids may work better if we mainly observe three cartel participants per tender.

Model 5, the only one also to include the number of bidders or the contract value as predictors, outperforms the other models and has a 5 to 10 percentage points more correct classification rate than model 1. The advantage of models 3 or 4 over model 1 is 3 to 5.7 percentage points. Therefore, the gain in calculating screens for subgroups is not quite as high as for the Ticino simulation (4.5 to 10.3 percentage points). This result suggests strategic reactions between competitive bidders and cartel participants, absent in the Ticino simulation. Finding less difference between models 2, 3, and 4 as opposed to model 1 compared to the Ticino simulation suggests that outsiders, aware of the existence of the bid-rigging cartel, have tried to benefit from an umbrella effect (Bos and Harrington, 2010). If the cartel did not enroll them, they have submitted bids nearer to collusive bids in some tenders, which have less disturbed the statistical patterns produced by bid rigging. However, even in the presence of strategic interactions, models 2 to 4 still outperform model 1 with a 3 to 5.7 percentage point decrease in the error rate by roughly more than 20% in some cases. Therefore, competition agencies should consider summary screens for subgroups to detect both complete and incomplete bid-rigging cartels.

Like for the Ticino simulation, the benchmarking method performs poorly when flagging incomplete bid-rigging cartels and does no better than tossing a coin. Specifically for truly collusive tenders, the correct classification rate is only between 8.7% and 14.7%.

Table 2.7: Correct classification rate in the Swiss data

Sample	Cart.F	Perc.Cart.F	Tenders	Bench.	M1	M2	M3	M4	M5
1	> 1	71%	All	0.524	0.612	0.637	0.642	0.645	0.673
			Comp.	0.901	0.612	0.607	0.612	0.62	0.646
			Coll.	0.147	0.615	0.671	0.677	0.677	0.706
2	> 2	75%	All	0.525	0.648	0.665	0.675	0.678	0.708
			Comp.	0.901	0.652	0.643	0.645	0.65	0.683
			Coll.	0.148	0.647	0.691	0.71	0.709	0.737
3	> 3	79%	All	0.511	0.706	0.722	0.748	0.745	0.759
			Comp.	0.901	0.705	0.688	0.705	0.707	0.719
			Coll.	0.121	0.708	0.758	0.792	0.784	0.800
4	> 4	83%	All	0.506	0.743	0.770	0.8	0.798	0.814
			Comp.	0.901	0.755	0.751	0.764	0.771	0.783
			Coll.	0.111	0.735	0.791	0.835	0.826	0.846
5	> 5	88%	All	0.494	0.766	0.805	0.813	0.818	0.841
			Comp.	0.901	0.771	0.769	0.786	0.788	0.813
			Coll.	0.087	0.763	0.844	0.842	0.849	0.871
Compl. Cartel	All	100%	All	0.617	0.826	0.813	0.82	0.823	0.833
			Comp.	0.900	0.83	0.818	0.819	0.827	0.833
			Coll.	0.334	0.823	0.808	0.823	0.82	0.834

Notes: "Sample", "Cartel.F", "Per.Cart.F", "Tenders", "Bench.", "M1", "M2", "M3", "M4" and "M5" denote the sample, the number of cartel firms in the collusive tenders, the percentage of cartel firms in the collusive tenders, the type of tenders, the results produced by the screening methods of the benchmarking approach, model 1, model 2, model 3, model 4 and model 5 respectively. For the type of tenders, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively.

When looking at the variable importance as reported in Table 2.8, we find for all models and samples that the Kolmogorov-Smirnov statistic (KS) is an important predictor. In many cases, it is among the three most important variables. Even if both collusive and competitive tenders generally do not follow a uniform distribution, the collusive ones usually are less uniform. Therefore, the Kolmogorov-Smirnov statistic for deviations from the uniform distribution tends to exhibit notably higher values in rigged tenders than in competitive tenders.

The random forest generally picks up a balanced set of screens for the variance and asymmetry along with the Kolmogorov-Smirnov statistic for model 1 in all samples. Specifically for the sample with complete cartels, we observe for models 2 to 5 that the random forest selects screens for the variance, mainly the coefficient of variation (CV) and the spread (SPD), along with the Kolmogorov-Smirnov statistic (KS). Screens for asymmetry in the distribution of bids remain unselected for models 2 and 5 when the cartel is complete, as in the Ticino case. However, when cartels are incomplete, the random forest selects for models 2 to 5 screens for asymmetry in the distribution of bids, mostly skewness (SKEW), relative distance (RD), percentage difference (DIFFP), and alternative distance (RDALT). Even though the results suggest that screens for asymmetry are less important than screens for variance and the Kolmogorov-Smirnov statistic (KS), as in the Ticino case, we find that screens for asymmetry help in detecting bid-rigging when a cartel is incomplete.

For all the samples with incomplete cartels, the minima and maxima of the summary screens are the most important variables, while complete cartels prevail the mean and the median. As for the Ticino case, the results suggest that a few competitive bids sufficiently disturb the statistical pattern produced by bid rigging that it becomes difficult to detect collusion by tender-based screens. In contrast, the use of the minimum or maximum of summary screens mitigates the distortion of competitive bids in the statistical patterns produced by bid rigging and makes possible the detection of incomplete and complete bid-rigging cartels in the Swiss data.

Table 2.8: Important predictors for the Swiss data

	Model 1		Model 2		Model 3		Model 4		Model 5	
	IV	MDG	IV	MDG	IV	MDG	IV	MDG	IV	MDG
1	SKEW	22.48	MIN3CV	9.98	MIN4CV	9.87	MIN4CV	5.8	STDBIDS	5.94
	RDNOR	21.73	MAX3KS	9.93	MAX4KS	9.55	MAX4KS	5.56	MEANBIDS	5.67
	SPD	21.67	MIN3SPD	9.66	MIN4SPD	9.34	MIN4SPD	5.39	MIN4CV	5.27
	KS	21.31	MAX3DIFFP	6.56	MIN4SKEW	8.14	MAX4RD	4.66	MAX4KS	5.11
	DIFFP	20.75	MIN3DIFFP	6.47	MAX4RD	8.08	MIN4SKEW	4.61	MIN4SPD	4.92
2	KS	19.64	MIN3CV	9.27	MIN4CV	9.26	MIN4CV	5.61	MIN4CV	5.23
	RDNOR	19.56	MAX3KS	9.23	MAX4KS	9.12	MAX4KS	5.54	STDBIDS	5.03
	SPD	19.51	MIN3SPD	8.98	MIN4SPD	8.87	MIN4SPD	5.32	MAX4KS	5.02
	SKEW	18.66	MIN3DIFFP	5.62	MAX4RD	7.71	MAX4RD	4.63	MEANBIDS	4.97
	CV	18.38	MAX3DIFFP	5.51	MIN4SKEW	7.67	MIN4SKEW	4.56	MIN4SPD	4.78
3	RDNOR	16.31	MAX3KS	8.42	MIN4CV	8.32	MIN4CV	5.46	NBRBIDS	5.88
	SPD	15.04	MIN3CV	8.32	MAX4KS	8.08	MAX4KS	5.43	MIN4CV	4.9
	KURTO	14.62	MIN3SPD	8.2	MIN4SPD	7.52	MAX4RD	5.01	MAX4KS	4.87
	KS	14.35	MAX3RD	5.09	MAX4RD	7.5	MIN4SPD	4.96	MAX4RD	4.53
	RDALT	14.02	MAX3RDALT	5.09	MIN4SKEW	7.4	MIN4SKEW	4.94	MIN4SKEW	4.49
4	RDNOR	13.28	MIN3SPD	8.15	MIN4CV	7.6	MIN4CV	5.15	NBRBIDS	7.49
	SPD	12.12	MAX3KS	8.13	MAX4KS	7.49	MAX4KS	5.09	MIN4CV	4.52
	KS	11.52	MIN3CV	8.09	MIN4SPD	7.46	MIN4SPD	4.99	MAX4KS	4.45
	RDALT	11.42	MAX3RD	4.51	MIN4SKEW	7.39	MIN4SKEW	4.96	MIN4SKEW	4.38
	CV	11.06	MIN3SKEW	4.49	MAX4RD	6.92	MAX4RD	4.66	MIN4SPD	4.33
5	KS	10.35	MIN3SPD	6.43	MAX4KS	6.05	MIN4CV	4.08	NBRBIDS	7.27
	SPD	10.14	MIN3CV	5.98	MIN4CV	6.05	MAX4KS	3.99	MIN4CV	3.51
	CV	9.83	MAX3KS	5.94	MIN4SPD	6.02	MIN4SPD	3.96	MIN4SPD	3.4
	RDNOR	9.53	MAX3RD	3.75	MIN4SKEW	5.09	MAX4RDALT	3.25	MAX4KS	3.39
	RDALT	8.03	MAX3RDALT	3.74	MAX4RDALT	4.96	MAX4RDNOR	3.25	MAX4RDALT	2.86
Compl. Cartel	KS	54.65	MEDIAN3CV	20.04	MEDIAN4KS	20.23	MEDIAN4KS	11.4	MEDIAN4KS	10.86
	CV	50.9	MEDIAN3SPD	18.68	MEDIAN4SPD	20.04	MEDIAN4SPD	11.27	MEDIAN4CV	10.82
	SPD	32.96	MEAN3CV	18.49	MEDIAN4CV	19.56	MEDIAN4CV	11.04	MEDIAN4SPD	10.75
	DIFFP	18.37	MEDIAN3KS	18.46	MEAN4KS	16.91	MEDIAN3CV	10.29	MEDIAN3CV	9.96
	RDNOR	17.1	MEAN3SPD	17.64	MEAN4CV	16.67	MEDIAN3KS	9.37	MEDIAN3SPD	9.22

Notes: "IV" and "MDG" denote the important variables selected by the random forest and the mean decrease in Gini index. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP" and "KURTO" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference and the kurtosis statistic respectively.

2.4.3 Robustness analysis

We investigate the robustness of our correct classification rates by discarding the most important predictors and applying the random forest to the remaining predictors. Since model 1 uses fewer predictors than the other models, we leave out the three most important variables, while for models 2 to 5, we drop the five best predictors. Table 2.9 reports the difference in percentage points in the correct classification rates when keeping vs. dropping the respective predictors.

Table 2.9: Differences between original random forest and random forest with discarded variables

Sample	Cart.F	Perc.Cart.F	Tenders	M1	M2	M3	M4	M5
1	> 1	71%	All	3.4	1.1	-0.2	0	1.8
			Comp.	3.6	-0.9	-1.2	-0.3	1.8
			Coll.	3.2	3.1	0.9	0.3	1.8
2	> 2	75%	All	1.8	1.3	-0.4	0	1.4
			Comp.	2.9	-0.4	-1.1	-0.3	1.4
			Coll.	0.6	3	0.3	0.3	1.3
3	> 3	79%	All	4.8	0.1	0.1	0.1	0.5
			Comp.	6.3	-0.4	-0.2	-0.2	0.1
			Coll.	3.2	0.6	0.6	0.5	0.9
4	> 4	83%	All	3.4	1.2	0.2	0.9	1.8
			Comp.	4.1	0.9	-0.2	0.5	0.9
			Coll.	2.6	1.6	0.6	1.3	2.7
5	> 5	88%	All	0.6	2	0.2	0.2	1.7
			Comp.	1.3	0.9	0.5	0.2	1.5
			Coll.	-0.2	3.1	-0.4	0	1.7
Compl. Cartel	All	100%	All	1	-0.2	0.4	0	0
			Comp.	0.7	0.5	0.6	0.2	-0.2
			Coll.	1.3	-0.8	0.3	-0.1	0.2

Notes: "Sample", "Cartel.F", "Per.Cart.F", "Tenders", "M1", "M2", "M3", "M4" and "M5" denote the sample, the number of cartel firms in the collusive tenders, the percentage of cartel firms in the collusive tenders, the type of tenders, model 1, model 2, model 3, model 4 and model 5 respectively. For the type of tenders, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively.

The overall correct classification rate of model 1 in samples 1, 3, and 4, keeping all variables, predominates when dropping the three best predictors by 3.4 to 4.8 percentage points. Considering the other models and samples, we observe more or less the same predictive power when discarding the most important variables. Therefore, the remaining predictors seem to be suitable substitutes for the discarded ones. Other variables become more important when the most important predictors are omitted, and the correct classification rate is hardly affected.

Furthermore, we investigate robustness for the type of contract. For both the cartel and post-cartel periods, we subsequently only consider contracts for road construction and asphalting. We exclude contracts for civil engineering and mixed contracts combining civil engineering with road construction or asphalting. This, because certain specific characteristics of contracts in civil engineering might affect the screens and, therefore, the correct classification rate. Dropping mixed contracts and contracts for civil engineering permits us to verify whether this importantly affects the correct classification rate among the remaining contracts for road construction and asphalting. Table 2.10 reports the difference in percentage points in the correct classification rates when using all contracts vs. using contracts for road construction and asphalting only.

In samples 1 and 2, we find the correct classification rates of the random forest for road construction and asphalting contracts to be superior to the classification rate of the random forest with all types of contracts. For example, the difference in the (overall) classification rate of model 1 in samples 1 and 2 accounts for 6.2 and 2.8 percentage points, respectively. A possible explanation could be that we implicitly suppress some competitors when we keep only the road construction and asphalting contracts. For example, in sample 1, the average percentage of collusive bidders is 80.9%, which is considerably higher, as is the situation with all types of contracts (71.1%, see Table 2.7). Therefore, the cartel percentage is higher for this restricted sample of road construction and asphalting contracts alone and explains the higher performance in samples 1 and 2. In sample 3, the situation begins to change for both types, the correct classification rates being quasi identical. Noticeably for all models in samples 3 and 4, the differences increase again. However, not as strong as before and in the opposite direction. Therefore, for an almost identical average percentage of cartel participants, the correct classification rates of the random forest for all types of contracts are slightly superior to those for road construction and asphalting.

To investigate the robustness of the correct classification rate across different machine learning algorithms, we also assess the performance of lasso regression and an ensemble method (including bagged trees, random

Table 2.10: Differences between original random forest and random forest using only contracts for road construction and asphalting

Sample	Cart.F	Perc.Cart.F	Tenders	M1	M2	M3	M4	M5
1	> 1	81%	All	-6.2	-5.2	-6.7	-5.5	-3.6
			Comp.	-7.2	-4.6	-6.9	-5.7	-3.8
			Coll.	-5.2	-5.8	-6.2	-5.2	-3.3
2	> 2	82%	All	-2.8	-3	-4.3	-3.6	-1.2
			Comp.	-3.6	-3.1	-4.7	-3.9	-1.4
			Coll.	-1.9	-2.9	-3.8	-3.2	-1
3	> 3	84%	All	0.1	-0.6	-0.9	-0.6	0.2
			Comp.	-1.9	-1.5	-2.7	-2.1	-1.9
			Coll.	2	0	0.5	0.5	1.9
4	> 4	86%	All	1.5	2.2	2	2.4	3.4
			Comp.	2.2	2.7	1.9	2.5	3.1
			Coll.	1	1.9	1.9	2.2	3.6
5	> 5	88%	All	2.6	3.3	2	2.1	2.7
			Comp.	3.1	3.3	2.2	1.8	3
			Coll.	1.9	3.2	1.9	2.4	2.3
Compl. Cartel	All	100%	All	0.7	0.3	0	0.3	0.3
			Comp.	0.9	-0.3	-0.4	0	-0.2
			Coll.	0.4	0.9	0.4	0.5	0.8

Notes: "Sample", "Cartel.F", "Per.Cart.F", "Tenders", "M1", "M2", "M3", "M4" and "M5" denote the sample, the number of cartel firms in the collusive tenders, the percentage of cartel firms in the collusive tenders, the type of tenders, model 1, model 2, model 3, model 4 and model 5 respectively. For the outcome classification, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively.

forest, and neural networks) for all models and samples. We explain these algorithms, also outlined by Huber and Imhof (2019), in more detail in Appendix 2.B. Table 2.11 reports the difference in percentage points in the correct classification rates of the random forest minus the correct classification rates of the lasso and ensemble method.

Table 2.11: Differences between original random and the lasso and the ensemble method

		Sample 1		Sample 2		Sample 3		Sample 4		Sample 5		Compl. Cart.	
		lasso	ens.	lasso	ens.	lasso	ens.	lasso	ens.	lasso	ens.	lasso	ens.
	All	-1.5	-2	-1.6	-2.3	-3.7	-2.7	-5.6	-6.7	-6	-4.3	0.9	0.4
M1	Comp.	-3.5	-1.4	-3.1	-0.7	-4.3	0.4	-5.9	0.9	-4	2.2	4.7	0.6
	Coll.	0.1	-2.9	-0.4	-3.9	-3.1	-5.9	-5.4	-14.1	-8.3	-11	-3	0.1
	All	-1.3	-1.1	-0.8	-0.7	-1.8	-1.3	-1.6	-1.8	-0.8	0.4	0.2	-0.8
M2	Comp.	-13	-9.9	4.8	1.5	2.2	1.4	2.7	1.3	2	2.4	4	-0.8
	Coll.	10.3	7.5	-6.5	-2.9	-5.9	-4	-5.7	-4.7	-3.7	-2	-3.6	-0.9
	All	-1.4	-0.9	-1.5	-1.2	-1.5	-1	-0.9	0.2	-0.3	0.6	0.1	0.1
M3	Comp.	0.9	-0.8	0.5	-0.6	1.9	0.7	0.3	1.9	2.6	-6.1	4.6	1.2
	Coll.	-3.8	-1.1	-3.5	-1.8	-4.9	-2.6	-2.2	-1.7	-3.4	7.5	-4.4	-1
	All	-0.7	-0.7	-0.7	-0.6	-1.4	-0.8	-0.9	-0.5	-0.3	1	0.2	-0.1
M4	Comp.	2.7	0.3	2.9	0.2	2.2	0.9	1.5	1.2	2.6	2.8	4.7	0.9
	Coll.	-4.3	-1.8	-4.4	-1.5	-5	-2.4	-3.2	-2	-3.4	-0.9	-4.3	-1.1
	All	-1.5	-2.9	-1.3	-1.4	-2.5	-1.1	-1.6	-1.8	1	-2.4	0.6	-0.1
M5	Comp.	-3.4	-3.4	-2.4	-2.2	-2.2	-1.4	-3	-0.1	2.4	5.9	4.4	0.6
	Coll.	0.3	-2.2	-0.1	-0.6	-2.7	-0.8	-0.3	-3.3	-3.5	-13.8	-3.2	-0.9

Notes: "lasso", "ens.", "M1", "M2", "M3", "M4" and "M5" denote the lasso, the ensemble of method, model 1, model 2, model 3, model 4 and model 5 respectively. For the type of tenders, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively.

Considering samples 1 and 2 in Table 2.11, we find that the lasso and ensemble method slightly outperform the random forest. The maximum difference in (overall) correct classification rates across models and samples amount to 2.9 percentage points. While the somewhat lower rates speak against the random forest, performance is more uniform. Therefore, there is less divergence across both the competitive and collusive periods, which may be important to practitioners. For samples 3, 4, and 5, in general, the lasso and ensemble method slightly outperform the random forest, in two cases even more profoundly, with higher correct classification rates of 4.3 to 6.7 percentage points for model 1 in samples 4 and 5. This implies that in samples 4 and 5 (with a high amount of collusive bidders), considering summary screens does not significantly improve the predictive power of the lasso and ensemble method, in contrast to the random forest. On the other hand, and as for samples 1 and 2, the random forest shows a more uniform performance (e.g. correct classification rates are not too different for competitive and collusive tenders). We find a similar performance of (overall) correct classification rates between the random forest and the ensemble method for complete cartels. However, the random forest slightly dominates the lasso regression. Considering the deviation across prediction of the competitive and collusive periods, the random forest and the ensemble method show a less divergent performance than the lasso.

To conclude, in Table 2.11 the random forest shows a somewhat lower correct classification rate than the lasso and the ensemble method. Still, it exhibits a more homogeneous correct classification rate across both the competitive and collusive tenders. All in all, this robustness check shows the stability of our preliminary results.

2.5 Conclusion

In this paper, we have suggested a robust method for flagging bid rigging in tenders that is likely to be more powerful for detecting incomplete cartels than previously suggested methods. Our approach combined screens, e.g. statistics derived from the distribution of bids in a tender, with machine learning to predict the probability of collusion. As a methodological innovation, we calculated the screens for all possible subgroups of three or four bids within a tender and considered summary statistics as the mean, median, maximum, and minimum for each screen (so-called summary screens) as predictors in the machine learning algorithm. We tackled the issue that competitive bids in incomplete cartels distort the statistical signals produced by bid rigging using these summary screens.

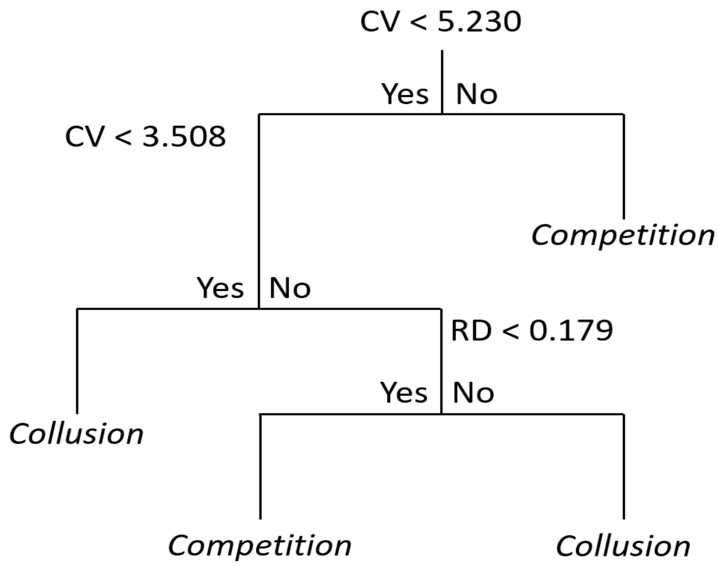
We first applied our approach to the Ticino bid-rigging cartel and found a correct out-of-sample classification rate of 77% to 86% even in the presence of simulated competitive bids. Our approach increasingly outperformed other methods using tender-based screens as the number of competitive bids per tender increased. In this simulation, there was no strategic reaction by design or interaction between competitive and collusive bidders. We also applied our method to data from the investigations involving partial cartels in the regions See-Gaster and Graubünden in Switzerland to allow for such reactions. The out-of-sample performance of machine learning using summary screens (calculated for all possible subgroups of three and four bids) as predictors again outperformed other screening methods. However, the performance of all machine learning-based methods in all models still decreased concerning the relative number of competitive bids in the data of the investigations involving incomplete cartels. This decrease indicates that cartel participants anticipated competition from non-cartel bidders. However, the less divergent classification rate between models indicates that it is likely that competitive bidders bid closer to collusive bids trying to benefit from the umbrella effect (Bos and Harrington, 2010), usually if there are a large number of cartel bidders.

Compared to tender-based, summary screens increased the correct classification rate by 5 to 10 and 3 to 5.7 percentage points for incomplete cartels in the Ticino simulation and the Swiss data from See-Gaster and Graubünden, respectively. This implies a substantial decrease in the error rate (one minus the correct classification rate) of 42.6% and 22.2% for the Ticino simulation and the Swiss data respectively. As screening by competition agencies can trigger investigations with legal consequences for potential cartel members, such decreases in the error rate appear more desirable. Thus, our results demonstrate the usefulness of combining machine learning with an improved set of statistical screens to reduce distortions of competitive bids in incomplete cartels. Moreover, the approach appears promising for detecting collusion in other industries or countries, being on the agenda for future research.

Appendices

2.A Adjusting the benchmarking rule

Figure 2.A.1: Adjusted classification tree



2.B Details about lasso regression and the ensemble method

Here discuss in more detail the machine learning approaches of the lasso regression and ensemble method. Similar to the random forest, the lasso regression and ensemble method require randomly splitting the data into training (used for estimating the model parameters) and test data (used for out-of-sample prediction and performance evaluation). Again, our training and test samples contain 75% and 25% of the observations respectively. Lasso regression corresponds to a penalized logit regression, where the penalty term restricts the sum of absolute coefficients on the regressors. Coefficients of less predictive variables shrink towards or even exactly to zero depending on the penalty term. Therefore, lasso regression may perform predictor selection. Estimating lasso logit coefficients is based on the following optimization problem:

$$\max_{\beta_0, beta} \left\{ \sum_{i=1}^{n} \left[y_i \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \right) - log \left(1 + e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}} \right) \right] - \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \quad (2.B.1)$$

where β_0 denotes intercept and slope coefficients on the predictors, β the slope coefficients on the predictors, x the vector of predictors, i indexes an observation in our data set (with n being the number of observations), j indexes a predictor (with p being the number of predictors), and λ a penalty term larger than zero. We use the same predictors as described in the main text for the different models. In our application, we use the hdm package by Chernozhukov, Hansen, and Spindler (2016) for the statistical software R. We apply 15-fold cross-validation to select the penalty term λ based on minimizing the mean squared error of prediction.

For the ensemble method, as by Huber and Imhof (2019) we apply the "Super-Learner" package for R by van der Laan, Polley, and Hubbard (2008) with default values for bagged regression tree, random forest and neural network algorithms in the "ipredbagg", "cforest" and "nnet" packages respectively. The ensemble method also relies on training data to estimate the model parameters and test data for prediction and performance evaluation. However, any estimation step now consists of a weighted average of bagged classification trees, random forest and neural networks. Bagged trees involve estimating single trees (rather than random forest) repeatedly using the outcome residuals of the respective previous tree as outcome. Rather than splitting the predictor space, neural networks aim at fitting a system of nonlinear functions that models the influence of the predictors of collusion in a flexible way. To do so, we model the association between the predictors and the outcomes using a network of non-linear intermediate functions, so-called hidden notes. Several layers of hidden nodes allow modelling associations and interactions between the predictors in a flexible way, with more nodes and layers increasing the variance but reducing the bias.

2.C Results for the statistical tests between the simulated bids and the competitive bids in the Ticino case

In the following, we test whether the simulated competitive bids are similar to the competitive bids of the post-cartel period. We calculate the screens for the five simulated competitive bids for each collusive tender. We test whether the screens differ from the screens calculated with the competitive bids in the post-cartel period. Since the screens are not normally distributed, we apply non-parametric tests to our data, which do not assume any particular distribution in the test procedures (see also Imhof, Karagök, and Rutz, 2018, Imhof, 2019). First, we apply the Mann-Whitney test (also called the Wilcoxon rank sum test).[28] Second, to ensure the robustness of the results, we use the Kolmogorov-Smirnov test, a more general test examining any kind of difference between the samples.[29]

Table 2.12: Statistical tests for the screens calculated with the simulated competitive bids against the competitive bids of the post-cartel period

Screens	z-statistic	p-value MW	KSa	p-value KS
CV	-1.14	0.2525	1.24	0.0934
KURTO	-0.93	0.3545	1.04	0.2311
SPD	-0.45	0.6541	1.12	0.1623
DIFFP	-1.64	0.1014	1.90	0.0015
SKEW	-0.06	0.9524	0.87	0.4377
RD	-0.10	0.9215	0.78	0.5820
RDNOR	0.1290	0.9874	0.83	0.4913
RDALT	0.1179	0.9061	0.77	0.5901
KS	1.31	0.1890	1.21	0.1084

Notes: "Screens", "z-statistic", "p-value MW" denote the screens tested, the z-statistic of the Mann-Whitney test and the p-value of the Mann-Whitney test respectively. "KSa" and "p-value KS" denote the asymptotic Kolmogorov-Smirnov statistic and the p-value of the Kolmogorov-Smirnov test respectively.

Table 2.12 indicates the test results. We find no rejection for all the tests (at the 5% significance level), except for the Kolmogorov-Smirnov test applied to the percentage difference (DIFFP). To sum up, the screens calculated with the simulated competitive bids do not significantly differ from the screens calculated with the "real" competitive bids in the post-cartel period. Therefore, the simulated competitive bids exhibit more or less the same statistical pattern as the "real" competitive bids. This result indicates that our simulation procedure adequately generates competitive bids for the purposes of our analyses.

2.D Descriptive statistics for the Ticino Cartel

Figure 2.D.1 visualizes the distribution of tenders for a predetermined number of bids. In either period, most tenders had four to eight bids (see Table 2.13). Table 2.14 shows the empirical distribution of the bids for each period. Both periods contain contracts of different values varying from several tens of Swiss francs to up

[28] See Rice (2007) chapter 11, page 435 ff.; Hollander, Wolfe, and Chicken (2014) chapter 4, page 115 ff.

[29] See Hollander, Wolfe, and Chicken (2014) chapter 5, page 190 ff.

three to five million Swiss francs. The mean and the median of the cartel period are superior to those of the post-cartel period. In either periods, the contract values exhibit higher means than medians, indicating a right-skewed distribution with outliers of comparably high contract values.

Figure 2.D.1: Distribution of tenders for a predetermined number of bids for the Ticino data

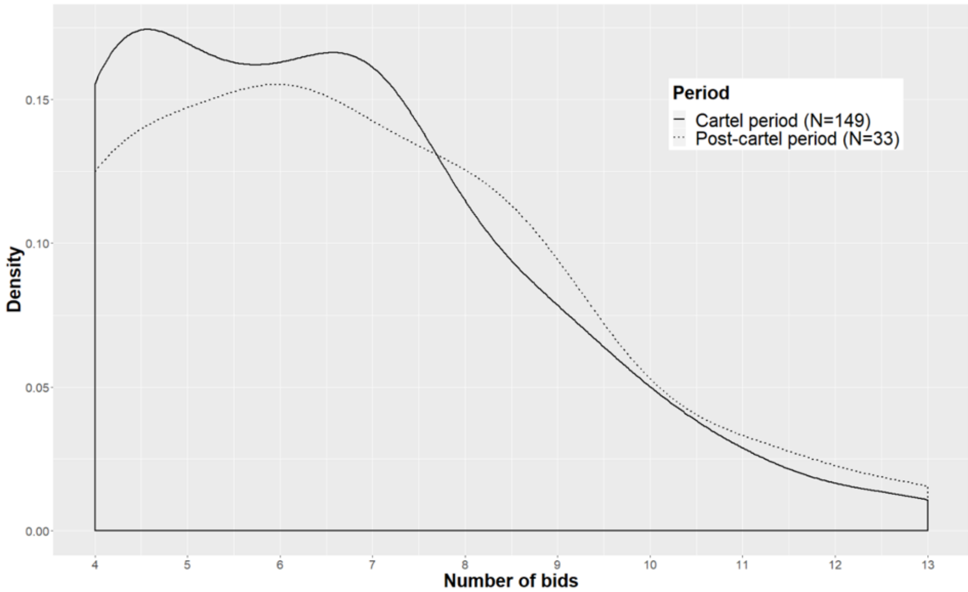


Table 2.13: Numbers of bids in a tender for the cartel and post-cartel periods in Ticino

Number of bids in a tender	4	5	6	7	8	9	10	10+
Cartel Period	32	24	23	28	15	12	7	8
Post-cartel Period	8	2	8	3	5	4	2	1

Table 2.14: Empirical distributions of collusive and competitive bids in the Ticino data (in million CHF)

	Cartel period	Post-cartel Period
Mean	1.08	0.69
Std	1.01	0.75
Min	0.02	0.04
Lower Q.	0.36	0.25
Median	0.78	0.44
Upper Q.	1.47	0.68
Max	4.85	2.95
N	149	33

Notes: "Mean", "Std", "Min", "Lower Q.", "Median", "Upper Q.", "Max", and "N" denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively.

2.E Descriptive statistics for the Swiss data

Figure 2.E.1 depicts the distribution of the number of bids per tender for complete cartels, incomplete cartels and competitive tenders respectively. While tenders with four to seven bids dominate, there is also a sufficient number of tenders with eight or more bids (see Table 2.15). Table 2.16 depicts the empirical distribution of the bids for each type of tender. The empirical distributions for tenders with complete cartels and with incomplete cartels are similar. However, this is not the case for competitive tenders, which have many more contracts, varying in value from one thousand Swiss francs to 148 million Swiss francs. As for the data from the Ticino cartel, all the bids' empirical distributions are right-skewed, such that the mean is higher than the median, but more strongly so for competitive tenders than for complete and incomplete cartels.

Table 2.15: Numbers of bids in a tender in the Swiss data

Number of bids	4	5	6	7	8	9	10	10+
Complete cartels	94	50	29	24	33	33	23	24
Incomplete cartels	56	36	38	40	27	28	24	38
Competitive tenders	786	559	365	257	158	129	74	70

Note: 'Complete cartels', 'Incomplete Cartels' and 'Competitive tenders' denote tenders with complete cartels, incomplete cartels and non-colluding firms, respectively.

Figure 2.E.1: Distribution of tenders for a predetermined number of bids in the Swiss data

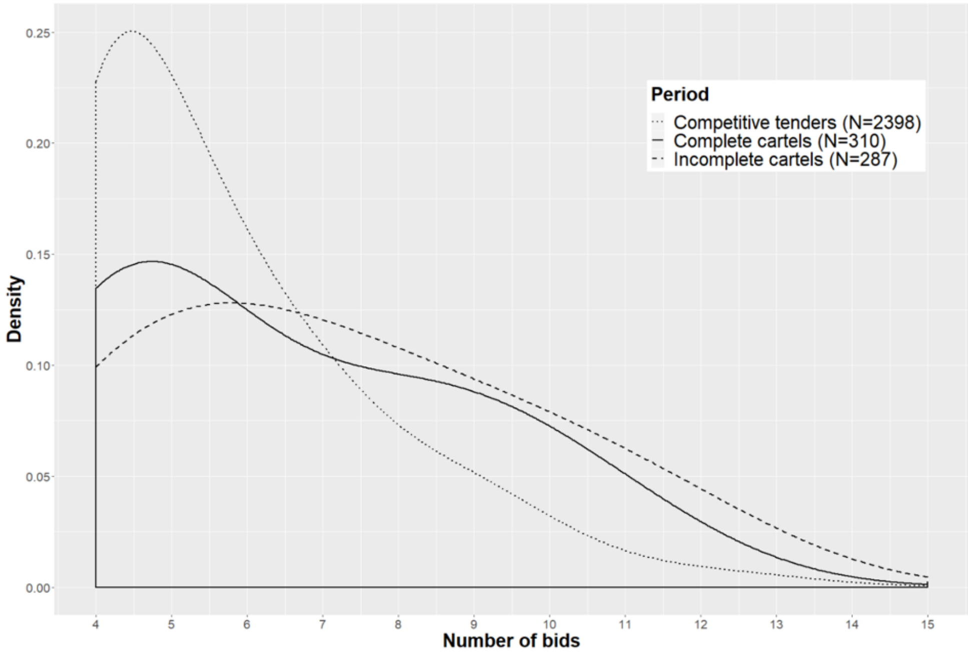


Table 2.16: Empirical distributions of bids in the Swiss data (in million CHF)

	Complete cartels	Incomplete cartels	Competitive tenders
Mean	0.36	0.4	0.72
Std	0.36	0.47	3.81
Min	0.03	0.02	0.001
Lower Q.	0.16	0.12	0.13
Median	0.29	0.25	0.31
Upper Q.	0.44	0.50	0.66
Max	3.45	3.46	147.73
N	310	287	2,398

Notes: "Mean", "Std", "Min", "Lower Q.", "Median", "Upper Q.", "Max", and "N" denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively.

2.F Descriptive statistics for predictors

In the following tables, we present tables of descriptive statistics for all the different samples used in the empirical analyses for both the Ticino simulation and the Swiss data. We review here the most important key information drawn from the descriptive statistics for the coefficient of variation (CV) and the normalized distance (RDNOR). Similar interpretations can be made for other screens.

For the Ticino cartel, the coefficient of variation exhibits a mean of 3.25 and a median of 2.97 with a low standard deviation of 1.18 (see Table 2.17). This contrasts with the post-cartel period, in which the mean and the median of the coefficient of variation amount to 9.51 and 8.49 respectively, with a larger standard deviation of 5.38 compared to the collusive period (see Table 2.18). If a large majority of the observations in the cartel period are below 3.83, we only find a few CVs below 5.65 in the post-cartel period, considering the upper and lower quartile respectively. For the Swiss data, we find similar values for the cartel period with a mean of 3.66, a median of 3.29 and a standard deviation of 2.09 (see Table 2.20). They all contrast with the values found for the post-cartel period (competitive tenders) in the Swiss data with a mean of 10.12, a median of 8.45 and a standard deviation of 7.89 (see Table 2.21). Note that in the following empirical analyses we select only competitive tenders with an anonymized year superior or equal to 8. Since collusive tenders superior or equal to 8 are absent in the anonymized years, we conclude that both bid-rigging cartels collapsed in this post-cartel period. This ensures that a competitive tender in the post-cartel period is really a ”competitive” one.

If we look at the coefficient of variation for the incomplete bid-rigging cartel in sample 1 of the Swiss data (collusive tenders characterized by incomplete cartels with at least two colluding firms), the CV is affected by the presence of competitive bids with a mean of 7.79, a median of 6.79 and a standard deviation of 3.89 (see Table 2.22). Looking more precisely at the minimum of all coefficients of variation calculated for subgroups of four bids in a tender (MIN4CV), we find a mean of 3.16, a median of 2.26 and a standard deviation of 2.97 for the incomplete bid-rigging cartels in sample 1 (see Table 2.22). However, the MIN4CV for the competitive tenders exhibits higher values with a mean of 6.24, a median of 4.49 and a standard deviation of 6.77 (see Table 2.21). Noteworthy, the differences are weaker for the maxima of all coefficients of variation calculated for subgroups of four bids (MAX4CV), between incomplete cartels in sample 1 (mean of 10.63, median of 9.43 and a standard deviation of 5.46 in Table 2.22) and competitive tenders (mean of 12.14, median of 10.13 and a standard deviation of 9.73 in Table 2.21). This example is crucial to understand the benefit delivered by summary statistics of the screens. Even if the maxima of the coefficient of variation is high in both cases of incomplete bid-rigging cartels and competition, the minima diverge notably and could be used to differentiate between competition and collusion.

The normalized distance (RDNOR) assumes higher values in collusive periods than in competitive periods. For example, the RDNOR exhibits a mean of 2.93 and a median of 2.72 with a standard deviation of 1.35 for the Ticino cartel (see Table 2.17). In the post-cartel period, the values of the RDNOR are lower with a mean of 1.02, a median of 0.74 and a standard deviation of 0.80 (see Table 2.18). Although less notable, we find a divergence in the Swiss data between collusive tenders (with a mean of 1.38, a median of 1.24 and a standard deviation of 0.79 in Table 2.20) and competitive tenders (with a mean of 1.04, a median of 0.87 and a standard deviation of 0.82 in Table 2.21). We find similar values for the minima of the normalized distance (MIN4RDNOR) between incomplete bid-rigging cartels in sample 1 (mean of 0.37, median of 0.15, standard deviation of 0.54 in Table 2.22) and competitive tenders in the Swiss data (with a mean of 0.51, a median of 0.29 and a standard deviation of 0.56 in Table 2.21). The values are more divergent for the maxima (MAX4RDNOR) between the two types of tender. For the incomplete bid-rigging cartels in sample 1, we observe a mean of 2.18, a median of 2.37 and a standard deviation of 0.68 in Table 2.22, contrasting with competitive periods, which exhibit a mean of 1.62, a median of 1.74 and a standard deviation of 0.81 in Table 2.21. The result indicates that the maxima of the RDNOR could be used to discriminate between incomplete bid-rigging cartels and competition.

Table 2.17: Descriptive statistics for the collusive tenders in Ticino (without simulated bids)

Predictors	Mean	Std	Min	Median.	Max	N
NBRBIDS	6.54	2.16	4	6	13	149
MEANBIDS	1134.58	1048.71	23.16	836.59	4967.5	149
STDBIDS	34.08	30.13	0.82	25.84	136.9	149
CV	3.25	1.18	1.52	2.97	10.2	149
KURTO	2.71	2.12	-3.08	2.84	8.14	149
SKEW	-1.13	0.96	-2.76	-1.34	2.21	149
SPD	0.1	0.04	0.04	0.09	0.37	149
D	49.15	43.01	1.64	39.1	272.45	149
RD	4.09	3.37	0.31	3.14	23.02	149
RDNOR	2.93	1.35	0.53	2.72	6.95	149
RDALT	7.06	6.4	0.43	5.47	40.02	149
DIFFP	5.09	1.98	1.03	5.13	21.74	149
KS	34.27	9.93	9.77	34.26	66.05	149
MIN3CV	0.7	0.8	0	0.41	4.36	149
MAX3CV	5.03	1.87	1.99	4.72	16.64	149
MEAN3CV	2.89	1.12	1.24	2.63	8.61	149
MEDIAN3CV	2.87	1.29	0.4	2.94	6.58	149
MIN3SKEW	-1.67	0.17	-1.73	-1.73	-0.29	149
MAX3SKEW	1.33	0.72	-1.38	1.7	1.73	149
MEAN3SKEW	-0.44	0.48	-1.56	-0.39	0.94	149
MEDIAN3SKEW	-0.68	0.69	-1.69	-0.74	1.26	149
MIN3D	7.67	18.44	0	1.17	106.92	149
MAX3D	74.23	60.37	1.73	60.09	319.13	149
MEAN3D	37.58	36.04	1.03	28.73	186.08	149
MEDIAN3D	36.6	41.83	1.1	21.92	272.45	149
MIN3RD	0.43	0.75	0	0.1	4.86	149

See next page

Predictors	Mean	Std	Min	Median.	Max	N
MAX3RD	3177.54	31051.89	1.72	73.49	376725.28	149
MEAN3RD	48.87	279.28	0.75	8.67	3060.42	149
MEDIAN3RD	4.47	14.02	0.49	2.6	171.76	149
MIN3RDNOR	0.32	0.38	0	0.14	1.55	149
MAX3RDNOR	1.89	0.15	1.1	1.96	2	149
MEAN3RDNOR	1.2	0.22	0.61	1.17	1.74	149
MEDIAN3RDNOR	1.27	0.28	0.5	1.27	1.85	149
MIN3RDALT	0.3	0.53	0	0.07	3.44	149
MAX3RDALT	2246.86	21957	1.22	51.97	266385	149
MEAN3RDALT	34.56	197.48	0.53	6.13	2164.04	149
MEDIAN3RDALT	3.16	9.92	0.34	1.84	121.45	149
MIN3DIFFP	0.51	0.76	0	0.16	4.03	149
MAX3DIFFP	8	3.34	1.87	7.18	34.62	149
MEAN3DIFFP	3.68	1.58	1.19	3.35	15.43	149
MEDIAN3DIFFP	3.42	2.12	0.35	3.23	10.15	149
MIN3SPD	0.01	0.02	0	0.01	0.09	149
MAX3SPD	0.1	0.04	0.04	0.09	0.37	149
MEAN3SPD	0.06	0.02	0.02	0.05	0.19	149
MEDIAN3SPD	0.06	0.03	0.01	0.06	0.14	149
MIN3KS	22.12	6.72	6.04	21.28	50.12	149
MAX3KS	851.31	3306.73	23.21	244.9	39320	149
MEAN3KS	88.19	88.53	17.01	64.31	863.76	149
MEDIAN3KS	49.42	38.86	15.83	34.73	247.55	149
MIN4CV	1.35	1.31	0.04	0.86	6.49	149
MAX4CV	4.21	1.6	1.69	3.92	13.92	149
MEAN4CV	3.06	1.16	1.35	2.81	9.2	149
MEDIAN4CV	3.29	1.46	0.54	2.99	12.88	149
MIN4SKEW	-1.67	0.58	-2	-1.95	1.47	149
MAX4SKEW	0.82	1.29	-1.92	1.47	2	149
MEAN4SKEW	-0.65	0.66	-1.92	-0.57	1.47	149
MEDIAN4SKEW	-0.82	0.76	-1.96	-0.9	1.47	149
MIN4D	18.48	35.07	0	1.54	171.35	149
MAX4D	61.85	49.83	1.7	52.45	307.84	149
MEAN4D	40.84	38.74	1.34	29.47	229.44	149
MEDIAN4D	43.7	43.33	0.98	31.25	272.45	149
MIN4RD	1.15	2.28	0	0.19	13.56	149
MAX4RD	46.81	160.33	0.4	16.33	1865.04	149
MEAN4RD	5.2	4.81	0.4	3.79	34.88	149
MEDIAN4RD	3.27	2.83	0.32	2.34	19.6	149
MIN4RDNOR	0.65	0.75	0	0.28	2.62	149
MAX4RDNOR	2.47	0.53	0.53	2.69	3	149
MEAN4RDNOR	1.59	0.41	0.53	1.55	2.62	149
MEDIAN4RDNOR	1.66	0.5	0.44	1.65	2.73	149
MIN4RDALT	1.2	2.33	0	0.2	13.86	149
MAX4RDALT	49.69	170.21	0.43	17.11	1966.14	149
MEAN4RDALT	5.48	5.21	0.43	4.02	40.06	149
MEDIAN4RDALT	3.41	2.88	0.34	2.53	20.01	149
MIN4DIFFP	1.28	1.87	0	0.25	6.99	149
MAX4DIFFP	6.84	3.12	1.03	6.26	34.1	149
MEAN4DIFFP	4.05	1.79	1.03	3.58	18.1	149
MEDIAN4DIFFP	4.4	2.46	0.35	4.6	21.74	149
MIN4KURTO	-2.38	3.14	-6	-3.19	3.76	149
MAX4KURTO	3.17	1.37	-3.08	3.83	4	149
MEAN4KURTO	1.47	1.18	-3.08	1.55	3.76	149

See next page

Predictors	Mean	Std	Min	Median.	Max	N
MEDIAN4KURTO	1.93	1.27	-3.08	2.09	3.87	149
MIN4SPD	0.03	0.03	0	0.02	0.17	149
MAX4SPD	0.1	0.04	0.04	0.09	0.37	149
MEAN4SPD	0.07	0.03	0.03	0.07	0.24	149
MEDIAN4SPD	0.08	0.04	0.01	0.07	0.34	149
MIN4KS	26.58	8.13	7.22	25.64	59.33	149
MAX4KS	267.41	363.93	15.73	116.15	2643.96	149
MEAN4KS	55.64	36.47	15.73	43.12	195.25	149
MEDIAN4KS	38.72	25.53	7.82	33.39	184.51	149

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov Statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.18: Descriptive statistics for the Ticino cartel in the post-cartel period

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	6.73	2.34	4	6	13	33
MEANBIDS	756.79	785.43	43.97	482.05	3191.78	33
STDBIDS	54.83	49.5	2.79	39.64	209.64	33
CV	9.51	5.38	1.71	8.49	21.12	33
KURTO	-0.08	1.78	-2.83	-0.16	6.06	33
SKEW	0.24	0.85	-1.46	0.31	2.36	33
SPD	0.31	0.2	0.04	0.26	0.84	33
D	29.67	36.08	0.68	17.42	149.28	33
RD	0.77	0.89	0.02	0.41	4.03	33
RDNOR	1.02	0.8	0.06	0.74	3.67	33
RDALT	1.22	1.23	0.05	0.72	4.84	33
DIFFP	5.16	5.02	0.23	3.73	20.6	33
KS	16.52	12.23	5.4	12.4	58.56	33
MIN3CV	2.58	2.8	0.07	1.73	12.23	33
MAX3CV	14.45	8.63	2.15	13.09	36.56	33
MEAN3CV	8.72	4.76	1.65	7.87	19.63	33
MEDIAN3CV	8.67	4.73	1.68	7.97	20.71	33
MIN3SKEW	-1.48	0.55	-1.73	-1.7	0.7	33
MAX3SKEW	1.48	0.47	-0.27	1.69	1.73	33
MEAN3SKEW	0.1	0.51	-1.05	0.21	1.2	33
MEDIAN3SKEW	0.14	0.79	-1.26	0.29	1.2	33
MIN3D	14.25	32.2	0.16	2.58	149.28	33
MAX3D	107.42	100.76	2.88	76.2	471.5	33
MEAN3D	47.3	47.2	1.75	31.57	221.08	33
MEDIAN3D	41	38.93	1.44	28.88	183.76	33

See next page

Predictors	Mean	Std	Min	Median	Max	N
MIN3RD	0.26	0.38	0.01	0.12	1.7	33
MAX3RD	39.4	59.34	0.86	17.33	274.21	33
MEAN3RD	4.37	4.62	0.51	3.4	25.42	33
MEDIAN3RD	1.65	1.17	0.53	1.16	4.24	33
MIN3RDNOR	0.25	0.28	0.01	0.16	1.09	33
MAX3RDNOR	1.74	0.29	0.76	1.85	1.99	33
MEAN3RDNOR	0.96	0.22	0.49	0.91	1.43	33
MEDIAN3RDNOR	0.96	0.29	0.53	0.9	1.49	33
MIN3RDALT	0.19	0.27	0.01	0.08	1.2	33
MAX3RDALT	27.86	41.96	0.61	12.25	193.89	33
MEAN3RDALT	3.09	3.27	0.36	2.4	17.97	33
MEDIAN3RDALT	1.17	0.83	0.37	0.82	3	33
MIN3DIFFP	1.53	2.45	0.05	0.58	13.25	33
MAX3DIFFP	21.91	14.6	2.6	17.43	62.88	33
MEAN3DIFFP	8.32	4.73	1.43	7.63	20.24	33
MEDIAN3DIFFP	7.1	4.23	1.3	6.12	20.6	33
MIN3SPD	0.05	0.06	0	0.03	0.28	33
MAX3SPD	0.31	0.2	0.04	0.26	0.84	33
MEAN3SPD	0.19	0.11	0.03	0.16	0.47	33
MEDIAN3SPD	0.19	0.11	0.03	0.17	0.49	33
MIN3KS	11.84	10.28	3.14	8.04	46.56	33
MAX3KS	113.2	230.7	8.41	58.13	1360.3	33
MEAN3KS	22.18	16.84	5.85	15.88	74.31	33
MEDIAN3KS	16.75	11.89	5.34	12.79	60.05	33
MIN4CV	4.4	4.32	0.5	2.93	20.66	33
MAX4CV	12.55	7.62	1.91	11.2	32.31	33
MEAN4CV	9.09	5	1.7	8.27	20.66	33
MEDIAN4CV	9.01	4.89	1.78	8.14	20.66	33
MIN4SKEW	-1.16	0.91	-2	-1.46	1.32	33
MAX4SKEW	1.14	0.99	-1.46	1.57	2	33
MEAN4SKEW	0.11	0.7	-1.46	0.22	1.32	33
MEDIAN4SKEW	0.19	0.73	-1.46	0.22	1.34	33
MIN4D	20.5	38.09	0.16	3.62	149.28	33
MAX4D	76.56	77.3	1.26	62.49	435.92	33
MEAN4D	37.81	35.59	1.26	25.7	149.28	33
MEDIAN4D	34.53	35.39	1.26	23.39	149.28	33
MIN4RD	0.49	0.92	0.01	0.17	4.03	33
MAX4RD	10.2	22.25	0.17	4.15	128.38	33
MEAN4RD	1.55	1.44	0.17	1.03	7.36	33
MEDIAN4RD	1.03	0.94	0.1	0.67	4.03	33
MIN4RDNOR	0.43	0.52	0.02	0.23	2.01	33
MAX4RDNOR	1.95	0.68	0.23	2.07	2.95	33
MEAN4RDNOR	0.97	0.41	0.23	0.87	2.01	33
MEDIAN4RDNOR	0.91	0.49	0.15	0.77	2.01	33
MIN4RDALT	0.51	0.92	0.01	0.17	4.03	33
MAX4RDALT	10.47	22.63	0.17	4.48	130.64	33
MEAN4RDALT	1.62	1.48	0.17	1.06	7.68	33
MEDIAN4RDALT	1.07	0.96	0.11	0.68	4.03	33
MIN4DIFFP	2.43	4.4	0.16	0.84	20.6	33
MAX4DIFFP	16.11	11.32	0.54	15.77	45.9	33
MEAN4DIFFP	6.78	4.45	0.54	6.06	20.6	33
MEDIAN4DIFFP	5.97	4.5	0.54	5.07	20.6	33
MIN4KURTO	-3.77	2.63	-6	-4.78	2.2	33
MAX4KURTO	2.65	1.73	-2.83	3.29	4	33

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEAN4KURTO	-0.08	1.17	-2.83	-0.08	2.2	33
MEDIAN4KURTO	0.17	1.52	-3.68	0.69	2.2	33
MIN4SPD	0.11	0.13	0.01	0.07	0.63	33
MAX4SPD	0.31	0.2	0.04	0.26	0.84	33
MEAN4SPD	0.24	0.15	0.04	0.2	0.63	33
MEDIAN4SPD	0.23	0.14	0.04	0.2	0.63	33
MIN4KS	13.88	12.22	3.79	9.27	52.64	33
MAX4KS	43.49	37.89	5.4	34.46	201.12	33
MEAN4KS	17.88	12.29	5.4	13.17	60.1	33
MEDIAN4KS	16.68	12.04	5.4	12.71	56.12	33

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.19: Descriptive statistics for collusive tenders of the Ticino cartel with five competitive bids

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	10.79	2.49	6	11	18	184
MEANBIDS	1417.6	1328.38	19	897.16	6080.35	184
STDBIDS	112.94	140.47	1.83	60.74	859.5	184
CV	7.38	3.23	2.77	6.89	23.83	184
KURTO	1.78	2.37	-1.91	1.24	9.87	184
SKEW	0.11	1.24	-2.6	-0.08	3.07	184
SPD	0.29	0.14	0.08	0.26	0.89	184
D	633.31	93.76	0.11	25.49	585.65	184
RD	0.93	1.04	0	0.64	5.77	184
RDNOR	1.84	1.55	0	1.52	8.19	184
RDALT	2.51	2.94	0	1.62	18.31	184
DIFFP	5.27	4.81	0.01	4.23	27.56	184
KS	16.68	5.95	5.44	15.44	37.34	184
MIN3CV	0.49	0.52	0	0.29	2.9	184
MAX3CV	13.83	6.03	4.57	12.54	37.62	184
MEAN3CV	6.2	2.41	2.4	5.9	19.05	184
MEDIAN3CV	5.6	2.23	1.48	5.08	13.65	184
MIN3SKEW	-1.72	0.03	-1.73	-1.73	-1.51	184
MAX3SKEW	1.71	0.06	1.22	1.73	1.73	184
MEAN3SKEW	-0.06	0.4	-1.19	-0.12	1.09	184
MEDIAN3SKEW	-0.1	0.77	-1.56	-0.25	1.73	184
MIN3D	4.61	10.12	0	0.66	70.32	184
MAX3D	239.79	245.6	4.9	146.83	1327.34	184

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEAN3D	82.36	91.89	1.86	45.74	391.88	184
MEDIAN3D	71.06	82.74	1.55	41.1	394.87	184
MIN3RD	0.05	0.08	0	0.01	0.52	184
MAX3RD	5367.51	51859.16	6.4	129.19	693080.61	184
MEAN3RD	43.17	245.44	0.95	7.71	2710.21	184
MEDIAN3RD	1.82	1.01	0.25	1.69	7.49	184
MIN3RDNOR	0.06	0.09	0	0.02	0.54	184
MAX3RDNOR	1.96	0.06	1.64	1.98	2	184
MEAN3RDNOR	1.04	0.17	0.6	1.06	1.56	184
MEDIAN3RDNOR	1.05	0.26	0.3	1.08	1.68	184
MIN3RDALT	0.03	0.06	0	0.01	0.37	184
MAX3RDALT	3795.41	36669.96	4.53	91.35	490082	184
MEAN3RDALT	30.53	173.55	0.67	5.45	1916.41	184
MEDIAN3RDALT	1.29	0.72	0.18	1.19	5.29	184
MIN3DIFFP	0.23	0.34	0	0.11	2.01	184
MAX3DIFFP	20.05	8.11	7.05	18.21	53.73	184
MEAN3DIFFP	6.33	2.43	2.48	5.8	13.45	184
MEDIAN3DIFFP	5.21	2.33	0.87	5	14.71	184
MIN3SPD	0.01	0.01	0	0.01	0.06	184
MAX3SPD	0.29	0.14	0.08	0.26	0.89	184
MEAN3SPD	0.13	0.05	0.05	0.12	0.45	184
MEDIAN3SPD	0.11	0.05	0.03	0.1	0.3	184
MIN3KS	8.83	3.51	3.06	8.1	22.3	184
MAX3KS	938.54	3130.28	34.68	339.85	39320	184
MEAN3KS	39.88	25.59	8.37	32.91	192.41	184
MEDIAN3KS	20.92	8.4	7.62	19.97	67.92	184
MIN4CV	0.94	0.95	0.04	0.73	7.67	184
MAX4CV	12.03	5.14	4.23	11.03	33.57	184
MEAN4CV	6.57	2.64	2.53	6.26	20.74	184
MEDIAN4CV	6.27	2.82	1.9	5.74	26	184
MIN4SKEW	-1.92	0.19	-2	-1.99	-0.67	184
MAX4SKEW	1.87	0.29	0.22	1.99	2	184
MEAN4SKEW	-0.08	0.53	-1.53	-0.17	1.2	184
MEDIAN4SKEW	-0.06	0.7	-1.54	-0.14	1.74	184
MIN4D	5.32	11.79	0	0.68	70.32	184
MAX4D	196.4	202.67	3.82	112.66	903.6	184
MEAN4D	75.31	87.29	1.35	41.48	449.88	184
MEDIAN4D	70.54	87.18	0.82	37.46	449.27	184
MIN4RD	0.06	0.11	0	0.02	0.74	184
MAX4RD	117.38	392.24	1.49	33.85	4710.41	184
MEAN4RD	3.59	4.07	0.37	2.58	44.88	184
MEDIAN4RD	1.31	0.78	0.12	1.21	4.63	184
MIN4RDNOR	0.09	0.14	0	0.03	0.81	184
MAX4RDNOR	2.74	0.29	1.32	2.84	3	184
MEAN4RDNOR	1.18	0.31	0.4	1.23	2.21	184
MEDIAN4RDNOR	1.13	0.41	0.17	1.18	2.12	184
MIN4RDALT	0.07	0.12	0	0.02	0.74	184
MAX4RDALT	124.1	415.85	1.56	34.79	4965.75	184
MEAN4RDALT	3.73	4.24	0.39	2.66	47.93	184
MEDIAN4RDALT	1.37	0.81	0.12	1.28	4.82	184
MIN4DIFFP	0.27	0.41	0	0.11	2.5	184
MAX4DIFFP	16.67	6.63	6	15.6	40.38	184
MEAN4DIFFP	5.99	2.63	1.71	5.49	14.87	184
MEDIAN4DIFFP	5.32	2.7	0.58	4.86	19.21	184

See next page

Predictors	Mean	Std	Min	Median	Max	N
MIN4KURTO	-5.57	0.76	-6	-5.84	-0.04	184
MAX4KURTO	3.9	0.2	2.88	3.98	4	184
MEAN4KURTO	0.64	0.77	-1.74	0.72	2.39	184
MEDIAN4KURTO	1.32	0.98	-2.78	1.51	3.24	184
MIN4SPD	0.02	0.02	0	0.02	0.18	184
MAX4SPD	0.29	0.14	0.08	0.26	0.89	184
MEAN4SPD	0.16	0.07	0.06	0.15	0.57	184
MEDIAN4SPD	0.15	0.08	0.05	0.13	0.75	184
MIN4KS	10.09	3.78	3.67	9.25	23.85	184
MAX4KS	288.31	349.99	13.07	137.96	2643.96	184
MEAN4KS	26.2	11.64	6.8	23.37	72.53	184
MEDIAN4KS	19.03	7.35	4.49	17.82	53.22	184

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.20: Descriptive statistics for the collusive tenders including only cartel participants in the Swiss data

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	6.57	2.44	4	6	13	308
MEANBIDS	379.88	376.09	34.42	305.81	3509.71	308
STDBIDS	13.15	13.67	0.49	9.75	109.94	308
CV	3.66	2.09	0.6	3.29	15.73	308
KURTO	0.16	1.65	-5.4	0.21	4.37	308
SKEW	0.08	0.81	-1.94	0.07	1.78	308
SPD	0.11	0.07	0.01	0.09	0.5	308
D	9.2	13.32	0.14	6.02	121.3	308
RD	1.16	1.36	0.01	0.75	13.66	308
RDNOR	1.38	0.79	0.02	1.24	5.03	308
RDALT	1.83	1.73	0.01	1.33	13.89	308
DIFFP	2.76	2.91	0.06	1.94	34.11	308
KS	36.54	20.15	6.59	31.2	167.93	308
MIN3CV	1.08	1.01	0.04	0.81	8.46	308
MAX3CV	5.49	3.41	0.73	4.72	21.17	308
MEAN3CV	3.41	1.94	0.57	3.04	15.32	308
MEDIAN3CV	3.46	2	0.63	3	16.87	308
MIN3SKEW	-1.48	0.44	-1.73	-1.7	0.69	308
MAX3SKEW	1.41	0.63	-1.49	1.68	1.73	308
MEAN3SKEW	0	0.49	-1.61	0.05	1.16	308
MEDIAN3SKEW	0.03	0.69	-1.7	0.07	1.53	308

See next page

Predictors	Mean	Std	Min	Median	Max	N
MIN3D	2.67	6.13	0	1.03	73.33	308
MAX3D	26.81	27.14	0.59	20.08	200.26	308
MEAN3D	11.62	13.4	0.43	8.44	118.06	308
MEDIAN3D	10.88	13.28	0.42	7.87	121.3	308
MIN3RD	0.38	0.73	0	0.14	6.13	308
MAX3RD	117.7	665.15	0.87	16.67	9157.71	308
MEAN3RD	7.53	23.86	0.55	3.04	291.4	308
MEDIAN3RD	1.95	2.79	0.29	1.37	39.01	308
MIN3RDNOR	0.3	0.33	0	0.17	1.63	308
MAX3RDNOR	1.73	0.28	0.76	1.85	2	308
MEAN3RDNOR	1	0.22	0.51	0.98	1.8	308
MEDIAN3RDNOR	0.99	0.26	0.33	0.98	1.85	308
MIN3RDALT	0.27	0.52	0	0.1	4.34	308
MAX3RDALT	83.23	470.33	0.62	11.79	6475.48	308
MEAN3RDALT	5.32	16.87	0.39	2.15	206.05	308
MEDIAN3RDALT	1.38	1.97	0.2	0.97	27.58	308
MIN3DIFFP	0.69	0.85	0	0.44	8.62	308
MAX3DIFFP	8.1	6.07	1.13	6.76	47.45	308
MEAN3DIFFP	3.37	2.32	0.77	2.77	22.76	308
MEDIAN3DIFFP	3.13	2.31	0.69	2.51	23.91	308
MIN3SPD	0.02	0.02	0	0.02	0.18	308
MAX3SPD	0.11	0.07	0.01	0.09	0.5	308
MEAN3SPD	0.07	0.04	0.01	0.06	0.37	308
MEDIAN3SPD	0.07	0.04	0.01	0.06	0.41	308
MIN3KS	26.02	17.06	4.76	21.48	137.18	308
MAX3KS	181.77	213.13	12.09	124.09	2751.23	308
MEAN3KS	48.26	25.64	7.44	42.03	199.62	308
MEDIAN3KS	38.19	20.28	6.19	33.58	158.95	308
MIN4CV	1.93	1.66	0.22	1.5	15.73	308
MAX4CV	4.74	2.95	0.6	4.07	19.16	308
MEAN4CV	3.54	2.01	0.6	3.13	15.73	308
MEDIAN4CV	3.6	2.07	0.6	3.21	15.73	308
MIN4SKEW	-1.07	0.98	-2	-1.5	1.78	308
MAX4SKEW	1.07	1.02	-1.93	1.48	2	308
MEAN4SKEW	0.04	0.68	-1.93	0.07	1.78	308
MEDIAN4SKEW	0.05	0.71	-1.93	0.03	1.78	308
MIN4D	4.47	11.64	0	1.36	121.3	308
MAX4D	20.38	20.98	0.36	14.66	121.3	308
MEAN4D	10.11	12.76	0.36	6.99	121.3	308
MEDIAN4D	9.73	13.27	0.17	6.4	121.3	308
MIN4RD	0.68	1.38	0	0.23	13.66	308
MAX4RD	9.14	19.83	0.15	4.31	298.7	308
MEAN4RD	1.66	1.53	0.15	1.26	13.66	308
MEDIAN4RD	1.26	1.37	0.07	0.87	13.66	308
MIN4RDNOR	0.53	0.58	0	0.31	2.62	308
MAX4RDNOR	1.9	0.72	0.22	2.09	2.98	308
MEAN4RDNOR	1.07	0.4	0.22	1.02	2.62	308
MEDIAN4RDNOR	1.02	0.44	0.1	0.93	2.62	308
MIN4RDALT	0.71	1.44	0	0.23	13.89	308
MAX4RDALT	9.67	22.08	0.15	4.57	340.47	308
MEAN4RDALT	1.74	1.62	0.15	1.33	13.89	308
MEDIAN4RDALT	1.31	1.42	0.07	0.9	13.89	308
MIN4DIFFP	1.19	1.93	0	0.67	23.91	308
MAX4DIFFP	6.2	5.28	0.64	4.98	39.3	308

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEAN4DIFFP	2.95	2.42	0.61	2.29	24.48	308
MEDIAN4DIFFP	2.82	2.74	0.2	2.08	34.11	308
MIN4KURTO	-2.96	2.98	-6	-4.23	3.75	308
MAX4KURTO	2.44	1.83	-5.4	3.14	4	308
MEAN4KURTO	0.14	1.36	-5.4	0.09	3.75	308
MEDIAN4KURTO	0.34	1.54	-5.4	0.61	3.75	308
MIN4SPD	0.05	0.04	0.01	0.03	0.47	308
MAX4SPD	0.11	0.07	0.01	0.09	0.5	308
MEAN4SPD	0.08	0.05	0.01	0.07	0.47	308
MEDIAN4SPD	0.09	0.05	0.01	0.08	0.47	308
MIN4KS	30.68	20.89	5.42	24.8	167.93	308
MAX4KS	88.22	69.07	6.59	67.21	458.02	308
MEAN4KS	39.97	20.41	6.59	35.08	167.93	308
MEDIAN4KS	36.66	19.93	6.59	31.38	167.93	308

Notes: “Mean”, “Std”, “Min”, “Median”, “Max”, and “N” denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for ”MEANBIDS”, ”STDBIDS”, ”D”, ”MIN3D”, ”MAX3D”, ”MEAN3D”, ”MEDIAN3D”, ”MIN4D”, ”MAX4D”, ”MEAN4D” and ”MEDIAN4D” are expressed in thousand CHF. ”KS”, ”CV”, ”SPD”, ”RD”, ”RDNOR”, ”RDALT”, ”SKEW”, ”DIFFP”, ”KURTO”, ”D”, ”STDBIDS”, ”MEANBIDS” and ”NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.21: Descriptive statistics for the competitive tenders in the Swiss data

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	5.73	1.86	4	5	13	1082
MEANBIDS	828.06	1803.84	13.63	423.97	37786.87	1082
STDBIDS	87.35	216.52	0.41	32.74	3996.24	1082
CV	10.12	7.89	0.76	8.45	128	1082
KURTO	0.25	2.27	-6	0.13	8.03	1082
SKEW	0.26	0.97	-2.68	0.28	2.47	1082
SPD	2.5	29.79	0.02	0.24	730.71	1082
D	54.94	223.18	0	14.19	4656.85	1082
RD	1.16	2.45	0	0.57	41.26	1082
RDNOR	1.04	0.82	0	0.87	6.95	1082
RDALT	1.61	3.19	0	0.84	47.49	1082
DIFFP	176.79	2246.36	0	4.36	50228.95	1082
KS	15.07	10.98	1.48	12.24	132.33	1082
MIN3CV	3.37	3.13	0.02	2.33	24.05	1082
MAX3CV	14.22	11.46	0.93	11.77	122.06	1082
MEAN3CV	9.37	6.9	0.73	7.92	91.8	1082
MEDIAN3CV	9.7	7.83	0.6	8.19	121.69	1082
MIN3SKEW	-1.38	0.62	-1.73	-1.67	1.68	1082
MAX3SKEW	1.47	0.53	-1.61	1.69	1.73	1082
MEAN3SKEW	0.13	0.58	-1.66	0.16	1.71	1082

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEDIAN3SKEW	0.22	0.83	-1.73	0.26	1.73	1082
MIN3D	13.03	30.15	0	3.79	364.58	1082
MAX3D	154.93	433.88	0.72	52.94	7506.55	1082
MEAN3D	69.05	178.7	0.44	25.66	2536.05	1082
MEDIAN3D	64.4	171.21	0.36	23.09	2640.97	1082
MIN3RD	0.32	0.68	0	0.12	8.83	1082
MAX3RD	69.28	265.38	0.13	12.27	5315.77	1082
MEAN3RD	8.2	25.53	0.08	2.78	421.57	1082
MEDIAN3RD	2.22	4.95	0.01	1.21	72.77	1082
MIN3RDNOR	0.27	0.3	0	0.16	1.72	1082
MAX3RDNOR	1.68	0.34	0.17	1.8	2	1082
MEAN3RDNOR	0.94	0.27	0.11	0.93	1.85	1082
MEDIAN3RDNOR	0.92	0.34	0.01	0.91	1.94	1082
MIN3RDALT	0.22	0.48	0	0.09	6.24	1082
MAX3RDALT	48.99	187.65	0.1	8.67	3758.82	1082
MEAN3RDALT	5.8	18.05	0.06	1.96	298.09	1082
MEDIAN3RDALT	1.57	3.5	0	0.85	51.46	1082
MIN3DIFFP	2.07	2.77	0	0.98	22.67	1082
MAX3DIFFP	226.26	2740.18	0.89	15.42	63802.49	1082
MEAN3DIFFP	92.86	1118.61	0.64	7.07	23734.93	1082
MEDIAN3DIFFP	60.05	1011.71	0.54	6.52	25112.51	1082
MIN3SPD	0.07	0.07	0	0.05	0.58	1082
MAX3SPD	2.5	29.79	0.02	0.24	730.71	1082
MEAN3SPD	1.14	12.52	0.01	0.16	279.23	1082
MEDIAN3SPD	0.76	10.69	0.01	0.17	264.32	1082
MIN3KS	10.96	8.72	1.1	8.72	107.96	1082
MAX3KS	82.65	179.47	4.21	43	4045.44	1082
MEAN3KS	21.45	21.35	2.62	16.51	463.06	1082
MEDIAN3KS	15.78	13.43	1.2	12.57	167.49	1082
MIN4CV	6.24	6.77	0.07	4.49	128	1082
MAX4CV	12.14	9.73	0.76	10.13	128	1082
MEAN4CVB	9.82	7.64	0.76	8.29	128	1082
MEDIAN4CV	10.24	8.21	0.76	8.55	128	1082
MIN4SKEW	-0.75	1.09	-2	-1.02	2	1082
MAX4SKEW	1.02	1	-1.99	1.4	2	1082
MEAN4SKEW	0.19	0.8	-1.99	0.22	2	1082
MEDIAN4SKEW	0.22	0.86	-1.99	0.19	2	1082
MIN4D	22.95	76.73	0	5.64	1226.06	1082
MAX4D	111.58	328.01	0.02	34.32	5764.43	1082
MEAN4D	59.42	181.05	0.02	19.23	3035.79	1082
MEDIAN4D	59.94	224.26	0	18.35	4656.85	1082
MIN4RD	0.71	2.05	0	0.2	41.26	1082
MAX4RD	6.73	17.51	0	2.55	266.72	1082
MEAN4RD	1.69	3.23	0	1	67.61	1082
MEDIAN4RD	1.24	2.28	0	0.74	41.26	1082
MIN4RDNOR	0.51	0.56	0	0.29	2.87	1082
MAX4RDNOR	1.62	0.81	0	1.74	2.98	1082
MEAN4RDNOR	0.95	0.51	0	0.9	2.87	1082
MEDIAN4RDNOR	0.9	0.55	0	0.84	2.87	1082
MIN4RDALT	0.75	2.18	0	0.21	43.5	1082
MAX4RDALT	7.07	18.5	0	2.69	267.55	1082
MEAN4RDALT	1.77	3.36	0	1.05	68.45	1082
MEDIAN4RDALT	1.3	2.4	0	0.78	43.5	1082
MIN4DIFFP	20.35	549.37	0	1.54	18073.22	1082

See next page

Predictors	Mean	Std	Min	Median	Max	N
MAX4DIFFP	206.69	2603.79	0.03	9.88	62175.31	1082
MEAN4DIFFP	116.56	1450.99	0.03	5.46	30433.18	1082
MEDIAN4DIFFP	140.96	2011.38	0	5.24	50228.95	1082
MIN4KURTO	-2.61	3.02	-6	-3.37	4	1082
MAX4KURTO	2.07	2.28	-6	2.94	4	1082
MEAN4KURTO	0.12	1.83	-6	0.1	4	1082
MEDIAN4KURTO	0.29	2.06	-6	0.6	4	1082
MIN4SPD	0.34	6.03	0	0.1	198.43	1082
MAX4SPD	2.5	29.79	0.02	0.24	730.71	1082
MEAN4SPD	1.54	17.19	0.02	0.2	388.02	1082
MEDIAN4SPD	1.78	23.58	0.02	0.21	621.75	1082
MIN4KS	12.89	10.47	1.38	10.24	132.33	1082
MAX4KS	36.64	66.05	1.48	22.53	1433.11	1082
MEAN4KS	16.65	11.9	1.48	13.55	132.33	1082
MEDIAN4KS	14.87	11.3	1.48	12.14	132.33	1082

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.22: Descriptive statistics for incomplete bid-rigging cartels in sample 1 (Swiss data)

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	7.49	2.54	4	7	13	252
MEANBIDS	435.9	476.67	18.01	296.62	3460.91	252
STDBIDS	35.12	50.17	1.54	19.62	362.86	252
CV	7.79	3.89	1.77	6.79	23.92	252
KURTO	0.41	2.12	-5.9	0.04	6.97	252
SKEW	-0.07	0.99	-2.59	-0.06	2.57	252
SPD	0.26	0.16	0.05	0.21	0.89	252
D	21.07	39.03	0.08	9.1	351.83	252
RD	1.39	2.67	0.01	0.58	28.37	252
RDNOR	1.41	1.11	0.01	1.14	5.48	252
RDALT	2.32	3.52	0.01	1.17	28.54	252
DIFFP	6.34	8.85	0.03	3.69	73.53	252
KS	16.62	8.04	4.15	15.24	57.54	252
MIN3CV	1.68	1.73	0	1.24	14.77	252
MAX3CV	12.26	6.39	2.9	10.66	38.48	252
MEAN3CV	7.12	3.48	1.55	6.32	22.96	252
MEDIAN3CV	7.06	3.81	1.01	6.28	29.82	252
MIN3SKEW	-1.6	0.36	-1.73	-1.72	0.97	252

See next page

Predictors	Mean	Std	Min	Median	Max	N
MAX3SKEW	1.53	0.48	-1.1	1.72	1.73	252
MEAN3SKEW	-0.04	0.5	-1.43	-0.03	1.47	252
MEDIAN3SKEW	-0.05	0.77	-1.71	-0.05	1.73	252
MIN3D	3.64	8.58	0	1.31	80.15	252
MAX3D	76.68	107.56	1.58	41.15	842.7	252
MEAN3D	29.56	44.27	0.62	16.57	437.36	252
MEDIAN3D	26.54	40.57	0.41	14.6	424.3	252
MIN3RD	0.23	0.45	0	0.07	3.35	252
MAX3RD	1203.97	9993.9	0.68	31.05	122393.82	252
MEAN3RD	27.54	191.86	0.31	4.5	2742.11	252
MEDIAN3RD	2.09	2.42	0.21	1.49	27.08	252
MIN3RDNOR	0.21	0.28	0	0.09	1.41	252
MAX3RDNOR	1.83	0.23	0.65	1.92	2	252
MEAN3RDNOR	1.02	0.22	0.33	1.02	1.73	252
MEDIAN3RDNOR	1.02	0.3	0.26	1.02	1.89	252
MIN3RDALT	0.16	0.32	0	0.05	2.37	252
MAX3RDALT	851.33	7066.76	0.48	21.95	86545.5	252
MEAN3RDALT	19.47	135.66	0.22	3.18	1938.96	252
MEDIAN3RDALT	1.48	1.71	0.15	1.06	19.15	252
MIN3DIFFP	0.97	1.58	0	0.47	12.6	252
MAX3DIFFP	19.61	12.81	2.29	16.77	87.65	252
MEAN3DIFFP	7.76	5.62	1.39	6.4	49.6	252
MEDIAN3DIFFP	7.11	6.59	1.14	5.72	73.53	252
MIN3SPD	0.03	0.04	0	0.02	0.32	252
MAX3SPD	0.26	0.16	0.05	0.21	0.89	252
MEAN3SPD	0.15	0.08	0.03	0.13	0.62	252
MEDIAN3SPD	0.15	0.09	0.02	0.13	0.83	252
MIN3KS	10.96	6.18	3	9.62	34.52	252
MAX3KS	357.52	3429.85	6.79	80.58	54476.99	252
MEAN3KS	27.25	34.52	6.03	21.55	495.16	252
MEDIAN3KS	18.68	11.16	3.36	16.1	99.02	252
MIN4CV	3.16	2.97	0.12	2.26	23.92	252
MAX4CV	10.63	5.46	2.38	9.43	33.94	252
MEAN4CV	7.43	3.65	1.64	6.54	23.92	252
MEDIAN4CV	7.65	3.91	1.69	6.76	25.38	252
MIN4SKEW	-1.42	0.79	-2	-1.78	1.87	252
MAX4SKEW	1.28	0.97	-1.98	1.79	2	252
MEAN4SKEW	-0.07	0.69	-1.98	-0.05	1.87	252
MEDIAN4SKEW	-0.07	0.79	-1.98	-0.03	1.87	252
MIN4D	4.98	11.53	0	1.63	102.18	252
MAX4D	60.49	87.82	0.18	33.62	771.65	252
MEAN4D	25.08	38.97	0.18	14.32	410.17	252
MEDIAN4D	24.93	42.34	0.14	13.17	424.3	252
MIN4RD	0.59	2.26	0	0.09	28.37	252
MAX4RD	25.25	166.8	0.04	7.14	2627.72	252
MEAN4RD	2.25	2.96	0.04	1.52	28.37	252
MEDIAN4RD	1.53	2.43	0.04	1.01	28.37	252
MIN4RDNOR	0.37	0.54	0	0.15	2.8	252
MAX4RDNOR	2.18	0.68	0.07	2.37	3	252
MEAN4RDNOR	1.11	0.46	0.07	1.08	2.8	252
MEDIAN4RDNOR	1.06	0.54	0.07	1.05	2.8	252
MIN4RDALT	0.61	2.29	0	0.1	28.54	252
MAX4RDALT	27.16	186.06	0.04	7.47	2933.56	252
MEAN4RDALT	2.35	3.08	0.04	1.62	28.54	252

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEDIAN4RDALT	1.59	2.46	0.04	1.08	28.54	252
MIN4DIFFP	1.83	4.55	0	0.58	57.15	252
MAX4DIFFP	16.09	11.96	0.3	13.53	77.43	252
MEAN4DIFFP	7.05	6.61	0.3	5.57	60.05	252
MEDIAN4DIFFP	6.8	7.49	0.24	5.09	73.53	252
MIN4KURTO	-4	2.64	-6	-5.28	3.94	252
MAX4KURTO	2.99	1.68	-5.9	3.68	4	252
MEAN4KURTO	0.16	1.38	-5.9	0.04	3.94	252
MEDIAN4KURTO	0.51	1.67	-5.9	0.64	3.94	252
MIN4SPRD	0.08	0.08	0	0.05	0.81	252
MAX4SPD	0.26	0.16	0.05	0.21	0.89	252
MEAN4SPD	0.19	0.11	0.04	0.16	0.81	252
MEDIAN4SPD	0.19	0.12	0.04	0.16	0.89	252
MIN4KS	12.65	7.28	3.64	10.91	42.15	252
MAX4KS	64.97	79.07	4.15	44.41	862.88	252
MEAN4KS	19.96	11.22	4.15	17.51	97.43	252
MEDIAN4KS	16.85	8.23	3.87	15.26	59.16	252

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.23: Descriptive statistics for incomplete bid-rigging cartels in sample 2 (Swiss data)

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	7.77	2.47	4	8	13	223
MEANBIDS	405.24	372.35	27.84	302.21	3002.37	223
STDBIDS	30.96	37.64	1.63	19.51	270.82	223
CV	7.6	3.77	1.77	6.66	23.92	223
KURTO	0.44	2.08	-5.75	0.04	6.97	223
SKEW	-0.07	0.97	-2.59	-0.08	2.57	223
SPD	0.26	0.16	0.05	0.21	0.89	223
D	20.13	36.04	0.08	8.9	351.83	223
RD	1.29	2.12	0.01	0.58	18.04	223
RDNOR	1.45	1.14	0.01	1.14	5.48	223
RDALT	2.3	3.24	0.01	1.17	20.74	223
DIFFP	6.3	9.13	0.03	3.69	73.53	223
KS	16.85	7.91	4.15	15.54	57.54	223
MIN3CV	1.47	1.22	0	1.15	7.05	223
MAX3CV	12.24	6.43	2.93	10.57	38.48	223
MEAN3CV	6.92	3.33	1.55	6.09	22.96	223

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEDIAN3CV	6.75	3.56	1.01	6	29.82	223
MIN3SKEW	-1.63	0.29	-1.73	-1.72	0.62	223
MAX3SKEW	1.58	0.38	-1.1	1.72	1.73	223
MEAN3SKEW	-0.04	0.45	-1.33	-0.04	1.31	223
MEDIAN3SKEW	-0.05	0.74	-1.69	-0.06	1.73	223
MIN3D	2.76	5.42	0	1.17	56.23	223
MAX3D	69.79	84.97	2.38	42.35	670.57	223
MEAN3D	26.13	30.47	1.42	16.72	235.67	223
MEDIAN3D	23.11	25.43	1.52	14.63	171.87	223
MIN3RD	0.18	0.37	0	0.06	3.35	223
MAX3RD	1355.45	10617.14	0.92	37.36	122393.82	223
MEAN3RD	30.39	203.82	0.43	4.66	2742.11	223
MEDIAN3RD	1.92	1.72	0.26	1.5	18.63	223
MIN3RDNOR	0.18	0.25	0	0.08	1.41	223
MAX3RDNOR	1.85	0.2	0.79	1.93	2	223
MEAN3RDNOR	1.02	0.2	0.41	1.02	1.66	223
MEDIAN3RDNOR	1.02	0.28	0.3	1.02	1.85	223
MIN3RDALT	0.13	0.26	0	0.04	2.37	223
MAX3RDALT	958.45	7507.45	0.65	26.42	86545.5	223
MEAN3RDALT	21.49	144.12	0.3	3.29	1938.96	223
MEDIAN3RDALT	1.36	1.21	0.18	1.06	13.18	223
MIN3DIFFP	0.81	1.25	0	0.43	10.47	223
MAX3DIFFP	19.73	12.93	2.29	16.81	87.65	223
MEAN3DIFFP	7.59	5.58	1.39	6.35	49.6	223
MEDIAN3DIFFP	6.92	6.67	1.14	5.71	73.53	223
MIN3SPD	0.03	0.02	0	0.02	0.15	223
MAX3SPD	0.26	0.16	0.05	0.21	0.89	223
MEAN3SPD	0.15	0.08	0.03	0.13	0.62	223
MEDIAN3SPD	0.14	0.09	0.02	0.12	0.83	223
MIN3KS	10.9	6.01	3	9.7	34.52	223
MAX3KS	394.81	3645.21	14.27	87.43	54476.99	223
MEAN3KS	27.97	36.27	6.25	21.99	495.16	223
MEDIAN3KS	19.22	11.35	3.36	16.97	99.02	223
MIN4CV	2.8	2.48	0.12	2.17	23.92	223
MAX4CV	10.63	5.48	2.48	9.36	33.94	223
MEAN4CV	7.22	3.5	1.64	6.32	23.92	223
MEDIAN4CV	7.41	3.75	1.69	6.51	25.38	223
MIN4SKEW	-1.51	0.71	-2	-1.82	1.84	223
MAX4SKEW	1.38	0.85	-1.96	1.82	2	223
MEAN4SKEW	-0.06	0.64	-1.96	-0.07	1.84	223
MEDIAN4SKEW	-0.07	0.74	-1.96	-0.04	1.84	223
MIN4D	3.82	8.45	0	1.43	96.23	223
MAX4D	58.39	72.08	0.18	37.52	535.26	223
MEAN4D	23.35	29.62	0.18	14.45	274.73	223
MEDIAN4D	23.01	32.93	0.14	13.49	351.83	223
MIN4RD	0.41	1.4	0	0.09	18.04	223
MAX4RD	27.52	177.13	0.04	7.86	2627.72	223
MEAN4RD	2.19	2.53	0.04	1.62	24.7	223
MEDIAN4RD	1.42	1.77	0.04	1.02	18.04	223
MIN4RDNOR	0.32	0.46	0	0.14	2.7	223
MAX4RDNOR	2.25	0.61	0.07	2.42	3	223
MEAN4RDNOR	1.11	0.42	0.07	1.09	2.7	223
MEDIAN4RDNOR	1.06	0.51	0.07	1.06	2.7	223
MIN4RDALT	0.42	1.42	0	0.09	18.05	223

See next page

Predictors	Mean	Std	Min	Median	Max	N
MAX4RDALT	29.63	197.6	0.04	8.35	2933.56	223
MEAN4RDALT	2.29	2.67	0.04	1.72	27.4	223
MEDIAN4RDALT	1.48	1.8	0.04	1.1	18.05	223
MIN4DIFFP	1.48	4.34	0	0.51	57.15	223
MAX4DIFFP	16.59	11.97	0.3	13.87	77.43	223
MEAN4DIFFP	7	6.68	0.3	5.51	60.05	223
MEDIAN4DIFFP	6.74	7.66	0.24	5.08	73.53	223
MIN4KURTO	-4.32	2.38	-6	-5.48	3.85	223
MAX4KURTO	3.16	1.43	-5.75	3.74	4	223
MEAN4KURTO	0.15	1.23	-5.75	0.04	3.85	223
MEDIAN4KURTO	0.52	1.56	-5.75	0.63	3.85	223
MIN4SPD	0.07	0.07	0	0.05	0.81	223
MAX4SPD	0.26	0.16	0.05	0.21	0.89	223
MEAN4SPD	0.18	0.11	0.04	0.15	0.81	223
MEDIAN4SPD	0.19	0.11	0.04	0.16	0.89	223
MIN4KS	12.55	7.02	3.64	11.06	40.78	223
MAX4KS	69.75	82.58	4.15	46.65	862.88	223
MEAN4KS	20.49	11.41	4.15	18.45	97.43	223
MEDIAN4KS	17.14	8.1	3.87	15.89	59.16	223

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.24: Descriptive statistics for incomplete bid-rigging cartels in sample 3 (Swiss data)

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	8.5	2.15	5	8	13	173
MEANBIDS	439.36	388.89	30.24	329.61	3002.37	173
STDBIDS	33.5	40.38	2.01	20.48	270.82	173
CV	7.54	3.26	1.77	6.76	21.7	173
KURTO	0.46	1.97	-2.69	-0.09	6.97	173
SKEW	-0.07	0.95	-2.59	-0.08	2.57	173
SPD	0.26	0.14	0.06	0.22	0.88	173
D	21.95	39.01	0.08	9.46	351.83	173
RD	1.14	1.76	0.01	0.58	12.47	173
RDNOR	1.54	1.17	0.02	1.17	5.48	173
RDALT	2.33	3.2	0.02	1.19	20.74	173
DIFFP	5.91	7.48	0.1	3.7	45.42	173
KS	16.55	7.64	6.19	15.48	57.54	173
MIN3CV	1.2	0.9	0	1.03	3.95	173

See next page

Predictors	Mean	Std	Min	Median	Max	N
MAX3CV	12.6	5.86	2.93	10.99	38.48	173
MEAN3CV	6.82	2.83	1.55	6.11	16.5	173
MEDIAN3CV	6.51	2.79	1.01	5.92	14.2	173
MIN3SKEW	-1.69	0.12	-1.73	-1.73	-0.88	173
MAX3SKEW	1.67	0.15	0.73	1.73	1.73	173
MEAN3SKEW	-0.03	0.39	-1.02	-0.04	0.88	173
MEDIAN3SKEW	-0.03	0.69	-1.48	-0.06	1.73	173
MIN3D	2.35	4.05	0	1.03	30.19	173
MAX3D	78.6	90.43	3.7	51.54	670.57	173
MEAN3D	28.43	32.44	1.58	17.96	235.67	173
MEDIAN3D	24.31	26.46	1.52	15.79	171.87	173
MIN3RD	0.11	0.19	0	0.04	1.76	173
MAX3RD	1737.96	12034.74	2.68	48.9	122393.82	173
MEAN3RD	37.63	231.03	0.92	5.15	2742.11	173
MEDIAN3RD	1.72	1.05	0.32	1.48	6.05	173
MIN3RDNOR	0.12	0.16	0	0.06	1	173
MAX3RDNOR	1.9	0.13	1.31	1.94	2	173
MEAN3RDNOR	1.02	0.18	0.61	1.03	1.46	173
MEDIAN3RDNOR	1.02	0.25	0.37	1.02	1.62	173
MIN3RDALT	0.08	0.14	0	0.03	1.25	173
MAX3RDALT	1228.92	8509.85	1.9	34.57	86545.5	173
MEAN3RDALT	26.61	163.36	0.65	3.64	1938.96	173
MEDIAN3RDALT	1.22	0.74	0.22	1.05	4.28	173
MIN3DIFFP	0.58	0.71	0	0.37	4.49	173
MAX3DIFFP	20.62	11.43	3.84	17.99	72.01	173
MEAN3DIFFP	7.33	3.89	1.55	6.42	22.8	173
MEDIAN3DIFFP	6.18	3.26	1.33	5.67	21.76	173
MIN3SPD	0.02	0.02	0	0.02	0.08	173
MAX3SPD	0.26	0.14	0.06	0.22	0.88	173
MEAN3SPD	0.14	0.07	0.03	0.13	0.38	173
MEDIAN3SPD	0.13	0.06	0.02	0.12	0.3	173
MIN3KS	10.03	5.06	3	9.18	34.52	173
MAX3KS	489.07	4136.34	25.56	97.45	54476.99	173
MEAN3KS	29.3	40.7	9.3	21.99	495.16	173
MEDIAN3KS	19.58	12.05	7.28	17.13	99.02	173
MIN4CV	2.14	1.45	0.12	1.76	8.25	173
MAX4CV	11.02	5.04	2.48	9.6	33.94	173
MEAN4CV	7.12	2.98	1.64	6.32	18.01	173
MEDIAN4CV	7.27	3.09	1.69	6.57	17.17	173
MIN4SKEW	-1.69	0.44	-2	-1.88	-0.04	173
MAX4SKEW	1.62	0.52	-0.4	1.84	2	173
MEAN4SKEW	-0.05	0.55	-1.42	-0.07	1.32	173
MEDIAN4SKEW	-0.06	0.67	-1.86	-0.05	1.72	173
MIN4D	2.86	4.84	0	1.13	30.28	173
MAX4D	66.68	76.42	1.56	43.87	535.26	173
MEAN4D	25.49	31.63	1.22	15.85	274.73	173
MEDIAN4D	24.96	35.2	0.96	15.12	351.83	173
MIN4RD	0.16	0.26	0	0.06	1.93	173
MAX4RD	33.76	200.71	1.09	10.34	2627.72	173
MEAN4RD	2.22	2.44	0.35	1.68	24.7	173
MEDIAN4RD	1.31	1.24	0.16	1.05	9.81	173
MIN4RDNOR	0.2	0.25	0	0.1	1.51	173
MAX4RDNOR	2.42	0.44	1.07	2.53	3	173
MEAN4RDNOR	1.11	0.35	0.43	1.1	2.06	173

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEDIAN4RDNOR	1.07	0.44	0.22	1.07	2.5	173
MIN4RDALT	0.16	0.26	0	0.07	2.03	173
MAX4RDALT	36.41	223.95	1.1	10.76	2933.56	173
MEAN4RDALT	2.31	2.61	0.37	1.75	27.4	173
MEDIAN4RDALT	1.37	1.27	0.16	1.11	9.91	173
MIN4DIFFP	0.71	0.9	0	0.41	6.52	173
MAX4DIFFP	17.69	10.69	3.03	15.55	70.09	173
MEAN4DIFFP	6.67	4.5	1.18	5.62	26.42	173
MEDIAN4DIFFP	6.33	5.07	0.98	5.23	34.35	173
MIN4KURTO	-4.9	1.6	-6	-5.55	1.44	173
MAX4KURTO	3.56	0.61	1	3.82	4	173
MEAN4KURTO	0.17	0.89	-1.78	-0.05	3.19	173
MEDIAN4KURTO	0.63	1.19	-4.42	0.63	3.57	173
MIN4SPD	0.05	0.03	0	0.04	0.2	173
MAX4SPD	0.26	0.14	0.06	0.22	0.88	173
MEAN4SPD	0.18	0.08	0.04	0.15	0.48	173
MEDIAN4SPD	0.18	0.09	0.04	0.16	0.48	173
MIN4KS	11.44	5.86	3.64	10.66	40.78	173
MAX4KS	81.53	89.76	12.72	57.17	862.88	173
MEAN4KS	20.95	12.1	8.31	18.47	97.43	173
MEDIAN4KS	16.96	7.85	5.9	15.89	59.16	173

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.25: Descriptive statistics for incomplete bid-rigging cartels in sample 4 (Swiss data)

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	9.08	1.87	6	9	13	135
MEANBIDS	448.31	384.31	41.46	337.02	3002.37	135
STDBIDS	31.74	34.92	2.26	21.49	270.82	135
CV	7.19	2.93	1.77	6.3	18.9	135
KURTO	0.47	1.97	-2.26	-0.18	6.97	135
SKEW	-0.09	0.94	-2.59	-0.12	2.57	135
SPD	0.26	0.13	0.06	0.21	0.76	135
D	22.85	42.7	0.26	9.34	351.83	135
RD	1.05	1.5	0.01	0.57	10.58	135
RDNOR	1.6	1.18	0.04	1.22	5.48	135
RDALT	2.33	3.02	0.04	1.27	20.74	135
DIFFP	5.82	7.82	0.1	3.36	45.42	135

See next page

Predictors	Mean	Std	Min	Median	Max	N
KS	16.91	7.21	6.62	16.29	57.54	135
MIN3CV	0.94	0.62	0	0.82	3.04	135
MAX3CV	12.33	5.32	3.36	10.64	32.54	135
MEAN3CV	6.47	2.56	1.55	5.85	15.38	135
MEDIAN3CV	6.07	2.49	1.01	5.67	13.93	135
MIN3SKEW	-1.71	0.07	-1.73	-1.73	-1.13	135
MAX3SKEW	1.69	0.1	0.93	1.73	1.73	135
MEAN3SKEW	-0.04	0.37	-1.02	-0.04	0.87	135
MEDIAN3SKEW	-0.03	0.66	-1.45	-0.07	1.73	135
MIN3D	1.96	3.7	0	0.87	30.19	135
MAX3D	81.04	93.45	5.67	55.61	670.57	135
MEAN3D	28.5	33.46	2.14	18	235.67	135
MEDIAN3D	23.43	25.63	2.24	16.39	171.87	135
MIN3RD	0.09	0.18	0	0.04	1.76	135
MAX3RD	2203.8	13597.91	3.46	62.53	122393.82	135
MEAN3RD	45.46	261.04	0.92	5.48	2742.11	135
MEDIAN3RD	1.69	0.96	0.45	1.5	5.61	135
MIN3RDNOR	0.1	0.14	0	0.06	1	135
MAX3RDNOR	1.92	0.1	1.42	1.96	2	135
MEAN3RDNOR	1.02	0.17	0.61	1.04	1.46	135
MEDIAN3RDNOR	1.02	0.24	0.49	1.02	1.6	135
MIN3RDALT	0.06	0.13	0	0.03	1.25	135
MAX3RDALT	1558.32	9615.17	2.44	44.22	86545.5	135
MEAN3RDALT	32.15	184.58	0.65	3.88	1938.96	135
MEDIAN3RDALT	1.2	0.68	0.32	1.06	3.97	135
MIN3DIFFP	0.44	0.47	0	0.27	2.32	135
MAX3DIFFP	20.71	11.13	3.84	17.81	63.38	135
MEAN3DIFFP	7.11	3.99	1.55	6.24	22.8	135
MEDIAN3DIFFP	5.68	2.93	1.33	5.29	21.76	135
MIN3SPD	0.02	0.01	0	0.02	0.06	135
MAX3SPD	0.26	0.13	0.06	0.21	0.76	135
MEAN3SPD	0.14	0.06	0.03	0.12	0.35	135
MEDIAN3SPD	0.12	0.05	0.02	0.11	0.3	135
MIN3KS	9.9	4.37	3.47	9.69	30.19	135
MAX3KS	603.24	4679.33	32.93	122.26	54477	135
MEAN3KS	30.89	43.76	10.24	23.63	495.16	135
MEDIAN3KS	20.34	11.39	7.28	17.92	99.02	135
MIN4CV	1.7	1.03	0.12	1.5	5.53	135
MAX4CV	10.83	4.53	2.8	9.4	28.01	135
MEAN4CV	6.76	2.69	1.64	6.08	16.54	135
MEDIAN4CV	6.88	2.88	1.69	6.17	17.17	135
MIN4SKEW	-1.81	0.29	-2	-1.92	-0.28	135
MAX4SKEW	1.72	0.4	-0.26	1.86	2	135
MEAN4SKEW	-0.06	0.52	-1.37	-0.1	1.18	135
MEDIAN4SKEW	-0.07	0.65	-1.69	-0.07	1.72	135
MIN4D	2.38	4.46	0	1.04	30.28	135
MAX4D	69.99	79.5	5.47	47.56	535.26	135
MEAN4D	25.93	33.56	1.39	16.57	274.73	135
MEDIAN4D	25.13	37.53	0.96	16.21	351.83	135
MIN4RD	0.12	0.19	0	0.05	1.47	135
MAX4RD	41.1	226.79	1.09	13.21	2627.72	135
MEAN4RD	2.27	2.46	0.35	1.72	24.7	135
MEDIAN4RD	1.25	0.95	0.16	1.06	5.95	135
MIN4RDNOR	0.16	0.19	0	0.09	1.02	135

See next page

Predictors	Mean	Std	Min	Median	Max	N
MAX4RDNOR	2.53	0.35	1.07	2.61	3	135
MEAN4RDNOR	1.11	0.33	0.43	1.11	2.05	135
MEDIAN4RDNOR	1.07	0.42	0.22	1.07	2.27	135
MIN4RDALT	0.12	0.17	0	0.06	1.04	135
MAX4RDALT	44.43	253.08	1.1	13.34	2933.56	135
MEAN4RDALT	2.37	2.65	0.37	1.79	27.4	135
MEDIAN4RDALT	1.31	1	0.16	1.11	6.3	135
MIN4DIFFP	0.53	0.59	0	0.37	3.21	135
MAX4DIFFP	18.02	10.4	3.03	15.55	56.74	135
MEAN4DIFFP	6.55	4.74	1.18	5.51	26.42	135
MEDIAN4DIFFP	6.13	5.32	0.98	4.82	34.35	135
MIN4KURTO	-5.26	1.06	-6	-5.69	-0.16	135
MAX4KURTO	3.71	0.39	2.18	3.86	4	135
MEAN4KURTO	0.16	0.8	-1.21	-0.07	2.41	135
MEDIAN4KURTO	0.72	0.98	-3.63	0.68	3.11	135
MIN4SPD	0.04	0.02	0	0.03	0.13	135
MAX4SPD	0.26	0.13	0.06	0.21	0.76	135
MEAN4SPD	0.17	0.08	0.04	0.15	0.45	135
MEDIAN4SPD	0.17	0.08	0.04	0.15	0.48	135
MIN4KS	11.21	5.03	4.25	10.99	36.12	135
MAX4KS	91.93	91.08	18.25	67.17	862.88	135
MEAN4KS	21.62	10.72	8.43	19.14	97.28	135
MEDIAN4KS	17.62	7.69	5.9	16.55	59.16	135

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Table 2.26: Descriptive statistics for incomplete bid-rigging cartels in sample 5 (Swiss data)

Predictors	Mean	Std	Min	Median	Max	N
NBRBIDS	9.42	1.81	7	9	13	104
MEANBIDS	434.09	290.66	86.12	345.62	1559.96	104
STDBIDS	29.66	26.44	4.58	21.89	191.92	104
CV	6.87	2.53	1.77	6.22	14.4	104
KURTO	0.52	2.06	-2.26	-0.15	6.97	104
SKEW	-0.1	0.94	-2.59	-0.16	2.57	104
SPD	0.25	0.12	0.06	0.21	0.66	104
D	23.61	46.99	0.26	8.88	351.83	104
RD	1.03	1.56	0.01	0.57	10.58	104

See next page

Predictors	Mean	Std	Min	Median	Max	N
RDNOR	1.65	1.24	0.04	1.25	5.48	104
RDALT	2.4	3.23	0.04	1.29	20.74	104
DIFFP	5.91	8.38	0.1	3.22	45.42	104
KS	17.25	7.04	6.93	16.66	57.54	104
MIN3CV	0.86	0.6	0	0.75	3.04	104
MAX3CV	11.97	4.77	3.36	10.54	26.24	104
MEAN3CV	6.19	2.19	1.55	5.74	12.47	104
MEDIAN3CV	5.71	2.04	1.55	5.55	11.38	104
MIN3SKEW	-1.71	0.08	-1.73	-1.73	-1.13	104
MAX3SKEW	1.71	0.05	1.48	1.73	1.73	104
MEAN3SKEW	-0.04	0.35	-1.02	-0.05	0.87	104
MEDIAN3SKEW	-0.05	0.62	-1.45	-0.07	1.32	104
MIN3D	1.47	2.09	0	0.81	13.14	104
MAX3D	76.95	75.24	9.36	56.91	530.84	104
MEAN3D	27.35	30.06	3.76	18.95	235.67	104
MEDIAN3D	21.88	21.69	2.24	16.45	144.14	104
MIN3RD	0.06	0.07	0	0.04	0.34	104
MAX3RD	2673.49	15390.37	3.46	75.98	122393.82	104
MEAN3RD	53.44	294.68	0.92	5.6	2742.11	104
MEDIAN3RD	1.66	0.88	0.45	1.48	5.61	104
MIN3RDNOR	0.08	0.08	0	0.06	0.39	104
MAX3RDNOR	1.93	0.09	1.42	1.96	2	104
MEAN3RDNOR	1.02	0.16	0.61	1.04	1.46	104
MEDIAN3RDNOR	1.02	0.23	0.49	1.02	1.6	104
MIN3RDALT	0.04	0.05	0	0.03	0.24	104
MAX3RDALT	1890.44	10882.63	2.44	53.73	86545.5	104
MEAN3RDALT	37.79	208.37	0.65	3.96	1938.96	104
MEDIAN3RDALT	1.17	0.62	0.32	1.05	3.97	104
MIN3DIFFP	0.37	0.4	0	0.24	2.32	104
MAX3DIFFP	20.4	11.4	3.84	17.28	63.38	104
MEAN3DIFFP	6.93	4.04	1.55	5.94	22.8	104
MEDIAN3DIFFP	5.4	2.68	1.33	5	14.62	104
MIN3SPD	0.02	0.01	0	0.01	0.06	104
MAX3SPD	0.25	0.12	0.06	0.21	0.66	104
MEAN3SPD	0.13	0.05	0.03	0.12	0.3	104
MEDIAN3SPD	0.12	0.04	0.03	0.11	0.25	104
MIN3KS	9.91	4.06	3.91	9.79	30.19	104
MAX3KS	748.1	5328.42	32.93	133.82	54476.99	104
MEAN3KS	32.57	48.85	11.03	24.65	495.16	104
MEDIAN3KS	20.44	8.91	9.05	18.19	65.09	104
MIN4CV	1.54	0.97	0.12	1.34	5.53	104
MAX4CV	10.53	4.01	2.8	9.12	21.92	104
MEAN4CV	6.46	2.3	1.64	5.96	13.1	104
MEDIAN4CV	6.52	2.52	1.69	6.07	17.17	104
MIN4SKEW	-1.82	0.31	-2	-1.94	-0.28	104
MAX4SKEW	1.78	0.31	0.11	1.88	2	104
MEAN4SKEW	-0.06	0.5	-1.37	-0.09	1.18	104
MEDIAN4SKEW	-0.07	0.62	-1.69	-0.07	1.72	104
MIN4D	1.95	3.57	0	0.93	30.28	104
MAX4D	68.42	69.32	8.37	50.56	495.98	104
MEAN4D	25.58	33.77	2.31	17.47	274.73	104
MEDIAN4D	24.81	39.62	0.96	15.7	351.83	104
MIN4RD	0.09	0.1	0	0.05	0.45	104
MAX4RD	49.4	257.96	1.09	15.98	2627.72	104

See next page

Predictors	Mean	Std	Min	Median	Max	N
MEAN4RD	2.33	2.67	0.35	1.76	24.7	104
MEDIAN4RD	1.24	0.92	0.16	1.07	5.95	104
MIN4RDNOR	0.13	0.14	0	0.09	0.58	104
MAX4RDNOR	2.56	0.37	1.07	2.67	3	104
MEAN4RDNOR	1.11	0.31	0.43	1.12	2.05	104
MEDIAN4RDNOR	1.07	0.4	0.22	1.08	2.27	104
MIN4RDALT	0.1	0.11	0	0.06	0.48	104
MAX4RDALT	53.65	287.9	1.1	16.19	2933.56	104
MEAN4RDALT	2.46	2.92	0.37	1.81	27.4	104
MEDIAN4RDALT	1.3	0.98	0.16	1.12	6.3	104
MIN4DIFFP	0.46	0.52	0	0.31	2.5	104
MAX4DIFFP	18.1	10.77	3.03	15.64	56.74	104
MEAN4DIFFP	6.47	4.92	1.18	5.34	26.42	104
MEDIAN4DIFFP	5.98	5.32	0.98	4.61	34.35	104
MIN4KURTO	-5.39	0.94	-6	-5.77	-1.12	104
MAX4KURTO	3.75	0.33	2.27	3.88	4	104
MEAN4KURTO	0.17	0.79	-1.07	-0.06	2.41	104
MEDIAN4KURTO	0.71	0.98	-3.63	0.62	3.11	104
MIN4SPD	0.03	0.02	0	0.03	0.13	104
MAX4SPD	0.25	0.12	0.06	0.21	0.66	104
MEAN4SPD	0.16	0.07	0.04	0.14	0.39	104
MEDIAN4SPD	0.16	0.07	0.04	0.14	0.48	104
MIN4KS	11.18	4.58	4.66	11.24	36.12	104
MAX4KS	101.52	99.91	18.25	75.09	862.88	104
MEAN4KS	22.21	10.63	9.43	19.85	97.28	104
MEDIAN4KS	18.1	7.29	5.9	17.06	59.16	104

Notes: "Mean", "Std", "Min", "Median", "Max", and "N" denote the mean, standard deviation, minimum, median, maximum, and number of observations respectively. The value for "MEANBIDS", "STDBIDS", "D", "MIN3D", "MAX3D", "MEAN3D", "MEDIAN3D", "MIN4D", "MAX4D", "MEAN4D" and "MEDIAN4D" are expressed in thousand CHF. "KS", "CV", "SPD", "RD", "RDNOR", "RDALT", "SKEW", "DIFFP", "KURTO", "D", "STDBIDS", "MEANBIDS" and "NBRBIDS" denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively.

Chapter 3

Detecting collusive coalitions

In different countries and auction formats

joint with **David Imhof***

Abstract

We propose an original application of screening methods using machine learning to detect collusive groups of firms in procurement auctions. As a methodical innovation, we calculate coalition-based screens by forming coalitions of bidders in tenders to flag bid-rigging cartels. Using Swiss, Japanese, and Italian procurement data, we investigate the effectiveness of our method in different countries and auction settings, in our cases, first-price sealed-bid and mean-price sealed-bid auctions. We correctly classify 90% of the collusive and competitive coalitions when applying four machine learning algorithms: lasso, support vector machine, random forest, and super learner ensemble method. Finally, we find that coalition-based screens for the variance and the uniformity of bids are, in all the cases, the most important predictors according to the random forest.

3.1 Introduction

Bid rigging conspiracies cost governments and taxpayers billions of dollars every year, given that OECD countries spend about 12% of their GDP on

*Chapter 3 is based on a manuscript accepted for publication at the *International Review of Law & Economics* and on a working paper. The working paper is published as Imhof and Wallimann (2021).

public procurement.[1] According to the OECD, the elimination of bid rigging could help reduce procurement prices by 20% or even more. Developing proactive methods for uncovering bid-rigging conspiracies is therefore of prime importance for competition and procurement agencies all over the world. Pro-active statistical methods to detect bid rigging in public procurement have initially been proposed by, for example, Harrington (2008) and Porter and Zona (1993). The more recent literature discusses the application of a wide range of methods to expose bid-rigging cartels in Brazil (Lima and Resende, 2021), Canada (Clark, Coviello, Gauthier, and Shneyerov, 2018), Japan (Chassang, Kawai, Nakabayashi, and Ortner, 2020), Sweden (Bergman, Lundberg, Lundberg, and Stake, 2020) and Switzerland (Huber and Imhof, 2019, Imhof, 2019).

In this paper, we add to this literature by proposing an original method of detection that focuses on coalitions formed by groups of firms. We apply our method to three different data sets from Japan, Switzerland, and Italy, for which the incidence of bid rigging is known. In all three countries, we find that, on average, our method correctly classifies nine coalitions out of ten as collusive or competitive. Moreover, the results remain robust in different auction formats, such as the first-price sealed-bid procurement mechanism in Japan and Switzerland and the mean-price sealed-bid auction in Italy. Our suggested method of detection is thus able to flag collusive groups of firms (collusive coalitions) from different bid-rigging cartels: (i) when all firms in a tender rig the contract, as in Japan and Switzerland (Ishii, 2014, Huber and Imhof, 2019, Imhof, 2019); (ii) when collusive firms face competitive firms, as in Italy and in Switzerland (see Conley and Decarolis, 2016, Wallimann, Imhof, and Huber, 2020); and (iii) when a cartel is active mostly in only one region of a market, and the firms rig only a subset of contracts (see Imhof, Karagök, and Rutz, 2018).

Our detection method is based on screens, that is, statistics derived from the distribution of bids in a tender. To derive screens for coalitions, we start by selecting three firms and isolate all the tenders in which those three firms submitted a bid. We calculate the screens based exclusively on the three bids of those firms obtaining the tender-based screens for a coalition for each tender. We then calculate the descriptive statistics of the tender-based screens, including the mean, median, minimum, and maximum for each coalition. These statistics, henceforth called 'coalition-based screens', synthesize the distributional features of bids for a specific coalition. Since we use data from different bid-rigging cases with complete information, we

[1] See https://www.oecd.org/competition/cartels/fightingbidrigginginpublicprocurement.htm (accessed 8 April 2021).

can identify a coalition as competitive and collusive to build the outcome variable. We focus on coalitions of three firms since we aim to detect even small bid-rigging cartels. Focusing on coalitions of two firms would impede the application of most of our screens, and with coalitions of four firms or more it would hinder the detection of the smallest bid-rigging cartels formed by three firms.

As in recent studies (Foremny, Kulejewski, Anysz, and Nicał, 2018, Rabuzin and Modrusan, 2019, García Rodríguez, Rodríguez Montequín, Ortega Fernández, and Villanueva Balsera, 2020, Silveira, Vasconcelos, Resende, and Cajueiro, 2021), we use machine learning to train and test models to flag bid-rigging cartels. For this purpose, machine learning is ideal since it focuses on developing predictive models to determine an outcome. Machine learning does not focus on the causal structural relationship, e.g., between collusion and the distributional pattern of bids. In other words, we remain agnostic about the effects of bid rigging on the distribution of bids when using machine learning techniques. However, we discuss the effects of bid rigging on coalition-based screens by illustrating some common important predictors in all the cases being considered. In our study, we combine the coalition-based screens described above with machine learning to predict whether a coalition of firms colluded in bidding or not. To train predictive models and evaluate their goodness of fit in independent test sets, we apply four widely used machine learning algorithms: the random forest (Breiman, 2001), the lasso (Frank and Friedman, 1993, Tibshirani, 1996), the support vector machines (Cortes and Vapnik, 1995), and the "super learner" ensemble method, including random forest, neural networks, gradient boosting, and least absolute shrinkage and selection operator (lasso) regression (van der Laan, Polley, and Hubbard, 2008).

We first apply our coalition-based approach to the Okinawa bid-rigging cartel from Japan (see also Ishii, 2014, Huber, Imhof, and Ishii, 2020). The four machine learning algorithms offer correct classification rates from 91.9% to 94.9% to classify a coalition as collusive or competitive. In addition, changing the perspective from a tender-based approach to a coalition-based approach increases the correct classification rate of three to six percentage points, corresponding to a decrease of between 27% and 55% in the error rate, defined as one minus the correct classification rate. Secondly, we implement our coalition-based approach on Swiss bid-rigging cartels (see also Huber and Imhof, 2019, Wallimann, Imhof, and Huber, 2020), and we find correct classification rates from 86.9% to 90.5%. The increase in the correct classification rates using a coalition-based approach amounts to four to seven percentage points when comparing the results of the various models applied to complete bid-rigging cartels, which in Wallimann, Imhof, and Huber (2020) amounts approximately to 83%. Our coalition-based approach, therefore, reduces the error rate by between 23% and 44%, inclusive. Finally, we apply our coalition-based approach to Italian bid-rigging cartels (see also Conley and Decarolis, 2016) and find correct classification rates from 84.8% to 90.1% for flagging collusive coalitions. We find that the medians of the coefficient of variation, the spread, and the KS-statistic are the most powerful predictors for flagging collusive coalitions for the three different countries. While the levels of the medians differ strongly between the cases, the effect of bid rigging on the screens goes in the same direction, and its magnitude is, to a certain extent, similar. Therefore, benchmarks in screening other markets in other countries should rely on the effect of bid rigging. For example, a decrease by a factor of two in the medians of the spread and the coefficient of variation would indicate potential competitive issues requiring further scrutiny.

We complement our analyses in three steps using the Swiss data. First, we add more summary statistics for the tender-based screens. With an enlarged set of coalitions-based screens, we find no significant improvement in the correct prediction rate, indicating that summary statistics based on the mean, median, minimum, and maximum are sufficient to deliver a good performance in predicting collusive and competitive coalitions. Second, we discuss why coalition-based screens for the variance and the uniformity of bids perform significantly better than those for the asymmetry of bids. We find that applying only screens for the asymmetry of bids to the Swiss data (omitting the coalition-based screens for the variance and uniformity of bids) produces a poor correct prediction rate. This might be due to the fact that, by forming a coalition (with few firms), the bid of the designated winner and thus the distance between the winning bid and the second lowest bid from the cartel is not systematically considered. Therefore, the asymmetry in the coalition's distribution of bids decreases. Finally, we investigate the number of bidders in coalitions formed with four firms. The result indicates an increase in the correct prediction rates, especially for the collusive coalitions. This might be explained by the increase in the predictive power of coalition-based screens for asymmetry. Including more firms in a coalition reduces the likelihood that the first bid in the tender will be omitted. Thus, coalition-based screens appear to be more asymmetric and thus more predictive.

Our paper relates to other studies using screens for uncovering cartels (see Abrantes-Metz, Froeb, Geweke, and Taylor, 2006, Esposito and Ferrero, 2006, Hueschelrath and Veith, 2014, Jiménez and Perdiguero, 2012, Abrantes-Metz, Kraten, Metz, and Seow, 2012, Huber and Imhof, 2019, Imhof, 2019). Calculating screens for subgroups as in our approach is also discussed by Conley and Decarolis (2016) and Chassang, Kawai, Nakabayashi, and Ortner (2020). First, Conley and Decarolis (2016) calculate subgroups to detect cartels in collusive auctions in Italy. In order to identify collusive bidders, we similarly rely on the bids observed in a tender. However, we do not consider firm-specific covariates, such as common owner, municipality, or country, to determine subgroups, as proposed in the study by Conley and Decarolis (2016). Different, we do not rely on firm-specific covariates, which could impede screening activity if firm-specific data are unavailable or there is not enough time to collect them in secrecy without attracting the attention of potential cartel participants. Chassang, Kawai, Nakabayashi, and Ortner (2020) show that winning bids tend to be isolated when bidders collude. They calculate the difference between a bidder's bid and the lowest bid submitted in a tender, therefore focusing on subgroups of two bids to calculate the distribution of differences. However, we do not focus solely on subgroups consisting of only the lowest bid in a tender and one of its opposing bidders.

More broadly, our study can be linked to papers on detecting bid-rigging cartels not relying on screens. One seminal paper by Bajari and Ye (2003) proposes two econometric tests for classifying pairs of firms as collusive. Subsequent papers apply and refine the econometric tests suggested by Bajari and Ye (2003) (see Jakobsson, 2007, Aryal and Gabrielli, 2013, Chotibhongs and Arditi, 2012a,b, Imhof, 2017, Bergman, Lundberg, Lundberg, and Stake, 2020). Imhof (2017), however, questions the performance of the econometric tests proposed by Bajari and Ye (2003) for detecting the Ticino cartel because econometric tests produce too many false negatives by failing to classify pairs of firms as collusiv. In contrast, the screens perform well in detecting the Ticino cartel. Our research is also associated with papers analyzing the effect of bid rigging (Pesendorfer, 2000, Ishii, 2009, Clark, Coviello, Gauthier, and Shneyerov, 2018) and with papers investigating the change in bidding patterns when bid rigging occurs (Porter and Zona, 1993, 1999).

The remainder of the paper is organized as follows. Section 3.2 outlines our method of detection. In Section 3.3, we apply our detection method to public procurement datasets from Italy, Japan, and Switzerland. We also discuss the observed variance screens and the Kolmogorov-Smirnov statistic, which are important in flagging bid-rigging cartels. Section 3.4 performs complementary analyses. In Section 3.5, we discuss the advantages and policy implications of our approach. Section 3.6 concludes the paper.

3.2 Detection method

In our study, we focus on supervised machine learning that entails a set of predictors (X), also features or covariates, to predict an outcome (Y). The outcome of our classification setting is given a value of 1 for a collusive coalition, which only includes cartel participants, and a value of 0 for a competitive coalition, which is formed only by competing firms. Machine learning requires the data to be randomly split into independent training and test datasets. In our applications, the training and test sets consist of 75% and 25% of the observations, respectively. We develop predictive models using all observations in our training set, where both features and outcomes are observed. The goal is to predict the outcomes in the test data based on their covariates. This is closely related to discrete choice analysis in econometrics, where statistical models specify a probability that an outcome takes a particular value conditional on the features (Athey and Imbens, 2019). However, machine learning aims to achieve goodness of fit in the independent test set by minimizing deviations between the predicted and the actual outcomes (Athey, 2019).

We assess the predictive performance of machine learning algorithms by comparing the algorithm's prediction with the actual outcome in the test set. The number of correct predictions divided by the total number of observations in the test set defines the 'accuracy' (also the correct classification rate) achieved by the algorithms. For every application, we create a dataset in which the binary outcome is balanced, i.e., with 50% collusive and 50% competitive coalitions. Balancing the dataset each time before splitting the sample enables the applied algorithms to build models predicting both coalition classes, collusive and competitive, equally well. After randomly balancing the dataset, we repeat the sample splitting into training and test data a hundred times. The correct classification rates of our applications are the average predictive performances of the hundred repetitions. In our study, we train machine learning algorithms with coalition-based screens (X) to flag collusive coalitions (Y) in the three countries of Italy, Switzerland, and Japan. In the following, we first discuss the machine learning algorithms used for training and testing our predictive models. We then describe the coalition-based approach and the screens entering into the algorithms as features.

3.2.1 Machine learning algorithms

The first machine learning algorithm we implement is the least absolute shrinkage and selection operator (lasso) regression, introduced by Frank and Frank and Friedman (1993) and Tibshirani (1996). In our case, a lasso regression is a type of logit regression using shrinkage. It includes a penalty term, restricting the sum of absolute coefficients on the regressors. Coefficients with a low predictive power shrink, depending on the penalty term, towards or exactly zero. As some coefficients become zero and the algorithm discards these variables from the model, the lasso regression can result in sparse models with only the most powerful predictive variables. Based on the mean squared error of prediction, we apply 15-fold cross-validation to select the penalty term. In our applications of the lasso regression, we use the *hdm* package by Chernozhukov, Hansen, and Spindler (2016) in the statistical software R.

Second, we use the random forest (see Breiman, 2001), an algorithm predicting the outcome by a majority rule across multiple individual decision trees. Therefore, this machine learning method draws random subsamples from the original training set and estimates the predictive model, in our case a decision tree, in each of the subsamples. A decision tree splits the feature space into a number of non-overlapping regions. Each split aims to maximize the homogeneity of the outcome according to a goodness of fit criterion. In the case of binary variables, the Gini coefficient is a popular criterion that measures the average gain in homogeneity of the outcome values. The splitting continues until the tree reaches a specific stopping rule, e.g., the minimum number of observations in a terminal node. The tree-based predictions of the outcome are based on whether collusive coalitions are present or absent in the region that contains the values of features for which the outcome is to be predicted. Using tree-based methods, there exists a bias-variance trade-off in the out-of-sample prediction. Through more splits, on the one hand, we reduce the bias and increase the flexibility of the model specification. On the other hand, more splits increase the variance in the test data due to regions being smaller. By repeatedly drawing many subsamples from the training set and estimating the decision tree, the random forest mitigates excessive variance in the test set. To reduce the correlation of tree structures across the subsamples and the prediction variation, the random forest considers each decision tree's splitting step only a random subsample of features. The subsample of features at each split amounts to the square root of potential predictors in our applications. To implement the random forest in the statistical software R, we use the *randomForest* package of Liaw and Wiener (2018), with growing a thousand trees. Implementing the random forest, we also present the most important variables according to the Gini Index as a measure of the best split selection, which measures the impurity of a given element with respect to the remaining classes.

Third, we implement support vector machines (Cortes and Vapnik, 1995). Support vector classifiers are based on the idea of finding a hyperplane that best segregates the training data into two categories. We can think of a hyperplane as a line separating the observed points in a two-dimensional space into two classes. We then map the observations of the test data into the space and predict them to belong to one class based on the side of the hyperplane on which they fall. In the training data, we want our data points to be as far away as possible from the hyperplane; as for these data points confidence in producing a correct classification will be high. The distance from the nearest data point in either of the two separated classes and the hyperplane is known as the margin. Giving a greater chance of new data being correctly classified, the algorithm chooses a hyperplane with the goal to achieve the greatest possible margin. However, the idea of the hyperplane as a line is a simplification, as a linear hyperplane might perform poorly when the data points are not separable with a line. Support vector machines offer an extension of the support vector classifier by enlarging the feature space using kernels and mapping the inputs into high-dimensional feature spaces. In our application of support vector machines, we use the *e1071* package Meyer (2015).

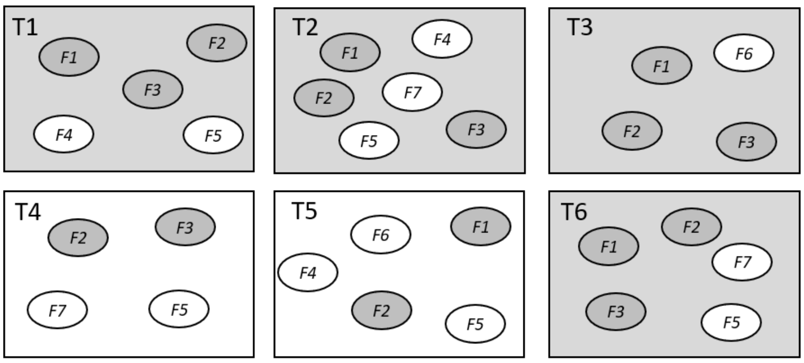
Fourth, we apply the *SuperLearner* package by van der Laan, Polley, and Hubbard (2008), which is an ensemble method. In our case, the super learner is a weighted average of four machine learning algorithms: gradient boosting, random forest, lasso and neural networks, using the *xgboost*, *cforest*, *glmnet* and *nnet* packages respectively. Gradient boosting resembles the random forest described above, as it grows a set of decision trees. However, unlike the later algorithm building each tree independently, gradient boosting is an additive model working in a forward stage-wise manner and therefore building only one tree at a time. While the random forest averages over all decision trees at the end, gradient boosting combines the results along the way. That is, building individual decision trees sequentially learning from mistakes made by previous ones. Neural networks aim at fitting a system of non-linear functions modeling the influence of the features on the outcome in a flexible way. The algorithm uses a network of non-linear intermediate functions, so-called hidden nodes, to model the association between the predictors and the outcomes.

3.2.2 Coalitions and predictors

Procurement markets are seldomly characterized by a complete cartel involving all firms bidding for a tender. In such cases, a suspicious group of bidders (a coalition) must be isolated by further statistical tests, as suggested, for example, by Imhof, Karagök, and Rutz (2018). As a methodical innovation, in our paper, we develop a coalition-based approach for flagging cartel participants. Our approach overcomes the isolation step and directly identifies collusive coalitions. By 'coalition' we mean a subgroup of three firms bidding together in at least three tenders. Coalition-based screens consisting of three bidders can detect bid-rigging cartels including more than three firms. However, coalition-based screens consisting of four bidders would not be able to flag bid-rigging cartels including with only three bidders. Therefore, we focus on three firms since our main aim is to create a method of detection that is capable of flagging small bid-rigging cartels and large ones.[2]

To prepare the predictors for an observation (a coalition), we extract all the tenders in which each of the three firms submitted a bid and discard all bids submitted by firms that are not part of the coalition. Figure 3.2.1 illustrates the procedure. The boxes represent tenders, the circles firms. In Figure 3.2.1, we have a sample with six tenders T1 to T6 and seven firms F1 to F7 applying for projects. We pick firms F1, F2, and F3 to form the first coalition, henceforth called coalition 123. Each of these firms submits a bid in each tenders T1, T2, T3, and T6 (grey). To form coalition 123, we extract this subgroup of tenders and discard all bids submitted by firms that are not part of the coalition. For example, in tender T1, we drop firms F4 and F5 (white circles).

Figure 3.2.1: The selection of coalition 123



In the next step, we calculate screens for the distribution of the three extracted bids in each tender. Screens are descriptive statistics describing the discrete distribution of bids in a tender (Abrantes-Metz, Froeb, Geweke, and Taylor, 2006, Abrantes-Metz, Kraten, Metz, and Seow, 2012, Harrington, 2008, Jiménez and Perdiguero, 2012, Imhof, 2019). Since screens summarize the behavior of the bidders in one tender, they refer to the category of behavioral screens as discussed by Harrington (2008). We make the simple hypothesis that bid rigging modifies the distribution of bids. There are two reasons for this: (i) the members of a bid rigging cartel know the bids of their competitors, and (ii) they coordinate the bids. Therefore, we can capture such distributional changes with the screens. This hypothesis is common to

[2]Using coalitions of only two firms does not allow us to calculate all possible screens, leaving us with a reduced set of predictors.

detection methods such as the econometric tests suggested by Bajari and Ye (2003).

Following Huber, Imhof, and Ishii (2020), Huber and Imhof (2019) and Wallimann, Imhof, and Huber (2020), we implement nine screens to uncover bid-rigging cartels. The screens can be assigned to three categories. The first category contains variance screens such as the coefficient of variation (see e.g. Abrantes-Metz, Froeb, Geweke, and Taylor, 2006, Abrantes-Metz, Kraten, Metz, and Seow, 2012, Imhof, 2019, Jiménez and Perdiguero, 2012) and the spread (see e.g. Wallimann, Imhof, and Huber, 2020). These screens capture the possible reduction in support of the distribution of bids or the convergence of bids when a cartel coordinates bids, and bidders exchange their bids before submitting them in the tendering process (Imhof, 2019). The second category contains the percentage difference, the absolute difference, the skewness, the relative distance, the alternative distance, and the normalized distance (see for these screens Huber and Imhof, 2019). Screens of this category measure whether the bids exhibit an asymmetrical distribution. Cartel participants can simultaneously affect both differences between losing bids and differences between the first and second lowest cartel bids. Empirical observations (see, e.g. Chassang, Kawai, Nakabayashi, and Ortner, 2020) have shown that the differences between the first and second lowest cartel bids increase, whereas the differences between losing bids decrease. This increases the asymmetry in the distribution of the bids and is explained by the necessity to ensure the contract is awarded to the winner designated by the cartel. The third category of predictors is based on the Kolmogorov-Smirnov statistic (hereafter the KS statistic), which is calculated to test whether the discrete distribution of bids follows a uniform distribution (see Wallimann, Imhof, and Huber, 2020). The KS statistic thus investigates how dissimilar the distribution of the bids is with a uniform probability distribution due to bid rigging.

By looking again at coalition 123, we illustrate the calculation of the screens with the coefficient of variation. For each tender in the extracted subgroup of tenders T1, T2, T3 and T6, we first calculate the coefficient of variation, that is, the standard deviation divided by the arithmetic mean of the three bids of firms F1, F2 and F3. We thus obtain four coefficients of variation, in other words, four tender-based screens. Thereafter, we calculate the mean, median, minimum, and maximum using these tender-based screens to obtain summary statistics for each coalition, the so-called coalition-based screens. Calculating the coalition-based screens for each screen presented above, we end up with 36 coalition-based screens for a coalition (observation) in the data. We then use these coalition-based screens as features (X) in our predictive models to determine the outcome (Y).

We repeat the building of coalitions and the calculation of the coalition-based screens for all possible coalitions of three firms if the three firms at least participate together in three tenders.[3]

3.3 Empirical analyses in different countries

We apply our original coalition-based approach to uncover collusive cartels in three countries: Japan, Switzerland, and Italy. These cases are discussed and screened in earlier studies (see Conley and Decarolis, 2016, Wallimann, Imhof, and Huber, 2020, Huber and Imhof, 2019, Ishii, 2014). While procurement agencies in Japan and Switzerland used first-price sealed-bid auctions, agencies in Italy used mean-price sealed-bid auctions. Therefore, we can train and evaluate our algorithms in different countries and different auction settings. We also present the ten most important variables ranked by the Gini Index according to the random forest for every application. Since the most important coalition-based screens for prediction remain stable across countries and auction settings, we briefly discuss them for further screening applications.

3.3.1 Okinawa cartel

For our first application, we use an empirical dataset from Japan originally introduced by Ishii (2014) and recently analyzed by Huber, Imhof, and Ishii (2020). The dataset contains construction contracts in Okinawa from April 2003 to March 2007. As the Okinawa Prefecture consists of 47 islands, the market is difficult to enter for firms outside this region. Thus, there is a natural geographical barrier to their entry into the construction market. Local agencies used a first-price sealed-bid auction to procure a contract and specified a reserve price and a lowest acceptable price. The lowest bid submitted between the lowest acceptable price and the reserve price won the contract.

During the whole period, the agency invited a set of qualified firms to submit a bid depending on the size of the tendered contract. In addition, agencies disclosed the identity of the invited firms prior to each tender procedure, a practice that ended in January 2006. The natural geographical barriers, the restricted number of competitors, and the disclosure of their identity notably simplified the emergence of bid rigging. Hence, the cartel participants communicated with each other before each tender and met to negotiate and agree on the firm that would win the contract and the winning

[3] We have chosen three projects, as this is the minimum for calculating summary statistics and allows us to achieve the most observations possible.

price. Thereafter, the other bidders not chosen to win the contract calculated a cover bid that was sufficiently higher than the winning price.

In June 2005, the Japanese Fair Trade Commission (hereafter JFTC) launched an investigation into bid-rigging conspiracies against a large number of firms involved in these tenders. In January 2006, to limit the risk of bid-rigging in the future, the Okinawa prefecture adapted its procurement system by inviting more firms and not revealing the identities of firms prior to the tendering procedure. At the same time, Japan's competition law was revised. Changes included increasing fines for conspiracies and introducing a leniency program granting a complete or partial exemption from financial penalties if a firm collaborates with the JFTC.

Table 3.1: The correct classification rates for the Okinawa cartel

Classifier	Prediction Results		
	CCR (%)	CCR coll (%)	CCR comp (%)
Lasso	92.3	93.5	91.2
Random forest	94.9	96.9	92.8
Super learner	93.9	94.7	93.1
Support vector machines	91.9	93.7	90.0

Note: 'CCR' denotes the correct classification rate, 'CCR coll' the correct classification rate of the collusive coalitions, and 'CCR comp' the correct classification rate of competitive coalitions.

To create the Japanese collusive coalitions, we consider contracts of type A+ in the pre-inspection period (see Huber, Imhof, and Ishii, 2020, for more details). For the competitive coalition, we use only contracts of type A+ in the post-amendment period, in which the JFTC sanctioned the cartel participants involved in light of Japanese competition law being revised and procurement rules in Okinawa reinforced. After recreating the coalitions, our final dataset contains 207 collusive and 1,793 competitive coalitions. The average number of projects per coalition amounts to 3.4 for both the pre- and post-amendment periods.

As stated in Table 3.1, we use the four algorithms presented above to achieve decent correct prediction rates from 91.9% to 94.9%. Therefore, the deviations between the predicted and actual outcomes are low. Our coalition-based approach outperforms the application of Huber, Imhof, and Ishii (2020) by about two to six percentage points depending on the algorithm.[4] This

[4]We do not compare our analysis with model 1 in Huber, Imhof, and Ishii (2020) but with

performance improvement may seem weak, but it is, in fact, notable if we consider the error rate (also the misclassification error), defined as one minus the correct prediction rate. In Huber, Imhof, and Ishii (2020), the error rate amounts to approximatively eleven percentage points for models using screens exclusively. Therefore, an improvement of two to six percentage points implies reducing the error rate of between 18% and 55% inclusive. Such a reduction in the error rate is valuable in light of the heavy legal consequences of a firm being flagged as a potential cartel participant and an investigation being opened against it. Furthermore, an investigation has procedural consequences in being costly for both the authority, i.e., the taxpayer, and the firms.

By comparing the accuracy of the machine learning algorithms, we see from Table 3.1 that the random forest achieves the highest correct classification rate. Moreover, for all algorithms, differences in the predictive performance between the collusive and competitive coalitions remain minor, despite slightly better prediction rates for the collusive coalitions. Nevertheless, the imbalance is the smallest for the super learner.

Figure 3.3.1: Variable importance plot for the Okinawa cartel. We compute the variable importance using the mean decrease in the Gini index and express it relative to the maximum.

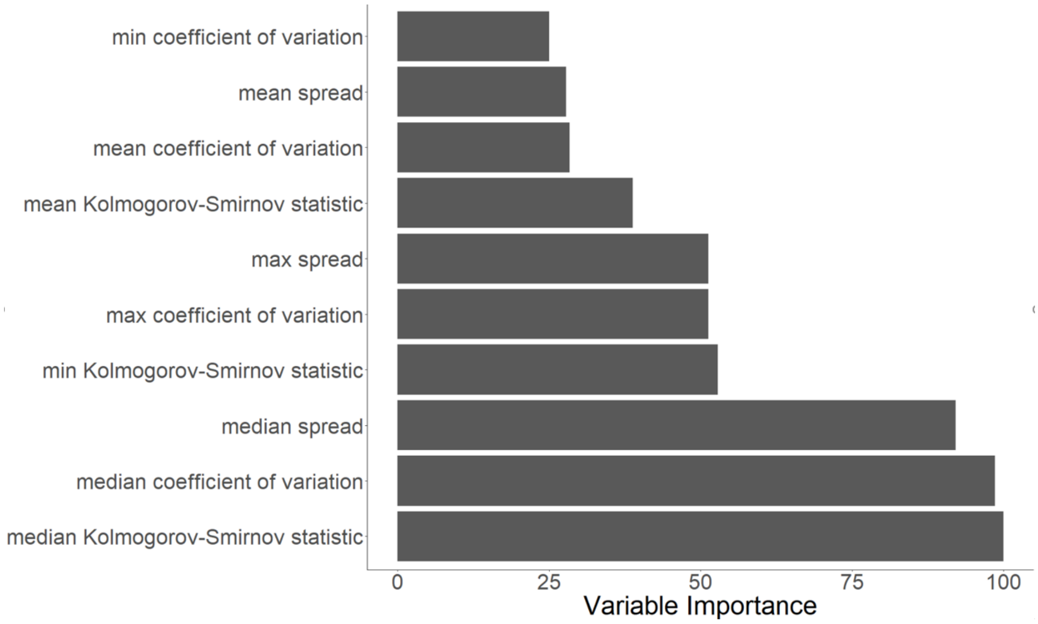


Figure 3.3.1 depicts the relative importance of the predictors according to the random forest. We notice that the median of the coefficient of variation,

model 2, which uses only screens, as in our approach.

the spread, and the KS statistic are the most important coalition-based screens for predicting bid rigging. Screens for asymmetry, however, appear to be unimportant in predicting bid rigging when also using screens for the variance or uniformity of bids to fit machine learning models.

3.3.2 Swiss cartels

Our second application considers the dataset from three bid-rigging cartels in Ticino, See-Gaster and Graubünden, discussed by Wallimann, Imhof, and Huber (2020). The Swiss Competition Commission (hereafter COMCO) convicted cartel participants in all three cases. However, COMCO only sanctioned cartel participants in two cases since the bid-rigging cartel in Ticino ceased its illegal activity before the revised competition law in Switzerland entered into force, including the possibility of sanctioning firms. The latter cartel was active in the period from January 1999 to March 2005 and included all firms in the road construction market in Switzerland's southernmost canton (see also Imhof, 2019). The firms rigged public and private contracts before stopping their anticompetitive activity. After the cartel came to an end, prices fell by roughly 30% (Imhof, 2019).

From 2004 to 2010, eight firms in the See-Gaster region (cantons of St. Gallen and Schwyz) participated in a bid-rigging conspiracy. The cartel participants met at least once a month to discuss future tenders for road construction, asphalting and civil engineering. The cartel members designated the winning firm, which then negotiated the price itself and the cover bids with the cartel participants in separate meetings.

The third cartel, which was active from 2004 to 2010, included most road construction firms in the canton of Graubünden, a canton characterized by valleys and mountains. The cartel was divided into two groups of cartel participants operating in the north and the south, respectively. As in the two latter investigations, the cartel participants discussed local and cantonal contracts for asphalting and construction tendered by the canton and the cities. COMCO estimated that the activities of the cartel pushed up prices by at least 10%.

In Switzerland, procurement agencies also take other criteria into account in awarding contracts and not just price, such as quality considerations and environmental aspects. Price, however, remains the most crucial criterion. Therefore, the procurement process in Switzerland is characterized by a first-price sealed-bid auction (for further explanations, see Wallimann, Imhof, and Huber, 2020).

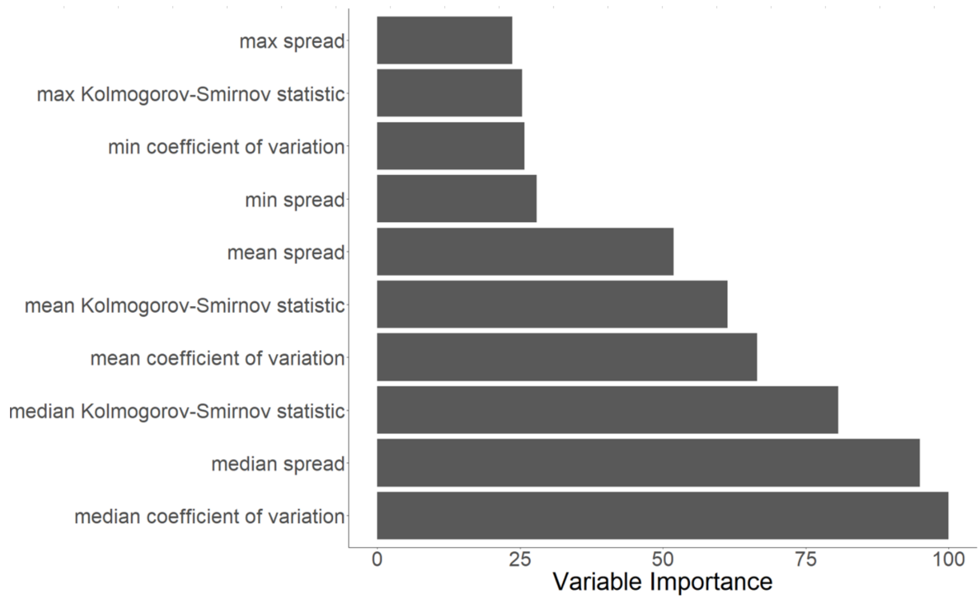
In this study, we pool the data of all three cartels. We use cartel participants to construct collusive coalitions. Then, competitive coalitions are

Table 3.2: The correct classification rates for the Swiss cartels

Classifier	Prediction Results		
	CCR (%)	CCR coll (%)	CCR comp (%)
Lasso	86.9	88.5	85.4
Random forest	89.7	88.3	91.1
Super learner	90.5	90.0	91.1
Support vector machines	87.2	88.1	86.3

Note: 'CCR' denotes the correct classification rate, 'CCR coll' the correct classification rate of the collusive coalitions, and 'CCR comp' the correct classification rate of competitive coalitions.

Figure 3.3.2: Variable importance plot for the Swiss cartels. We compute the variable importance using the mean decrease in the Gini index and express it relative to the maximum.



created with former cartel participants to investigate the changes between the collusive and competitive coalitions. At the end of the formation of all coalitions, we end up with 646 competitive and 896 collusive coalitions. The average number of projects per coalition amounts to 21.4 and 44.9 for competitive and collusive coalitions.

As shown in Table 3.2, the correct prediction rates amount to 86.9%, 89.7%, 90.5%, and 87.2% for the lasso, random forest, super learner, and support vector machines, respectively. The super learner reaches the lowest misclassification error in predicting collusive and competitive coalitions in the Swiss data. We improve the predictive performances of Wallimann, Imhof, and Huber (2020) by three to seven percentage points if we consider only the complete bid-rigging cartels (with no competition of firms that are not part of the cartel) in the various models applied. Such increases in the correct prediction rate imply a decrease of between 23% and 44% inclusive in the error rate.

Like the Okinawa application, we observe a convergence of the algorithms but a slightly better performance for the random forest and super learner. We also notice that the random forest and the super learner are slightly better at predicting competitive coalitions. Therefore, they produce fewer false positives (one minus the correct prediction rate for competitive tenders) than false negatives (wrongly flagging a collusive coalition as competitive). The reverse applies to the lasso and the support vector machines, which predict better collusive coalitions but with a lower overall correct classification rate. All four machine learners exhibit a similar correct classification rate for the collusive coalition, i.e., the same false negative results. Increasing false positive results for the lasso and the support vector machines explains the difference in the overall correct classification rates.

Figure 3.3.2 reports the most important coalition-based screens according to the random forest. We again observe that medians for the KS statistic and the spread and coefficient of variation are the three most important coalition-based screens in classifying collusive coalitions. However, again, in the ten most predictive coalition-based features, we do not find any screens for the asymmetry of bids.

3.3.3 Italian cartels

Our third application involves contracts for roadworks tendered in the Turin municipality of Italy between 2000 and 2003, first introduced by Conley and Decarolis (2016). They use two datasets: a validation dataset and a main dataset. In our application, we use the validation dataset because there are no court decisions in the main dataset (and we would have no prior knowledge of the existence of a cartel in this dataset).

The procurement agency in Turin used an average bid auction for tendering the roadwork contracts. First, it defined a reserve price for a contract and publicly announced it. Then, based on the reserve price, interested firms submitted a bid, which was a discount based on the reserve price. Having collected all the bids, the agency first ranked them and discarded the ten percent lowest and highest bids to calculate a trimmed mean. The agency then calculated a second mean for all bids (including discarded ones) higher than the trimmed mean in the first step. The firm with the highest bid lower than the mean of the second step won the contract (see Conley and Decarolis, 2016, for details).

In 2008, the Court of Justice in Turin identified eight cartels involving 95 firms as potential cartel participants and sentenced 27 firms for bid-rigging conspiracies. The firms mainly formed cartels with nearby companies. Overall, the coordination of bids paid off because the suspected cartel participants won 80% of the tendered contracts, though they accounted for only 10% of all the bidders.

Table 3.3: The correct classification rates for the Italian cartels

Classifier	Prediction Results		
	CCR (%)	CCR coll (%)	CCR comp (%)
Lasso	84.8	83.9	85.8
Random Forest	89.1	87.6	90.6
Super learner	90.1	89.9	90.3
Support Vector Machines	85.2	83.2	87.3

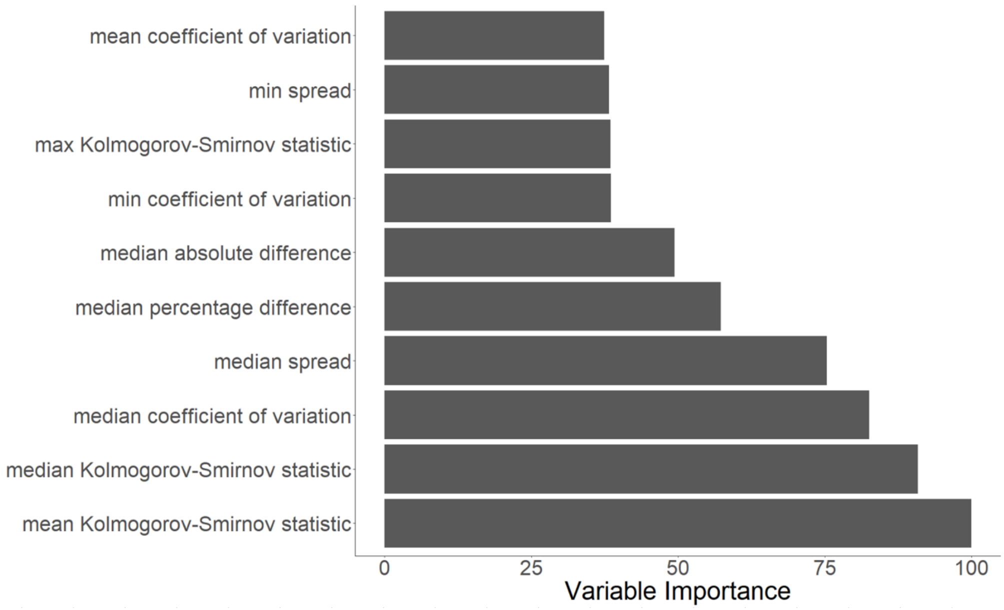
Note: 'CCR' denotes the correct classification rate, 'CCR coll' the correct classification rate of the collusive coalitions, and 'CCR comp' the correct classification rate of competitive coalitions.

By recreating the coalitions in the Italian data, we take 75 of the most frequent competitive bidders and obtain 21,340 competitive coalitions with an average of 20.7 contracts. Next, we calculate collusive coalitions within each of the eight cartels. We end up with 1,474 collusive coalitions with an average of 47.4 contracts.

Our coalition-based models reach correct classification rates from 84.8% to 90.1% in detecting the Italian cartels and therefore perform well in a different kind of auction procedure (see Table 3.3). Again, we find the super learner and the random forest to be the best performing algorithms compared to the lasso and the support vector machines. We notice that the lasso, the support vector machines, and the random forest perform better in predicting competitive coalitions and thus produce fewer false negatives than false positives.

Figure 3.3.3 presents the ten most predictive coalition-based screens in predicting Italian collusive coalitions. We again find similar important predictors as in the latter applications: the median for the KS statistic and the coefficient of variation and spread are the most predictive coalition-based screens with the mean of the KS statistic. Unlike the previous cases, we notice that two screens for the asymmetry of bids, the median of the percentage difference and of the absolute difference, also play a role in the top-ten predictors. Nevertheless, screens for the variance and uniformity of bids dominate the best predictors.

Figure 3.3.3: Variable importance plot for the Italian cartels. We compute the variable importance using the mean decrease in the Gini index and express it relative to the maximum.



3.3.4 The most predictive coalition-based screens

Applying our approach to three different countries, we find the same screens to be the most important predictors (X) for flagging collusive coalitions (Y): we mainly find coalition-based screens for the variance, i.e., medians of the coefficient of variation and the spread. Table 3.4 reports the mean values of these coalition-based screens. We find that the medians of the spread and of the coefficient of variation are, on average, considerably higher for competitive coalitions. If the level of the variance of bids differs across countries, the effect of bid rigging is similar in magnitude. Bid rigging affects the variance of the bids by decreasing them by two for Swiss bid-rigging cartels and by three for the Italian and Japanese bid-rigging cartels. Bid rigging also decreases the spread by a factor of two for Switzerland, by a factor of three for Japan, and by a factor of between three and four for Italy.

Alongside screens for variance, we find that the median of the KS statistic, calculated to test if a discrete bid distribution follows a uniform probability distribution law, is also a powerful coalition-based screen. Table 3.4 indicates that bid rigging notably increases the KS statistic in all cases. In other words, the results suggest that bid rigging and the related necessary bid coordination transform the distribution of bids in a much less uniform distribution. Again, the level of the KS statistic differs across countries, but the effect of bid rigging follows the same direction in all cases. For example, bid rigging on average doubles the KS statistic for coalitions in Japan and Switzerland compared to their competitive counterparts, whereas for the former in Italy, this screen increases by a factor of five.

Table 3.4: Mean and standard deviation of the coalitions' medians

	Okinawa		Italy		Switzerland	
	Coll.	Comp.	Coll.	Comp.	Coll.	Comp.
Coefficient of variation	1.06	3.19	10.13	30.73	3.38	6.80
	(2.48)	(2.86)	(16.17)	(20.68)	(1.58)	(3.01)
Spread	0.02	0.06	0.32	1.16	0.07	0.14
	(0.05)	(0.06)	(0.70)	(1.25)	(0.02)	(0.07)
KS statistic	286.53	143.60	35.57	7.13	34.22	17.81
	(192.43)	(559.20)	(40.85)	(9.72)	(15.17)	(7.03)

Note: 'Coll.' denotes collusive coalitions, 'Comp.' competitive coalitions. The figures in brackets are the standard deviations.

3.4 Complementary analyses

In this section, we outline the complementary analyses we perform using the Swiss data. First, we enlarge our set of predictors. Second, we investigate why coalition-based screens for the variance and uniformity of bids perform better than those for the asymmetry of bids. Finally, we form coalitions with four firms.

3.4.1 Using additional coalition-based screens

In the previous section, we calculate coalition-based screens (X) by taking into account the summary statistics mean, median, minimum, and maximum of the tender-based screens for each coalition. In this section, we investigate the robustness of these summary statistics, chosen using the Swiss data. Therefore, we calculate the 5th, 10th, 25th, 75th, 90th, and 95th percentiles from the tender-based screens for each coalition in the Swiss data. Then, we add them to the coalition-based screens we use in our original application, i.e., mean, median, minimum, and maximum. Thus, we fit our models with ninety coalition-based screens in our first complementary analysis.

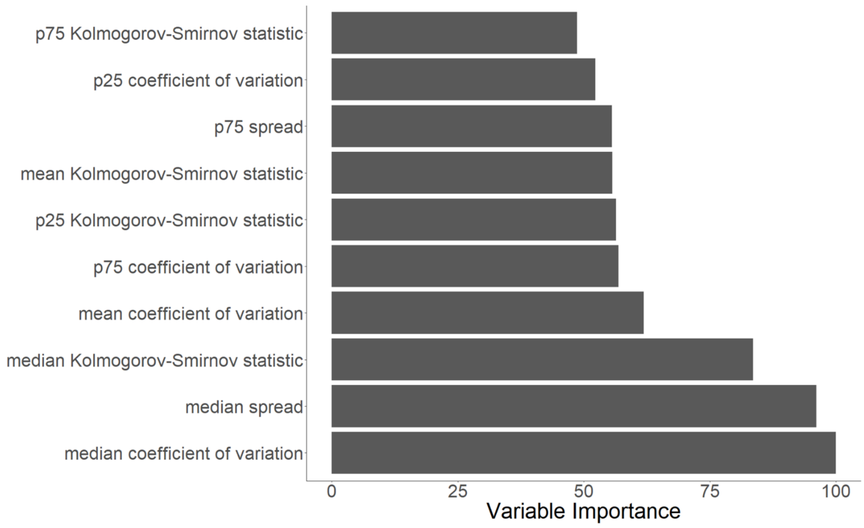
Table 3.5: Changes in accuracy when adding a new set of predictors in the application of the Swiss cartels

Classifier	Changes in percentage points		
	CCR	CCR collusion	CCR competition
Lasso	1.8	0.7	2.9
Random forest	-0.1	-0.8	0.6
Super learner	0.6	-0.1	1.3
Support vector machines	0.9	0.1	1.7

Note: 'CCR' denotes the correct classification rate, 'CCR collusion' the correct classification rate of the collusive coalitions, and 'CCR competition' the correct classification rate of competitive coalitions.

Table 3.5 shows the increase in percentage points of the correct classification rates when performing this analysis compared to the results obtained using the Swiss data in Section 3.3. We observe that the overall improvement in accuracy is relatively low, amounting to from -0.1 to 1.8 percentage points depending on the algorithm. The increase is the highest for the lasso but slightly negative for the random forest. Furthermore, we notice that the predictive performance increases more for the competitive coalitions (from 0.6 to 2.9 percentage points) while remaining stable for the collusive coalitions (from -0.8 to 0.7 percentage points). As the overall change in the goodness of fit for the four algorithms is quite low, we assume that the gain of additional coalition-based screens is negligible.

Figure 3.4.1: Variable importance plot for the Swiss cartels with more predictors. We compute the variable importance using the mean decrease in the Gini index and express it relative to the maximum.



By looking at the three most important predictors according to the random forest, we find the medians for the coefficient of variation, the spread, and the KS statistic remain the most predictive coalition-based screens (see Figure 3.4.1). The upper and lower quartiles of these descriptive statistics appear in the top-ten best predictors, but rather not at the top of the ranking.

3.4.2 The investigation of predictors measuring asymmetry

In the three different countries, predictors measuring asymmetry do not appear to be important (according to the random forest) in flagging collusive coalitions when also implementing coalition-based screens related to the variance and the uniformity of bids. This result might be puzzling when we remember that Imhof (2019), Huber, Imhof, and Ishii (2020) and Wallimann, Imhof, and Huber (2020) find screens for the asymmetry to be relevant in predicting the Japanese and Swiss bid-rigging cartels. In fact, asymmetry in the distribution of bids arises when we simultaneously analyze the bids from the winner designated by the cartel and the cover bids submitted by other cartel participants. However, in our coalition-based approach, we select only three bidders and thus not necessarily the designated winner. Therefore, the absence of the designated winner in calculating the tender-based screens with only three cartel participants can limit the predictive power of screens based on the asymmetry of bids.

Table 3.6: Changes in accuracy when only considering screens for the asymmetry as predictors

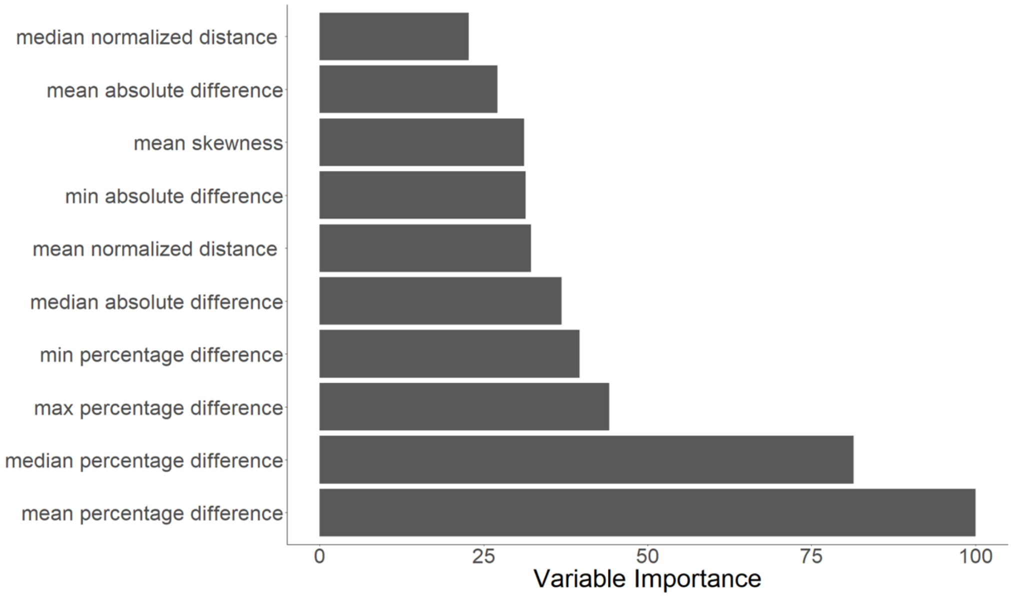
Classifier	Changes in percentage points		
	CCR	CCR collusion	CCR competition
Lasso	-2.9	-2.2	-1.3
Random forest	-2.7	-1.4	-4.1
Super learner	-2.3	-1.8	-2.7
Support vector machines	-3.2	-2.3	-3.4

Note: 'CCR' denotes the correct classification rate, 'CCR collusion' the correct classification rate of the collusive coalitions, and 'CCR competition' the correct classification rate of competitive coalitions.

To further investigate the importance of these screens, in a second complementary analysis, we discard screens for the variance and the uniformity of bids. We then repeat our estimation procedure for the Swiss data. Table 3.6 reports that correct classification rates decrease by 2.3 to 3.2 percentage points. Figure 3.4.2 shows that summary statistics for the percentage difference are the most predictive coalition-based screens. However, these coalition-based screens are related to the variance of the bids if they do not include the designated winner's bid. If the variance is reduced for the losing bids, and if one takes into account mainly the losing bids in calculating the tender-based screens, then the coalition-based screens for the percentage difference will be smaller for collusive coalitions than for competitive coalitions. In fact, a look at the descriptive statistics indicates that the mean of the Swiss cartels' medians of the percentage difference for the collusive coalitions amounts to 3.3, as opposed to 5.6 for the competitive coalitions.

Therefore, in a second step, we discard coalition-based screens for the percentage difference and the absolute difference since they might be related to the variance screens to analyze only screens for the asymmetry of the bids. Using sixteen predictors for the asymmetry of bids, we obtain a considerable decrease in the correct classification rates of from 17.3 to 20.5 percentage points (see Table 3.7). The decline is less for collusive coalitions (from 11.8 to

Figure 3.4.2: Variable importance plot for the Swiss cartels with only screens for the asymmetry of bids. We compute the variable importance using the mean decrease in the Gini index, and express it relative to the maximum.



20.0 percentage points) but still large. In conclusion, coalition-based screens for asymmetry do not seem to be important for flagging collusive coalitions. Therefore, the variance and the uniformity of bids remain the most important features for describing changes in the distributional pattern of the bids in collusive coalitions.

3.4.3 Coalition-based screens with four bidders

In the last step, we investigate the correct classification rate by forming coalitions of four firms, not three. However, it is more advisable to compute coalitions of three firms since it allows bid-rigging cartels formed with three bidders to be uncovered. Using coalitions with four firms makes it difficult to detect bid-rigging cartels formed with three bidders and therefore restricts the broader scope of applying our suggested method based on coalitions. Moreover, calculating coalitions based on four firms might be somewhat more intense computationally. For example, if we calculate all possible coalitions of three firms formed with 75 firms, we obtain 67,525 potential coalitions to calculate. With coalitions of four firms for 75 firms, the potential coalitions amount to 1,215,450, or eighteen times the number of potential coalitions with three firms.

Table 3.7: Changes in accuracy when only considering screens for the asymmetry as predictors and discarding screens for the percentage difference and the absolute difference

Classifier	Changes in percentage points		
	CCR	CCR collusion	CCR competition
Lasso	-17.7	-15.2	-18.7
Random forest	-21.2	-19.8	-23.9
Super learner	-20.5	-20.0	-23.0
Support vector machines	-18.0	-11.8	-24.2

Note: 'CCR' denotes the correct classification rate, 'CCR collusion' the correct classification rate of the collusive coalitions, and 'CCR competition' the correct classification rate of competitive coalitions.

Nonetheless, we calculate coalitions of four firms using the Swiss data, which from a computation perspective, is easier since there are three datasets for each cartel with a lower number of firms than the other cartels in Italy and Okinawa. We end up with a total of 3,207 coalitions of four firms, 2,097 collusive, and 1,110 competitive coalitions. We then use the same coalition-based screens as in Section 3.3 to recapitulate the changes in percentage points for the correct classification rates (correct classification rates for coalitions of four firms minus correct classification rates for coalitions of three firms from Table 3.2). We find that the overall correct classification rates for coalitions of four firms are a little higher, with an increase of from 2.1 to 4.0 percentage points (see Table 3.8), compared to the correct classification rates for three coalitions. It seems that the increase is mainly driven by an increase in the correct classification rates of collusive coalitions amounting to from 5.1 to 6.1 percentage points. The correct classification rates for competitive coalitions at the opposite fall by 1.8 to 2.3 percentage points except for the super learner (increase of 0.9 percentage points). We also observe that coalition-based screens for asymmetry in the form of the skewness of the bids appear in the top-ten predictors according to the random forest with the coefficient of variation. Including a higher number of firms in the coalition reduces the likelihood that the winning bids will be omitted and therefore includes a greater distance between the first and second lowest bids in the coalition, naturally leading to more asymmetry. The fact that coalition-based screens for the asymmetry of bids have a higher predictive power could explain the overall rise in the correct prediction rates, specifically those of the collusive coalitions. The increase in the correct classification rates for coalitions with four firms also appears unsurprising since there were four or more cartel participants in most tenders in the Swiss data.

Table 3.8: Changes in accuracy when using coalition-based screens with four bidders in the Swiss data

Classifier	Changes in percentage points		
	CCR	CCR collusion	CCR competition
Lasso	3.4	5.4	-1.8
Random forest	2.1	5.1	-2.3
Super learner	4.0	5.9	0.9
Support vector machines	3.6	6.1	-1.9

Note: 'CCR' denotes the correct classification rate, 'CCR collusion' the correct classification rate of the collusive coalitions, and 'CCR competition' the correct classification rate of competitive coalitions.

3.5 Policy recommendations

3.5.1 Advantages of a coalition-based detection method

The coalition-based approach proposed in this paper has several advantages in flagging bid-rigging cartels. We first reach correct classification rates of 90% with the super learner in Italy, Japan, and Switzerland. In other words, we classify nine coalitions out of ten correctly on average. This result remains stable while considering different auction formats, i.e., the first-price sealed-bid and the average bid auction. Super learner outperforms the other algorithms in two out of three cases and does not exhibit an imbalance in predicting both classes (collusive and competitive coalitions). Its greater performance derives from the use of multiple machine learning models, for which the algorithm creates an optimally weighted average.[5] Super learner is then advisable in our case. Besides, the machine learning literature is rapidly growing, and we assume that future research implementing novel machine learning algorithms will increase accuracy.

Moreover, our coalition-based approach directly flags firms as cartel participants and can detect complete and incomplete bid-rigging cartels. If correctly

[5] See also https://cran.r-project.org/web/packages/SuperLearner/vignettes/Guide-to-SuperLearner.html (accessed 30 April 2021).

calibrated, it can also flag partial cartels, that is, complete or incomplete bid-rigging cartels active in one specific area or one specific type of contract (see, for example, Imhof, Karagök, and Rutz, 2018, Abrantes-Metz, Froeb, Geweke, and Taylor, 2006). In our cases, the bid-rigging cartels in Japan and Switzerland are complete for most tenders, but the Italian bid-rigging cartels are not. Identifying sub-groups of firms as cartel participants is important because markets are not always characterized by bid-rigging conspiracies affecting all contracts or involving all the firms. Therefore, our approach is not only applicable to different countries or auction formats but also to different kind of bid-rigging cartels. Such possible broad applications render the coalition-based approach attractive for screening procurement markets and future research.

Finally, the data requirement is low, as we need only the bids and the identity of the firms to calculate coalition-based screens. Other tender-based screens, such as those dealt with in Huber and Imhof (2019) or Wallimann, Imhof, and Huber (2020), do not require the bidders to be identified. Such low data requirements contrast with other methods of detection, which need cost-related variables or firm-specific covariates to implement the econometric tests, as suggested in Bajari and Ye (2003) or more recently in Conley and Decarolis (2016). A low data requirement is crucial for two reasons. First, it allows the screening of large procurement datasets. If the data are available in a digital form, a competition or procurement agency can apply the detection method in a minimum amount of time. Second, it could be challenging to obtain information specific to firms without attracting the cartel's attention to a possible investigation. Indeed, in some cases, it could destabilize the cartel and have a preventive effect. However, cartel participants will undoubtedly take more precautions and destroy evidence impeding the success of a future investigation.

3.5.2 Ex-ante Screening

When screening procurement markets, we suggest two different possibilities. The first consists of using data from previous cartels to fit predictive models (with machine learning algorithms) to apply them to a new dataset for which no prior information on collusion exists. The second possibility is to use benchmarks to isolate groups of suspicious contracts or firms. For the latter possibility, Table 3.4 in Section 3.3 might offer a starting point for screening procurement markets.

However, for both possibilities, one should be aware that the institutional context of each country – for example, the choice of the auction format or other country-specific characteristics – largely influences the distribution of bids in each tender. Coalition-based screens thus exhibit dissimilar values across countries and classes. For example, the values of the coefficients of variation for the Swiss collusive coalitions exhibit slightly higher average values than the Japanese competitive coalitions (see Table 3.4). Therefore, training models in one country to be able to test them in another could, in such circumstances, be hazardous, as already noted by Huber, Imhof, and Ishii (2020). Nonetheless, the effects of bid rigging go in the same direction, and their magnitudes might be similar in some cases. Hence, if a competition agency intends to apply the method to a different market or country, we recommend using benchmarks based on the effect of bid rigging rather than benchmarks based on the level of the screens. For example, a decrease by two in the variance on a market could be suspicious and should be subjected to further statistical inquiry to confirm the initial diagnostic. Moreover, further research should investigate the possibility of normalizing bids or screens by country to enable predictive models to be transferred directly from one country to another.

A competition agency can implement both predictive tender-based and coalition-based screens to fit models or assess approximate benchmarks. If the amount of data to screen is small (e.g., fewer than a hundred firms bidding in the data), one can directly apply the coalition-based approach. However, if the amount of data to screen is large (e.g., more than a thousand firms bidding in the data), the tender-based approach of Huber and Imhof (2019) or Wallimann, Imhof, and Huber (2020) is simpler to apply (e.g., less computationally intensive) in order to identify markets for specific products or different geographical areas that are potentially suspicious.

To increase the confidence level, a competition agency could also combine both types of screens, i.e., tender-based and coalition-based. Once a bench of suspicious tenders with the tenders-based screens has been identified, one can apply the coalition-based screens to verify whether the firms participating in the suspicious tenders are sufficiently suspect to open an investigation. Such a double testing procedure increases the reasonable grounds for identifying bid-rigging conspiracies and offers greater confidence to competition agencies in screening procurement markets. Here the coalition-based approach will provide precious assistance because it allows the identification of potential cartel participants to be confirmed with a high degree of confidence: the correct prediction rates in the three different countries indicate that nine firms out of ten are correctly classified as being competitors or cartel participants. In other words, a firm flagged as potentially collusive using the coalition-based approach has a 90% likelihood of being a cartel participant. The level of likelihood should be sufficiently high to constitute reasonable grounds for opening an investigation.

3.6 Conclusion

Our paper contributes to the literature on cartel detection in manifold ways. We have developed an original detection method based on screens by focusing on coalitions. This approach allows a broader application by detecting complete and incomplete bid-rigging cartels and partial cartels in different auction formats. Coalition-based screens delivered more correct classification rates than previous methods using tender-based screens. Using the super learner, we correctly classified on average at least nine coalitions out of ten in Italy, Japan, and Switzerland. The performance of the super learner surpassed other algorithms and is balanced across collusive and competitive coalitions. It thus remained the most suitable algorithm for our application.

Although an increase in the performance of three to ten percentage points might appear low, the coalition-based screens reduced the error rate by half in some cases. Such falls in the error rate are desirable given the heavy legal and procedural consequences for firms that have been flagged as potential cartel participants. Furthermore, the coalition-based screens do not oppose the tender-based screens but constitute an interesting complement limiting the risk of false positives in screening procurement markets.

Furthermore, we found that the levels of the most important coalition-based screens differ considerably between countries, though the magnitude of the effects of bid rigging is similar. Thus, a decrease by a factor of two in the median of the coefficient of variation and the spread, as well as an increase by a factor of two in the median of the KS statistic, could indicate potential collusion. Future empirical research should investigate the possibility of normalizing screens or bids by country or market to continue developing a general screening method that is both the most reliable and has the broadest applicability. In addition, future theoretical research should focus on structural models explaining why bid rigging reduces the variance and renders the distribution of bids less uniform than in competitive tenders.

Chapter 4

Business analytics meets artificial intelligence

Assessing the demand effects of discounts on Swiss train tickets

joint with **Martin Huber** and **Jonas Meier***

Abstract

We assess the demand effects of discounts on train tickets issued by the Swiss Federal Railways, the so-called 'supersaver tickets', based on machine learning, a subfield of artificial intelligence. Considering a survey-based sample of buyers of supersaver tickets, we investigate which customer- or trip-related characteristics (including the discount rate) predict buying behavior, namely: booking a trip otherwise not realized by train, buying a first- rather than second-class ticket, or rescheduling a trip (e.g. away from rush hours) when being offered a supersaver ticket. Predictive machine learning suggests that customer's age, demand-related information for a specific connection (like departure time and utilization), and the discount level permit forecasting buying behavior to a certain extent. Furthermore, we use causal machine learning to assess the impact of the discount rate on rescheduling a trip, which seems relevant in the light of capacity constraints at rush hours. Assuming that (i) the discount rate is quasi-random conditional on our rich set of characteristics and (ii) the buying decision increases weakly monotonically in the discount rate, we identify the discount rate's effect among 'always buyers',

*Chapter 4 is based on a working paper. It is published as Huber, Meier, and Wallimann (2021). We are grateful to the SBB Research Fund for financial support.

who would have traveled even without a discount, based on our survey that asks about customer behavior in the absence of discounts. We find that on average, increasing the discount rate by one percentage point increases the share of rescheduled trips by 0.16 percentage points among always buyers. Furthermore, investigating effect heterogeneity across observables suggests that the effects are higher for leisure travelers and during peak hours when controlling several other characteristics.

4.1 Introduction

Organizing public transport involves a well-known trade-off between consumer welfare and provider revenue. Typically, consumers value frequency, reliability, space, and low fares (Redman, Friman, Gärling, and Hartig, 2013) while suppliers aim at operating with a minimum number of vehicles to maximize profits. In general, the allocation can be improved as providers do not account for the positive externalities on consumers (Mohring, 1972). In particular, service frequency reduces travelers' access and waiting costs. This so-called 'Mohring-effect' leads to economies of scale, implying the need for subsidies to achieve the first-best solution in terms of welfare. Consequently, it may be socially optimal to subsidize railway companies to reduce fares (Parry and Small, 2009). To assess such a measure's effectiveness on demand, policy-makers would need to know how individuals respond to lower fares. However, it is generally challenging to identify causal effects of discounts on train tickets (or goods and services in general) due to confounding or selection. For instance, discounts might typically be provided for dates or hours with low train utilization such that connections with and without discount are not comparable in terms of baseline demand. A naive comparison of sold tickets with and without discount would therefore mix the influence of the discount with that of baseline demand. In this context, we apply machine learning (a subfield of artificial intelligence) to convincingly assess how discounts on train tickets for long-distance connections in Switzerland, the so-called 'supersaver tickets', affect demand by exploiting a unique data set of the Swiss Federal Railways (SBB) that combines a survey of supersaver buyers with train utilization records.

More concisely, our study provides two use cases of machine learning for business analytics in the railway industry: (i) Predicting buying behavior among supersaver customers, namely whether customers booked a trip otherwise not realized by train (additional trip), bought a first-class rather than a second-class ticket (upselling), or rescheduled their trip e.g. away from rush hours (demand shift); (ii) analysing the causal effect of the discount on demand shifts among customers that would have booked the trip even without a discount. This is feasible because our unique survey contains information on how supersaver buyers would have decided in the absence of a discount, e.g., whether they are so-called 'always buyers' and would have booked the connection even at the regular fare. For both prediction and causal analysis, we make use of appropriately tailored machine learning techniques, which learn the associations between the demand outcomes of interest, the discount rate, and further customer or trip-related characteristics in a data-driven way and helps avoiding model misspecification. Such a targeted combination of predictive and causal machine learning can therefore improve demand forecasting and decision-making in companies and organizations. While predictive machine learning permits optimizing forecasts about demand and customer behavior as a function of observed characteristics, causal machine learning permits evaluating the causal effect of specific interventions like a discount regime for optimizing the offer of such discounts. Concerning the prediction task, we use the so-called random forest, see Breiman (2001), as machine learner to forecast the supersaver customers' behavior and obtain accuracy or correct (out of sample) classification rate of 58% (demand shift), 65% (additional trip), and 82% (upselling), respectively. Trip-related characteristics like seat capacity, utilization, departure time, and the discount rate, but also customer's age turn out to be strong predictors.

Concerning the causal analysis (which is more challenging than mere prediction), we impose (i) a selection-on-observables assumption implying that the discount rate is as good as randomly assigned when controlling for our rich set of trip- and demand-related characteristics and (ii) weak monotonicity of any individual's decision to purchase an additional trip (otherwise not realized) in the discount rate, implying that a higher (rather than lower) discount does either positively or not affect any customer's buying decision. As a methodological contribution, we formally show how these assumptions permit tackling the selectivity of discount rates and survey responses to identify the discount rate's effect on demand shifts (rescheduling away from rush hours) for the subgroup 'always buyers', based on the survey information on how customers would have behaved in the absence of a discount. Furthermore, we discuss testable implications of monotonicity, namely that among all survey respondents, the share of additional trips must increase in the discount rate, and the selection on observables assumptions, requiring that conditional on trip- and demand-related characteristics, the discount must not be associated with personal characteristics (like age or gender) among always buyers. Hypothesis tests do not point to the violation of these implications.

Based on our causal identification strategy, we estimate the marginal effect of slightly increasing the (continuously distributed) discount rate based on the causal forest (CF), see Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019), and find that on average, increasing the discount rate by one percentage point increases the share of rescheduled trips by 0.16 percentage points among always buyers. In a second approach, we binarize the discount rates by splitting them into two discount categories of less than 30% (relative to the regular fare) and 30% or more. Applying double machine learning (DML), see Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018), we find that discount rates of 30% and more on average increase the share of rescheduled trips 3.6 percentage points, which is in line with the CF-based results. Therefore, our paper provides the first empirical evidence (at least for Switzerland) that such discounts can help balance out train utilization across time and reduce overload during peak hours, albeit the magnitude of the impact on always buyers appears limited.

When investigating the heterogeneity of effects across all of our observed characteristics using the CF, our results suggest that demand-related trip characteristics (like seat capacity, utilization, departure time, and distance) have some predictive power for the size of the discounts' impact on shifting demand. Such information on heterogeneous effects appears interesting for optimizing the allocation of discounts for the purpose of shifting demand, as the SBB has (due to its monopoly in the Swiss long-distance passenger rail market) agreed with the Swiss price monitoring agency to provide a fixed amount of discounted tickets per year, but is free to chose the timing and connections for discounts. In a second heterogeneity analysis, we investigate whether effects differ systematically across a pre-selected set of characteristics, namely: age, gender, possession of a half fare travel card, travel distance, whether the purpose is business, commute, or leisure, and whether the departure time is during peak hours. Using the regression approach of Semenova and Chernozhukov (2020), we find that conditional on the other characteristics, the effects of increasing the discount by one percentage point on rescheduling are by more than 0.2 percentage points higher during peak hours and for leisure travelers, differences that are statistically significant at the 10% level when, however, not controlling for multiple hypothesis testing. These effects appear plausible as leisure travelers are likely more flexible and discounts during peak hours make trips at times of increased demand even more attractive. We do not find statistically significant effect differences for the other pre-selected characteristics, which could, however, be due to the (for the purpose of investigating effect heterogeneity) limited sample of several thousand observations.

Our paper is related to a growing literature applying statistical and machine learning methods for analyzing transport systems, as well as to methodological studies on causal inference for so-called principal strata, see

Frangakis and Rubin (2002), i.e., endogenous subgroups like the always buyers. Typically, it is hard to identify the causal effect of some treatment (or intervention) like a discount on such a non-randomly selected subgroup defined in terms of how a post-treatment variable (e.g. buying decision) depends on the treatment (e.g. treatment). One approach is to give up on point identification and instead derive upper and lower bounds on a set of possible effects for groups alike the always buyers based on the aforementioned monotonicity assumption (and possibly further assumptions about the ordering of outcomes of always buyers and other individuals), see for instance Zhang and Rubin (2003), Zhang, Rubin, and Mealli (2009), Imai (2008), Lee (2009), and Blanco, Flores, and Flores-Lagunes (2011). Alternatively, the treatment effect on always buyers is point-identified when invoking a selection-on-observables or instrumental variable assumption for selection into the survey, see for instance Huber (2014), which requires sufficiently rich data on both survey participants and non-participants for modeling survey participation. In contrast to these previous studies, the approach in this paper point-identifies the treatment effect by exploiting the rather unique survey feature that customers were asked about their behavior in the absence of the discount, which under monotonicity permits identifying the principal stratum of always buyers directly in the data.

Furthermore, our work is related to conceptual studies on transport systems, considering, for instance, the previously mentioned positive externalities of an increased service for customers that are not accounted for by transportation providers. Such externalities typically arise from economies of scale due to fixed costs and a 'Mohring effect', implying that service frequency reduces waiting costs (Mohring, 1972). The study by Parry and Small (2009) suggests that lower fares can boost overall welfare by increasing economies of scale (off-peak) and decreasing pollution and accidents (at peaks). Similarly, De Palma, Lindsey, and Monchambert (2017) argue that time-dependent ticket prices may increase overall welfare as overcrowding during peak hours is suboptimal for both consumers and providers. As public transport is usually highly subsidized, governments may directly manage the trade-off mentioned above. As this involves taxpayer money, it is a question of general interest how the subsidies should be designed. Based on their results, Parry and Small (2009) conclude that even substantial subsidies are justified due to lower fares' positive welfare effect. In contrast, Basso and Silva (2014) find that the contribution of transit subsidies to welfare diminishes once congestion is taxed and alternatives are available, i.e., bus lanes. Irrespective of the specific policy instrument, the consumer's willingness to shift demand drives these policies' effectiveness. While many factors affect this willingness, most studies conclude that consumers are price sensitive (Paulley, Balcombe, Mackett, Titheridge, Preston, Wardman, Shires, and White, 2006). In this context, we aim at contributing to a better understanding of how time-dependent pricing translates to consumer decisions.

More broadly, our paper relates to the literature on policies targeting demand shifts. Among these, the setting of car parking costs, fiscal regulations, or even free public transport has been analyzed (e.g. Batty, Palacin, and González-Gil, 2015, Rotaris and Danielis, 2014, Zhang, Lindsey, and Yang, 2018, De Witte, Macharis, Lannoy, Polain, Steenberghen, and Van de Walle, 2006). Another stream of literature applies machine learning algorithms in the context of public transport. Examples are short-term traffic flow forecasts for bus rapid transit (Liu and Chen, 2017) or metro (Liu, Liu, and Jia, 2019) services. Further, Hagenauer and Helbich (2017) and Omrani (2015) implement machine learning algorithms to predict travel mode choices. Yap and Cats (2020) predict disruptions and their passenger delay impacts for public transport stops. In other research fields, also applications of causal (rather than predictive) machine learning are on the rise (see for instance Yang, Chuang, and Kuan, 2020, Knaus, 2021). This is, to the best of our knowledge, the first study using causal machine learning in the context of public transport. Finally, a growing literature discusses the opportunities of data-driven business decision-making (Brynjolfsson and McElheran, 2016) by assessing the relevance of predictive and causal machine learning. Ascarza (2018) and Hünermund, Kaminski, and Schmitt (2021) show that companies may gain by designing their policies based on causal machine learning. For instance, firms can target the relevant consumers much more effectively when accounting for their heterogeneity in terms of reaction to a treatment. Our study provides a use case of how the machine learning-based assessment of discounts could also be implemented in other businesses and industries facing capacity constraints.

This paper proceeds as follows. Section 4.2 presents the institutional setting of passenger railway transport in Switzerland. Section 4.3 describes our data, coming from a unique combination of a customer survey and transport utilization data. Section 4.4 discusses the identifying assumptions underlying the causal machine learning approach as well as testable implications. Section 4.5 outlines the predictive and causal machine learning methods. Section 4.6 presents the empirical results. Section 4.7 concludes.

4.2 Institutional background

The railway system in Switzerland is known for its high quality of service. Examples include the high level of system integration with frequent services, synchronized timetables, and comprehensive fare integration, see Desmaris (2014). In Switzerland, a country of railway tradition, the state-owned incumbent Swiss Federal Railways (SBB) operates the long distance passenger rail market as monopolist (Thao, von Arx, and Frölicher, 2020). Furthermore, nationally operating long-distance coaches may only be approved if they do not 'substantially' compete with existing services. Thus, the SBB competes exclusively with motorized private transport in Swiss long-distance traffic. The company also owns most of the rail infrastructure, which the Federal Government funds. However, since the end of 2020, the companies Berne-Lötschberg-Simplon Railways (BLS) and Southeast Railways (SOB) operate a few links on behalf of the SBB. Different to regional public transport that Swiss taxpayers subsidize with approximately CHF 1.9 bn per year, the operation of the long distance public transport itself has to be self-sustaining (Wegelin, 2018).

Because of the monopoly position of the SBB in long distance passenger transport, the prices are screened by the Swiss 'price watchdog' (or price monitoring agency) to prevent abuse. Based on the price monitoring act, the watchdog keeps a permanent eye on how prices and profits develop. By the end of 2014, the watchdog concluded that the SBB charged too high prices. As a consequence, and through a mutual agreement, the SBB and the Swiss price watchdog agreed on a significantly higher supply of supersaver tickets, which were first offered in 2009. Using a supersaver ticket, customers can travel on long distance public transport routes with a discount of up to 70%. Thereafter, additional agreements were regularly reached regarding the number and scope of the supersaver tickets. While only a few thousand supersaver tickets were sold in 2014, sales increased to about 8.8 million in 2019, see Lüscher (2020).

From the SBB's perspective, these tickets can serve two purposes. First, the tickets might be used as means to balance out the utilization of transport services. For instance, supersaver tickets could reduce the high demand during peak hours which is a key challenge for public transport. Thus, balancing the demand may reduce delays and increase the number of free seats, which is valued by the consumers. The average load of SBBs' seats amounts to 30% in the long distance passenger transport.[1] For this reason, there is, in the literal sense, room for improving the allocation. Second, price sensitive customers can be acquired during off-peak hours at rather negligible marginal costs.

Despite the increasing interest in the supersaver tickets in recent years, many users of the Switzerland public transport network purchase a so-called 'general abonnement' travel ticket (GA). This (annually renewed) subscription

[1] See https://reporting.sbb.ch/verkehr(assessed on March 24, 2021).

provides free and unlimited access to the public transport network in Switzerland. In 2019, about 0.5 million individuals owned a GA in Switzerland, roughly 6% of the Swiss population. The GA's cost amounts to 3,860 and 6,300 Swiss francs for the second and first class, respectively. In the same year, about 2.7 million individuals held a relatively cheap half fare travel ticket amounting to 185 Swiss francs. The latter implies a price reduction of 50% for public transport tickets in Switzerland. Overall, discounts provided through supersaver tickets are slightly lower for owners of half fare tickets, as the SBB aims to attract non-regular public transport users. In our causal analysis, we therefore also control for the possession of a half fare ticket.

4.3 Data

To investigate supersaver tickets' effect, we use a unique cross-sectional data set provided by the SBB. Our sample consists of randomly surveyed buyers of supersaver tickets that purchased their tickets between January 2018 and December 2019. These survey data are matched with data on distances between any two railway stops as well as utilization-related information relevant for the supply and calculation of discounts. In section 4.6, we provide descriptive statistics for these data.

4.3.1 Survey data

The customer survey is our primary data source. It, for instance, includes the outcome variable 'demand shift', a binary indicator of whether an interviewee rescheduled her or his trip due to buying a supersaver ticket. 'Yes' means that the departure time has been advanced or postponed because of the discount. A second variable characterizing customer behavior is an indicator for upselling, i.e., whether someone purchased a first rather than a second-class ticket as a reaction to the discount. Another question asks whether an interviewee would have bought the train trip in the same or a higher class even without being offered a discount, which permits judging whether an additional trip has been sold through offering the discount and allows identifying the subgroup of always buyers under the assumptions outlined further below. Our continuously distributed treatment variable is the discount rate of a supersaver ticket relative to the standard fare, which may take positive values of up to 70%.

Furthermore, we observe two kinds of covariates, namely trip- or demand-related factors and personal characteristics of the interviewee. The former are important control variables for our causal identification strategy outlined below and include the difference between the days of purchase and travel, the weekday, month, and year, an indicator for buying a half fare ticket, departure time, peak hour,[2] number of tickets purchased per person, class (first or second), indicators for leisure trips, commutes, or business trips, the number of companions (by children and adults if any) and a judgment of how complicated the ticket purchase was on a scale from 1 (complicated) to 10 (easy). Furthermore, it consists of indicators for the point of departure, destination, and public holidays. The personal characteristics include age, gender, migrant status, language (German, French, Italian), and indicators for owning a half fare travel ticket or other subscriptions like those of regional tariff associations, specific connections, and Gleis 7 ('rail 7'). The latter is a travelcard for young adults not older than 25, providing free access to public transport after 7pm.

4.3.2 Factors driving the supply of supersaver tickets

In addition to the survey, we have access to factors determining the supply of supersaver tickets with various discounts. This is crucial for our causal analysis that hinges on controlling for all characteristics jointly affecting the discount rate and the demand shift outcome. While the information on the distances between railway stops in Switzerland is publicly available,[3] the SBB provides us for the various connections with information on utilization data, the number of offered seats, and contingency schemes, which define the quantity of offered discounts. This allows us to account for travel distance, offered seats, capacity utilization, and quantities of offered supersaver tickets for various discount levels, as well as quantities of supersaver tickets already sold (both quantities at the time of purchase). Furthermore, we create binary indicators for the 27 different contingency schemes of the SBB present in our data, which change approximately every month.

The variables listed in the previous paragraph are important, as the SBB calculates the supply of supersaver tickets based on an algorithm considering four types of inputs: Demand forecasts, advance booking deadlines, number of supersaver tickets already sold, and contingency schemes defining the amount and the size of offered discounts based on the three previous inputs.

[2] Peak hour is defined as a departure time between 6am and 8:59am or between 4pm and 18:59pm, from Monday to Friday. These time windows are chosen on the base of the SBB's train-path prices. For further details, see https://company.sbb.ch/en/sbb-as-business-partner/services-rus/onestopshop/services-and-prices/the-train-path-price.html (assessed on March 24 2021).

[3] See the Open Data Platform of the SBB: https://data.sbb.ch/explore/dataset/linie-mitbetriebspunkte (accessed on March 24, 2021).

The schemes are set as a function of the SBB's self-imposed goals, such as customer satisfaction but also depend on the requirements imposed by the price watchdog. The algorithm calculates a journey's final discount as a weighted average of all discounts between any two adjacent railway stops along a journey. The weights depend on the distances of the respective subsections of the trip. To approximate the (not directly available) demand forecasts of the SBB, we consider the quarterly average of capacity utilization and the number of offered seats for any two stops, which are available by (exact) departure time, workday, class, and weekend. In addition, we make use of indicators for place of departure, destination, month, year, weekday, and public holidays. We use this information to reconstruct the amount and size of offered discounts by taking values from the contingency schemes that correspond to our demand forecast approximation combined with the difference between buying and travel days. Comparing this amount and size of offered discounts with a buyer's discount, we estimate the number of supersaver tickets already sold for the exact date of purchase.

4.3.3 Sample construction

Our initial sample contains 12,966 long-distance train trips that cover 61,469 sections between two adjacent stops. For 12.2% of these sections, there is no information on the capacity utilization available, which can be due to various reasons. First, for some cases, capacity utilization data is missing. Second, passengers traveling long-distance may switch to regional transport in exceptional cases causing problems for determining utilization. A further reason could be issues in data processing. Altogether, missing information occurs in 3,967 trips of our initial sample. We tackle this problem by dropping all journeys with more than 50% of missing information, which is the case for 320 trips or 2.5% of our initial sample. After this step, our evaluation sample consists of 12,646 trips. For the remaining 3,647 trips with missing information (which now account for a maximum of 50% of all sections of a journey), we impute capacity utilization as the average of the remaining sections of a trip. In our empirical analysis, we include an indicator for whether some trip information has been imputed as well as the share of imputed values for a specific trip as control variables. Finally, we note that our causal analysis makes (in contrast to the predictive analysis) only use of a subsample, namely observations identified as always buyers who would have traveled even without a discount, all in all, 6,112 observations.

4.4 Identification

We subsequently formally discuss the identification strategy and assumptions underlying our causal analysis of the discounts among always buyers.

4.4.1 Definition of causal effects

Let D denote the continuously distributed treatment 'discount rate' and Y the outcome 'demand shift', a binary indicator for rescheduling a trip due to being offered a discount. More generally, capital letters represent random variables in our framework, while lower case letters represent specific values of these variables. To define the treatment effects of interest, we make use of the potential outcome framework, see for instance Rubin (1974). To this end, $Y(d)$ denotes the potential outcome hypothetically realized when the treatment is set to a specific value d in the interval $[0, Q]$, with 0 indicating no discount and Q indicating the maximum possible discount. For instance, $Q = 0.7$ would imply the maximum discount of 70% of a regular ticket fare. The realized outcome corresponds to the potential outcome under the treatment actually received, i.e. $Y = Y(D)$, while the potential outcomes under discounts different to one received remain unknown without further statistical assumptions.

A further complication for causal inference is that our survey data only consists of individuals that purchased a supersaver ticket, a decision that is itself an outcome of the treatment, i.e. the size of the discount. Denoting by S a binary indicator for purchasing a supersaver ticket and by $S(d)$ the potential buying decision under discount rate d, this implies that we only observe outcomes Y for individuals with $S = 1$. In general, making the survey conditional on buying introduces Heckman-type sample selection (or collider) bias, see Heckman (1976) and Heckman (1979), if unobserved characteristics affecting the buying decision S also likely affect the inclination of shifting the timing of the train journey Y. Furthermore, it is worth noting that $S = S(D)$ implies that buying a supersaver ticket is conditional on receiving a non-zero discount. For this reason, non-treated subjects paying regular fares (with $D = 0$) are not observed in our data. Yet, the outcome in our sample is defined relative to the behavior without treatment, as Y indicates whether a has passenger has changed the timing of the trip because of a discount. This implies that $Y(0) = 0$ by definition, such that the causal effect of some positive discount d vs. no discount is $Y(d) - Y(0) = Y(d)$ is directly observable among observations that actually received d. However, it also appears interesting to investigate whether the demand shift effect varies across different (non-zero) discount rates $d \in (0, Q]$ to see whether the size matters. This is complicated by the fact that supersaver customers with different discount rates that are observed in our data might in general differ importantly in terms of background characteristics also affecting the outcome, exactly because they bought their trip and were selected into the survey under non-comparable discount regimes. Our causal approach aims at tackling exactly this issue to establish customer groups that are comparable across discount rates in order to identify the effect of the latter.

Based on the potential notation, we can define different causal parameters of interest. For instance, the average treatment effect (ATE) of providing discount levels d vs. d' (for $d \neq d'$) on outcome Y, denoted by $\Delta(d, d')$, corresponds to

$$\Delta(d, d') = E[Y(d) - Y(d')]. \quad (4.4.1)$$

Furthermore, the average partial effect (APE) of marginally increasing the discount level at $D = d$, denoted by $\theta(d)$, is defined as

$$\theta(d) = E\left[\frac{\partial Y(D)}{\partial D}\right]. \quad (4.4.2)$$

Accordingly, $\theta(D)$ corresponds to the APE when marginally increasing the actually received discount of any individual (rather than imposing some hypothetical value d for everyone).

The identification of these causal parameters based on observable information requires rather strong assumptions. First, it implies that confounders jointly affecting D and Y can be controlled for by conditioning on observed characteristics. In our context, this appears plausible, as treatment assignment is based on variables related to demand (like weekdays or month), contingency schemes, capacity utilization, and supersaver tickets already sold - all of which is available in our data, as described in section (4.3). Second, identification requires that selection S is as good as random (i.e., not associated with outcome Y) given the observed characteristics and the treatment, an assumption known as missing at random (MAR), see for instance Rubin (1976) and Little and Rubin (1987). However, the latter condition appears unrealistic in our framework, as our data lack important socio-economic characteristics likely affecting preferences and reservation prices for public transport, namely education, wealth, or income. For this reason, we argue that the ATE and APE among the individuals selected for the survey ($S = 1$), i.e. conditional on buying a supersaver ticket, which are defined as

$$\Delta_{S=1}(d, d') = E[Y(d) - Y(d')|S = 1], \theta_{S=1}(D) = E\left[\frac{\partial Y(D)}{\partial D}\right], \quad (4.4.3)$$

cannot be plausibly identified either. The reason is that if an increase in the discount rate induces some customers to buy a super saver ticket, then buyers with lower and higher discounts will generally differ in terms of their average reservation prices and related characteristics (as education or income), which likely also affect the demand-shift outcome Y.

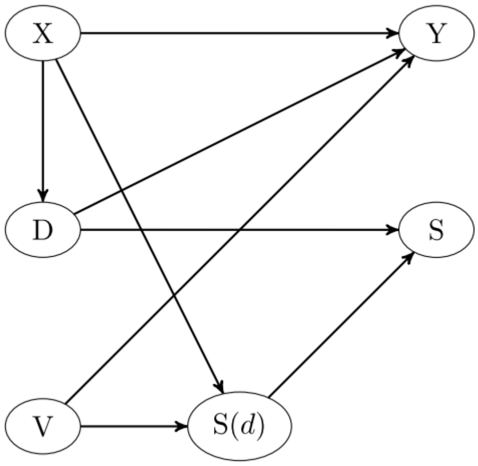
To tackle this sample selection issue, we exploit the fact that our data provide information on whether the supersaver customers would have purchased a ticket for this specific train trip also in the absence of any discount. Provided that the interviewees give accurate responses, we thus have information on $S(0)$, the hypothetical buying decision without treatment. Under the assumption that each customer's buying decision is weakly monotonic in the treatment in the sense that anyone purchasing a trip in a specific travel class (e.g., second class) without discount would also buy it for that class in the case of any positive discount, this permits identifying the group of always buyers. Importantly, we therefore define always buyers as those that would buy the trip not in a lower travel class (namely second rather than first class) without discount. For alway buyers, $S(0) = S(d) = 1$ for any $d > 0$, such that their buying decision is always one and thus not affected by the treatment, implying the absence of the selection problem. In the denomination of Frangakis and Rubin (2002), the always buyers constitute a so-called principal stratum, i.e., a subpopulation defined in terms of how the selection reacts to different treatment intensities. Therefore, sample selection bias does not occur within such a stratum, in which selection behavior is by definition homogeneous. For this reason, we aim at identifying the ATE and APE on the always buyers:

$$\begin{aligned}\Delta_{S(0)=1}(d,d') &= E[Y(d)-Y(d')|S(0)=1] \\ &= E[Y(d)-Y(d')|S(0)=S(d'')=1] \text{ for } d'' \in (0,Q], \\ \theta_{S(0)=1}(D) &= E\left[\frac{\partial Y(D)}{\partial D}\Big|S(0)=1\right] \\ &= E\left[\frac{\partial Y(D)}{\partial D}\Big||S(0)=S(d'')=1\right] \end{aligned} \tag{4.4.4}$$

where the second equality follows from the monotonicity of S in D that is formalized further below.

Figure 4.4.1 provides a graphical illustration of our causal framework based on a directed acyclic graph, with arrows representing causal effects. Observed covariates X that are related to demand are allowed to jointly affect the discount rate D and the demand-shift outcome Y. X may influence the potential purchasing decision under a hypothetical treatment $S(d)$, implying that buying a ticket given a specific discount depends on observed demand

Figure 4.4.1: Causal framework



drivers like weekday, month, etc. Furthermore, unobserved socio-economic characteristics V (like the reservation price) likely affect both $S(d)$ and Y. This introduces sample selection when conditioning on S, e.g. by only considering survey respondents ($S = 1$). We also note that S is deterministic in D and $S(d)$ (as $S = S(D)$), even when controlling for X. This is the case because conditional on $S = 1$, D is associated with V, which also affects Y, thus entailing confounding of the treatment-outcome relation. A reason for this is for instance that buyers under higher and lower discounts are generally not comparable in terms of their reservation prices. In the terminology of Pearl (2000), S is a collider that opens up a backdoor path between D and Y through V. Theoretically, this could be tackled by jointly conditioning on the potential selection states under treatment values d vs. d' considered in the causal analysis, namely $S(d), S(d')$, as controls for the selection behavior. This is typically not feasible in empirical applications when only the potential selection corresponding to the actual treatment assignment is observed, $S = S(D)$. In our application, however, we do have information on $S(0)$ and can thus condition on being an always buyer under the mentioned monotonicity assumption.

4.4.2 Identifying assumptions

We now formally introduce the identification assumptions underlying our causal analysis.

Assumption 1 (identifiability of selection under non-treatment):
$S(0)$, is known for all subjects with $S = 1$.

Assumption 1 is satisfied in our data in the absence of misreporting, as subjects have been asked whether they would have bought the train trip even in the absence of discount.

Assumption 2 (conditional independence of the treatment):
$Y(d), S(d) \perp D|X$ for all $d \in (0, Q]$.

By Assumption 2, there are no unobservables jointly affecting the treatment assignment on the one hand and the potential outcomes or selection states under any positive treatment value on the other hand conditional on covariates X. This assumption is satisfied if the treatment is quasi-random conditional on our demand-related factors X. Note that the assumption also implies that $Y(d) \perp D|X, S(0) = 1$ for all $d \in (1, Q]$.

Assumption 3 (weak monotonicity of selection in the treatment):
$\Pr(S(d) \geq S(d')|X) = 1$ for all $d > d'$ and $d, d' \in (1, Q]$.

By Assumption 3, selection is weakly monotonic in the treatment, implying that a higher treatment state can never decrease selection for any individual. In our context, this means that a higher discount cannot induce a customer to not buy a ticket that would have been purchased under a lower discount. An analogous assumption has been made in the context of nonparametric instrumental variable models, see Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996), where, however, it is the treatment that is assumed to be monotonic in its instrument. Note that monotonicity implies the testable implication that $E[S - S(0)|X, S = 1, D = d] = E[(1 - S(0)|X, S = 1, D = d]$ weakly increases in treatment value d. In words, the share of customers that bought the ticket because of the discount must increase in the discount rate in our survey population when controlling for X.

Assumption 4 (common support):
$f(d|X, S(0) = 1) > 0$ for all $d \in (1, Q]$.

Assumption 4 is a common support restriction requiring that $f(d|X, S(0) = 1)$, the conditional density of receiving a specific treatment intensity d given X and $S(0) = 1$ (or conditional probability if the treatment takes discrete values), henceforth referred to as treatment propensity score, is larger than zero among always buyers for the treatment doses to be evaluated. This implies that the demand-related covariates X do not deterministically affect the discount rate received such that there exists variation in the rates conditional on X.

Our assumptions permit identifying the conditional ATE given X (CATE), denoted by $\Delta_{X,S(0)=1}(d, d') = E[Y(d) - Y(d')|X, S(0) = 1]$ for $d \neq d'$ and

$d, d' \in (1, Q]$. To see this, note that

$$\begin{aligned}\Delta_{X,S(0)=1}(d,d') &= E[Y|D=d,X,S(0)=1] - E[Y|D=d',X,S(0)=1],\\ &= E[Y|D=d,X,S(0)=1,S=1]\\ &\quad - E[Y|D=d',X,S(0)=1,S=1],\end{aligned} \tag{4.4.5}$$

where the first equality follows from Assumption 2 and the second from Assumption 3, as monotonicity implies that asymptotically, $S=1$ if $S(0)=1$. Together with Assumption 1, which postulates the identifiability of $S(0)$, it follows that the causal effect on always buyers is nonparametrically identified, given that common support (Assumption 4) holds. If follows that the ATE among always buyers is identified by averaging over the distribution of X given $S(0)=1, S=1$:

$$\begin{aligned}\Delta_{S(0)=1}(d,d') &= E[E[Y|D=d,X,S(0)=1,S=1]\\ &-E[Y|D=d',X,S(0)=1,S=1]|S(0)=1,S=1].\end{aligned} \tag{4.4.6}$$

Furthermore, considering (4.4.5) and letting $d-d' \to 0$ identifies the conditional average partial effect (CAPE) of marginally increasing the treatment at $D=d$ given $X, S(0)=1$, denoted by $\theta_{X,S(0)=1}(D) = E\left[\frac{\partial Y(D)}{\partial D}\Big||X,S(0)=1\right]$:

$$\theta_{X,S(0)=1}(d) = \frac{\partial E[Y|D=d,X,S(0)=1,S=1]}{\partial D}. \tag{4.4.7}$$

Accordingly, the APE among always buyers that averages over the distributions of X and D is identified by

$$\theta_{S(0)=1}(D) = E\left[\frac{\partial E[Y|D,X,S(0)=1,S=1]}{\partial D}\right]. \tag{4.4.8}$$

Our identifying assumptions yield a testable implication if some personal characteristics (like customer's age) that affect $S(d)$ are observed, which we henceforth denote by W. In fact, D must be statistically independent of W conditional on $X, S(0)=1, S=1$ if X is sufficient for avoiding any cofounding of the treatment-outcome relation. To see this, note that personal characteristics must by Assumption 2 not influence the treatment decision conditional on X. This statistical independence must also hold within subgroups (or principal strata) in which sample selection behavior (and thus sample selection/collider bias) is controlled for like the always buyers, i.e. conditional on $S(d), S=1$.

4.5 Estimation based on machine learning

In this section, we outline the predictive and causal machine learning approaches used in our empirical analysis of the evaluation sample.

4.5.1 Predictive machine learning

Let $i \in \{1, ..., n\}$ be an index for the different interviewees in our sample of size n and $\{Y_i, D_i, X_i, W_i, S_i(0)\}$ denote the outcome, treatment, the covariates related to the treatment and the outcome, the observed personal characteristics, and the buying decision without discount of these interviewees that by the sampling design all satisfy $S_i = 1$ (because they are part of the survey). Therefore, Y_i represents customer i's demand shift (rescheduling behavior) under customer i's received discount rate D_i relative to no discount. We in a first step investigate which observed predictors among the covariates X, W as well as the size of the discount D are most powerful for predicting demand shifts by machine learning algorithms. We point out that this analysis is of descriptive nature as it does not yield the causal effects of the various predictors, but merely their capability of forecasting Y. In particular, our approach averages the predictions of Y over different levels of treatment intensity D and thus different customer types in terms of reservation price (related to $S(0)$) and unobserved background characteristics that likely vary with the treatment level.

Therefore, we also perform the prediction analysis within subgroups defined upon the treatment level to see whether the set of important predictors is affected by the treatment intensity. To this end, we binarize the treatment such that it consists of two categories, namely (non-zero) discounts below 30%, i.e. covering the treatment range $d \in (0, 0.3)$, and more substantial discounts of 30% and more, $d \in [0.3, 0.7]$, as 70% is the highest discount observed in our data. In the same manner, we also assess the predictive power when considering the decision to buy a trip that would not have been realized without discount (additional trip), i.e. $S_i - S_i(0)$, as outcome. As $S_i = 1$ is equal to one for everyone, the outcome corresponds to $1 - S_i(0)$ and indicates whether someone has been induced purchase the ticket because of the discount, i.e. is not an always buyer. As a further consumer behavior-related outcome to be predicted, we also consider buying a first-class rather than second-class ticket because of the discount (upselling).

Prediction is based on the random forest, a nonparametric machine learner suggested by Breiman (2001) for predicting outcomes as a function of covariates. Random forests rely on repeatedly drawing subsamples from the original data and averaging over the predictions in each subsample obtained by a decision tree, see Breiman, Friedman, Olshen, and Stone (1984). The idea of decision trees is to recursively split the covariate space, i.e. the set of possible values of X, W, into a number of non-overlapping subsets (or nodes). Recursive splitting is performed such that after each split, a statistical goodness-of-fit criterion like the sum of squared residuals, i.e. the difference between the outcome and the subset-specific average outcome, is minimized across the newly created subsets. Intuitively, this can be thought of as a regression of the outcome on a data-driven choice of indicator functions for specific (brackets of) covariate values. At each split of a specific tree, only a random subset of covariates is chosen as potential variables for splitting in order to reduce the correlation of tree structures across subsamples, which together with averaging predictions overall subsamples reduces the estimation variance of the random forest when compared to running a single tree in the original data. Even when using an excessive number of splits (or indicator functions for covariate values) such that some of them do not importantly predict the outcome, averaging over many samples will cancel out those non-predictive splits that are only due to sampling noise. Forest-based predictions can be represented by smooth weighting functions that bear some resemblance with kernel regression, with the important difference that random forests detect predictive covariates in a data-driven way. We use the *randomforest* package by Liaw and Wiener (2018) for the statistical software R to implement the random forest based on growing 1,000 decision trees.

4.5.2 Causal machine learning

Our second part of the analysis assesses the causal effect of increasing discount rates on demand shifts among always buyers while controlling for the selection into the survey and the non-random assignment of the treatment based on Assumptions 1 to 4 of section 4.4. We apply the causal forest (CF) approach of Wager and Athey (2018), and Athey, Tibshirani, and Wager (2019) to estimate the CAPE and APE of the continuous treatment, as well as the double machine learning (DML) approach of Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) to estimate the ATE of a binary treatment of a discount $\geq 30\%$ vs. $< 30\%$ in the sample of always buyers.

The CF adapts the random forest to the purpose of causal inference. It is based on first running separate random forests for predicting the outcome Y and the treatment D as a function of the covariates X using leave-one-out cross-fitting. The latter implies that the outcome or treatment of each observation is predicted based on all observations in the data but its own, in order to safeguard against overfitting bias. Second, the predictions are used for computing residuals of the outcomes and treatments, in which the influence of X has been partialled out. Finally, a further random forest is applied to average over so-called causal trees, see Athey and Imbens (2016), in order to estimate the CAPE. The causal tree approach contains two key modifications when compared to standard decision trees. First, instead of an outcome variable, it is the coefficient of regressing the residual of Y on the residual of D, i.e. the causal effect estimate of the treatment, that is to be predicted. Recursive splitting aims to find the largest effect heterogeneities across subsets defined in terms of X to estimate the CAPE accurately. Secondly, within each subset, different parts of the data are used for estimating (a) the tree's splitting structure (i.e., the definition of covariate indicator functions) and (b) the causal effect of the treatment to prevent spuriously large effect heterogeneities due to overfitting.

The CAPE estimate obtained by CF can be thought of as a weighted regression of the outcome residual on the treatment residual. The random forest-determined weight reflects the importance of a sample observation for assessing the causal effect at specific values of the covariates. After estimating the CAPE given X, the APE is obtained by appropriately averaging over the distribution of X among the always buyers. For implementing CAPE and APE estimation, we use the *grf* package by Tibshirani, Athey, Friedberg, Hadad, Hirshberg, Miner, Sverdrup, Wager, and Wright (2020) for the statistical software R. We set the number of trees to be used in a forest to 1000. We select any other tuning parameters like the number of randomly chosen covariates considered for splitting or the minimum number of observations per subset (or node) by the built-in cross-validation procedure.

We also estimate the ATE among always buyers in our sample based on DML for a binary treatment defined as $\tilde{D} = I\{D \geq 0.3\}$, with $I\{\cdot\}$ denoting the indicator function that is equal to one if its argument is satisfied and zero otherwise. Furthermore, let $\mu_d(X) = E[Y|\tilde{D} = d, X, S(0) = 1, S = 1]$ denote the conditional mean outcome and $p_d(X) = \Pr(\tilde{D} = d|X, S(0) = 1, S = 1)$ the propensity score of receiving treatment category d (with $d = 1$ for a discount $\geq 30\%$ and $d = 0$ otherwise) in that population. Estimation is based on the sample analog of the doubly robust identification expression for the ATE, see Robins, Rotnitzky, and Zhao (1994) and Robins and Rotnitzky (1995):

$$\begin{aligned}\Delta_{S(0)=1}(1,0) &= E\left[\mu_1(X) - \mu_0(X) + \frac{(Y - \mu_1(X)) \cdot \tilde{D}}{p_1(X)}\right. \\ &- \left.\frac{(Y - \mu_0(X)) \cdot (1 - \tilde{D})}{p_0(X)}\big|S(0) = 1, S = 1\right]. \end{aligned} \tag{4.5.1}$$

We estimate (4.5.1) using the *causalweight* package for the statistical software R by Bodory and Huber (2018). As machine learners for the conditional mean outcomes $\mu_D(X)$ and the propensity scores $p_D(X)$ we use the random forest with the default options of the *SuperLearner* package of van der Laan, Polley, and Hubbard (2008), which itself imports the *ranger* package by Wright and

Ziegler (2017) for random forests. To impose common support in the data used for ATE estimation, we apply trimming threshold of 0.01, implying that we drop observations with estimated propensity scores smaller than 0.01 (or 1%) and larger than 0.99 (or 99%) from our sample.

4.6 Empirical results

4.6.1 Descriptive statistics

Before discussing the results of our machine learning approaches, we first present some descriptive statistics for our data in Table 4.1, namely the mean and the standard deviation of selected variables by always buyer status and binary discount category ($\geq 30\%$ and $< 30\%$). We see that discounts and regular ticket fares of always buyers are on average lower than those of other customers. Another interesting observation is that in either discount category, we observe less leisure travelers among the always buyers than among other customers, which can be rationalized by business travelers responding less to price incentives by discounts. This is also in line with the finding that always buyers tend to purchase more second-class tickets. More generally, we see non-negligible variation in demand-related covariates across the four subsamples defined in terms of buying behavior and discount rates. For instance, among always buyers, the total amount of supersaver tickets offered is on average larger in the higher discount category, while it is lower among the remaining clients. This suggests that neither the treatment nor being an always buyer is quasi-random, a problem we aim to tackle based on our identification strategy outlined in section 4.4. Concerning the demand-shift outcome, we see that always buyers change the departure time less frequently than others. With regard to upselling,we recognise that the relative amount of individuals upgrading their second-class to a first-class ticket is the same for both discount categories, i.e. $\geq 30\%$ and $< 30\%$.

4.6.2 Predicting buying decisions

We subsequently present our predictive analysis based on the random forest and investigate which covariates importantly predict three outcomes, namely whether customers booked a trip otherwise not realized by train (additional trip), bought a first-class rather than a second-class ticket (upselling), or rescheduled their trip e.g. away from rush hours (demand shift). For this purpose, we create three distinct datasets in which the values of the respective binary outcome are balanced, i.e. 1 (for instance, upselling) for 50% and 0 (no upselling) for 50% of the observations. We balance our data because we aim to train a model that predicts both outcome values equally well. Taking the demand shift outcome as an example, our data with non-missing covariate or outcome information contain 3481 observations with $Y = 1$ and 9576 observations with $Y = 0$. We retain all observations with $Y = 1$ and randomly draw 3481 observations with $Y = 0$ to obtain such a balanced data set. In the next step, we randomly split these 6962 observations into a training set consisting of 75% of the data and a test set (25%). In the training set, we train the random forest using the treatment D and all covariates X, W as predictors. In the test set, we predict the outcomes based on the trained forest, classifying e.g. observations with a demand shift probability ≥ 0.5 as 1. We then compare the predictions to the actually observed outcomes to assess model performance based on the correct classification rate (also known as accuracy), i.e. the share of observations in the test data for which the predicted outcome corresponds to the actual one.

For each of the outcomes, Table 4.2 presents the 30 most predictive covariates in the training set ordered in decreasing order according to a variable importance measure. The latter is defined as the total decrease in the Gini index (as a measure of node impurity in terms of outcome values) in a tree when including the respective covariate for splitting, averaged over all trees in the forest. The results suggest that trip- and demand-related characteristics like seat capacity, utilization, departure time, and distance are important predictors. Concerning personal characteristics, also customer's age appears to be relevant. Furthermore, also the treatment intensity D has considerable predictive power. Interestingly, specific connections (defined by indicators for points of departure and destination) turn out to be less important characteristics conditional on the other covariates already mentioned.

At the bottom of Table 4.2 we also report the correct classification rates for the three outcomes. While the accuracy in predicting a demand shift amounts to 58%, which is somewhat better than random guessing but not particularly impressive, the performance is more satisfactory for predicting decisions about additional trips with an accuracy of 65% and quite decent for upselling (82%). We note that when predicting upselling, we drop the variables 'class', which indicates whether someone travels in the first or second class, and 'seat capacity', which refers to the capacity in the chosen class, from the predictors. The reason is that upselling is defined as switching from second to first class, and therefore, the chosen class and the related seat capacities are actually part of the outcome to be predicted. Tables 4.8 and 4.7 in the Appendix present the predictive outcome analysis separately for subsamples with discounts $\geq 30\%$ and $< 30\%$, respectively. In terms of which classes of variables are most predictive (trip- and demand-related characteristics, age, discount rate) and also in terms of accuracy, the findings are rather similar to those in Table 4.2. In general, machine learning appears useful for forecasting customer behavior in the context of demand for train trips, albeit not equally well for all aspects of interest. Such forecasts may for instance serve as a base for customer segmentation, e.g. into customer groups more and less inclined to book an additional trip or switch classes or departure times.

Table 4.1: Mean and standard deviation by discount and type

discount	< 30%		≥ 30%	
always buyers	No	Yes	No	Yes
discount	0.21	0.19	0.57	0.53
	(0.07)	(0.08)	(0.12)	(0.13)
regular ticket fare	44.36	36.14	47.19	32.91
	(29.38)	(25.47)	(30.14)	(23.78)
age	47.22	47.68	45.59	48.77
	(15.36)	(16.14)	(15.80)	(16.49)
gender	0.51	0.55	0.53	0.59
	(0.50)	(0.50)	(0.50)	(0.49)
diff. purchase travel	3.42	3.23	7.72	7.19
	(6.96)	(6.76)	(11.23)	(10.30)
distance	136.49	127.86	126.15	116.76
	(77.38)	(71.49)	(69.98)	(66.04)
capacity utilization	35.51	39.19	26.46	33.15
	(14.16)	(14.31)	(13.24)	(13.75)
seat capacity	328.28	429.57	303.83	445.14
	(196.19)	(196.10)	(185.42)	(188.54)
offer total	33.95	44.10	70.97	98.34
	(42.57)	(50.68)	(69.57)	(84.45)
sold total	28.04	37.29	13.70	25.75
	(41.92)	(50.31)	(36.37)	(53.67)
half fare travel ticket	0.74	0.79	0.62	0.74
	(0.44)	(0.40)	(0.49)	(0.44)
leisure	0.77	0.69	0.82	0.76
	(0.42)	(0.46)	(0.39)	(0.43)
class	1.38	1.65	1.33	1.73
	(0.48)	(0.48)	(0.47)	(0.44)
Swiss	0.89	0.92	0.88	0.88
	(0.31)	(0.28)	(0.33)	(0.32)
demand shift	0.31	0.19	0.31	0.23
	(0.46)	(0.40)	(0.46)	(0.42)
upselling	0.49	0.00	0.49	0.00
	(0.50)	(0.00)	(0.50)	(0.00)
obs.	1151	2221	5529	3745

Notes: Regular ticket fare is in Swiss francs. 'diff. purchase travel' denotes the difference between purchase and travel day. 'Offer total' and 'sold total' denote the total amount of supersaver tickets offered and the total amount of supersaver tickets sold respectively.

Table 4.2: Predictive outcome analysis

demand shift		upselling		additional trip	
variable	importance	variable	importance	variable	importance
departure time	142.694	capacity utilization	295.924	seat capacity	147.037
seat capacity	121.42	offer level B	188.861	D	128.086
age	119.846	offer level C	149.911	age	123.948
capacity utilization	119.606	D	132.095	departure time	123.516
D	112.474	age	100.258	capacity utilization	113.160
distance	112.143	departure time	98.909	distance	101.730
offer level B	84.142	offer level A	93.303	offer level B	84.989
diff. purchase travel	81.167	distance	87.319	diff. purchase travel	80.236
offer level C	76.238	offer level D	85.408	offer level A	78.507
offer level A	75.971	diff. purchase travel	62.841	offer level C	77.097
number of sub-journeys	73.096	number of sub-journeys	55.978	number of sub-journeys	69.443
offer level D	61.763	rel. sold level A	44.505	ticket purchase complexity	64.498
ticket purchase complexity	57.071	ticket purchase complexity	41.819	offer level D	56.888
rel. sold level A	51.377	offer level E	37.159	class	51.456
rel. amount imputed values	42.222	rel. sold level B	34.462	rel. sold level A	46.969
rel. sold level B	38.144	rel. amount imputed values	30.747	rel. amount imputed values	38.869
adult companions	34.176	rel. sold level C	28.635	rel. sold level B	36.484
rel. sold level C	28.201	adult companions	25.115	half fare	35.785
offer level E	25.714	2019	18.88	adult companions	34.446
gender	23.707	gender	18.47	halfe fare travel ticket	28.465
amount purchased tickets	19.575	rush hour	17.448	gender	25.419
German	18.659	Saturday	16.173	rel. sold level C	24.679
travel alone	18.605	German	15.457	offer level E	22.556
2019	18.082	leisure	15.304	leisure	20.438
French	17.906	amount purchased tickets	15.112	no subscriptions	19.793
saturday	17.487	travel alone	14.792	amount purchased tickets	19.283
Friday	17.272	half fare	14.306	German	19.119
peak hour	17.064	French	14.161	travel alone	18.139
class	16.973	Thursday	13.413	2019	17.192
leisure	16.892	scheme 20	13.411	French	17.026
correct prediction rate	0.581		0.817		0.653
balanced sample size	6962		6738		7000

Notes: 'Offer level A', 'offer level B', 'offer level C', 'offer level D' and 'offer level E' denote the amount of supersaver tickets with discount A, B, C, D and E respectively. 'Diff. purchase travel' denotes the difference between purchase and travel day. 'Rel. sold level A', 'rel. sold level B', 'rel. sold level C' and 'rel. sold level D' denote the relative amount of supersaver tickets offered with discount A, B, C and D respectively. The relative amounts are in relation to the seats offered. 'No subscriptions' indicates not possessing any subscription. For predicting upselling, the covariates 'class' and 'seat capacity' are dropped.

4.6.3 Testing the identification strategy

Before presenting the results for the causal analysis, we consider two different methods to partially test the assumptions underlying our identification strategy. First, we test Assumption 3 (weak monotonicity) by running the CF and DML procedures as well as a conventional OLS regression in which we use buying an additional trip $(1 - S(0))$, i.e. not being an always buyer, as outcome variable and X as control variables in our sample of supersaver customers. The CF permits estimating the conditional change in the share of surveyed customers induced to buy an additional trip by modifying the discount rate D given X, i.e. $\frac{\partial E[(1-S(0))|D,X,S=1]}{\partial D}$, as well as the average thereof across X conditional on sample selection, $E\left[\frac{\partial E[(1-S(0))|D,X,S=1]}{\partial D}\middle|S=1\right]$. DML, on the other hand, yields an estimate of the average difference in the share of additional trips across the high and low treatment categories conditional on sample selection, $E[E[(1-S(0))|D<0.3,X,S=1] - E[(1-S(0))|D \geq 0.3,X,S=1]|S=1]$. Finally, the OLS regression of $(1 - S(0))$ on D and all X in our sample tests monotonicity when assuming a linear model.

Table 4.3 reports the results that do not provide any evidence against the monotonicity assumption. When considering the continuous treatment D, the CF-based estimate of $E\left[\frac{\partial E[(1-S(0))|D,X,S=1]}{\partial D}\middle|S=1\right]$ is highly statistically significant and suggests that augmenting the discount by one percentage point increases the share of customers otherwise not buying the ticket by 0.56 percentage points on average. Furthermore, any estimates of the conditional change $\frac{\partial E[(1-S(0))|D,X,S=1]}{\partial D}$ are positive, as displayed in the histogram of Figure 4.6.1, and 82.2% of them are statistically significant at the 10% level, 69.1% at the 5% level. Furthermore, the OLS coefficient of 0.544 is highly significant. Likewise, the statistically significant DML estimate points to an increase in the share of additional trips by 18.4 percentage points when switching the binary treatment indicator from $D < 0.3$ to $D \geq 0.3$.

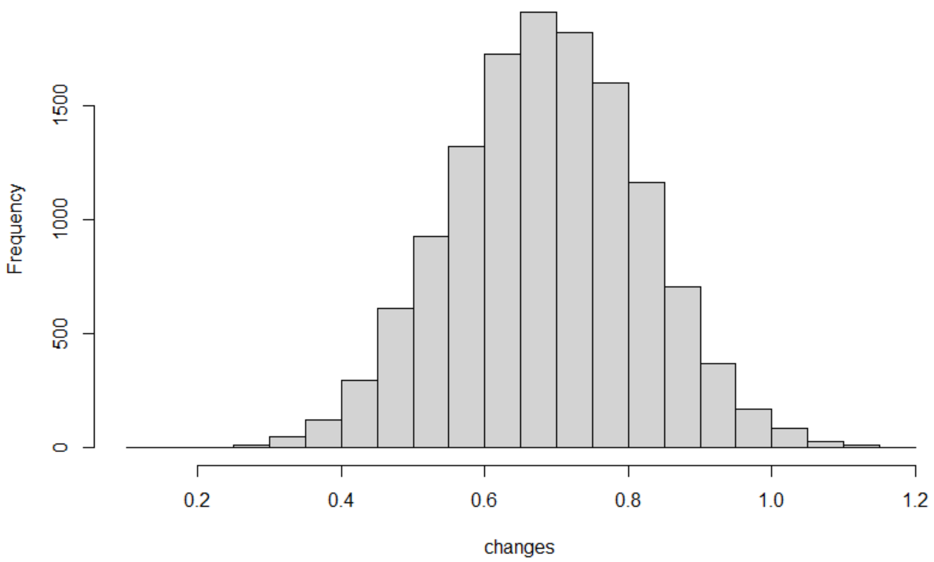
We also test the statistical independence of D and W conditional on X in our sample of always buyers, as implied by our identifying assumptions, see the discussion at the end of section 4.4. To this end, we randomly split the evaluation data into a training set (25% of observations) and a test set (75% of all observations). In the training data set, we run a linear lasso regression (Tibshirani, 1996) of D on X in order to identify important predictors by means of 10-fold cross-validation. In the next step, we select all covariates in X with non-zero lasso coefficients and run an OLS regression of D on the selected covariates in the test data. Finally, we add W to that regression in the test data and run a Wald test to compare the predictive power of the models with and without W. We repeat the procedure of splitting the data,

Table 4.3: Monotonicity tests

	CF: av. change	OLS: coef.	DML: $D \geq 0.3$ vs $D < 0.3$
change in $(1 - S(0))$	0.564	0.544	0.184
standard error	0.060	0.031	0.009
p-value	0.000	0.000	0.000
trimmed observations			1760
number of observations	12924		

Notes: 'CF', 'OLS', and 'DML' stands for estimates based on causal forests, linear regression, and double machine learning, respectively. 'trimmed observations' is the number of trimmed observations in DML when setting the propensity score-based trimming threshold to 0.01*. Control variables consist of* X*.*

Figure 4.6.1: Monotonicity given X



performing the lasso regression in the training set, and running the OLS regressions and the Wald test in the test set 100 times. This yields an average p-value of 0.226, with 15 out of 100 p-values being smaller than 5%. These results do not provide compelling statistical evidence that W is associated with D conditional on X, even though the training sample is relatively small and thus favors selecting too few predictors in X (due to the cross-validation that trades off bias due to including fewer predictors and variance due to including more predictors).

We note that performing lasso-based variable selection and OLS-based testing in different (training and test) data avoids correlations of these steps that could entail an overestimation of the goodness of fit. Nonetheless, our findings remain qualitatively unchanged when performing both steps in all of the evaluation data. Repeating the cross-validation step for the lasso-based covariate selection 100 times and testing in the total sample yields an even higher average p-value of 0.360. Finally, we run a standard OLS regression of D on all elements of X (rather than selecting the important ones by lasso) in the total sample and compare its predictive power to a model additionally including W. Also in this case, the Wald test entails a rather high p-value of 0.343. In summary, we conclude that our tests do not point to the violation of our identifying assumptions.

4.6.4 Assessing the causal effect of discounts

Table 4.4 presents the main results of our causal analysis, namely the estimates of the discount rate's effect on the demand shift outcome, which is equal to one if the discount induced rescheduling the departure time and zero otherwise. We note that all covariates, i.e. both the trip- or demand-related factors X and the personal characteristics W, are used as control variables, even though we have previously claimed that X is sufficient for identification. There are, however, good reasons for including W as well in the estimations. First, conditioning on the personal characteristics available in the data may reduce estimation bias if X is - contrarily to our assumptions and to what our tests suggest - not fully sufficient to account for confounding. Second, it can also reduce the variance of the estimator, e.g. if some factors like age are strong predictors of the outcome. Third, having W in the CF allows for a more fine-grained analysis of effect heterogeneity based on computing more 'individualized' partial effects that (also) vary across personal characteristics.

Considering the estimates of the CF, we obtain an average partial effect (APE) of 0.161, suggesting that increasing the current discount rate among always buyers by one percentage point increases the share of rescheduled trips by 0.16 percentage points. This effect is statistically significant at the 5% level. As a word of caution, however, we point out that the standard error is non-negligible such that the magnitude of the impact is not very precisely estimated. When applying DML, we obtain an average treatment effect (ATE) of 0.038 that is significant at the 1% level, suggesting that discounts of 30% and more on average increase the number of demand shifts by 3.8 percentage points compared to lower discounts, which is qualitatively in line with the CF. Furthermore, we find a decent overlap or common support in

Table 4.4: Effects on demand shift

	CF: APE	DML: ATE $D \geq 0.3$ vs $D < 0.3$
effect	0.161	0.038
standard error	0.072	0.010
p-value	0.025	0.000
trimmed observations		151
number of observations	5903	

Notes: 'CF' and 'DML' stands for estimates based on causal forests, linear regression, and double machine learning, respectively. 'trimmed observations' is the number of trimmed observations in DML when setting the propensity score-based trimming threshold to 0.01. *Control variables consist of both X and W.*

most of our sample in terms of the estimated propensity scores across lower and higher discount categories considered in DML, see the propensity score histograms in Appendix 4.A. This is important as ATE evaluation hinges on the availability of observations with comparable propensity scores across treatment groups. Only 151 out of our 5903 observations are dropped due to too extreme propensity scores below 0.01 or above 0.99 (pointing to a violation of common support).[4] In summary, our results clearly point to a positive average effect of the discount rate on trip rescheduling among always buyers, which is, however, not overwhelmingly large.

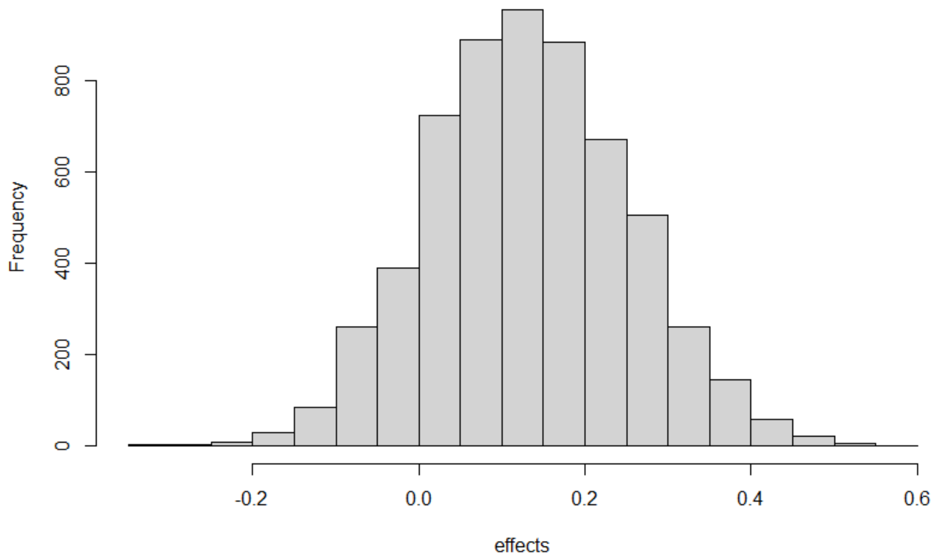
4.6.5 Effect heterogeneity

In this section, we assess the heterogeneity of the effects of D on Y across interviewees and observed characteristics. Figure 4.6.2 shows the distribution the CF-based conditional average effects (CAPE) of marginally increasing the discount rate given the covariates values of the always buyers in our sample (which are also the base for the estimation of the APE). While the CAPEs are predominantly positive, they are quite imprecisely estimated. Only 2.9% and 0.8% of the positive ones are statistically significant at the 10% and 5% levels, respectively. Further, only 0.1% of the negative ones are statistically significant at the 10% level. Yet, the distribution points to a positive marginal effect for most always buyers and also suggests that the magnitude of the effects varies non-negligibly across individuals.

Next, we assess the effect heterogeneity across observed characteristics based on the CF results. First, we run a conventional random forest with

[4]Our findings of a positive ATE remain robust when setting the propensity score-based trimming threshold to 0.02 (ATE: 0.042) or 0.05 (ATE: 0.045).

Figure 4.6.2: CAPEs



the estimated CAPEs as the outcome and the covariates as predictors to assess the covariates' relative importance for predicting the CAPE, using the decrease in the Gini index as importance measure as also considered in section 4.6.2. Table 4.5 reports the 20 most predictive covariates ordered in decreasing order according to the importance measure. Demand-related characteristics (like seat capacity, utilization, departure time, and distance) turn out to be the most important predictors for the size of the effects, also customer's age has some predictive power. Similarly as for outcome prediction in section 4.6.2, specific connections (characterized by points of departure or destination) are less important predictors of the CAPEs given the other information available in the data.

While Table 4.5 provides information on the best predictors of effect heterogeneity, it does not give insights on whether effects differ importantly and statistically significantly across specific observed characteristics of interest. For instance, one question relevant for designing discount schemes is whether (marginally) increasing the discounts is more effective among always buyers so far exposed to rather small or rather large discounts. Therefore, we investigate whether the CAPEs are different across our binary treatment categories defined by $\tilde{D}$ (30% or more and less than 30%). To this end, we apply the approach of Semenova and Chernozhukov (2020) based on (i)

Table 4.5: Most important covariates for predicting CAPEs

covariate	importance
seat capacity	11.844
offer level C	11.164
capacity utilization	5.144
departure time	5.122
distance	4.287
offer level D	4.015
class	3.434
saturday	2.933
age	2.429
number of sections	2.373
diff. purchase travel	2.110
offer level A	1.634
offer level B	1.610
half fare	1.524
scheme 17	1.496
half fare travel ticket	1.373
rel. sold level B	0.901
ticket purchase complexity	0.847
leisure	0.773
rel. sold level A	0.770

Notes: 'Offer level A', 'offer level B', 'offer level C' and 'offer level D' denote the amount of supersaver tickets with discount A, B, C and D respectively. 'Rel. offer level A', 'rel. offer level B' and 'rel. offer level C' denote the relative amount of supersaver tickets offered with discount A, B and C. The relative amounts are in relation to the seats offered.

plugging the CF-based predictions into a modified version of the doubly robust functions provided within the expectation operator of (4.5.1) that is suitable for a continuous D and (ii) linearly regressing the doubly robust functions on the treatment indicator $\tilde{D}$. The results are reported in the upper panel of Table 4.6. While the point estimate of -0.104 suggests that the demand shifting effect of increasing the discount is on average smaller when discounts are already quite substantial (above 30%), the difference is far from being statistically significant at any conventional level.

Using again the method of Semenova and Chernozhukov (2020), we also investigate the heterogeneity among a limited and pre-selected set of covariates that appears interesting for characterizing customers and their travel purpose, namely age, gender, and travel distance, as well as indicators for leisure trip and commute (with business trip being the reference category), traveling during peak hours, and possession of a half fare travel tickets. As displayed in the lower panel of Table 4.6, we find no important effect heterogeneities across the age or gender of always buyers or as a function of travel distance conditional on the other information included in the regression, as the coefficients on these

Table 4.6: Effect heterogeneity analysis

	effect	st. err.	p-value
Discounts categories ($D \geq 0.3$ vs $D < 0.03$)			
APE for $D < 0.3$ (constant)	0.209	0.089	0.019
Difference APE $D \geq 0.3$ vs $D < 0.3$ (slope coef.)	-0.104	0.122	0.395
Customer and travel characteristics			
constant	-0.154	0.295	0.602
age	-0.002	0.004	0.556
gender	-0.022	0.129	0.866
distance	-0.000	0.001	0.697
leisure trip	0.297	0.165	0.072
commute	0.241	0.241	0.316
half fare travel ticket	0.228	0.142	0.109
peak hours	0.222	0.133	0.094

Notes: Business trip is the reference category for the indicators 'leisure trip' and 'commute'.

variables are close to zero. In contrast, the effect of demand shift is (given the other characteristics) substantially larger among always buyers with a half fare travel tickets and among commuters, however, neither coefficient is statistically significant at the 10% level (even though the half fare coefficient is close).

For leisure trips, the coefficient is even larger (0.297), suggesting that all other included variables equal, a one percentage point increase in the discount rate increases the share of rescheduled trips by 0.29 percentage points more among leisure travelers than among always buyers traveling for business. The coefficient is statistically significant at the 10% level, even though we point out that the p-value does not account for multiple hypothesis testing of several covariates. This finding can be rationalized by leisure travelers being likely more flexible in terms of timing than business travelers. Also the coefficient on peak hours is substantially positive (0.222) and statistically significant at the 10% level (again, without controlling for multiple hypothesis testing). This could be due to peak hours being the most attractive travel time, implying that costumers are more willing to reschedule their trips when being offered a discount within peak hours. We conclude that even though several coefficients appear non-negligible, statistical significance in our heterogeneity analysis is overall limited, which could be due to the (for the purpose of investigating effect heterogeneity) limited sample of several thousand observations.

4.7 Conclusion

In this study, we applied causal and predictive machine learning to assess the demand effects of discounts on train tickets issued by the Swiss Federal Railways (SBB), the so-called 'supersaver tickets', based on a unique data that combines a survey of supersaver customers with rail trip- and demand-related information provided by the SBB. In a first step, we analyzed which customer- or trip-related characteristics (including the discount rate) are predictive for three outcomes characterizing buying behavior, namely: booking a trip otherwise not realized by train (additional trip), buying a first- rather than second-class ticket (upselling), or rescheduling a trip (e.g. a demand shift away from rush hours) when being offered a supersaver ticket. The random forest-based results suggested that customer's age, demand-related information for a specific connection (like seat capacity, departure time, and utilization), and the discount level permit forecasting buying behavior to a certain extent, with correct classification rates amounting to 58% (demand shift), 65% (additional trip), and 82% (upselling), respectively.

As predictive machine learning cannot provide the causal effects of the predictors involved, we, in a second step, applied causal machine learning to assess the impact of the discount rate on the demand shift among always buyers (who would have traveled even without a discount), which appears interesting in the light of capacity constraints at rush hours. To this end, we invoked the identifying assumptions that (i) the discount rate is quasi-random conditional on our covariates and (ii) the buying decision increases weakly monotonically in the discount rate and exploited survey information about customer behavior in the absence of discounts. We also considered two approaches for partially testing our assumptions, which did not point to a violation of the latter. Our main results based on the causal forest suggested that increasing the discount rate by one percentage point entails an average increase of 0.16 percentage points in the share of rescheduled trips among always buyers. This finding was corroborated by double machine learning with just two discount categories, suggesting that discount rates of 30% and more on average increase the share of rescheduled trips by 3.6 percentage points when compared to lower discounts. Furthermore, when investigating effect heterogeneity across a pre-selected set of characteristics, we found the causal forest-based effects to be higher (with marginal statistical significance when not controlling for multiple hypothesis testing) for leisure travelers and during peak hours when also controlling for customer's age, gender, possession of a half fare travel card, and travel distance. Finally, our effect heterogeneity analysis also revealed that demand-related information is most predictive for the size of the effect of the discount rate.

Using state-of-the-art machine learning tools, our study appears to be the first (at least for Switzerland) to provide empirical evidence on how discounts on train tickets affect customers' willingness to reschedule trips - an important information for designing discount schemes aiming at balancing out train utilization across time and reducing overload during peak hours. Even though the overall impact on the demand shifts on always buyers might not be as large as one could hope for, the causal forest pointed to the existence of customer segments that are likely more responsive and could be scrutinized when collecting a larger amount of data than available for our analysis. Furthermore, our empirical approach may also be applied to other countries or transport industries facing capacity constraints. For instance, we would expect that in a setting with higher competition from alternative public transport modes like long distance bus services (not present in Switzerland), the impact of train discounts may well be different. More generally, our study can be regarded as a use case for how predictive and, in particular, causal machine learning can be fruitfully applied for business analytics and as decision support for optimizing specific interventions like discount schemes based on impact evaluation.

Appendices

4.A Propensity score plots

Figure 4.A.1: Propensity score estimates in the higher discount category ($D \geq 0.3$)

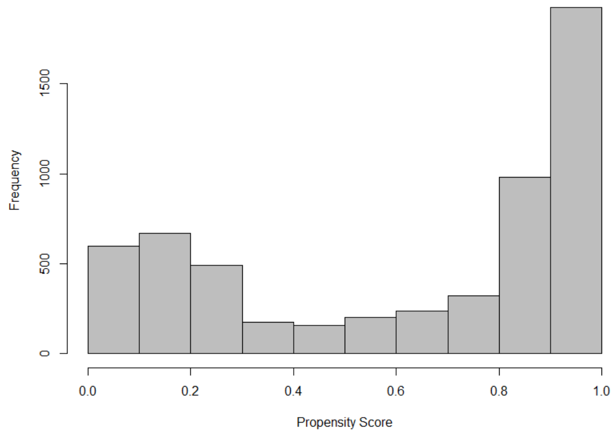
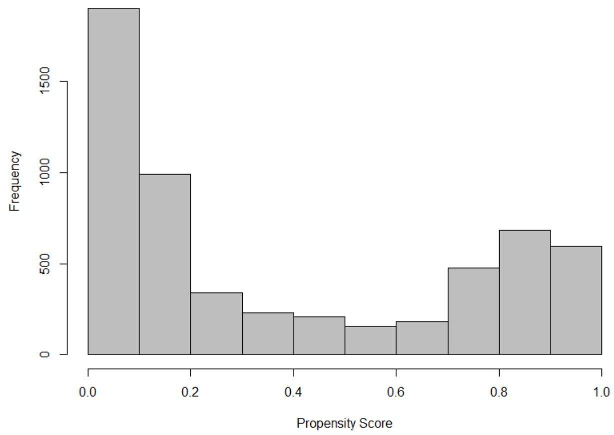


Figure 4.A.2: Propensity score estimates in the lower discount category ($D < 0.3$)



4.B Further tables

Table 4.7 and Table 4.8 present the predictive outcome analysis separately for subsamples with discounts $< 30\%$ and $\geq 30\%$, respectively.

Table 4.7: Predictive outcome analysis, $D < 0.3$

demand shift		upselling		additional trip	
variable	importance	variable	importance	variable	importance
departure time	37.33	capacity utilization	41.387	seat capacity	25.342
seat capacity	27.871	offer level D	27.669	age	21.639
capacity utilization	26.508	age	22.145	capacity utilization	20.168
distance	26.31	D	19.077	distance	18.970
age	26.223	offer level C	17.324	departure time	18.527
D	25.08	departure time	16.538	D	18.076
number of sub-journeys	17.403	distance	15.897	ticket purchase complexity	16.637
offer level C	15.643	offer level B	15.696	offer level C	12.085
diff. purchase travel	15.299	rel. sold level B	10.992	rel. sold level B	11.709
ticket purchase complexity	15.116	number of sub-journeys	10.019	offer level D	11.641
rel. sold level B	15.03	diff. purchase travel	9.573	number of sub-journeys	11.347
offer level D	15.012	rel. sold level C	9.367	diff. purchase travel	10.328
rel. sold level C	14.625	offer level A	7.857	offer level B	10.185
offer level B	14.413	rel. sold level D	7.33	rel. sold level C	8.993
rel. sold level A	11.856	ticket purchase complexity	7.319	rel. sold level A	8.162
offer level A	11.329	offer level E	7.183	offer level A	7.643
rel. amount imputed values	10.079	rel. sold level A	6.769	class	7.381
rel. sold level D	9.625	rel. amount imputed values	5.422	rel. sold level D	6.964
adult companions	8.503	adult companions	4.881	rel. amount imputed values	6.189
offer level E	6.511	rush hour	4.143	adult companions	5.785
gender	5.214	leisure	3.602	leisure	5.398
leisure	5.154	gender	3.599	offer level E	4.866
destination Geneva Airport	4.83	2019	3.597	gender	4.692
departure Zuerich	4.736	travel alone	3.047	German	3.801
class	4.598	Friday	2.92	halfe fare travel ticket	3.686
travel alone	4.59	German	2.825	travel alone	3.540
peak hour	4.545	French	2.479	French	3.493
Friday	4.524	departure Zuerich	2.429	Friday	3.419
German	4.522	destination Zuerich Airport	2.427	half fare	3.163
amount purchased tickets	4.349	scheme 20	2.427	2019	3.136
correct prediction rate	0.555		0.772		0.605
balanced sample size	1642		1140		1202

Notes: 'Diff. purchase travel' denotes the difference between purchase and travel day. 'Rel. offer level A', 'rel. offer level B', 'rel. offer level C' and 'rel. offer level D' denote the relative amount of supersaver tickets offered with discount A, B, C and D respectively. The relative amounts are in relation to the seats offered. 'Offer level A', 'Offer level B', 'Offer level C', 'Offer level D' and 'Offer level E' denotes the amount of supersaver tickets with discount A, B, C, D and E respectively. 'No subscription' indicates not possessing any subscription. For predicting upselling, the covariates 'class' and 'seat capacity' are dropped.

Table 4.8: Predictive outcome analysis, $D \geq 0.3$

demand shift		upselling		additional trip	
variable	importance	variable	importance	variable	importance
departure time	114	seat capacity	246.396	capacity utilization	133.936
seat capacity	95.799	offer level B	178.212	age	105.107
age	95.209	offer level C	127.327	departure time	100.091
capacity utilization	89.422	D	100.618	capacity utilization	97.889
distance	85.503	offer level A	88.947	distance	85.647
D	80.447	Tageszeitinmin	82.658	D	83.671
diff. purchase travel	69.276	age	78.886	offer level B	73.399
offer level B	68.75	distance	73.885	offer level A	69.936
offer level A	65.766	offer level D	72.452	diff. purchase travel	67.823
offer level C	60.513	diff. purchase travel	55.321	offer level C	64.689
number of sub-journeys	57.767	number of sub-journeys	48.622	number of sub-journeys	58.100
offer level D	44.626	rel. sold level A	38.997	class	54.857
ticket purchase complexity	44.434	ticket purchase complexity	31.991	ticket purchase complexity	49.348
rel. sold level A	39.796	offer level E	27.041	offer level D	46.685
rel. amount imputed values	32.144	rel. amount imputed values	25.586	rel. sold level A	42.589
adult companions	25.925	adult companions	24.73	rel. amount imputed values	35.411
rel. sold level B	20.536	rel. sold level B	17.099	half fare	31.308
gender	18.629	gender	15.835	adult companions	28.732
offer level E	17.521	2019	15.416	half fare travel ticket	23.900
travel alone	15.15	Saturday	14.256	gender	20.562
amount purchased tickets	14.939	amount purchased tickets	13.396	rel. sold level B	20.036
German	14.855	rush hour	12.947	offer level E	18.595
French	14.415	German	12.878	German	16.257
2019	14.387	leisure	12.344	amount purchased tickets	15.847
Sunday	13.821	travel alone	12.28	leisure	15.500
destination Zuerich Airport	13.387	half fare	11.882	no subscription	15.434
Saturday	13.378	Friday	11.646	travel alone	15.132
class	13.27	scheme 20	11.559	Swiss	14.951
half fare	13.258	French	11.477	Saturday	14.613
rel. amount imputed values	13.048	Sunday	11.39	2019	14.279
correct prediction rate	0.589		0.809		0.629
balanced sample size	5320		5598		5798

Notes: 'Diff. purchase travel' denotes the difference between purchase and travel day. 'Rel. offer level A', 'rel. offer level B', 'rel. offer level C' and 'rel. offer level D' denote the relative amount of supersaver tickets offered with discount A, B, C and D respectively. The relative amounts are in relation to the seats offered. 'Offer level A', 'Offer level B', 'Offer level C', 'Offer level D' and 'Offer level E' denotes the amount of supersaver tickets with discount A, B, C, D and E respectively. 'No subscription' indicates not possessing any subscription. For predicting upselling, the covariates 'class' and 'seat capacity' are dropped.

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program," *Journal of the American Statistical Association*, 105(490), 493–505.

——— (2011): "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies," *Journal of Statistical Software*, 42(13), 1–17.

——— (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59(2), 495–510.

Abadie, A., and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *The American Economic Review*, 93(1), 113–132.

Abrantes-Metz, R. M., L. M. Froeb, J. F. Geweke, and C. T. Taylor (2006): "A Variance Screen for Collusion," *International Journal of Industrial Organization*, 24(3), 467–486.

Abrantes-Metz, R. M., M. Kraten, A. D. Metz, and G. Seow (2012): "Libor Manipulation?," *Journal of Banking and Finance*, 36(1), 136–150.

Abrate, G., G. Fraquelli, and G. Viglia (2012): "Dynamic pricing strategies: Evidence from European hotels," *International Journal of Hospitality Management*, 31(1), 160–168.

Addessi, W., B. Biagi, and M. G. Brandano (2019): "Evaluating the effect of the introduction of the euro on tourist flows: A synthetic control approach," *The World Economy*, 42(5), 1554–1575.

Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of Causal Effects using Instrumental Variables," *Journal of American Statistical Association*, 91(434), 444–472.

Aryal, G., and M. F. Gabrielli (2013): "Testing for collusion in asymmetric first-price auctions," *International Journal of Industrial Organization*, 31(1), 26–35.

Ascarza, E. (2018): "Retention futility: Targeting high-risk customers might be ineffective," *Journal of Marketing Research*, 55(1), 80–98.

Asker, J. (2010): "A study of the internal organization of a bidding cartel," *American Economic Review*, 100(3), 724–762.

Athey, S. (2019): "The Impact of Machine Learning on Economics," *University of Chicago Press.*

Athey, S., and G. W. Imbens (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360.

——— (2017): "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31(2), 3–32.

——— (2019): "Machine learning methods that economists should know about," *Annual Review of Economics*, 11, 685–725.

Athey, S., J. Tibshirani, and S. Wager (2019): "Generalized random forests," *The Annals of Statistics*, 47(2), 1148–1178.

Baggio, R. (2011): "Collaboration and cooperation in a tourism destination: A network science approach," *Current Issues in Tourism*, 14(2), 183–189.

Bajari, P., and L. Ye (2003): "Deciding between competition and collusion," *Review of Economics and Statistics*, 85(4), 971–989.

Basso, L. J., and H. E. Silva (2014): "Efficiency and substitutability of transit subsidies and other urban transport policies," *American Economic Journal: Economic Policy*, 6(4), 1–33.

Batty, P., R. Palacin, and A. González-Gil (2015): "Challenges and opportunities in developing urban modal shift," *Travel Behaviour and Society*, 2(2), 109–123.

Becerra, M., J. Santaló, and R. Silva (2013): "Being better vs. being different: Differentiation, competition, and pricing strategies in the Spanish hotel industry," *Tourism Management*, 34, 71–79.

Bergman, M. A., J. Lundberg, S. Lundberg, and J. Y. Stake (2020): "Interactions across firms and bid rigging," *Review of Industrial Organization*, 56(1), 107–130.

Beritelli, P. (2011): "Cooperation among prominent actors in a tourist destination," *Annals of Tourism Research*, 38(2), 607–629.

——— (2017): "Umfrage - Nachfrageanalyse Wintersport für Schweiz Tourismus," Discussion paper, Auftraggeber: Schweiz Tourismus.

Biagi, B., M. G. Brandano, and M. Pulina (2017): "Tourism taxation: A synthetic control method for policy evaluation," *International Journal of Tourism Research*, 19(5), 505–514.

Blanco, G., C. A. Flores, and A. Flores-Lagunes (2011): "Bounds on Quantile Treatment Effects of Job Corps on Participants' Wages," *Working paper.*

Bodory, H., and M. Huber (2018): "The causalweight package for causal inference in R," *SES Working Paper 493, University of Fribourg.*

Bos, I., and J. E. Harrington (2010): "Endogenous cartel formation with heterogeneous firms," *RAND Journal of Economics*, 41(1), 92–117.

Bouttell, J., P. Craig, J. Lewsey, M. Robinson, and F. Popham (2018): "Synthetic control methodology as a tool for evaluating population-level health interventions," *Journal of Epidemiology and Community Health*, 72(8), 673–678.

Bramwell, B., and B. Lane (2000): "Tourism Collaboration and Partnerships: Politics, Practice and Sustainability," *Channel View Publications*, 2, 198–229.

Breiman, L. (2001): "Random forests," *Machine Learning*, 45, 5–32.

Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984): *Classification and Regression Trees.* Wadsworth.

Brynjolfsson, E., and K. McElheran (2016): "The rapid adoption of data-driven decision-making," *American Economic Review*, 106(5), 133–39.

Castillo, V., L. Figal Garone, A. Maffioli, and L. Salazar (2017): "The causal effects of regional industrial policies on employment: A synthetic control approach," *Regional Science and Urban Economics*, 67, 25–41.

Cavallo, E., S. Galiani, I. Noy, and J. Pantano (2013): "Catastrophic natural disasters and economic growth," *The Review of Economics and Statistics*, 95(5), 1549–1561.

Chassang, S., K. Kawai, J. Nakabayashi, and J. Ortner (2020): "Data Driven Regulation: Theory and Application to Missing Bids," *NBER working paper series*.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21, C1–C68.

Chernozhukov, V., C. Hansen, and M. Spindler (2016): "hdm: High-Dimensional Metrics," *R Journal*, 8(2), 185–199.

Chotibhongs, R., and D. Arditi (2012a): "Analysis of collusive bidding behavior," *Construction Management and Economics*, 30(3), 221–231.

——— (2012b): "Detection of Collusive Behavior," *Journal of Construction Engineering and Management*, 138(11), 1251–1258.

Clark, R., D. Coviello, J.-F. Gauthier, and A. Shneyerov (2018): "Bid rigging and entry deterrence in public procurement: Evidence from an investigation into collusion and corruption in Quebec," *The Journal of Law, Economics, and Organization*, 34(3), 301–363.

Conley, T. G., and F. Decarolis (2016): "Detecting bidders groups in collusive auctions," *American Economic Journal: Microeconomics*, 8(2), 1–38.

Cortes, C., and V. Vapnik (1995): "Support-vector networks," *Machine learning*, 20(3), 273–297.

Crede, C. J. (2019): "A structural break cartel screen for dating and detecting collusion," *Review of Industrial Organization*, 54(3), 543–574.

De Palma, A., R. Lindsey, and G. Monchambert (2017): "The economics of crowding in rail transit," *Journal of Urban Economics*, 101, 106–122.

De Witte, A., C. Macharis, P. Lannoy, C. Polain, T. Steenberghen, and S. Van de Walle (2006): "The impact of "free" public transport: The case of Brussels," *Transportation Research Part A: Policy and Practice*, 40(8), 671–689.

Decarolis, F., M. Goldmanis, and A. Penta (2020): "Marketing agencies and collusive bidding in online ad auctions," *Management Science*, 66(10), 4433–4454.

Desmaris, C. (2014): "The reform of passenger rail in Switzerland: More performance without competition," *Research in Transportation Economics*, 48, 290–297.

Doerr, L., F. Dorn, S. Gaebler, and N. Potrafke (2020): "How new airport infrastructure promotes tourism: evidence from a synthetic control approach in German regions," *Regional Studies*, 54, 1402–1412.

Eppen, G., W. Hanson, and K. Martin (1991): "Bundling - New Products, New Markets, Low Risk," *Sloan Management Review*, 32(4), 7–14.

Esposito, F., and M. Ferrero (2006): "Variance screens for detecting collusion: an application to two cartel cases in Italy," *Working paper.*

Falk, M. (2010): "A dynamic panel data analysis of snow depth and winter tourism," *Tourism Management*, 31(6), 912–924.

Falk, M., and M. Scaglione (2018): "Effects of ski lift ticket discounts on local tourism demand," *Tourism Review*, 73(4), 480–491.

——— (2021): "Effects of price model copycats in the ski industry," *Tourism Analysis*, 26(1), 71–75.

Falk, M., and R. Steiger (2020): "Size facilitates profitable ski lift operations," *Tourism Economics*, 26(7), 1197–1211.

Feinstein, J., M. Block, and F. Nold (1985): "Asymmetric Information and Collusive Behavior in Auction Markets," *The American Economic Review*, 75(3), 441–460.

Firgo, M., and A. Kügler (2018): "Cooperative pricing in spatially differentiated markets," *Regional Science and Urban Economics*, 73, 51–67.

Flagestad, A., and C. A. Hope (2001): "Strategic success in winter sports destinations: A sustainable value creation perspective," *Tourism Management*, 22(5), 445–461.

Foremny, A., J. Kulejewski, H. Anysz, and A. Nicał (2018): "Collusion and bid rigging in the construction industry: case studies from Poland," *Proceedings of the Creative Construction Conference.*

Frangakis, C., and D. Rubin (2002): "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29.

Frank, L. E., and J. H. Friedman (1993): "A statistical view of some chemometrics regression tools," *Technometrics*, 35(2), 109–135.

Froeb, L. M., D. Sibley, M. Doane, and B. Pinto (2014): "Screening for collusion as a problem of inference," *Oxford handbook of international antitrus economics*, pp. 523–553.

Galiani, S., and B. Quistorff (2017): "The Synth_runner Package: Utilities to Automate Synthetic Control Estimation Using Synth," *The Stata Journal*, 17(4), 834–849.

García Rodríguez, M. J., V. Rodríguez Montequín, F. Ortega Fernández, and J. M. Villanueva Balsera (2020): "Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain," *Complexity, 2020*.

Garrod, B., and A. Fyall (2017): "Collaborative destination marketing at the local level: benefits bundling and the changing role of the local tourism association," *Current Issues in Tourism*, 20(7), 668–690.

Gonseth, C. (2013): "Impact of snow variability on the Swiss winter tourism sector: Implications in an era of climate change," *Climatic Change*, 119(2), 307–320.

Hagenauer, J., and M. Helbich (2017): "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Systems with Applications*, 78, 273–282.

Harrington, J. E. (2008): *Detecting cartels*.

Heckman, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables, and a Simple Estimator for such Models," *Annals of Economic and Social Measurement*, 5(4), 475–492.

——— (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153–161.

HES-SO Wallis (2016): "Wertschöpfung des Tourismus im Wallis," Discussion paper.

Ho, T. K. (1995): "Random Decision Forests," *Proceedings of the 3rd International Conference on Document Analysis.*

Hollander, M., D. A. Wolfe, and E. Chicken (2014): *Nonparametric Statistical Methods*, Wiley Series in Probability and Statistics. Wiley, third edition edn.

Huber, M. (2014): "Treatment evaluation in the presence of sample selection," *Econometric Reviews*, 33(8), 869–905.

Huber, M., and D. Imhof (2019): "Machine learning with screens for detecting bid-rigging cartels," *International Journal of Industrial Organization*, 65, 277–301.

Huber, M., D. Imhof, and R. Ishii (2020): "Transnational machine learning with screens for flagging bid-rigging cartels," *Working papers SES 519, Faculty of Economics and Social Sciences of Fribourg (Switzerland).*

Huber, M., J. Meier, and H. Wallimann (2021): "Business analytics meets artificial intelligence: Assessing the demand effects of discounts on Swiss train tickets," *arXiv working paper.*

Hueschelrath, and Veith (2014): "Cartel Detection In Procurement Markets," *Managerial and Decision Economics*, 35(6), 404–422.

Hünermund, P., J. Kaminski, and C. Schmitt (2021): "Causal Machine Learning and Business Decision Making," *Working paper.*

Imai, K. (2008): "Sharp bounds on the causal effects in randomized experiments with 'truncation-by-death'," *Statistics & Probability Letters*, 78(2), 144–149.

Imbens, G. W., and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Imhof, D. (2017): "Econometric Tests to Detect Bid-rigging Cartels: Does It Work?," *Working Papers SES 483, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland).*

——— (2019): "Detecting Bid-Rigging Cartels With Descriptive Statistics," *Journal of Competition Law & Economics*, 15(4), 427–467.

Imhof, D., Y. Karagök, and S. Rutz (2018): "Screening for Bid Rigging—Does It Work?," *Journal of Competition Law & Economics*, 14(2), 235–261.

Imhof, D., and H. Wallimann (2021): "Detecting bid-rigging coalitions in different countries and auction formats," *arXiv working paper.*

Ishii, R. (2009): "Collusion in Repeated Procurement Auction: a Study of Paving Market in Japan," *International Journal of Industrial Organization*, 27(2), 137–144.

——— (2014): "Bid Roundness Under Collusion in Japanese Procurement Auctions," *Review of Industrial Organization*, 44(3), 241–254.

Jakobsson, M. (2007): "Bid Rigging in Swedish Procurement Auctions," *Working Paper.*

Jiménez, J. L., and J. Perdiguero (2012): "Does Rigidity of Prices Hide Collusion?," *Review of Industrial Organization*, 41(3), 223–248.

Kaul, A., S. Klössner, G. Pfeifer, and M. Schieler (2017): "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates," *MPRA Working paper, Munich University Library.*

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015): "Prediction policy problems," *American Economic Review*, 105(5), 491–95.

Knaus, M. C. (2021): "A double machine learning approach to estimate the effects of musical practice on student's skills," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(1), 282–300.

Koenig, U., and B. Abegg (1997): "Impacts of climate change on winter tourism in the swiss alps," *Journal of Sustainable Tourism*, 5(1), 46–58.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76(3), 1071–1102.

Liaw, A., and M. Wiener (2018): "Package randomForest: Breiman and Cutler's Random Forest for Classificaction and Regression," *R package version 4.6-14.*

Lima, R. C., and G. M. Resende (2021): "Using the Moran's I to detect bid rigging in Brazilian procurement auctions," *The Annals of Regional Science*, 66(2), 237–254.

Little, R., and D. Rubin (1987): *Statistical Analysis with Missing Data.* Wiley, New York.

LIU, L., AND R.-C. CHEN (2017): "A novel passenger flow prediction model using deep learning methods," *Transportation Research Part C: Emerging Technologies*, 84, 74–91.

LIU, Y., Z. LIU, AND R. JIA (2019): "DeepPF: A deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, 101, 18–34.

LOHMANN, C., AND N. CRASSELT (2012): "Investments in ski areas: Effects of compensation payments in a complementary value network," *Tourism Economics*, 18(2), 339–358.

LÜSCHER, R. (2020): "10 Jahre Sparbillette – Attraktive Preise und höhere Nachfrage für den öV.," Discussion paper.

MACHIAVELLI, A. (2001): "Tourist destinations as integrated systems," *Tourism Review*, 56(3/4), 6–11.

MALASEVSKA, I., AND E. HAUGOM (2018): "Optimal prices for alpine ski passes," *Tourism Management*, 64, 291–302.

MCAFEE, R. P., AND J. MCMILLAN (1992): "Bidding rings," *The American Economic Review*, 82(3), 579–599.

MEYER, D. (2015): "Support vector machines: The Interface to libsvm in package e1071," *R News*.

MOHRING, H. (1972): "Optimization and scale economies in urban bus transportation," *The American Economic Review*, 62(4), 591–604.

MORENO-GENÉ, J., L. SÁNCHEZ-PULIDO, E. CRISTOBAL-FRANSI, AND N. DARIES (2018): "The economic sustainability of snow tourism: The case of ski resorts in Austria, France, and Italy," *Sustainability (Switzerland)*, 10(9), 1–20.

NUSAIR, K., H. J. YOON, S. NAIPAUL, AND H. G. PARSA (2010): "Effect of price discount frames and levels on consumers' perceptions in low-end service industries," *International Journal of Contemporary Hospitality Management*, 22(6), 814–835.

OECD (2013): "Roundtable on Ex Officio Cartel Investigations and the Use of Screens to Detect Cartels," *SSRN Electronic Journal*.

OMRANI, H. (2015): "Predicting travel mode of individuals by machine learning," *Transportation Research Procedia*, 10, 840–849.

PARRY, I. W., AND K. A. SMALL (2009): "Should urban transit subsidies be reduced?," *American Economic Review*, 99(3), 700–724.

PAULLEY, N., R. BALCOMBE, R. MACKETT, H. TITHERIDGE, J. PRESTON, M. WARDMAN, J. SHIRES, AND P. WHITE (2006): "The demand for public transport: The effects of fares, quality of service, income and car ownership," *Transport policy*, 13(4), 295–306.

PEARL, J. (2000): *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

PECHLANER, H., A. PRESENZA, AND M. CIPOLLINA (2010): "Analysing tourism stakeholders networks," *Tourism Review*, 65(4), 17–30.

PERDUE, R. R. (2002): "Perishability, yield management, and cross-product elasticity: A case study of deep discount season passes in the Colorado ski industry," *Journal of Travel Research*, 41(1), 15–22.

PESENDORFER, M. (2000): "A Study of Collusion in First-Price Auctions," *Review of Economic Studies*, 67(3), 381–411.

PIKKEMAAT, B., AND K. WEIERMAIR (2007): "Innovation through cooperation in destinations: First results of an empirical study in Austria," *Anatolia: An International Journal of Tourism and Hospitality Research*, 18(1), 67–83.

PORTER, R. H., AND J. D. ZONA (1993): "Detection of Bid Rigging in Procurement Auctions," *Journal of Political Economy*, 101(3), 518–538.

——— (1999): "Ohio school milk markets: an analysis of bidding," *RAND Journal of Economics*, 30(2), 263–288.

POWELL, D. (2017): "Synthetic Control Estimation Beyond Case Studies: Does the Minimum Wage Reduce Employment?," *RAND working paper*.

RABUZIN, K., AND N. MODRUSAN (2019): "Prediction of Public Procurement Corruption Indices using Machine Learning Methods.," *Proceedings of the 9th International Conference on Data Science, Technology and Applications*.

REDMAN, L., M. FRIMAN, T. GÄRLING, AND T. HARTIG (2013): "Quality attributes of public transport that attract car users: A research review," *Transport policy*, 25, 119–127.

RICE, J. A. (2007): *Mathematical Statistics and Data Analysis*. BROOKS/COLE CENGAGA Learning, third edition edn.

Rigall-I-Torrent, R., and M. Fluvià (2007): "Public goods in tourism municipalities: Formal analysis, empirical evidence and implications for sustainable development," *Tourism Economics*, 13(3), 361–378.

——— (2010): "Managing tourism products and destinations embedding public good components: A hedonic approach," *Tourism Management*, 32(2), 244–255.

Robins, J. M., and A. Rotnitzky (1995): "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90, 122–129.

Robins, J. M., A. Rotnitzky, and L. Zhao (1994): "Estimation of Regression Coefficients When Some Regressors Are not Always Observed," *Journal of the American Statistical Association*, 90, 846–866.

Rotaris, L., and R. Danielis (2014): "The impact of transportation demand management policies on commuting to college facilities: A case study at the University of Trieste, Italy," *Transportation Research Part A: Policy and Practice*, 67, 127–140.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

——— (1976): "Inference and Missing Data," *Biometrika*, 63, 581–592.

Rütter, H., U. Rütter-Fischbacher, A. Berwert, and M. Landolt (2001): "Der Tourismus im Wallis: Wertschöpfungsstudie," Discussion paper.

Saastal Bergbahnen AG (2016): "Geschäftsbericht 2015/16," Discussion paper.

——— (2017): "Geschäftsbericht 2016/17," Discussion paper.

——— (2018): "Geschäftsbericht 2017/18," Discussion paper.

Saastal Tourismus AG (2019): "Geschäftsbericht 2017/18," Discussion paper.

——— (2020): "Geschäftsbericht 2018/19," Discussion paper.

Saló, A., A. Garriga, R. Rigall-I-Torrent, M. Vila, and M. Fluvià (2014): "Do implicit prices for hotels and second homes show differences in tourists' valuation for public attributes for each type of accommodation facility?," *International Journal of Hospitality Management*, 36, 120–129.

Samuelson, P. A. (1954): "The Pure Theory of Public Expenditure," *The Review of Economics and Statistics*, 36(4), 387.

Seilbahnen Schweiz, S. (2017): "Fakten und Zahlen zur Schweizer Seilbahnbranche," Discussion paper.

Semenova, V., and V. Chernozhukov (2020): "Debiased machine learning of conditional average treatment effects and other causal functions," *forthcoming in the Econometrics Journal.*

Silveira, D., S. Vasconcelos, M. Resende, and D. O. Cajueiro (2021): "Won't Get Fooled Again: A Supervised Machine Learning Approach for Screening Gasoline Cartels," *CESifo Working Paper Series.*

Steiger, R., D. Scott, B. Abegg, M. Pons, and C. Aall (2019): "A critical review of climate change risk for ski tourism," *Current Issues in Tourism*, 22(11), 1343–1379.

Stettler, J., M. Zemp, and A. Steffen (2015): "Alpine Smart Tourism Destination: Überblick über den Smart Tourism Ansatz sowie über geplante Initiativen.," *ITW Working Paper.*

Svensson, B., S. Nordin, and A. Flagestad (2005): "A governance perspective on destination development-exploring partnerships, clusters and innovation systems," *Tourism Review*, 60(2), 32–37.

Thao, V. T., W. von Arx, and J. Frölicher (2020): "Swiss cooperation in the travel and tourism sector: long-term relationships and superior performance," *Journal of Travel Research*, 59(6), 1044–1060.

Tibshirani, J., S. Athey, R. Friedberg, V. Hadad, D. Hirshberg, L. Miner, E. Sverdrup, S. Wager, and M. Wright (2020): "grf: Generalized Random Forests," *R package version 1.2.0.*

Tibshirani, R. (1996): "Regression Shrinkage and Selection via The Lasso," *Journal of the Royal Statistical Society, Series B*, 58(1), 267–288.

Tkalec, M., I. Zilic, and V. Recher (2017): "The effect of film industry on tourism: Game of Thrones and Dubrovnik," *International Journal of Tourism Research*, 19(6), 705–714.

VAN DER LAAN, M., E. C. POLLEY, AND A. E. HUBBARD (2008): "Super Learner," *Statistical Applications of Genetics and Molecular Biology*, 6(1).

VARIAN, H. R. (2014): "Big data: New tricks for econometrics," *Journal of Economic Perspectives*, 28(2), 3–28.

WAGER, S., AND S. ATHEY (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113(523), 1228–1242.

WALLIMANN, H. (2020): "A complementary product of a nearby ski lift company," *Tourism Economics*, published ahead of print. doi: 10.1177/1354816620968329.

WALLIMANN, H., D. IMHOF, AND M. HUBER (2020): "A Machine Learning Approach for Flagging Incomplete Bid-rigging Cartels," *Working papers SES 513, Faculty of Economics and Social Sciences of Fribourg (Switzerland).*

WEGELIN, P. (2018): "Is the mere threat enough? An empirical analysis about competitive tendering as a threat and cost efficiency in public bus transportation," *Research in Transportation Economics*, 69, 245–253.

WILKE, E. P., B. K. COSTA, O. B. D. L. FREIRE, AND M. P. FERREIRA (2019): "Interorganizational cooperation in tourist destination: Building performance in the hotel industry," *Tourism Management*, 72, 340–351.

WRIGHT, M. N., AND A. ZIEGLER (2017): "ranger: A fast implementation of random forests for high dimensional data in C++ and R," *Journal of Statistical Software*, 77, 1–17.

WYSS, R., T. LUTHE, AND B. ABEGG (2015): "Building resilience to climate change – the role of cooperation in alpine tourism networks," *Local Environment*, 20(8), 908–922.

YANG, J.-C., H.-C. CHUANG, AND C.-M. KUAN (2020): "Double machine learning with gradient boosting and its application to the Big N audit quality effect," *Journal of Econometrics*, 216(1), 268–283.

YANG, W., L. ZHANG, AND A. S. MATTILA (2016): "Luxe for Less: How Do Consumers React to Luxury Hotel Price Promotions? The Moderating Role of Consumers' Need for Status," *Cornell Hospitality Quarterly*, 57(1), 82–92.

YAP, M., AND O. CATS (2020): "Predicting disruptions and their passenger delay impacts for public transport stops," *Transportation*, pp. 1–29.

Zehrer, A., and K. Hallmann (2015): "A stakeholder perspective on policy indicators of destination competitiveness," *Journal of Destination Marketing and Management*, 4(2), 120–126.

Zhang, J., R. Lindsey, and H. Yang (2018): "Public transit service frequency and fares with heterogeneous users under monopoly and alternative regulatory policies," *Transportation Research Part B: Methodological*, 117, 190–208.

Zhang, J., and D. B. Rubin (2003): "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by 'Death'," *Journal of Educational and Behavioral Statistics*, 28(4), 353–368.

Zhang, J., D. B. Rubin, and F. Mealli (2009): "Evaluating The Effects of Job Training Programs on Wages through Principal Stratification," *Journal of the American Statistical Association*, 104(485), 166–176.